EQS 6 Structural Equations Program Manual

Peter M. Bentler
University of California, Los Angeles

Multivariate Software, Inc.
15720 Ventura Blvd., Suite 306
Encino, California USA 91436-2989

Tel: (818) 906 0740 · Fax: (818) 906 8205
Email: sales@mvsoft.com · Web: www.mvsoft.com

The correct bibliographic citation for this document is as follows:

Bentler, P. M. (2006). **EQS 6 Structural Equations Program Manual**.
Encino, CA: Multivariate Software, Inc.

ISBN 1-885898-03-7

Copyright © 1989-2006 by Peter M. Bentler
All rights reserved.

Version 6
Printed in the United States of America

TABLE OF CONTENTS

1.	OVERVIEW	1
2.	BASIC STRUCTURAL EQUATION MODELS	21
3.	PROGRAM INPUT	59
4.	PROGRAM OUTPUT	103
5.	STATISTICAL METHODS	125
6.	LAGRANGE MULTIPLIER AND WALD TESTS	159
7.	MULTISAMPLE COVARIANCE AND CORRELATION STRUCTURES	187
8.	MEAN AND COVARIANCE STRUCTURES	203
9.	MULTISAMPLE MEAN AND COVARIANCE STRUCTURES	223
10.	GROWTH CURVE MODELS	241
11.	MULTILEVEL METHODS	267
12.	MISSING DATA	275
13.	SIMULATION AND RESAMPLING	301
14.	RELIABILITY AND PRACTICAL FIT	337
15.	REFERENCES	371

PREFACE

Linear structural equation modeling has become an indispensable methodology for specifying, estimating, and testing hypothesized interrelationships among a set of substantively meaningful variables. For many years, an arcane matrix language was used by researchers to set up and run models, and to describe them in publications. While technically accurate, this language unnecessarily hid the important first principles that can be mastered quite easily by anyone with a reasonable scientific background, and suggested the need for complicated specialized procedures to handle anything other than standard applications. Happily for the research community, many programs in the field have given up these arcane approaches and recently have become much easier to use. Unfortunately, this ease of use is often obtained at the expense of hiding the basic principles that underlie the methodology. Ever since its first public distribution via BMDP Statistical Software (Bentler, 1985), EQS verified that it was possible to take many complications out of structural modeling by providing a simple, consistent, yet technically advanced and accurate approach to structural modeling. This history continues today with the release of EQS 6. Even though the range of models, data, and statistical methods available for analysis has expanded substantially, EQS maintains its tradition of delivering the utmost in simplicity with the greatest possible statistical accuracy.

This program *Manual* provides an overall description of the EQS program, as it will be encountered in any setting. The procedures for setting up models, running them under a wide variety of options, and interpreting the program output is described herein for EQS on mainframes, a wide variety of Unix and Linux platforms, as well as for Windows and, hopefully as in the past, Macintosh computers. While Windows and Mac users have available the text interface described herein, these versions of the program are substantially expanded as compared to the mainframe and Unix versions. They contain even simpler to use graphical interfaces as well as a wide variety of general purpose statistical methods and plots that might be useful in research. Only the Windows program is available immediately. Its features are described in Bentler, P. M., & Wu, E. J. C. (2002). *EQS 6 for Windows User's Guide.* Encino, CA: Multivariate Software.

There are literally dozens of new features in this release of EQS. These include user interface improvements, greater program stability, and a wide variety of new statistical methods, many of which are outgrowths of my own structural modeling research program with colleagues from around the world as well as with current and former students. The theoretical statistics, algorithmic experiments, and applied data analysis that form a conceptual and experimental basis for EQS were developed in part with support by research grants DA00017 and DA01070 from the National Institute on Drug Abuse. The original contributions of Ke-Hai Yuan to this field, many of which are incorporated in EQS 6, have been unparalleled in the recent history of statistical psychometrics. Equally invaluable programming contributions to this research were made by Eric J. -C. Wu and, more recently, David Sookne, who also updated several chapters of this *Manual*. Valuable drafts of Chapters 10-13 were provided by Seongeun Kim. The advice on statistical, computational, human engineering, and a variety of other issues provided by various colleagues and students, especially, in recent years, R. Abbott, G. E. Barnes, M.. Berkane, B. Byrne, W. Chan, C. –P. Chou, P. Dudgeon, D. Enzmann, M. S. Gold, S. Green, G. –S. Guo, G. Hancock, K. Hayashi, G. Hellemann, L. Hu, M. Jamshidian, Y. Kano, K. Kim, S. –Y. Lee, L. Li, J. –J. Liang, M. D. Newcomb, M. Otten, W. –Y. Poon, T. Raykov, A. Satorra, V. Savalei, R. Schumacker, J. Stein, M. L. Tang, J. Ullman, J. Weng, S. -T. Wu, J. Xie, Y. –F. Yung, and W. Zhu is gratefully acknowledged. E. L. Houck provided invaluable quality control advice on the *Manual* as well as the program. The helpful encouragement of M. Kozelchick, I. Nuñez, T. Stroud, and K. Monterrosa of Multivariate Software is gratefully acknowledged. Final editing of this *Manual* was done by Virginia Lawrence of CogniText. The cover was designed by Brandon Morino.

The release of EQS 6 was delayed due to a variety of professional and personal circumstances. The most difficult problem was Microsoft's termination of support for some key utilities, distributed by 3rd party vendors, that had been used in EQS for over a decade. As a result, Eric J. –C. Wu decided to rewrite huge portions of code. The assistance of Y. He, Y. de la Peña Ay, and especially, G. –S. Guo, in achieving this code renewal are gratefully acknowledged.

As always, suggestions for improvement of the program and its documentation are actively solicited.

1. OVERVIEW

The computer program EQS was developed to meet two major needs in statistical software. At the theoretical level, applied multivariate analysis based on methods that are more general than those stemming from multinormal distribution theory for independent observations have not been available to statisticians and researchers for routine use. At the applied level, powerful and general methods have also required extensive knowledge of matrix algebra and related topics that are often not routinely available among researchers. EQS is meant to make advanced multivariable analysis methods accessible to the applied statistician and practicing data analyst. This document provides a brief overview of the steps required to run EQS. It is hoped that the reader is already familiar with the basic concepts of structural modeling, but a brief introduction to the topic is provided in the next chapter. This section outlines some features of the program to allow the knowledgeable reader to skip to Chapter 3 on program input. It also provides a guide to recent applications of structural equation modeling -- *SEM* -- in a variety of fields, as well as references to the statistical literature relevant to EQS. The references provided in this *Manual* – the largest collection of SEM references ever collected – are meant to supplement the excellent texts, published bibliographies, and recent reviews (e.g., Bentler & Dudgeon, 1996; MacCallum & Austin, 2000; Wolfle, 2003). As you read this chapter, you can skip the references unless you are looking for source material. If you are a beginner in this field, you should skim or skip the more technical material, which is not required as a background to Chapter 2. To keep the tone somewhat informal, as you see, you, the reader, may be called "you."

A linear structural model is often motivated or generated by causal hypotheses (e.g., Bentler & Speckart, 1981; Perugini & Bagozzi, 2001). Hence a popular older name for the methodology is "causal modeling" (Bagozzi, 1980; Bentler, 1980, 1998; Faulbaum & Bentler, 1994; Gollob & Reichardt, 1987; James, Mulaik, & Brett, 1982; Kenny, 1979; Saris, & Stronkhurst, 1984). This terminology and its implications have come under attack (Cliff, 1983; Holland, 1986, 1988; Marini & Singer, 1988; Sobel, 1994, 1995). Some good recent discussions of causality in this context, including causal discovery using methods related to structural modeling, is given by Glymour, Scheines, Spirtes, and Kelly (1987), Greenland (2000a), Lauritzen (2001), Pearl (2000, 2001), Robins (2001), Shadish, Cook, and Campbell (2002), Shipley (2000a), Snijders and Hagenaars (2001), Wegener and Fabrigar (2000), and West, Biesanz, and Pitts (2000). Luckily, hypotheses about means and covariances can be tested with EQS whether or not one truly has a "causal" theory in mind, especially since critiques (e.g., Freedman, 1987) have not found technical flaws in the methods (Bentler, 1987). Nonetheless, in applications things can – and easily do – go wrong, and a lot of care is needed to assure that results are free of artifacts and that reports do not imply misleading conclusions. As noted below, for example, the simple regression equation is a building-block of SEM, yet artifacts can and do appear in this simple model (Campbell & Kenny, 1999). Hence no doubt artifacts can appear in models with several regression-like equations, as well as in more complex models with equations at several levels (Marsh & Hau, 2002). Similarly, even though one is interested in a given model, others may exist that are equivalent in terms of fit (Hershberger, 1994; Lee & Hershberger, 1990; MacCallum, Wegener, Uchino, & Fabrigar, 1993; Markus, 2002; Raykov & Marcoulides, 2001; Raykov & Penev, 2001; Stelzl, 1986; Williams, Bozdogan, & Aiman-Smith, 1996). Yet, this point is often ignored by researchers (Steiger, 2001). Clearly, continuously upgraded training is needed to assure that inferences drawn from modeling are appropriate (e.g., Muthén, 1987). Nonetheless, consensus does not exist on a number of rather basic issues, and SEM and closely related fields are continually changing. Hence your knowledge is bound to be incomplete -- witness such unresolved issues as the role of 1-stage versus multi-stage approaches to testing (Anderson & Gerbing, 1988; Bentler, 2000; Bollen, 2000; Hayduk & Glaser, 2000; Herting & Costner, 2000; Mulaik & Millsap, 2000; Yuan & Chan, 2002), the possible value of limited information estimation and testing (e.g., MacCallum, Browne, & Preacher, 2002; Meehl & Waller, 2002), the controversy about how and when to use significance tests (Harlow, Mulaik, & Steiger, 1997; Nickerson, 2000), the possibility of substituting meta-analysis (Arthur, Bennett, & Huffcutt, 2001; Hedges & Pigott, 2001; Rosenthal & DiMatteo, 2001; Yuan & Bushman, 2002) for traditional tests (Howard, Maxwell, & Fleming, 2000), new approaches to contrasts and effect sizes (Rosenthal, Rosnow, & Rubin, 2000; Thompson, 2002), and so on.

Bentler-Weeks Model

EQS implements a general mathematical and statistical approach to the analysis of linear structural equation systems. The mathematical model is by now quite old (Bentler & Weeks, 1979, 1980, 1982, 1985; Bentler, 1983a,b). It has stood the test of time because it subsumes a variety of covariance and mean structure models, including multiple regression, path analysis, simultaneous equations, first- and higher-order confirmatory factor analysis, as well as regression and structural relations among latent variables. The model can handle all the linear data structures of the Jöreskog-Keesling-Wiley model (Cudeck, du Toit, & Sörbom, 2001; Jöreskog, 1973, 1977; Jöreskog & Sörbom, 1994; Keesling, 1972; Wiley, 1973) and related models (e.g., Browne & Arminger, 1995; McArdle & McDonald, 1984; McDonald, 1980), but it has a particular simplicity and coherence that makes it easy to learn and apply to new modeling situations. And it can be easily specified in a program like EQS.

Linear latent variable models have wide applicability (e.g., Anderson, 1990). Nonetheless, being a linear model, the Bentler-Weeks model cannot effectively deal with nonlinear latent variable models such as are currently being developed (e.g., Amemiya, 1985; Arminger & Muthén, 1998; Bollen, 1995; Carroll, Ruppert, & Stefanski, 1995; Lee & Zhu, 2000, 2002; Schumacker & Marcoulides, 1998; Wall & Amemiya, 2000; Yang-Wallentin & Jöreskog, 2001; Yalcin & Amemiya, 2001; Zhu & Lee, 1999). Nonetheless, linearity is usually a good approximation, especially in the social and behavioral sciences where a priori nonlinear functional forms are not typical, and where errors of measurement are not trivial. Witness, for example, the extensive use of linear regression and the rare use of nonlinear regression in these fields. Actually, models that at first glance appear to require special handling often can be brought into the EQS framework. One example is data from a data cube rather than a data matrix (Bentler & Lee, 1979; Oort, 1999, 2001), which can sometimes be handled as a generalized multimode latent variable model (Bentler, Poon, & Lee, 1988). Another example involves ordinal categorical data. Although a linear model may not best describe the relations among the observed categorical variables, if a transformation is made to express the category scores on a variable as a function of an underlying continuous normally distributed variable, linear relations again can be used to describe the relations among the underlying variables. In EQS, we use the theory of Lee, Poon, and Bentler (1995) involving polychoric and polyserial correlations to achieve this, and the Bentler-Weeks setup again becomes relevant. Other nonlinear models can be implemented with linear models using nonlinear constraints (e.g., du Toit & Cudeck, 2001; Kenny & Judd, 1984; Li, Duncan, Duncan, Yang-Wallentin, Acock, & Hops, 2001; Wood, 2000). Such constraints are planned (Jamshidian & Bentler, 1993) for EQS 6, but are not implemented in its initial release.

The matrix algebra that underlies the Bentler-Weeks model is actually not needed to specify a model correctly. Only a few critical concepts are needed to understand the model, and all these concepts will be clearly explained – several times, in appropriate places, without resorting to matrix details. However, a short summary is given here to take the mystery out this model. If you are acquainted with regression models such as

$$y = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + e \qquad (1.1)$$

you know that the *parameters* of the model are the β ("beta") coefficients, the variances and covariances of the x_i (often taken as known, or fixed), and the variance of the error term e. The Bentler-Weeks model contains a series of such equations relating dependent variables to independent variables. Here, y is a dependent variable, and the x_i and e are independent variables. Note that the Bentler-Weeks definition of each and every variable in a model as either an *independent* or a *dependent* variable is unusual, because the x_i need not be experimentally manipulated, and residual terms such as e are rarely considered important enough to have the "independent" variable designation. But if e is never on the left side of an equation, in this framework, it is an independent variable. Dependent variables can be expressed as a structural regression function of other variables, while independent variables are never structurally regressed on other variables. In EQS, models are specified by regression-like equations; "regression-like" because the error term e may not necessarily be independent of the other predictors as in regression (see e.g., McDonald & Ho, 2002, for a discussion of the distinction). The only other things to specify are variances and covariances, and possibly means or intercepts, of the independent variables.

A model will contain many equations. Dependent variables are collected in the column vector variable η ("eta"), which has as many entries as there are variables on the left side of equations. Independent variables are collected in the vector variable ξ ("ksi"), which contains all the other variables that may be considered in a model, except for the dependent variables. In this setup, observed variables, latent factors, errors in variables, or disturbances in factors are all treated equally and go either into ξ or η. If there is a path diagram for the model, each and every variable in that diagram is placed into either η or ξ. Although this seems mysterious, actually it is very simple. If a variable has a one-way arrow aiming at it, it is one of the η variables. Otherwise it is a ξ variable! The structural matrix equation that relates these variables is given in the Bentler-Weeks approach by

$$\eta = \beta\eta + \gamma\xi \tag{1.2}$$

where β contains the coefficients for the regression of η variables on each other and γ ("gamma") contains the coefficients for the regression of η variables on ξ variables. The equation is a form of *simultaneous equation* model. What is different here is that the beta weights, above, may be placed into either β or γ, depending on what type of predictor is involved. You do not need to know this – the program figures it out. All you need to know is that the coefficient represented by a one-way arrow in a diagram will be placed by EQS into either β or γ. Only independent variables have variances and covariances as parameters of the model. The covariance matrix Φ ("phi") of the independent variables is given as

$$\Phi = E(\xi - \mu_\xi)(\xi - \mu_\xi)' \tag{1.3}$$

where μ_ξ ("mu") is the vector of means of the independent variables. You simply have to know that a two-way arrow in a diagram represents a covariance (correlation) that will go into the appropriate part of Φ. If there is a *constant* (called V999 in EQS) in the diagram, a one way arrow from the constant to a variable will be translated into μ_ξ. Thus the parameter matrices of the model show up easily in a diagram but are given abstractly as β, γ, Φ and μ_ξ. These can be used to generate the model-implied means μ and covariance matrix Σ (cap "sigma") of the *observed* variables. To obtain this, we arrange all the variables into a vector $v=(\eta',\xi')'$ ("nu"), and select all the observed variables $z = Gv$ from these using a known [0,1] selection matrix G. (G is known because the diagram will show which of the η and ξ are observed variables.) Then we can equivalently write the Bentler-Weeks linear relation as $v=Bv+\Gamma\zeta$, where B contains β and some known zeros, and Γ (cap "gamma") contains γ and some known [0,1] entries. A bit of algebra, and the algebra of expectations, shows that the covariance matrix and mean vector of the z variables under the Bentler-Weeks model are given as

$$\Sigma = G(I-B)^{-1}\Gamma\Phi\Gamma'(I-B)^{-1}'G' \text{ and } \mu = G(I-B)^{-1}\Gamma\mu_\xi, \tag{1.4}$$

These equations are given for reference purposes. You should know they exist, though there is no special reason to remember them because in EQS you will not work with these matrices directly. However, your input will allow EQS to create these matrices and work with them internally.

The Bentler-Weeks model is only one approach to linear structural equation modeling. The main alternative model was developed by Jöreskog (1973), Keesling (1972), and Wiley (1973). It does not distinguish between independent and dependent variables in the sense above. Its key distinction is between measured and latent variables, and has separate linear relations for different sets of variables. A factor analytic measurement model relates measured and latent variables, and a simultaneous equation model describes the relations between latent variables. The name "FASEM" to describe the **f**actor **a**nalytic **s**imultaneous **e**quation **m**odel has been proposed (Bentler, 1986b) as more accurate and descriptive than "LISREL," the name of a particular computer program (Jöreskog & Sörbom, 1994) that is sometimes used as a generic name for SEM (e.g., Hayduk, 1996). Both FASEM and Bentler-Weeks model can handle the same types of models, though the setup is more consistent across types of models for the latter.

Because the population means μ and covariances Σ are structured in terms of more basic parameters, this methodology is often called *structural modeling*. The structure is on the right hand side of the above equations. In order to write the Bentler-Weeks parameters in a simpler notation, we may let the vector θ ("theta") represent all the free parameters in $\{\beta, \gamma, \Phi, \mu_\xi\}$. When the means of the observed variables are not relevant, we just do covariance structure modeling, and we can write $\Sigma = \Sigma(\theta)$, that is, the covariance matrix of the observed variables depends on the parameters θ. Similarly, a structure for the observed means is $\mu = \mu(\theta)$. Mean structures are almost always done when there is also a covariance structure. In practice structural models are estimated by comparing the vector of

sample means $\bar{X}$ to the model $\mu(\theta)$, and the sample covariance matrix S to the model $\Sigma(\theta)$, and finding those parameter estimates $\hat{\theta}$ that minimize some measure of discrepancy between data and model.

Statistical Methods

To provide a coherent way to describe estimation and testing in SEM, we may arrange all the data to be modeled ($\bar{X}$ and S), and ignoring the nonredundant parts of S, in a vector s. (The covariance of two variables V1 and V2 is the same as the covariance of V2 and V1, so both are not needed.) Similarly, we may arrange the implied population means μ and nonredundant covariances from Σ to be modeled into a vector as σ ("sigma"), which, of course, depends on the parameters $\sigma = \sigma(\theta)$. We will use this abstract notation even in more complicated situations, when, for example, s may contain data from several populations and σ contains the corresponding model.

The statistical theory used in EQS is closely related to computational theory. The computations depend on the *distribution assumed* for the observed variables, as well as on a *weight matrix* used during the computations. The function minimized is given by

$$Q = (s - \sigma(\theta))' W (s - \sigma(\theta)) \qquad (1.5)$$

where, as just noted, s is the vector of data to be modeled, i.e. the variances and covariances of the observed variables, and possibly the means, and σ is a model for the data. Among possible choices of estimates for the basic parameters θ, the optimal estimates $\hat{\theta}$ are those that minimize the function Q, yielding $\hat{Q}$. Corresponding to each estimate, there is a standard error estimate that can be used to evaluate the null hypothesis that $\theta_i = 0$ in the population. Also, the model can be evaluated. In the simplest situations, with a sample size N, the model test statistic T is usually computed as $T = (N-1)\hat{Q}$ (but note: *some important test statistics in EQS are not of this form*). T is used to evaluate the null hypothesis $\sigma = \sigma(\theta)$, i.e., whether the model holds in the population. The model will be accepted (not rejected) when the estimated model $\sigma(\hat{\theta})$ closely reproduces the data s, i.e., when the discrepancy ($s - \sigma(\hat{\theta})$) between the data and estimated model is small. You should be sure to understand that statistical acceptance of a model does not prove that the model is true!

The weight matrix W can be specialized in several ways to yield a number of different estimators that depend on the continuous distribution assumed (normal, elliptical, heterogeneous kurtosis, arbitrary) and the computations invoked ($W = I$, W optimal and fixed, W optimal and reweighted). The following table of methods is thus generated:

Case Weights = Identity or Fixed				
Weight Matrix	**Distribution Theory**			
	Normal[1]	**Elliptical**[1]	**Heterogeneous Kurtosis**[1]	**Arbitrary**
Identity	LS	ELS		
Fixed	GLS	EGLS	HKGLS	AGLS
Reweighted	RLS=ML	ERLS	HKRLS	ARLS[2]

1. Can be used with arbitrary distributions by using ME = xx, ROBUST
2. Test is equivalently available with AGLS as Yuan-Bentler corrected chi-square

Although the table organizes a variety of methods, it deals only with the situation in which all cases (subjects, observations) are independent, and each case contributes equally to the computation of means and covariances for

continuous variables. This is the standard situation, covering the vast majority of applications. Below we will consider several special situations when the above table does not describe the optimal procedure: missing data (for which ordinary sample means and covariances cannot even be computed), models for correlations (where standard covariance structure methods are inadequate), categorical variable models (in which the ordinary covariances are not the best data to be modeled), multilevel models (in which observations within a level may be correlated), and case-robust methods (in which observations or cases are weighted to achieve a special objective).

In this table LS means least squares, and the qualifiers "E" (elliptical), "HK" (heterogeneous kurtosis) and "A" (arbitrary) are used to designate the relevant non-normal distributions. (Normal theory needs no qualifier.) In estimation with a fixed weight matrix, the matrix is chosen to be calculable once from the input data, subject to belonging to the class of optimal weights that yield minimum chi-square or best generalized least squares (GLS) estimates. In estimation with a reweighted (R) matrix, the weight matrix not only belongs to the class of optimal matrices, but also is updated iteratively. That is, W is considered to be expressible in terms of the parameters of the model $W = W(\theta)$, and, as the parameter estimates are improved, so is W. Ordinary normal theory RLS yields maximum likelihood (ML) estimates.

Although RLS and ML yield the same estimates and standard errors (Lee & Jennrich, 1979), in addition to the RLS function based on Q, a separate ML function is computed and printed. The ML method is the typical default used by most researchers. It is appropriate when the data are multivariate normally distributed, and in this case yields estimates that are most precise in the sense of having the smallest standard errors when sample size is large. Hence the ML method has much to recommend it. Under the same conditions as ML, GLS also yields optimal estimates. LS may perform better in certain small sample situations, but its advantages are not yet established. Possibly it can be useful when ML has so much power that trivial discrepancies between model and data are associated with a test that rejects the model (e.g., Browne, MacCallum, Kim, Anderson, & Glaser, 2002).

The elliptical and heterogeneous kurtosis estimators and tests are unique to EQS. Elliptical theory assumes that variables are symmetrically distributed, but perhaps may have excess (or less) kurtosis as compared to a normal distribution. For example, there may be more cases in the tails than a normal curve would permit. However, all variables are assumed to have the same degree of excess kurtosis, i.e., there is an assumption of homogeneous marginal kurtoses. Elliptical distribution theory is described in such sources as Berkane and Bentler (1986, 1987a,b), Berkane, Kano, and Bentler (1994), Cambanis, Huang, and Simons (1981), Fang and Anderson (1990), Fang, Kotz, and Ng (1990), Gupta and Varga (1993), Kano, Berkane, and Bentler (1993), Muirhead (1982), Tyler (1983), and Waternaux (1984). It was developed for structural models primarily by Browne (1982, 1984), Bentler (1983a, 1985), Bentler and Berkane (1985), Bentler and Dijkstra (1985), and Shapiro and Browne (1987). A promising new approach was given by Berkane, Oden, and Bentler (1997), but this is not yet implemented in EQS. The more general heterogeneous kurtosis estimators allow different variables to have different degrees of kurtosis. HK methodology (HK=AVE) was developed Kano, Berkane, and Bentler (1990), while an alternative implementation of the method (HK=GEO) was given by Bentler, Berkane, and Kano (1991). When data are normal, both elliptical and HK methods reduce to normal theory statistics. When data are elliptical, the HK method reduces to it. There has been little study of elliptical estimators (e.g., Boomsma & Hoogland, 2001; Harlow, 1985; Hoogland, 1999; Sharma, Durvasula, & Dillon, 1989), and almost none of the heterogeneous kurtosis methods (e.g., Hu, Bentler, & Kano, 1992). An advantage of these methods is that they can handle models with a large number of badly distributed variables. However, they do not naturally deal with skewed variables. A new theory for skewed and kurtotic variables has recently been developed (Yuan & Bentler, 1997b, 1999b,c, 2000a). Its methodology will be incorporated into future versions of EQS.

In contrast, the arbitrary distribution methods are limited to small and intermediate sized models. This theory was developed by Browne (1982, 1984) and Chamberlain (1982), and is called asymptotically distribution-free (ADF) by Browne and optimal minimum distance by Chamberlain. Evidence shows that this methodology requires really large sample sizes to perform in a trustworthy manner (e.g., Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992; Yuan & Bentler, 1998a). The Yuan-Bentler (1997d) statistic, mentioned in footnote 2 as generated conceptually by ARLS, generates better performance with small samples. Possibly an even better all-around option is the Yuan-Bentler (1999a) F test. In EQS, these new tests are accompanied by an improved standard error estimator as well (Yuan & Bentler, 1997c).

As footnote 1 to the table shows, the methods based on specific distributional assumptions can be adjusted to deal with arbitrary distributions. Usually, this is done with a command such as METHOD = ML, ROBUST. What this does is accept the normal theory ML *estimates* $\hat{\theta}$, but scale the test statistic T to yield a robust test statistic using a theory of Satorra and Bentler (1986, 1988, 1994). It also improves standard error estimates using a so-called robust covariance matrix or sandwich triple product matrix (e.g., Bentler & Dijkstra, 1985; Browne, 1984; Kauermann & Carroll, 2001). In addition, extending the work of Browne (1984), EQS computes several residual-based test statistics based on Yuan and Bentler (1998a) and Bentler and Yuan (1999b). The residual-based test statistics $T_{YB(RES)}$ and $T_{F(RES)}$ seem to be promising in small samples (Nevitt, 2000). At this time, the Satorra-Bentler methodology and the residual based tests are probably the most accurate methods for dealing with nonnormal data. Actually, in EQS 6, the theory of these statistics has been extended to deal with various other analysis methods such as correlation structures.

Missing Data

Missing data are inevitable, and EQS has several methods for dealing with this problem. Ad hoc methods have been used in data analysis for decades and certain of these can be used in confidence under specialized circumstances. Ad hoc methods include mean imputation, listwise deletion, pairwise computations, stochastic regression imputation, hot deck imputation as well as more recently developed methods such as similar response pattern imputation or person mean imputation (Bernaards & Sijtsma, 2000). In these approaches, a modified data set or a covariance matrix is created that subsequently can be analyzed by any existing standard method designed for complete data. An advantage of these approaches is that they are relatively practical to implement; indeed several of these methods for dealing with incomplete data can be found in most well-known statistical program packages. Furthermore, nonnormality can be routinely handled when an imputed data matrix is analyzed with a distribution-free method. These methods are all appropriate when the amount of missing data is extremely small. In fact under some conditions there may be only marginal loss of accuracy or efficiency when compared to maximum likelihood; see Gold & Bentler (2000). In EQS, you have easy access to mean imputation, and listwise and pairwise present computations. However, these simple methods also can be inappropriate to use. For example, listwise deletion can render a longitudinal study with few cases left, resulting in grossly inefficient estimates (e.g., Brown, 1994). When the missing data mechanism is so-called missing at random (MAR), existing simulation results indicate that listwise deletion causes parameter estimates to be biased even for normal data (Little & Rubin, 1987; Schafer, 1997).

EQS for the first time adopts and extends an old theory to provide appropriate statistics to resurrect the pairwise-present method in covariance structure analysis. Marsh (1998) showed that this method, as typically implemented, cannot be trusted for evaluating models. However, many years ago, van Praag, Dijkstra, and Van Velzen (1985) had shown that there is a statistically adequate way to use this method with nonnormal data, at least with missing completely at random (MCAR) data. Their method appears not to have been studied further, nor to be implemented in any program. However, it is implemented in EQS 6 as a generalization of the AGLS method to missing data (the ordinary ADF method is a special case). But since AGLS tends to break down in small samples, it is likely that the same result may be observed here (Gold, Bentler, & Kim 2003). Hence for EQS 6 we have developed a new way to use the van Praag et al. theory. We developed it for ML, ROBUST. That is, ML estimators based on the pairwise covariances are obtained, and the resulting statistics are subsequently corrected in the conceptual way described by Satorra and Bentler (1994) using the van Praag et al. theory. This approach has also been extended to the case of known normal data, requiring less heavy computations and hence allowing its use with larger models; this will become available in EQS.

The ML approach to missing normal data has recently become very popular (Collins, Schafer, & Kam, 2001; Schafer & Graham, 2002; West, 2001). Pairwise correlations are not part of this approach. There are at least two ways to use ML, direct maximum likelihood (e.g., Neale, 2000; Wothke, 2000) and a two-stage approach based on the unstructured mean and covariance estimates obtained by the so-called EM (expectation/maximization) algorithm (e.g., Graham & Hofer, 2000). See also Jamshidian and Bentler (1999), and the summary provided by Enders (2001a). Typical assumptions under these two methods are ignorable nonresponse or MAR, and multivariate normality of data. Unfortunately, there is no effective way of verifying whether data are MAR, and data are often non-normal. Under nonnormality, normal theory methods generally lead to incorrect model and parameter

evaluation even with complete data, so it is unlikely that one can avoid such incorrectness with an added missing data problem. As an improvement over current methods, Yuan and Bentler (2000d) built on Arminger and Sobel (1990) and dropped the normal distribution assumption, and thus were able to develop several more accurate procedures for model inference in the direct ML approach. Also, based on the theory of generalized estimating equations (see Yuan & Jennrich, 1998, 2000), they provided a way to obtain consistent standard errors of the two-stage estimates. The direct ML approach with Yuan-Bentler corrections is implemented in EQS 6, along with some new residual-based tests. Preliminary independent evaluations are positive (Enders, 2001b). Their minimum chi-square approach, which leads to an estimator that is asymptotically at least as efficient as the two likelihood based estimators for either normal data or nonnormal data, will be implemented in EQS in the future. Multiple imputation (see e.g., Graham & Hofer, 2000; Longford, 2001; Schafer, 2001; Sinharay, Stern, & Russell, 2001) is not currently implemented. Specialized procedures, such as adding auxiliary variables correlated with missingness to improve estimation (Collins, Shafer, & Kam, 2001; Graham, 2003) can be accomplished in EQS following their developers' approaches.

A problem raised by incomplete data is that there are lots of missing and complete data patterns, and hence also several sets of means and covariance matrices of different sizes. Potentially, these various samples may be missing completely at random (MCAR) samples from a single population with a given mean vector and covariance matrix. If this can be established, a single SEM model for the population can be considered; while if means and covariances are not homogeneous, this may not be advisable. There have been developments to test MCAR in several different areas (e.g., generalized estimating equations, Chen & Little, 1999). In the area of multivariate normal data, there are currently two proposed test statistics for analyzing whether incomplete data patterns are MCAR (Little, 1988; Tang & Bentler, 1998). Little, as well as Tang-Bentler, use the EM algorithm to obtain MLE of free parameters under the MCAR assumption, and propose a test statistic to evaluate an MCAR null hypothesis. Little's (1988) MCAR test is based on evaluating the homogeneity of available means for different patterns of incomplete data. Little also mentioned, but did not study, a test based on both means and covariances, in which the homogeneity of available covariance matrices is simultaneously studied with homogeneity of means. He expected that this test might not perform well due to its typically large degrees of freedom. In quite a different context, Tang and Bentler (1998) studied covariance structures, such as factor analysis models, for incomplete data. When their test is specialized to that of an unstructured but common covariance matrix for all patterns of incomplete data, it provides a test of the MCAR assumption. In fact, their test can be shown to specialize to a test that can be constructed based on a chi-square difference rationale applied to Little's two proposed tests. A serious problem with these likelihood approaches is that they break down when the number of subjects for a given pattern of incomplete data is very small. Kim and Bentler (1999) developed a new approach, included uniquely in EQS, that uses a GLS rationale to develop tests of homogeneity of means and covariances that is much more stable in small samples. More details on missing data methods are given in Chapter 12.

Normal, Polychoric and Polyserial Correlations

The standard statistical theory is that for means and covariances arising from continuous variables. Although it is possible to analyze correlations as if they are covariances, without special handling this can lead to technical errors (Cudeck, 1989; Fouladi, 2000; Jöreskog, 1978; Lee, 1985a; Shapiro & Browne, 1990; Yuan & Bentler, 2000a). When the data vector *s* is not the symmetric part of a covariance matrix *S*, but rather represents the lower triangle of a correlation matrix *R*, ANALYSIS=CORRELATION tells EQS to use a correlation structure statistical methodology that uses a weight matrix appropriate to correlations to get correct results. Other methods also would be possible (Bentler & Lee, 1983), but remain to be implemented in EQS. You should recognize, though, that statistical correctness is not the only issue to be concerned with. Analysis of correlations implies throwing out information about the scale and thus variances of the variables. This may or may not be appropriate. For example, some phenomena, such as growth, imply changing variances (and, most likely, means as well), which get lost with correlation structures. But if variances can be ignored, correlations are fine.

If your data are normal, you can use normal theory correlation structure methods. These are based on Jennrich's (1970) statistical theory, also given by Shapiro and Browne (1990), and extended in EQS to the usual LS, GLS, and ML (really, RLS – true ML is not available) methods. Of course, if your data are not normal, you should use our extension of the Satorra-Bentler (1994) robust methodology and robust standard errors (Bentler & Dijkstra, 1985),

applied to correlation structures. That is, you use a command such as METHOD=ML, ROBUST and the test statistic and standard errors of the normal theory estimator are corrected. Furthermore, you get the Yuan-Bentler residual-based test statistics $T_{YB(RES)}$ and $T_{F(RES)}$ that were developed in the covariance structure context and have been extended here. It is logical to expect these tests to perform well in correlation structures also. EQS also provides an arbitrary distribution alternative. This uses a weight matrix *W* based on the distribution-free approach to correlation structures developed by de Leeuw (1983) and Steiger and Hakstian (1982).

Similarly, the vector *s* may represent special kinds of correlations, so-called polychoric and polyserial correlations. Polychoric correlations relate ordered categorical variables to each other, and polyserial correlations related ordered categorical and continuous variables. These correlations require special computations based on the assumption that a categorical variable is a categorized version of an underlying truly continuous, normally distributed, variable. EQS computes these correlations automatically by simply designating the categorical variables. These correlations are described, for example, by Olsson (1979a), Olsson, Drasgow, and Dorans (1982), and Poon and Lee (1987). See also Poon, Lee, Afifi, and Bentler (1990), Poon, Lee, and Bentler (1990), and Poon, Lee, Bentler and Afifi (1993). EQS follows the tradition of Muthén (1984; Muthén & Satorra, 1995b), Jöreskog (1994), and Lee, Poon and Bentler (1990b, 1992) of estimating a structural model subsequent to estimating these "poly" correlations. As above, the recommended method for modeling with these correlations is to use our new extension of the Satorra-Bentler (1994) and Bentler-Dijkstra (1985) robust methodology applied to this situation. You use some standard normal theory correlation estimator xx, and then the test statistic and standard errors of the estimator are corrected using the command METHOD=xx, ROBUST. See also Muthén (1993). We further provide EQS's new extension to these correlations of the Yuan and Bentler (1998a) and Bentler and Yuan (1999b) residual-based test statistics $T_{YB(RES)}$ and $T_{F(RES)}$. In addition, EQS provides an arbitrary distribution alternative, using the weight matrix *W* based on the theory of Lee, Poon, and Bentler (1995; see also 1989, 1990a,b, 1992). Issues related to the use of this methodology, as well as some alternatives (e.g., Jöreskog & Moustaki, 2001), are discussed in Chapter 5.

Multilevel Models

As stated above, the basic statistical theory underlying EQS models uses the assumption that observations (cases, subjects) are independent. This is a standard assumption of multivariate analysis, but it may be false in some important situations. In particular, when the data are obtained under a complex sampling design or a hierarchical sampling design, such as students nested within schools, it is possible to observe data at several levels (e.g., at the level of students, or at the level of schools) and both the model and statistical methodology should take the sampling structure and the data level into account. This is done with multilevel models having regression structures (e.g., Bryk & Raudenbush, 1992; Goldstein, 1995; Greenland, 2000b, 2002; Raudenbush & Bryk, 2002) and multilevel structural equation models (e.g., Bentler & Liang, 2002; Goldstein & McDonald, 1988; Hox, 1995; Hox & Maas, 2001; Kaplan & Elliott, 1997; Kreft & de Leeuw, 1998; Lee, 1990; Lee & Poon, 1998; McDonald, 1994; Muthén, 1989, 1994; Muthén & Satorra, 1995a). Recent overviews are given by de Leeuw and Kreft (in press), Reese and Duan (2002), Heck and Thomas (2000), Hox (2002), and Snijders and Bosker (1999). EQS handles these kinds of models in several ways. First, a hierarchical linear modeling (HLM) approach is available (Chou, Bentler & Penz, 2000). For SEM, Muthén's ML method (MUML) is provided due to its popularity. This is an approximation to maximum likelihood and equivalent to it when the data are balanced. While its "within" parameters are well behaved, the "between" parameters and model tests with MUML may have some bias (Hox & Maas, 2001). A practical, true ML approach developed by Bentler and Liang (2002, 2003; Liang & Bentler, 1999a, 2001) is also incorporated into EQS. See du Toit and du Toit (in press) for another approach. Other types of models that allow dependence by modeling the covariances among observations (e.g., Weng & Bentler, 1987, 1997) have not been implemented. While time series models also deal with dependent observations, these are generally not framed as SEM models; however, when this can be done, programs like EQS can handle them (du Toit & Browne, 2001; Hamagami & McArdle, 2001; Hershberger, 1998; Hershberger, Molenaar, & Corneal, 1996; Kaplan, Harik, & Hotchkiss, 2001; McArdle, 2001; Molenaar, 1985, 1994, 1999; Molenaar, de Gooijer, & Schmitz, 1992). Of course, the range of models having a hierarchical structure is quite wide, and far exceeds SEM type models (see e.g., Raudenbush & Bryk, 2002; Reise, 2000; Reise & Duan, 2001, and articles therein). They include random effects models, random coefficient models, mixed effects models, and covariance component models (e.g., Agresti, Booth, Hobert, & Caffo, 2000; du Toit & Cudeck, 2001; Goldstein, 1986; Hedecker & Gibbons, 1997; Hedecker & Rose, 2000; Laird & Ware, 1982; Longford, 1987, 1993ab; Singer, 1998; Muthén, 2000; Teachman, Duncan, Yeung, &

Levy, 2001). But the SEM approach, i.e., EQS, can handle certain variants of these models (e.g., Rovine & Molenaar, 2000, 2001).

Statistics in multilevel SEM models with continuous variables currently require a multivariate normality assumption for the hierarchical data or its error structure. Since appropriately taking non-normality into account is one of the key features of EQS, corrections for multilevel nonnormality also are part of EQS. The literature discusses how to do distribution-free analysis (Lee & Poon, 1994), but this is mainly applicable to data with a small number of variables and huge sample sizes. Yuan and Bentler (2002bc, 2003) have recently developed statistics that are robust to violation of the normality assumption. These use a scaling correction akin to the Satorra-Bentler (1994) correction, and more adequate or robust standard error estimates with a triple product "sandwich-type" matrix. As preliminary evaluations are encouraging, and since these have been found to work well in the standard situation of independent observations, one of their methods is now incorporated into EQS. Chapter 11 discusses multilevel models in more detail.

Case Weighting

There are circumstances when individual observations are weighted differentially, or should be weighted differentially due to the special nature of a sampling design (e.g., Kaplan & Ferguson, 1999; Johnson & Elliott, 1998; Winship & Radbill, 1994). When sampling units have different probabilities of being selected, it is appropriate to weight cases differentially in order to have the computed sample statistics more accurately estimate their population counterparts. Similarly, when individual observations have different variances, it may be useful to control for such heterogeneity by weighting cases according to their precision. These goals can be achieved by computing a weighted mean and weighted covariance matrix. In turn, these can be modeled using a variety of methods, including for the first time methods that allow nonnormal distributions. EQS only needs the weight variable to be designated in the data file.

A different reason for case weighting is to downweight observations that might inappropriately impact the sample means and covariances. Standard linear modeling methods such as ANOVA and SEM are liable to catastrophic breakdown in nonregular situations. As noted by Wilcox (1995, p. 57), "It should be stressed that outliers are not the only reason for considering robust methods. Small shifts in a distribution can have a large impact on the mean which might render it a potentially misleading measure of the typical individual." He also raises the provocative question of whether discoveries have been lost due to nonuse of robust methods (Wilcox, 1998), and provides an introduction to basic methods for "substantially improving power and accuracy" (Wilcox, 2001). It is known that the influence function associated with the sample covariance is quadratic, so that a few influential cases or outliers can lead to inappropriate solutions for virtually all standard statistical methods that rely on sample covariances (Hampel, Ronchetti, Rousseeuw, & Stahel, 1986; Poon & Poon, 2002; Yuan & Bentler, 2001a; Zimmerman & Williams, 2000). SEM methods, of course, rely heavily on means and covariances as their basic data to be modeled; alternatives to standard methods were considered necessary quite early (Huba & Harlow, 1987). Methods for handling data in nonstandard situations, such as methods that have bounded influence functions and can tolerate a high proportion of bad data before breaking down, have existed for a long time (e.g., Hoaglin, Mosteller, & Tukey, 1983). However, in the past, these methods have been presented primarily as exploratory and graphical methods, with little attention paid to standard problems of inference in the multivariate case (Huber, 1981). These issues were solved in a series of papers by Yuan and his colleagues (Yuan & Bentler, 1998b,c, 2000c; Yuan, Chan, & Bentler, 2000). In general, these methods seek to weight cases or observations differentially. By finding the appropriate weight to give to each case, the influence of outliers on a robust procedure is minimal. Since there are in principle a lot of potential weight functions (see e.g., Table 11-1 of Hoaglin, et al., 1983), there are also potentially a lot of different case weight vectors that could be used. Examples are Maronna's (1976) M-estimator weights, Huber (1977) type weights, multivariate-t weights (Lange, Little, & Taylor, 1989), and Hampel's (1974) bounded influence weights, as recommended by Campbell (1980). Based on good simulation performance, case-robust weighting is available in EQS by specifying CROBUST=#,#. Along with this are some new test statistics developed for EQS. This is discussed in Chapter 5.

EQS also provides simpler ways of dealing with outliers. In the graphical versions of EQS, outliers can be viewed in plots, highlighted, and deleted from the data file. In all versions, a procedure is included to locate outliers by marking those cases that contribute maximally to Mardia's (1970) normalized coefficient of multivariate kurtosis.

While there are many approaches to locating outliers or influential cases (see, e.g., Berkane & Bentler, 1988; Bollen, 1987b; Comrey, 1985; Devlin, Gnanadesikan, & Kettenring, 1981; Gallini & Casteel, 1987; Leese & Brown, 1997; Poon, Lew, & Poon, 2000; Poon & Poon, 2002; Rasmussen, 1988; Rensvold & Cheung, 1999), this method is very effective. It is also possible to specify that EQS should delete cases during computations if, for example, they are outliers.

Sample Size

The statistical theory described above is *asymptotic*, that is, it is assumed that N approaches infinity. Yet, "About 18% of the studies we reviewed used samples of fewer than 100 individuals" (McCallum & Austin, 2000, p. 215). N=100 is not asymptotic, but is it good enough? Empirical research regarding the relevance of asymptotic theory to practical data analysis, to yield recommendations about sample sizes that might be appropriate under various circumstances, has been going on for about two decades (e.g., Anderson & Gerbing, 1984; Balderjahn, 1985; Bandalos, 1997; Bearden, Sharma, & Teel, 1982; Boomsma, 1983; Breivik & Olsson, 2001; Finch, West, & MacKinnon, 1997; Fouladi, 1999; Gallini & Mandeville, 1984; Gerbing & Anderson, 1985; Geweke & Singleton, 1980; Hancock, 2001; Harlow, 1985; Hoyle, 1999; Hu, Bentler, & Kano, 1992; Jackson, 2001, 2003; LaDu & Tanaka, 1989; Marsh, Hau, Balla, & Grayson, 1998; Muthén & Kaplan, 1985, 1992; Tanaka, 1984, 1987). The quality of results that may occur from a given study with small samples will depend on the features of the model of interest -- parameter estimates, standard errors, *z*-tests, test statistics, mediational effects, with good results achieved in the approximate order listed. That is, estimates settle down at the smallest sample sizes, standard errors at still larger Ns, and so on. Unfortunately, it is hard to come up with good rules of thumb, because other features such as the communality level of the variables (MacCallum, Widaman, Zhang, & Hong, 1999; Velicer & Fava, 1998), the degree and kind of nonnormality that might exist (e.g., Boomsma & Hoogland, 2001; Finch, West, & MacKinnon, 1997; Hoogland, 1999; Yuan & Bentler, 1999b), the extent of missing data (Muthén & Muthén, 2002), and certainly the estimation method used (e.g., Hu, Bentler & Kano, 1992) affect any conclusions. Ideally, SEM methods would perform well across all the design features just mentioned, but this is not the case (e.g., Bentler & Yuan, 1999b). Luckily, in some situations sample size may not matter much: "if communalities are high, recovery of population factors in sample data is normally very good, almost regardless of sample size, level of overdetermination, or the presence of model error" (MacCallum, Widaman, Preacher, & Hong, 2001, p. 636). On the other hand, for testing models, the Browne (1982, 1984) and Chamberlain (1982) distribution-free method is a disaster in most modeling contexts at all but the largest sample sizes (e.g., Curran, West, & Finch, 1996; Yuan & Bentler, 1997d). A development that may help in small samples is that of two-stage least squares. While old as a method, new implementations are distribution-free and imply some robustness to misspecification that is not available in full information methods (Bollen, 1996, 2001; Bollen & Biesanz, 2002).

Virtually all programs compute standard errors based on the expected second derivatives under the model, the Fisher information or expected information matrix, or their analogues for various fit functions. EQS for the first time additionally provides standard errors for a wide variety of statistical methods based on the second partial derivatives, Hessian, or observed information of the fitting function. These standard errors are sometimes considered to be the gold standard for evaluating sampling variability of estimates, especially in small samples (e.g., Dolan & Molenaar, 1991). They are generally not provided due to complexity of computations. In EQS, when the observed information matrix is specified, it is substituted for the Fisher information matrix or its analogue in all relevant formulas, including those of the sandwich estimator that provides robust standard errors. This option should improve accuracy on parameter tests at small sample sizes.

So, what minimal N must you have? In small samples, many models may be accepted, i.e., not rejected statistically. Thus, as a practical matter, in small samples it is imperative to test multiple models, some of which may be logically appropriate even though they may not be your favorite models. If some of these models are rejected, your N is probably large enough. That is, N should be large enough so that power exists to reject alternative models (Hancock, Lawrence, & Nevitt, 2000; MacCallum, Browne, & Sugawara, 1996; Matsueda & Bielby, 1986; Satorra, 2003; Satorra & Saris, 1985). However, determining power is not straightforward. The estimation of power depends on having a test statistic that can be relied upon under the given circumstances, i.e., typically, that correctly can be referred to the hypothesized noncentral chi-squared distribution. Since various test statistics that are presumed to have central χ^2 distributions do not behave this way under realistic data gathering conditions (e.g.,

violation of normality in the case of normal maximum likelihood), it is likely that the noncentral distributions used to calculate power also may not suffice (see Curran, Bollen, Paxton, Kirby, & Chen, 2002; Mooijaart, 2003; Satorra, 2003; Yuan & Hayashi, in press). Clearly better small sample methods are needed.

Given the model characteristics, some methods are better able to cope with small samples than other methods. For example, Bentler and Yuan (1999b) studied the normal theory based likelihood ratio statistic T_{ML}, the Satorra-Bentler rescaled statistic $\bar{T}$, the Yuan and Bentler version T_{YB} of Browne's (1984) residual based ADF statistic, and a variant of their F-statistic (Yuan & Bentler, 1999a) derived from the residual-based ADF statistic. They found the F-statistic to perform best of all these mentioned at the smallest sample sizes. The F-test also performed very well in Yuan and Bentler's (1999a) and Nevitt's (2000) study. It is available in EQS. However, Fouladi (1999) recommended using a Bartlett (1950) correction, and applying it to the Satorra-Bentler (1994) scaled and adjusted statistics. She found that the Bartlett-corrected adjusted statistic performed best. In a related paper (2000) she found that Bartlett and Swain rescaling was best with small samples and very mild non-normality, but that S-B scaled and adjusted procedures were better with more severe nonnormality. Nevitt (2000) also found that a Bartlett-corrected statistic performed best in his study, which was more extensive than that of Fouladi. Nevitt's final conclusion was that the k-factor-corrected Satorra-Bentler scaled (not adjusted) test statistic is best used to evaluate model fit in small samples. Yuan and Marshall (2000) propose a new approach based on the bootstrap. This recent work has not yet been incorporated into EQS.

Interactions, Mixtures, Nonlinear, and Conditional Models

As noted above, EQS currently has no automated features for dealing with interactions, mixtures, and nonlinear relations among variables. This is a drawback since, for example, the idea of interactions is a standard one in linear models and ANOVA. Yet, easy SEM implementations continue to be out of reach. Although introduced many years ago (e.g., Kenny & Judd, 1984; Mooijaart & Bentler, 1986), only recently has a huge and technical literature on this topic developed (e.g., Schumacker & Marcoulides, 1998; Yang-Wallentin, 2001). In spite of useful simplifications (e.g., Jaccard & Wan, 1995; Ping, 1996, 1998; Wood, 2000), and important technical advances (Klein & Moosbrugger, 2000), this methodology is still not ready to use with a few simple commands. In most implementations, you need to specify constraints, and getting this right can be problematic (Li, Duncan, & Acock, 2000; Schumacker, 2002; Wen, Marsh, & Hau, 2002).

Even though nonlinear models with polynomial relations go back several decades (e.g., Etezadi-Amoli & McDonald, 1983; McDonald, 1967), today "the most challenging problems are generalizations of structural equation modeling that involve nonlinear functions of latent variables" (Browne, 2000, p. 663). According to Wall and Amemiya (2000), the challenge is basic: the Kenny-Judd approach and its extension by others produces inconsistent parameter estimates unless the latent variables are normally distributed. Although the Bollen (1995, 1996; Bollen & Paxton, 1998) approach is consistent, it seems to lead to biased estimates (Moulder & Algina, 2002) and, according to Wall and Amemiya, it "cannot be extended to general polynomial models" such as they have developed. Wall and Amemiya (2001) propose another approach as an extension of the Kenny and Judd (1984) methodology, and Yalcin and Amemiya (2001) develop a method for nonlinear factor analysis. These approaches seem to be general and practical enough to implement in EQS in the future.

Latent variable mixture models, in which unknown individuals in a sample actually belong to two or more latent groups with unknown probabilities, also are not part of EQS 6. Such models are becoming quite popular in the theoretical literature (e.g., Arminger, Stein, & Wittenberg, 1999; Dolan & van der Maas, 1998; Hoshino, 2001; Jedidi, Jagpal, & DeSarbo, 1997; Jedidi, Ramaswamy, DeSarbo, & Wedel, 1996; Muthén, 2001ab, 2002; Muthén & Shedden, 1999; Yung, 1997). While this methodology certainly is interesting, the ability to determine the number of latent groups with any precision in real data seems not yet to be established. Rindskopf (2001) gives a simple example where two group means are separated by four standard deviations (real groups are almost never separated this much), yet the existence of two groups is quite hard to discern. Bauer and Curran (2003, 2004) discuss advantages as well as the serious limitations of the categorical approach. Hopefully, the technical issues in correctly identifying the number of latent groups will soon be solved, at which point practical implementations certainly will be incorporated into EQS.

There is another recent methodological trend that is not part of EQS 6. This is a literature on nonlinear models that uses advanced statistical methods based on Bayesian theory using the Gibbs sampler and the Metropolis-Hastings algorithm (e.g., Ansari, Jedidi, & Dube, 2002; Arminger & Muthén, 1998; Jedidi & Ansari, 2001; Lee & Zhu, 2000; Scheines, Hoijtink, & Boomsma, 1999; Song & Lee, 2002b; Zhu & Lee, 1999). A typical approach is to estimate individual specific parameters, such as factor scores, while also taking into account the uncertainty in such estimates. So far these methods have been applied only to small problems, and require special programming. Since Bayesian theory is not generally known to applied researchers, a simple way to produce a practical implementation of these methods for the nonspecialist has not yet been achieved.

A final trend that is not incorporated into EQS 6 is that of graphical modeling (e.g., Cox & Wermuth, 1996; Edwards, 1995; Lauritzen, 1996; Shipley, 2000ab, 2002), which uses figures something like path diagrams to describe and model multivariate dependencies in nonexperimental data. McDonald (2002) gives details on how graphical models and path models relate in specialized models, and Koster (1999) discusses a Markov property of normal SEM models. One way to think about graphical models is that they are conditional independence models, where, for example, two variables are independent if conditioned on a third variable. Another way to think about them is that they are models for the inverse covariance matrix rather than the covariance matrix. Actually, they are even more general, in fact, an advantage of this class of models over SEM is that they do not necessarily need to assume linearity. However, in spite of continuous technical and computational development in statistics (e.g., Geiger & Heckerman, 2002; Lauritzen, 2001; Levitz, Perlman, & Madigan, 2001; Richardson & Spirtes, 2002) and computer science (Pearl, 2000), at the present time the methodology does not easily permit the use of a priori latent variable structures as are typical in SEM. Thus, though the conditional independence ideas lead to interesting new model test statistics (Shipley, 2000b), so far they are not applicable to general latent variable models. But, like SEM, this methodology and its relatives (Gill & Robins, 2001) provides another approach to causal inference in observational research that you should know also exists.

This concludes the overview of methods in EQS. How to use the basic methods is shown in Chapter 2. The use of various program options meaningfully requires a discussion of program input in Chapter 3. The resulting program output is discussed in Chapter 4. The various statistical methods are further discussed in Chapter 5. Statistical methods for modifying models are discussed in Chapter 6. Multisample covariance structures, useful in evaluating the equivalence of models for several groups, are set up in EQS in the same simple way as models for a single group, though cross-group constraints typically are also specified. See Chapter 7. Models with intercepts, called mean and covariance structure models, are a bit more complicated. As will be seen in Chapter 8, specification of mean structures involves the use of a constant (called V999 in EQS) and is thus somewhat more involved than the basic procedures reviewed in Chapter 2. Multiple group models with mean structures are discussed in Chapter 9. The specialized but popular growth curve models are discussed in Chapter 10. As noted above, Chapter 11 deals with multilevel models, while Chapter 12 deals with missing data. Resampling methods, useful to get better standard errors or to study the empirical performance of test statistics, are discussed in Chapter 13. Reliability coefficients and practical fit indices are reviewed in Chapter 14. Chapter 15 concludes with references.

Model Specification with EQS

EQS makes use of a simple and straightforward specification language to describe the model to be analyzed, and provides extensive syntax error checking in order to make the program as easy and error-free to use as possible. Several major sections of input information are used in EQS, most of which are not required in simple problems. A brief description of each section is given below. More detailed information follows in Chapter 3.

Each section is begun by a slash and a keyword, capitalized below. Those sections that are not always needed are noted as optional.

/TITLE (Optional)

EQS permits you to supply as much title information as desired to help identify the program output. All title information is reprinted once in the output. Subsequent output pages print only the first line.

/SPECIFICATIONS

This section describes the problem to be analyzed, including the number of cases (observations), the number of variables and type of input matrix, the method(s) of estimation desired, type of matrix to be analyzed, input data file name, cases to be deleted, etc.

/LABELS (Optional)

Labels may be provided for measured latent variables and latent factors.

/EQUATIONS (Optional)

This section is used to describe the equations involved in the model under consideration, and in typical models is mandatory. There is one equation for each dependent variable in the system. A dependent variable is one that is a structured regression function of other variables; it is recognized in a path diagram by having one or more arrows aiming at it. The /EQUATIONS section also provides information for the automatic selection of variables from the input matrix. This section can be ignored when /MODEL and /RELIABILITY are used.

/VARIANCES (Optional)

Every independent variable in the model must have a variance, which is specified in this section. The fixed constant V999 used in structured means models does not have a variance. Variances can be ignored only with /MODEL and /RELIABILITY.

/COVARIANCES (Optional)

If independent variables are correlated, their covariances are specified in this section. In addition to the usual covariances, such as those between the residuals of observed variables or between independent latent variables, the covariances between any pair of independent variables may be estimated, subject only to the requirement that the model be identified.

/MODEL (Optional)

This section can be used instead of the three sections just given. In many cases, the /MODEL section is easier to use, in part because E and D variables are generated automatically.

/RELIABILITY (Optional)

This section can be used instead of the four sections above in the special case when there is one factor (which must be independent), and all measured Vs are dependent.

/CONSTRAINTS (Optional)

Parameters that are constrained to be equal are indicated in this section. In addition, any parameter may be related to other parameters by a particular linear constraint. Cross-groups constraints in multisample analysis are specified here.

/INEQUALITIES (Optional)

An upper and/or lower bound for any free parameter to be estimated can be specified in this section.

/MATRIX (Optional)

The /MATRIX section signals the beginning of the input covariance or correlation matrix. When raw data is used, it will be read from another file.

/STANDARD DEVIATIONS (Optional)

If the input provided in /MATRIX is a correlation matrix, providing standard deviations will cue the program to transform the correlation matrix into a covariance matrix prior to estimation and testing.

/MEANS (Optional)

Models with intercepts structure the means of latent and observed variables. This section provides the sample means of the observed variables. If raw data is analyzed, this section is not needed, as the means can be computed.

/KURTOSIS (Optional)

If raw data is not analyzed, this command allows for the provision of marginal kurtoses of variables, which are used in the new Heterogeneous Kurtosis (HK) statistical methodology. With raw data, this section is not needed.

/TECHNICAL (Optional)

This section is used to override the program's default values for various technical constants, e.g., the convergence criterion and number of iterations. It also enables the use of arbitrary start values to override the program's defaults.

/PRINT (Optional)

The results of an effect decomposition, the model-reproduced covariance and correlation matrices, including model-based estimates of means in structured means models, and the correlations among parameter estimates are not printed unless you specify, in this section, that this output is desired.

/LMTEST (Optional)

Inadequacies in models can be evaluated by the /LMTEST. It can suggest fixed parameters to free and constraints to release.

/WTEST (Optional)

Unnecessary free parameters can be evaluated by the /WTEST. It can suggest free parameters that can be fixed at zero without much degradation in fit.

/SIMULATION (Optional)

The empirical behavior of statistics produced by EQS can be determined by repeated sampling from a population or by bootstrapping or jackknifing. The design of such a study is specified in this section.

/OUTPUT (Optional)

Technical output from EQS can be saved in a compact form and with greater precision, as specified in this section. This would usually be done in a simulation study.

/SAVE (Optional)

EQS can save a covariance matrix, or raw data (with or without factor scores), on a data file which can be read by EQS or by other programs.

/DEFINE (Optional)

In a hierarchical linear model, this section permits parameters of the first level to be specified for analysis at the second level.

/END

This keyword marks the termination of the program input.

The output from EQS includes an echo of the input, giving you a record of the job submitted. If the input contains information that the program finds problematic, error messages will be printed. If the errors appear to be trivial, the program will continue into the computational section, but if the errors are major the program will not proceed. You will then have to correct such errors and resubmit the job.

This introduction to EQS will have made it apparent that EQS is a powerful program that performs many types of SEM analyses. It will become clear in the next chapter that EQS is also extremely easy to use. Although there is great ease and flexibility in the program, EQS can do no more than provide for an appropriate analysis of a set of data with the best available statistics. It is up to the investigator to be sure that the questions being asked of the analysis are appropriate to structural modeling, and that the data are of sufficient quality to potentially answer the theoretical questions under analysis. Appropriate use of the statistical techniques made available in EQS does not necessarily imply that correct inferences will be drawn from the results of an analysis. Such inferences typically require additional assumptions about the context of a study and its data. We believe that with care, reasonable scientific inferences can be drawn from a SEM analysis. Some writers, in contrast, are not convinced: "When we come to models for relationships between latent variables we have reached a point where so much has to be assumed that one might justly conclude that the limits of scientific usefulness have been reached if not exceeded" (Bartholomew & Knott, 1999, p. 190). This warning is particularly apt when "causal" interpretations of the results are desired. The meaning and role of the philosophical concept of "causality" as applied to structural modeling is not discussed in this *Manual*, although it is recognized that analyses are often undertaken in order to evaluate a causal hypothesis regarding the influences that may exist among the variables in a system. To repeat, however admirable it may be to model a causal process, such a motivation does not provide adequate justification for uncritical use of causal language in describing some particular result or application of a linear statistical model.

Illustrative Applications

In the last few years, the use of covariance structure analysis and structural equation has exploded (see e.g., Hershberger, 2003). One of the best ways to learn about this methodology is to study practical applications in your field. Some recent references are the following: *ability growth and decline* (Raykov, 2000), *absenteeism* (Geurts, Schaufeli, & Rutte, 1999), *academic self concept and achievement* (Marsh, Byrne, & Yeung, 1999; Marsh, Hau, & Kong, 2002), *acculturation* (Gil, Wagner, & Vega, 2000), *achievement goals* (Elliot & McGregor, 2001), *achievement inequality* (Carbonaro & Gamoran, 2002), *achievement motivation* (Barron & Harackiewicz, 2001), *achievement tests* (Muraishi & Toyoda, 1998), *action and coping self-efficacy* (Schwarzer, & Renner, 2000), *action-control beliefs* (Malmberg, Wanner, Sumra, & Little, 2001), *ADHD* (Rapport, Scanlan, & Denney, 1999), *adolescent dating* (Wolfe, Scott, Reitzel-Jaffe, Wekerle, Grasley, & Straatman, 2001), *affect* (Remington, Fabrigar, & Visser, 2000), *affect and alcohol problems* (McCreary & Sadava, 2000), *affirmative action* (Iyer, Leach, & Crosby, 2003; Federico & Sidanius, 2002), *aging and health* (Pedersen, Steffensson, Berg, Johansson, & McClearn,1999), *aging and memory* (Della-Maggiore, Sekuler, Grady, Bennett, Sekuler, & McIntosh, 2000), *aging and perception* (Salthouse, Toth, Daniels, Parks, Pak, Wolbrette, & Hocking, 2000), *AIDS and HIV counseling* (Nyamathi & Stein, 1997), *AIDS risk behaviors* (Schröder, Hobfoll, Jackson, & Lavin, 2001; Stein, Nyamathi, & Bentler, 2001), *alcohol expectancies* (Wood, Read, Palfai, & Stevenson, 2001), *alcohol prevention* (Taylor, Graham, Cumsille, & Hansen, 2000), *alcohol use* (Barnes, Murray, Patton, Bentler, & Anderson, 2000; Harford, & Muthén, 2001; Lonczak, Huang, Catalano, Hawkins, Hill, Abbott, Ryan, & Kosterman, 2001; Migneault, Velicer, Prochaska, & Stevenson, 1999), *alcohol warning labels* (MacKinnon, Nohre, Cheong, Stacy, & Pentz, 2001), *alexithymia* (Loas, Corcos, Stephan, Pellet, Bizouard, Venisse, Perez-Diaz, Guelfi, Flament, & Jeammet, 2001), *allergies* (Jacobsen, Herskind, Nielsen, & Husby, 2001), *altruism* (Maner, Luce, Neuberg, Cialdini, Brown, & Sagarin, 2002), *antisocial and prosocial activities* (Duncan, Duncan, Strycker, & Chaumeton, 2002), *anxiety symptoms* (White & Farrell, 2001), *arterial disease* (Carmelli, Fabsitz, Swan, Reed, Miller, & Wolf, 2000), *atherosclerosis* (Whiteman, Deary, & Fowkes,

2000), *attachment* (Oxford, Harachi, Catalano, Haggerty, & Abbott, 2000), *attention* (Strauss, Thompson, Adams, Redline, & Burant, 2000), *battered women* (Campbell & Weber, 2000), *behavior problem help seeking* (Lau, & Takeuchi, 2001), *beliefs about depression interventions* (Jorm, Christensen, Medway, Korten, Jacomb, & Rogers, 2000), *big 5 personality factors* (Barbaranelli & Caprara, 2000), *binge drinking* (Tucker, Orlando, & Ellickson, 2003), *binge eating* (Pratt, Telch, Labouvie, Wilson, & Agras, 2001), *bioaging* (Anstey & Smith, 1999), *birth outcomes* (Rini, Dunkel-Schetter, Wadhwa, & Sandman, 1999; Sheehan, 1998; Shenkin, Starr, Pattie, Rush, Whalley, & Deary, 2001), *blood pressure* (Batista-Foguet, Coenders, & Ferragud, 2001; Llabre, Spitzer, Saab, & Schneiderman, 2001), *body image* (Marsh, 1999), *body image and mood* (Holsen, Kraft, & Røysamb, 2001), *borderline personality disorder* (Maffei, & Fossati, 1999), *botany* (Escudero, Iriondo, Olano, Rubio, & Somolinos, 2000), *brain disease* (Barbarotto, Capitani, & Laiacona, 2001), *brain imaging* (Taylor, Krause, Shah, Horwitz, & Mueller-Gaertner, 2000), *brain structure* (Posthuma, de Geus, Neale, Pol, Baare, Kahn, & Boomsma, 2000), *brand personality* (Aaker, Benet-Martínez, & Garolera, 2001), *breast cancer* (Schnoll, Harlow, Stolbach, & Brandt, 1998), *breast feeding* (Pérez-Escamilla, Cobas, Balcazar, & Holland Benin, 1999), *bronchial asthma* (Woeller, Kruse, Schmitz, & Richter, 1998), *bullying* (Roland & Idsoe, 2001), *burnout* (Okabayashi, Sugisawa, Takanashi, Nakatani, & Shibata, 1999), *business ethics* (Patterson, 2001), *buying behavior* (Granzin, & Olsen, 1998), *cancer and depression* (Ybema, Kuijer, Buunk, DeJong, & Sanderman, 2001), *cancer and quality of life* (Nuamah, Cooley, Fawcett, & McCorkle, 1999), *cardiovascular reactivity* (Gump, Matthews, & Raeikkoenen, 1999), *career success* (Seibert, Kraimer, & Liden, 2001), *caregiving* (Miltiades & Pruchno, 2002), *cellular phone loyalty* (Gerpott, Rams, & Schindler, 2001), *cerebral palsy* (Bartlett & Palisano, 2000), *CES-D* (Grayson, Mackinnon, Jorm, Creasey, & Broe, 2000), *child attachment* (Doyle, Markiewicz, Brendgen, Lieberman, & Voss, 2000), *child behavior problems* (Linares, Heeren, Bronfman, Zuckerman, Augustyn, & Tronick, 2001), *child care effects* (NICHD Early Child Care Research Network, 2002), *child emotion regulation* (Belsky, Friedman, Hsieh, 2001), *child health* (Wills & Cleary, 2000), *child leukemia outcomes* (Chen, Zelter, Bentler, Byrne et al., 1998), *child molestation* (Hunter & Figueredo, 2000), *child neglect* (Kiriski, Dunn, Mezzich, & Tarter, 2001), *child sexual abuse* (Ferrara, 1999; Shapiro & Levendosky, 1999), *child vocabulary competence* (Bornstein & Haynes, 1998; Bornstein, Haynes, & Painter, 1998), *chronic pain* (Riley, Wade, Robinson, & Price, 2000), *cigarette smoking* (Windle & Windle, 2001), *circumplex structure* (Fabrigar, Visser, & Browne, 1997), *clinical trials* (Frosch, Stein, & Shoptaw, 2002; Siddiqui & Ali, 1999), *close relationships* (Campbell, Simpson, Kashy, & Fletcher, 2001), cocaine use (Stacy, Newcomb & Bentler, 1995), *cocaine and fetal growth* (Bandstra, Morrow, Anthony, Churchill, Chitwood, Steele, Ofir, & Xue, 2001), *cognitive abilities* (Carlstedt, 2001; Maitland, Intrieri, Schaie, & Willis, 2000), *cognitive neuroscience* (Petersson, Reis, Askeloef, Castro-Caldas, & Ingvar, 2000), *cognitive screening* (Anderson, Burton, Parker, & Godding, 2001), *combat and sexual harassment* (Fontana, Litz, & Rosenheck, 2000), *communications* (Scheufele & Shah, 2000), *commuter travel* (Kuppam & Pendyala, 2001), *competence skills* (Epstein, Griffin, & Botvin, 2000a; Griffin, Botvin, Scheier, Epstein, & Doyle, 2002), *compliance in randomized trials* (Dunn,1999), *conduct problems* (Goldstein, Prescott, & Kendler, 2001), *conflict resolution in couples* (Zuroff & Duncan, 1999), *conservation behavior* (Corral-Verdugo & Figueredo, 1999), *consumer behavior* (Chan, 2001), *context-specific self-efficacy* (Gwaltney, Shiffman, Norman et al., 2001), *coping skills intervention* (Rotheram-Borus, Stein, & Lin, 2001), *corpus callosum* (Hines, Chiu, McAdams, Bentler, & Lipcamon, 1992), *counseling psychology* (Quintana & Maxwell, 1999), *couples* (Wendorf, 2002), *creative and wise people* (Helson & Srivastava, 2002), *cross-cultural measurement equivalence* (Byrne & Campbell, 1999; Cheung & Rensvold, 1999; Little, 1997; Steenkamp, & Baumgartner, 1998), *cross-cultural research* (Cheung & Rensvold, 2000; Little, 2000), *cross-generation child behavior* (Cohen, Kasen, Brook, & Hartmark, 1998), *cross-nation health* (Haugland, Wold, Stevenson, Aaroe, & Woynarowska, 2001), *culture* (van de Vijver & Leung, 2000), *cytology* (Kulkarni, Kamal, & Arjune, 1999), *deafness* (Bavelier, Tomann, Hutton, Mitchell, Corina, Liu, & Neville, 2000), *decision making* (Singhapakdi, Vitell, & Franke, 1999), *delinquency and genetics* (Taylor, McGue, Iacono, & Lykken, 2000), *dentistry* (Fenlon, Sherriff, & Walter, 2000; Hakeberg, Hagglin, Berggren, & Carlsson, 2001), *depersonalization* (Lea, Spears, & de Groot, 2001), *depression and anxiety* (Christensen, Jorm, Mackinnon, Korten, Jacomb, Henderson, & Rodgers, 1999), *depression and psychosis* (Mueller, Wetzel, Szegedi, & Benkert, 1999), *deviance* (Chen & Kaplan, 1997; Ullman & Newcomb, 1999), *diabetes care* (Watkins, Connell, Fitzgerald, Klem, Hickey, & Ingersoll-Dayton , 2000), *differential item functioning* (Chan, 2000), *disability* (Tait & Purdie, 2000), *drinking and driving* (Grube & Voas, 1996; Labouvie & Pinsky, 2001), *drug problems and work adjustment* (Galaif, Newcomb, & Carmona, 2001), *drug treatment outcomes* (Aiken, Stein, & Bentler, 1994; Grella, Hser, Joshi, & Anglin, 1999), *drug use consequences* (Newcomb & Bentler, 1988), *drug use progression* (Duncan, Duncan, & Hops, 1998), *drug use vulnerability* (Brook, Whiteman, Finch, & Cohen, 2000; Felix-Ortiz & Newcomb, 1999), *drug visibility* (Saxe, Kadushin, Beveridge, Livert, Tighe, Rindskopf, Ford, & Brodsky, 2001), *dyslexia* (Schulte-Koerne, Deimel, Bartling & Remschmidt, 1999), *eating disorders* (Thompson,

Coovert, & Stormer, 1999), *ecological predation* (Peckarsky, Taylor, McIntosh, McPeek, & Lytle, 2001), *EEG asymmetry* (Hagemann, Naumann, Thayer, & Bartussek, 2002), *egocentrism* (Murray, Holmes, Bellavia, Griffin, & Dolderman, 2002), *electronic commerce* (Ramamurthy, Premkumar, & Crum, 1999), *empathy and action* (Batson, Chang, Orr, & Rowland, 2002), *employee turnover* (Mak & Sockel, 2001; Sager, Griffeth & Hom, 1998), *enrollment in intervention trials* (Spoth, Redmond, & Shin, 2000), *environmental monitoring* (Malaeb, Summers, & Pugesek, 2000), *epilepsy* (Mitchell, Scheier, & Baker, 2000), *European attitudes* (Pepermans & Verleye, 1998), *evaluation research* (Ullman, Stein, & Dukes, 2000), *evolutionary ecology* (Gomez & Zamora, 2000), *evolutionary psychology* (Budaev, 1999), *exercise* (Blue, Wilbur, & Marston-Scott, 2001), *expectancies* (Weary & Reich, 2001), *expectancy confirmation in marriage* (NcNulty & Karney, 2002), *exposure to drinking* (Cumsille, Sayer, & Graham, 2000), *expressed emotion* (Chambless, Bryan, Aiken, Steketee, & Hooley, 1999), *family conflict* (Duncan, Strycker, Duncan, & Okut, 2002), *family research* (Volk & Flori, 1996), *fetal growth* (Feldman, Dunkel-Schetter, Sandman, & Wadhwa, 2000), *fetal lead exposure* (Chuang, Schwartz, Gonzales-Cossio, Lugo, Palazuelos, Aro, Hu, & Hernandez-Avila, 2001), *fMRI* (Bullmore, Horwitz, Honey, Brammer, Williams, & Sharma, 2000), *fruit and vegetable consumption* (Kratt, Reynolds, & Shewchuk, 2000), *functional impairment* (Leon, Solomon, Mueller, Turvey, Endicott, & Keller, 1999), *gambling* (Turner, Ialomiteanu, & Room, 1999), *gender differences in pain* (Riley, Robinson, Wade, Myers, & Price, 2001), *gender discrimination* (Foster, 2000), *genetics* (Abrahamson, Baker, & Caspi, 2002; Borkenau, Riemann, Angleitner & Spinath, 2001; Kendler, Neale, Sullivan, Corey, Gardner & Prescott, 1999), *goal shielding* (Shah, Friedman, & Kruglanski, 2002), *gratitude* (McCullough, Emmons, & Tsang, 2002), *group differences* (Millsap, 1998), *group membership and similarity* (Chen & Kenrick, 2002), *Gulf war syndrome* (Haley, Luk, & Petty, 2001), *health care attitudes and behavior* (Auerbach & Pegg, 2002; Levin, 1999), *health promoting behavior* (Kagee & Dixon, 2000), *hearing loss* (Shriberg, Friel-Patti, Flipsen & Brown, 2000), *heritability of attitudes* (Olson, Vernon, Jang, & Harris, 2001), *highway congestion* (Golob & Regan, 2001), *HIV risk* (Burkholder & Harlow, 1996), *home environment and child cognition* (Whiteside-Mansell & Bradley, 1996), *homelessness* (Galaif, Nyamathi, & Stein,1999), *human factors* (Ellis & Allaire, 1999), *hyperactive and antisocial behavior* (Patterson, DeGarmo, & Knutson, (2000), *hypertension* (McCaffery, Pogue-Geile, Debski, & Manuck, 1999), *identity-management* (Kessler & Mummendey, 2002), *ideology and prejudice* (Duckitt, Wagner, du Plessis, and Birum, 2002), *immigrants* (Santos, Bohon, Sanchez-Sosa, 1998), *income satisfaction* (Solberg, Diener, Wirtz, Lucas, & Oishi, 2002), *industrial medicine* (Punnett, & van der Beek, 2000), *infant strange situation behavior* (Kroonenberg, van Dam, van Ijzendoorn, & Mooijaart, 1997), *infant temperament* (Lemery, Goldsmith, Klinnert, & Mrazek, 1999), *information processing* (Trumbo, 1999), *information resources* (Kearns & Lederer, 2000), *initiation of smoking* (Stein, Newcomb, & Bentler, 1996), *intercultural contact* (Nauck, 2001), *internalizing symptoms* (Crawford, Cohen, Midlarsky, & Brook, 2001), *intervention modeling* (Jo & Muthén, 2001), *intimate relationships* (Frank & Brandstätter, 2002), *job applicant reactions* (Schmitt, 2002), *job control* (Smith, Tisak, Hahn, & Schmieder, 1997), *job resources and mental health* (Taris, 1999), *job satisfaction* (Law & Wong, 1999), *jobs and health* (de Jonge, Mulder, & Nijhuis, 1999), *joint ventures* (Calantone & Zhao, 2001), *judgment based SEM* (Hyman & Leibowitz, 2001), *justice theory* (Heuer, Penrod, Hafer, & Cohn, 2002; Skitka & Mullen, 2002), *language maintenance* (Yeung, Marsh, & Suliman, 2000), *language testing* (Bachman, 2000), *Latino school aspirations* (Goldenberg, Gallimore, Reese, & Garnier, 2001), *lead and pregnancy* (Rothenberg, Khan, Manalo, Jiang, Cuellar, Reyes, Acosta, Jauregui, Diaz, Sanchez, Todd, & Johnson, 2000), *learning* (Tremblay, Gardner, & Heipel 2000), *learning styles* (Loo, 1999), *leukemia treatment* (Chen, Zeltzer, Bentler, Byrne, Nicholson, et al., 1998), *life management* (Freund & Baltes, 2002), *life satisfaction* (Schimmack, Radhakrishnan, Oishi, Dzokoto, & Ahadi, 2002), *Likert scales* (Weng & Chen, 2000), *literature review bias* (Bushman & Wells, 2001), *loneliness* (Larose, Guay, & Boivin, 2002), *longitudinal models* (Scott & Handcock, 2001; Plewis, 2001), *love* (Murray, Holmes, Griffin, Bellavia, & Rose, 2001), *malnutrition* (Pollitt, Durnin, Husaini, & Jahari, 2000; Pollitt, Jahari, & Walka, 2000), *manufacturing* (Cua, Junttila, & Schroeder, 2002; Fawcett & Myers, 2001; Koufteros, Vonderembse, & Doll, 2002; Papke-Shields & Malhotra, 2001; Tu, Vonderembse, & Ragu-Nathan, 2001), *marijuana use* (Brook, Whiteman, Finch, Morojele, & Cohen, 2000), *marital distress* (Grote & Clark, 2001), *marital satisfaction* (Cobb, Davila, & Bradbury, 2001), *marketing* (Kempf & Smith, 1998), *maternal-child health care* (Ahmed & Mosley, 2002), *media influence* (Thompson, & Heinberg, 1999), *medical school selection* (Shen, Bentler, & Comrey, 1995), *memory* (Bowden, Dodds, Whelan, Long, Dudgeon, Ritter, & Clifford, 1997; Christensen, Wood, & Barrett, 2003), *memory and brain traffic* (Taylor, Horwitz, Shah, Fellenz, Mueller-Gaertner, & Krause, 2000), *mental and physical health* (Shifren, Park, Bennett, & Morrell, 1999), *meta-analysis* (Verhaeghen & Salthouse, 1997; Viswesvaran & Ones, 1995), *methadone maintenance* (Avants, Margolin, & McKee, 2000), *military downsizing* (Evans, 1999), *MIS* (Swanson & Dans, 2000), *mixed-sex interactions* (Sadler & Woody, 2003), *molecular psychiatry* (Rowe, den Oord, Stever, Giedinghagen, Gard, Cleveland, Gilson, Terris, Mohr, Sherman, Abramowitz, & Waldman, 1999), *money and well-being* (Srivastava, Locke, & Bartol, 2001), *moral identity* (Aquino & Reed, 2002), *mother-child interaction*

(Whiteside-Mansell, Bradley, Little, Corwyn, & Spiker, 2001), *mother's caregiving* (Baydar, Greek, & Gritz, 1999), *multinational employees* (Ryan, Chan, Ployhart, & Slade, 1999), *multiple sclerosis* (Bentler & Houck, 1996), *multitrait-multimethod models* (Coenders & Saris, 2000; Corten, Saris, Coenders, van der Veld, Aalberts, & Kornelis, 2002; Dumenci, 2000; Eid, 2000; Lance, Noble, & Scullen, 2002; Scherpenzeel & Saris, 1997; Tomas, Hontangas, & Oliver, 2000; Wothke, 1996), *multitrait-multioccasion models* (Hernández & González-Romá, 2002), *muscle velocity* (Surakka, Alanen, Aunola, & Karppi, 2001), *music perception* (Magill & Pressing, 1997), *nausea in pregnancy* (Zhou, O'Brien, & Soeken, 2001), *neighborhood and aggression* (Colder, Mott, Levy, & Flay, 2000), *neighborhood and drugs* (Duncan, Duncan, & Strycker, 2002), *network analysis* (van Duijn, van Busschbach, & Snijders, 1999), *neural modeling* (Horwitz, Friston, & Taylor, 2000), *neural interactions and memory* (McIntosh, 1999), *neuropsychology* (Burgess, Veitch, de Lacy Costello, & Shallice, 2000), *nursing research* (Trinkoff, Zhou, Storr, & Soeken, 2000), *nutrition* (Birch & Fisher, 2000), *nutrition intervention* (Reynolds, Yaroch, Franklin, & Maloy, 2002), *occupational health* (Van Yperen & Snijders, 2000), *occupational therapy* (Mulligan, 1998), *organizational citizenship* (Cardona, Lawrence, & Bentler, 2004), *organizational research* (Larwood, Wright, Desrochers, & Dahir 1998), *pain* (Conant, 1998; Davis, Reeves, Hastie, Graff-Radford, & Naliboff, 2000; Hasegawa, Mishima, Matsumoto, Sasaki, Kimura, Baba, Senami, Kanemura, Takano, Shibata, 2001), *panic attacks* (Stewart, Taylor, Jang, Cox, Watt, Fedoroff, & Borger, 2001), *parent-child affect* (Spoth, Redmond, Shin, & Huck, 1999), *parenting* (Gutman & Eccles, 1999; Metsapelto, Pulkkinen, & Poikkeus, 2001; Oyserman, Bybee, Mowbray, & MacFarlane, 2002; Sieving, Maruyama, Williams, & Perry, 2000), *parenting intervention* (Martinez & Forgatch, 2001; Reid, Wesbster-Stratton, & Beauchaine, 2001), *parenting stress* (Ostberg & Hagekull, 2000), *patient needs and quality of life* (Slade, Leese, Taylor, & Thornicroft, 1999), *peer dynamics* (Patterson, Dishion, & Yoerger, 2000), *perceived locus of causality* (Chatzisarantis, Hagger, Biddle, & Karageorghis, 2002), *perception* (Olzak, & Wickens, 1999), *perfectionism* (Cox, Enns, & Clara, 2002; Dunkley, Zuroff, & Blankstein, 2003), *performance perceptions* (Ehrlinger & Dunning, 2003), *perinatal stressors* (DeMier, Hynan, Hatfield, Varner, Harris, & Manniello, 2000), *periodontology* (Kawamura, Tsurumoto, Fukuda, & Sasahara, 2001), *peritraumatic dissociation* (Marshall, Orlando, Jaycox, Foy, & Belzberg, 2002), *person fit* (Reise & Widaman, 1999), *personality, affectivity and self esteem* (Watson, Suls, & Haig, 2002), *personality assessment* (Crowley & Fan, 1997), *personality disorders* (Cruz, Joiner, Johnson, Heisler, Spitzer, & Pettit, 2000; Crawford, Cohen, & Brook, 2001ab), *personality, mood, and stress* (Hemenover, 2001), *personality and psychopathology* (O'Connor, 2002), *personnel decisions* (Struthers, Weiner, & Allred, 1998), *persuasive communication* (Albarracín & Wyer, 2001), *PET scans* (Rajah, McIntosh, & Grady, 1999; Nezafat, Shadmehr, & Holcomb, 2001), *physical activity* (Causgrove, 2000; Dishman, Motl, Saunders, Dowda, Felton, Ward, & Pate, 2002; Motl, Dishman, Saunders, Dowda, Felton, Ward, & Pate, 2002), *physical fitness* (Maia, Lefevre, Claessens, Renson, Vanreusel, & Beunen, 2001), *physical stature* (Beunen, Thomis, Maes, Loos, Malina, Claessens, & Vlietinck, 2000), *physician militancy* (Burke, 2001), *planned behavior* (Sideridis, Kaissidis, & Padeliadu, 1998), *political behavior* (Barker & Carman, 2000), *posthumous reputation* (Simonton, 1991), *poverty and child development* (Guo, & Harris, 2000), *pregnancy among teens* (Goodyear, Newcomb, & Allison, 2000), *prejudice* (Agnew, Thompson, & Gaines, 2000), *prenatal cocaine effects* (Accornero, Morrow, Bandstra, Johnson, & Anthony, 2002; Bandstra, Morrow, Anthony, Accornero, & Fried, 2001; Bendersky & Lewis, 1999), *prevention curriculum effects* (Park, Kosterman, Hawkins, Haggerty, Duncan, Duncan, & Spoth, 2000), *preventive intervention* (Spoth & Redmond, 2002; Scheier, Botvin, & Griffin, 2001), *problem behavior* (Ary, Duncan, Biglan, Metzler, Noell, & Smolkowski, 1999; Cheong & Raudenbush, 2000; Turbin, Jessor, & Costa, 2000), *product development* (Koufteros, Vonderembse, & Doll, 2001), *psychological distress* (Manne & Schnoll, 2001), *psychological disturbances and memory* (Adams, Stanczak, Leutzinger, Waters, & Brown, 2001), *psychopathy* (Cooke & Michie, 2001; Windle & Dumenci, 1999; Kosson, Cyterski, Steuerwald, Neumann, & Walker-Matthews, 2002), *psychosis* (Dudgeon, Mackinnon, Bell, & McGorry, 2001), *psychosomatics* (Berntsson, Kohler, & Gustafsson, 2001), *psychotherapy* (Burns, & Spangler, 2000), *PTSD* (Erickson, Wolfe, King, King, & Sharkansky, 2001; Orlando & Marshall, 2002; Vuksic-Mihaljevic, Mandic, Bensic, & Mihaljevic, 2000), *quality of care* (Smith, Brown, Borowsky, Davis, Center, & Lurie, 2001), *quality of life* (Power, Harper, & Bullinger, 1999; Sullivan, Kempen, Van Sonderen, & Ormel, 2000), *quantitative trait loci* (van den Oord, 2000), *racial attitudes* (Stephan, Boniecki, Ybarra, Bettencourt, Ervin, Jackson, McNatt, & Renfro, 2002), *racial segregation* (Postmes & Branscombe, 2002), *radiation therapy distress* (Mose, Budischewski, Rahn, Zander-Heinz, Bormeth, & Bottcher, 2001), *rape recovery* (Boeschen, Koss, Figueredo, & Coan, 2001), *reading disability* (Gayán & Olson, 1999), *reasoned action* (O'Callaghan, Callan, & Baglioni, 1999), *reasoning* (Holyoak & Simon, 1999), *recidivism* (Krauss, Sales, Becker, & Figueredo, 2000), *rehabilitation* (Peek, 2000), *religious attitudes* (Macdonald, 1998), *resistance to persuasion* (Sagarin, Cialdini, Rice, & Serna, 2002), *retrieval in persuasion* (Tormala, Petty, & Briñol, 2002), *risk analysis* (Siegrist, 2000), *risk and protective factors in substance use* (Epstein, Griffin, & Botvin, 2001; Griffin, Scheier,

Botvin, & Diaz, 2000, 2001), *safer sexual behavior* (Bryan, Fisher, & Fisher, 2002), *schisms in groups* (Sani & Todman, 2002), *schizophrenia* (King, 2000), *school dropout* (Battin-Pearson, Newcomb, Abbott, Hill, Catalano, & Hawkins, 2000), *school misbehavior* (Bryant, Schulenberg, Bachman, O'Malley, & Johnston, 2000), *school violence* (Mayer & Leone, 1999), *selection procedures* (Smither, Millsap, Stoffey, & Reilly, 1996), *self concept* (Marsh, & Yeung, 1998), *self-control* (Wills & Stoolmiller, 2002), *self-determination and work* (Deci, Ryan, Gagné, Leone, Usunov, & Kornazheva, 2001), *self efficacy* (Caprara, Scabini, Barbaranelli, Pastorelli, Regalia, & Bandura, 1998; Gibbons & Weingart, 2001; Rimal, 2001), *self esteem* (Motl & DiStefano, 2002), *self medication* (Damphousse & Kaplan, 1998), *sexual abuse* (Merrill, Thomsen, Sinclair, Gold, & Milner, 2001); *sexual activity* (Pillai, & Gupta, 2000), *sexual assault* (Merrill, Thomsen, Gold, & Milner, 2001), *sexual assertiveness* (Morokoff, Quina, Harlow, Whitmire, Grimley, Gibson, & Burkholder, 1997), *sexual harassment* (Harned, 2000; Harned & Fitzgerald, 2002), *sexual offender treatment* (Hunter & Figueredo, 1999), *sexual orientation* (Kirk, Bailey, Dunne, & Martin, 2000), *sexual risk behavior* (Duncan, Strycker, & Duncan, 1999; Ramirez-Valles, Zimmerman, & Newcomb, 1998), *SF36* (Keller, Ware, Bentler, Aaronson, Alonso, et al., 1998), *sleep* (Carmelli, Bliwise, Swan, & Reed, 2001), *smoking* (Epstein, Griffin, & Botvin, 2000bc; Orlando, Ellickson, & Jinnett, 2000), *smoking and drinking* (Murray, Istvan, Cribbie, & Barnes, 2002), *smoking cessation* (Williams, Gagné, Ryan, & Deci, 2002), *social benefit expectancies* (Griffin, Epstein, Botvin, Spoth, 2001), *social class* (Muntaner, Oates, & Lynch, 1999), *socialization* (Collins & Bagozzi, 1999), *social psychology* (Ebreo, Linn, & Vining, 1996), *social relations* (Snijders & Kenny, 2000), *social support* (Dormann & Zapf, 1999; Mason & Windle, 2001), *speaker identification* (Yuo & Wang, 1999), *sport psychology* (Kelley, Eklund, & Ritter-Taylor, 1999), *spousal distress in cancer* (Fang, Manne, & Pape, 2001), *stages in drug use* (Bentler, Newcomb, & Zimmerman, 2002), *status attainment* (Bond & Saunders, 1999), *stress* (Skinner & Brewer, 2002; Roesch, 1999), *student perceptions* (Griffith, 1999), *substance abuse treatment outcome* (Latimer, Newcomb, Winters, & Stinchfield, 2000), *substance use* (Duncan, Duncan, Alpert, Hops, Stoolmiller, & Muthén, 1998; Rose, Chassin, Presson, & Sherman, 2000), *suicide ideation* (Prinstein, Boergers, Spirito, Little, & Grapentine, 2000), *sun protection* (Jackson & Aiken, 2000), *survival data* (Posner & Baker, 2000), *Tai Chi exercise* (Li, Duncan, Duncan, McAuley, Chaumeton, & Harmer, 2001; Li, Harmer, McAuley, Fisher, Duncan, & Duncan, 2001), *teacher burnout* (Byrne, 1999), *teaching evaluation* (Marsh & Roche, 2000), *temperament & substance use* (Wills, Cleary, Filer, Shinar, Mariani, & Spera, 2001), *test bias* (Millsap, 1997), *TQM* (Curkovic, Melnyk, Handfield, & Calantone, 2000), *trait and state* (Kenny & Zautra, 1995, 2001; Steyer & Partchev, 2001; Tisak & Tisak, 2000), *trait similarity* (Judge, Erez, Bono, & Thoresen, 2002), *transition to parenthood* (Rholes, Simpson, Campbell, & Grich, 2001), *twins* (Prescott, Aggen, & Kendler, 2000), *union structure* (Mellor & Mathieu, 1999), *venture capital success* (Schefczyk, 2001), *vision* (Peterzell, & Teller, 2000), *WAIS-R structure* (Ward, Axelrod, & Ryan, 2000), *war and child adjustment* (Smith, Perrin, Yule, & Rabe-Hesketh, 2001), *weight control* (Neumark-Sztainer, Wall, Story, & Perry, 2003), *welfare outcomes* (Zaslow, Hair, Dion, Ahluwalia, & Sargent, 2001), *well-being* (McAuley, Blissmer, Marquez, Jerome, Kramer, & Katula, 2000), *wellness as protection* (Epstein, Griffin, & Botvin, 2002), *wetland management* (La Peyre, Mendelssohn, Reams, Templet, & Grace, 2001), and *working memory and fluid intelligence* (Engle, Tuholski, Laughlin, & Conway, 1999). These publications are illustrative, rather than exhaustive, and no claim is made that the summary label for each study is the most appropriate one, or that the most significant works in each category have been provided.

Further Readings

This *Manual* provides a self-contained presentation of the basics of structural modeling and enough advanced topics to enable the reader to understand and use EQS in a meaningful fashion. It must be recognized, however, that this manual is not a self-contained text, since a course in structural equation modeling would need to cover in greater detail such topics as path analysis, confirmatory factor analysis, simultaneous equations, and various aspects of multivariate analysis including a variety of technical topics such as nonlinear programming. The concept of latent variable, and the related idea of errors of measurement, is a crucial aspect of structural modeling. A few additional introductory and survey materials on latent variable models, either in the context of EQS, or consistent with concepts in EQS, are given in the following sources.

Two books that focus specifically on introductions to EQS are: Byrne, B. M. (1994). *Structural Equation Modeling with EQS and EQS/Windows.* Thousand Oaks, CA: Sage, and Dunn, G. Everitt, B. & Pickles, A. (1993). *Modeling Covariances and Latent Variables using EQS.* London: Chapman & Hall. Overviews that are compatible with the approach used in EQS include Kline (1998), Loehlin (1998), Mueller (1996), Raykov and Marcoulides (2000),

Schumacker and Lomax (1996), Ullman (2000) and Ullman and Bentler (2003), as well as works in French (Rousel, Durrieu, Campoy, & Akremi, 2002), German (Bentler, Wu, & Houck, 1996), and Japanese (Kano, 1997). These sources do not cover many of the new features in EQS 6. Several, however, are being updated to include EQS 6.

Introductions to the broader literature on structural modeling, as well as more technical reviews and specialized topics, can be found Aigner and Goldberger (1977), Aigner, Hsiao, Kapteyn, and Wansbeek (1984), Austin and Calderón (1996), Austin and Wolfle (1991), Bartholomew and Knott (1999), Bentler and Dudgeon (1996), Bentler and Yuan (1999a), Bollen (1989b, 2002), Browne and Arminger (1995), Cudeck, du Toit, and Sörbom (2001), Duncan, Duncan, Strycker, Li, and Alpert (1999), Everitt (1984), Goldberger and Duncan (1973), Hoyle (1995, 1999), Kaplan (2000), Little, Schnabel, and Baumert (2000), MacCallum and Austin (2000), Marcoulides and Moustaki (2002), Marcoulides and Schumacker (1996, 2001), Moskowitz and Hershberger (2002), Muthén (2002), Pugesek, Tomer, and von Eye (2003), Rigdon (1998), Rose, Chassin, Presson, and Sherman (2000), Schumacker and Marcoulides (1998), Shipley (2000), Thompson (2000), Wolfle (2003), and Yuan and Bentler (2001b).

2. BASIC STRUCTURAL EQUATION MODELS

A minimal set of key concepts are needed to translate a substantive theory into the form of a model that can be statistically estimated and tested. Models are specified in EQS using a specially developed language that is intended to be easy and straightforward to use. The language is concerned primarily with the unambiguous specification of the variables involved in a structural model. Four types of variables are allowed. The variables in the input data file to be analyzed are observed or measured **v**ariables; they are called Vs in EQS. Hypothetical constructs are unmeasured latent variables, as in factor analysis, and are called **f**actors or Fs. **E**rrors or Es generate residual variation in measured variables, and the corresponding residuals in factors are called **d**isturbances or Ds. Thus V, F, E, and D variables are introduced as needed in the examples below. In addition to specifying variables, a model must contain information on the hypothesized effects of variables on each other (regression coefficients) and the specification of variables that are not explained by other variables (variances and covariances). This latter information is implemented in the program in accord with the representation system of Bentler and Weeks (1979, 1980) given in Chapter 1 as equations (eqs.) (1.2) – (1.4). This chapter introduces the relevant and nonmathematical part of the Bentler-Weeks approach, and shows how to use the concepts and the EQS language to set up regression, path, factor, and general structural models. The reader is urged to study each setup and run the corresponding model, or some variant of it, with EQS. *The experienced researcher may skip this chapter*.

The most popular versions of EQS (e.g., for EQS 6 for Windows) contain a graphical user interface. In such versions, models can be specified directly by creating a path diagram, or by following a series of menus in the *Build EQS* module. Both diagrams and menus are directed towards building a *model file*, typically called a **.eqs* file, which contains commands that specify the details of a model run. In the file name, the "*" is a mnemonic name you specify, while the "eqs" part is a recommended shorthand to designate model files. In this chapter, the commands themselves are introduced and discussed. A text editor can be used to build the *.eqs file directly, that in turn can be submitted to an EQS run on any computer system for estimation, testing, and other results. That is, *.eqs files can be transported across computer systems.

An old empirical study will provide a concrete illustration of how the basic Bentler-Weeks concepts are implemented in EQS. The example is the model of the stability of alienation taken from the work of Wheaton, Muthén, Alwin, and Summers (1977). In this study, 932 individuals' responses to a series of attitude items on several occasions were modeled in accord with a particular concept of what the items measured. An assumption (here untested) is that items could be meaningfully aggregated into scales of "anomie" and "powerlessness." The anomie and powerlessness variables were actually measured at several time points, given here for the years 1967 and 1971. A question of interest involved the stability of these measures across time. Since it is possible that the stability of these measures could be influenced by irrelevant sources of variance, especially socioeconomic status, two measures of such status, "education" and Duncan's occupational status index ("SEI"), were obtained as well.

The six variables (V), and their covariances based on a sample of size 932 respondents, are:

V1 = Anomie (1967)	V2 = Powerlessness (1967)
V3 = Anomie (1971)	V4 = Powerlessness (1971)
V5 = Education	V6 = Occupational status index, SEI.

Stability of Alienation: Covariance Matrix

	V_1	V_2	V_3	V_4	V_5	V_6
V_1	11.834					
V_2	6.947	9.364				
V_3	6.819	5.091	12.532			
V_4	4.783	5.028	7.495	9.986		
V_5	-3.839	-3.889	-3.841	-3.625	9.610	
V_6	-21.899	-18.831	-21.748	-18.775	35.522	450.288

The 6(7)/2 = 21 elements of this matrix represent the data vector *s*, given in eq. (1.5), that is to be modeled according to a theory that proposes to explain how the data were generated. Before turning to a meaningful structural theory for these data, let us consider a model for these data associated with standard multivariate analysis. In the absence of a structural theory, the variances and covariances of the variables V1-V6 are themselves parameters to be estimated. For example, they are parameters in the multivariate normal distribution that might be hypothesized to have generated the sample data. In this case the vector σ in (1.5) is simply the vector of population variances and covariances, which is not considered to be a function of structural parameters θ. Rather obviously, (1.5) can thus be minimized for any nonnegative definite choice of *W* by setting the estimate $\hat{\sigma} = s$, causing the function *Q* to attain its lower bound of zero. But there are as many parameters in σ (here, 21) as there are data points (sample variances and covariances), so this unstructured or "saturated" model has no degrees of freedom (df) and will always fit any set of data. As a consequence, unless a further hypothesis is imposed, the saturated model is very uninteresting. In the absence of a structural theory, traditional hypotheses might be that $\sigma_{13} = 0$ or that $\sigma_{24} = \sigma_{56}$. Each such restriction gains a degree of freedom. Also, then $\hat{\sigma} \neq s$, and hence *Q* would be nonzero, and a nonstructural but potentially interesting hypothesis could be tested.

The following is a path diagram of the saturated model:

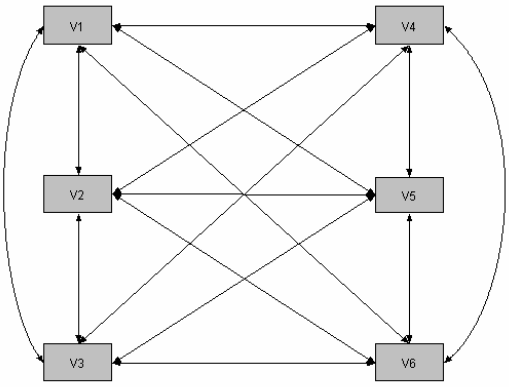


Figure 2.1

In this diagram, each measured variable is shown in a rectangle, as is traditional. The variance of each of these variables is not shown, but it is a parameter of the model, so each rectangle has been shaded as a reminder. Each variable's variance is a parameter here, but in subsequent models this will not be true and some variables will not be shaded. The covariances between variables are shown as two-way arrows, as is traditional. Each of these pairwise associations is a parameter in the model. There are thus six variances and 15 covariances in the model. The traditional hypothesis $\sigma_{13} = 0$ would be shown in the diagram by removing the two-way arrow that connects V1 and V3. Removing such a path would introduce a degree of freedom, and yield a testable hypothesis that may or may not be true. (Note: The number following a variable should be read as a subscript, but it is written here in the format required by EQS. Most computers cannot read subscripts in text files.)

In the absence of a structural model, only a relatively limited range of hypotheses on σ or on functions of σ can be evaluated. In this particular study, where variables V1 and V3, as well as V2 and V4, are the same variables measured at two time points, one might consider a hypothesis such as $\sigma_1^2 = \sigma_3^2$, which can be tested by EQS if desired. The range of such hypotheses involving parametric restrictions on the variances and covariances is generally rather limited, so that a more generic approach to multivariate analysis would impose a structure on the measured variables based on fewer parameters (elements of θ) than data (elements of *s*). Ideally, the model should contain very few parameters, so that the difference between data points and number of parameters, the degrees of freedom, is quite large. Then, even the optimal choice of parameters $\hat{\theta}$ will yield $\hat{\sigma} = \sigma(\hat{\theta})$ that may not exactly reproduce the data. Thus the residual $(s - \sigma(\hat{\theta}))$ will be nonzero, $\hat{Q}$ (= *Q* evaluated at $\hat{\theta}$) will be nonzero, and a

major statistical question is whether $\hat{Q}$ is sufficiently small, given the sample size involved and the df associated with the model. If $\hat{Q}$ is statistically small, the model gives a good representation of the data. Note: in SEM, the word "residual" has two major meanings: (1) there are *residuals in fitting the model to data*, namely, $(s - \sigma(\hat{\theta}))$, and (2) there are *residual variables*, such as errors in variables (called Es in EQS) or disturbances in factors (called Ds in EQS). So a model may fit perfectly, with no residuals, i.e., $(s - \sigma(\hat{\theta}))$ = 0; but almost surely, there will be some residual variables in the model! The context usually makes the specific meaning clear. To make things yet more complicated, there is actually a third meaning of "residual" which is something in between (1) and (2): namely, (3) model fit evaluated for each individual instead of the entire sample. See e.g., Raykov and Penev (2001) or Reise and Widaman (1999). We will not emphasize this third meaning in this *Manual*.

Because the modeling residual $(s - \sigma(\hat{\theta}))$ is intrinsic to our modeling enterprise, we can say that we are fitting the model $\sigma(\hat{\theta})$ to the data *s*. However, we are not testing whether the model fits the data; this is not really the null hypothesis. The null hypothesis concerns population parameters only: $\sigma = \sigma(\theta)$. In doing modeling we will always initially assume that the null hypothesis is true. That is, we assume $\sigma = \sigma(\theta)$, but then, of course, we allow the data to tell us if this is a bad hypothesis. After all, if *s* is a good sample from σ, we should have $s \cong \sigma(\theta)$ for a correct model. But if *s* is drawn from a different population, or if our model is wrong, our model will not approximate the sample data very well.

Stated differently, if were able to analyze the *population* covariance matrix with elements in the vector σ, under the null hypothesis, our structural model would give estimates $\hat{\theta}$ that are identical to θ. Hence, the fit would be perfect, and the residuals $(\sigma - \sigma(\hat{\theta}))$ would be zero. On the other hand, if our null hypothesis, or model, is not correct, then even in the population the residuals $(\sigma - \sigma(\hat{\theta}))$ are nonzero, and hence $\hat{Q}$ (a weighted sum of squares of residuals) would be nonzero. With an incorrect model, the larger $\hat{Q}$ is in the population, the more *power* our test has to reject the null hypothesis. Browne, MacCallum, Kim, Anderson, and Glaser (2002) discuss a situation where power is large even though the residuals are small on average. Cudeck and Henly (1991) provide a good discussion of the various forms of discrepancy involved in fitting models, and Browne and Cudeck (1993) discuss ways to evaluate those discrepancies. Power was introduced to SEM by Satorra and Saris (1985) and Matsueda and Bielby (1986), and is an important issue in small samples. Useful power tables are provided by MacCallum, Browne, and Sugawara (1996). The bootstrap with downweighting of outlier cases may be needed to provide adequate power analyses in nonstandard situations (Yuan & Hayashi, in press).

This chapter will approach the problem of developing and testing a structural theory for the alienation data in stages, allowing a number of rather simple concepts to be introduced in turn. Several small models for these data, or subsets of the data, will be developed, and the problem of specifying the models in EQS will be described as a way of giving you familiarity with the program. The chapter will begin with a diagrammatic introduction to the familiar topic of multiple regression.

A Simple Regression Model

The ordinary multiple regression equation $y = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + e$ serves as the basic building block of all linear structural models. This was noted in the discussion surrounding eq. (1.1). In most single-population covariance structure models, including regression, the intercept β_0 is typically not of interest and the variables are taken to be deviations from means (i.e., β_0 is excluded from the equation, as above). See Chapter 8 for a discussion of models that involve means. As is well known, the parameters of the equation are the regression coefficients. These can be estimated by some method such as least squares, given a sample of observations on the variables. It is somewhat less well known among data analysts that the variance σ^2 of the residual *e* is also a parameter to be estimated, but this is usually considered to be a minor point because the estimate falls out readily via the standard

approach to regression. From the structural modeling perspective, however, even this specification is incomplete. In particular, from the perspective of the Bentler-Weeks model:

The parameters of any linear structural model are the regression coefficients and the variances and covariances of the independent variables.

Hence, the variances of the variables x_1 - x_p and each of the pairwise covariances among these variables are also parameters of the model. These are typically considered to be known numbers when the regression model is based on the concept of "fixed" regressors. In virtually all structural modeling applications, however, these variables are best considered to be "random" variables whose variances and covariances are not known a priori before sample data is obtained, and that must be estimated as well. Of course, in the regression model, the optimal estimates turn out to be the sample values of these variances and covariances, but this convenience should not obscure the point that they are parameters that need to be estimated.

The Bentler-Weeks concept of "independent" variables is broader than the typical one. Any variable that is not a dependent variable in any regression equation is an independent variable, including, in this case, the residual variable *e*. But in regression models the variables x_i are all taken to be orthogonal or uncorrelated with *e*, so that these zero covariances are not parameters of the model while the variance of *e* is a parameter. And, of course, a "dependent" variable is any variable that is expressed as a structural regression function of other variables. In a single equation, as in this case, there is thus only one dependent variable. In models with many equations, there are as many dependent variables as there are equations, and typically the dependent variables are on the left side of the "=" sign. Note in particular that a corollary of the Bentler-Weeks approach is the following:

Dependent variables do not have variances or covariances with other variables as parameters of a structural model.

Such variances and covariances are to be explained by the parameters. Thus *y* indeed has a variance, and *y* covaries with x_i, but these numbers are a function of, i.e., generated by, the model parameters. Models with means also will have means and intercepts of independent variables as parameters. And further, though no program implements this concept, the parameters may include the higher-order multivariate product-moments such as skewnesses and kurtoses of the independent variables. This point will not be discussed further; see Bentler (1983a).

A simple regression problem with the previously given data can serve to fix ideas as well as to introduce the EQS language. Consider the stability of Powerlessness. One could index this by the regression of V4 on V2, that is, by the size of the coefficient β in V4 = βV2 + E4. But if one suspects that stability is influenced by Education, it may be more desirable to consider the regression V4 = β_1V2 + β_2V5 + E4. The coefficient β_1, obtained when V5 (Education) is included in the equation, may differ from β. EQS uses the convention that measured variables are indexed sequentially V1, V2, ... as read in; all six variables of the example will be read (some can be discarded internally). Residuals in measured variables, or errors in variables, are denoted E and are followed by the number of the measured variable being predicted. A path diagram for the problem, using *b* instead of β, is as follows.

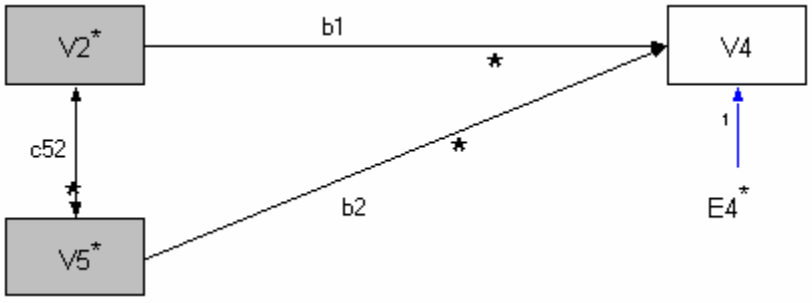


Figure 2.2

This uses the usual convention that measured variables (Vs) are represented in rectangles while the unmeasured residual is not in a rectangle. The independent V variables, V2 and V5, are shaded, while the dependent variable, V4, is not shaded. Thus the independent variables are the shaded variables.

According to the Bentler-Weeks concepts, the parameters of the model are:

Regression coefficients: β_1 (=b1), β_2 (=b2), 1.0
Variances of independent variables: $\sigma^2_{V2}, \sigma^2_{V5}, \sigma^2_{E4}$
Covariances of independent variables: $\sigma_{V5,V2}$ (= σ_{52} = c_{52}).

It will be apparent that the path diagram represents regression coefficients as unidirectional arrows, and covariances between independent variables as two-way arrows. Note also that the path from E4 to V4 is implicit; it is actually another regression coefficient, but it is known and fixed at 1.0 (because it is implicit in the equation V4 = β_1V2 + β_2V5 + 1.0E4). Most path diagrams do not clearly show that the variances of the independent variables are also parameters. One could put this information in the diagram, but since it is typically absent, the user who translates a diagram into equations must be quite careful. *In principle, there is a one-to-one correspondence between a diagram and a model specified via equations*, but lack of a standard convention makes this point obscure. Furthermore, just as diagrams typically do not tell the whole story of a model, neither do equations; the variances and covariances of independent variables are sometimes forgotten or presented as an afterthought in presentations of linear structures.

Structural models represented in diagram form must also be carefully scanned for what is *not* present in the model, as is true even in this simplest of all possible models. In Figure 2.2, for example, we note two main points. Most obviously, no other variables are included in the model; hence such other variables should be eliminated from any program input. Less obviously, the model has imposed the typical, and almost always necessary, assumption that the predictor variables V2 and V5 are not correlated with E4, the residual error variable. If a correlation were specified, the diagram would include a two-way arrow connecting these variables. The absence of a two-way arrow says, "This covariance is fixed at zero." Similarly, the absence of a one-way arrow would say that a regression coefficient is known and fixed at zero.

A few more words need to be said about the fixed 1.0 path from E4 to V4, because a generic principle of *identification* is involved. Parameter, equation, and model identification is a complex topic that deals with the issue of whether a structural model has been specified so that the parameters of the model are unique. If the parameters were subject to any arbitrariness, it would be difficult to speak of them as true parameters that are to be estimated, since a wandering target would be involved. In the simple case of regression, we can always write the equation in two different but equivalent forms. The first involves fixing the path and then considering the variance of the residual as a free parameter, as was done above. But it would also be possible to write the equation as V4 = β_1V2 + β_2V5 + β_3E4, considering β_3 as a free parameter to be estimated and the variance of E4 as fixed (say, at 1.0). While regression is rarely developed this way, it certainly could be, since it is a trivial variation of the identical model. Note that it is *not* possible to free both parameters β_3 *and* the variance of E4, since the product (β_3E4) must be a single and unique number given the other parameters (=V4 - β_1V2 - β_2V5). Similarly, it is not possible to fix both parameters (e.g., β_3 = 1.0, σ^2_{E4} = 1.0) because that would imply that the residual variance is not to be minimized by optimal choice of β_1 and β_2. More generically, then, in all structural models including multiple equation models:

Every unmeasured variable in a structural model must have its scale determined. This can always be done by fixing a path from that variable to another variable at some known value (usually 1.0). An alternative method for determining the scale of an <u>independent</u> unmeasured variable is to fix its variance at some known value (usually 1.0).

Note that the scale of a dependent unmeasured variable cannot be determined by fixing its variance, because dependent variables do not have variances as parameters. The above simple rule will help you to avoid identification difficulties most of the time, though there are many nuances in practice (e.g., Bollen & Jöreskog, 1985; Shapiro & Browne, 1983; Steiger, 2002). It also applies to latent variable models that involve F and D variables. As already indicated, EQS allows V, F, E, and D types of variables. F variables are typically taken as factors or substantive

latent variables, while D variables are residuals or disturbances in equations. Thus measured variables are always V variables, and all F, E, and D variables must have their scales fixed for a model to have a chance of being identified.

The simple regression model is intended to be compared to the second-moment summary statistics, specifically, the variances and covariances of the measured variables. Thus the data vector *s* for the regression model consists of all of the variances and covariances of the measured variables. With p variables, there are p(p + 1)/2 such elements. Here p = 3, so there are 3(4)/2 = 6 data points (the sample variances of V2, V4, and V5, and their three pairwise covariances). The number of parameters to be estimated is the sum of the unknown regression coefficients (2), the variances of the independent variables (3), and their covariances (1). Thus, there are six parameters to be estimated, based on six data points. It is a *saturated* model. Evidently, the regression model has as many free parameters as data points, so it is basically an uninteresting model considered alone since there are no degrees of freedom to test any hypothesis in the model. If one of the parameters, most typically a regression coefficient, were fixed at a certain value, e.g., $\beta_2 = 0$, then there would be only five parameters and the model would have one df that could be used to test the null hypothesis of model fit $\sigma = \sigma(\theta)$; this hypothesis is identical to the hypothesis $\beta_2 = 0$.

The EQS program can be used to perform a regression analysis, although when the data are multivariate normal, programs specifically developed for regression are more appropriate because of the many special features they contain. However, there do not appear to be many programs to perform such estimation for variables that are elliptically or arbitrarily distributed as allowed in EQS. The regression problem discussed previously, considered in a simple normal theory least squares context, is submitted to EQS as shown on the left. Comments to the right of the exclamation marks are ignored by EQS.

```
/TITLE                                          ! Optional - can be omitted
 STABILITY OF POWERLESSNESS                     !     "
/SPECIFICATIONS                                 ! Necessary
 CAS=932; VAR=6; ME=LS;                         ! Number of cases and variables and
                                                ! Least squares estimation method
/EQUATIONS                                      ! Must have EQUATIONS, MODEL or RELIABILITY
 V4 = 1*V2 - 1*V5 + E4;                         ! section.  Free parameters shown by stars
                                                ! +1, -1 are initial guesses or estimates
/VARIANCES                                      ! Necessary for independent variables
 V2 = 9*; V5 = 9*; E4 = 2*;                     ! Arbitrary start values given
/COVARIANCES                                    ! Needed here to show that the covariance
 V5,V2 = -4*;                                   ! is a free parameter
/MATRIX                                         ! Not needed if data are on external file
 11.834                                         ! Data given in free format
  6.947   9.364                                 ! VAR=6 needs lower triangle 6-by-6 matrix
  6.819   5.091  12.532                         !
  4.783   5.028   7.495   9.986                 ! Compare last row to original matrix
 -3.839  -3.889  -3.841  -3.625  9.610         !
 -2.1899 -1.8831 -2.1748 -1.8775 3.5522 4.50288 ! No semicolon here
/END                                            ! Last input statement
```

The input to EQS evidently uses the Bentler-Weeks concept of independent and dependent variables, since as discussed above, variables V2, V5, and E4 are considered to be independent variables while V4 is a dependent variable.

In essentially every model, any equation specifying a dependent variable (here, V4) *will contain an independent variable* (here, E4) *that is the residual or unpredicted part of the dependent variable.*

As the regression diagram shows, only one covariance is a parameter of the model since the other covariances among independent variables are fixed at known, zero values. In EQS, free parameters such as the regression coefficients β_i are not denoted by a Greek symbol; rather they are indicated by an asterisk or star ("*") with an associated number that represents the researcher's guess as to the value of that parameter. Thus the input stream can be scanned for the number of stars, which represent the free parameters to be estimated by the program. Here, there are six.

In the Windows environment, the input matrix need not be given in the input file. It will simply be computed from the active data file. However, you need to become acquainted with various ways of running EQS. It may seem

peculiar to say in /SPEC that there are 6 variables, or to read an input MATRIX that is 6 by 6 corresponding to VAR=6 when the problem under consideration is only concerned with three variables. Actually, EQS has an automatic variable selection feature that performs the necessary reduction to the relevant variables: nothing needs to be explicitly stated. After all, the model is completely given by the EQUATIONS and VARIANCES sections. If a variable does not appear in these sections, it will not be included in the model. (Note that independent variables cannot have covariances without having variances. Thus, COV does not help to define which variables are to be selected.) The advantage of this automatic selection device is that a researcher may have dozens of models to evaluate on various parts of a given data set. In such a case it may be desirable to have the variable numbers of the large data set have a constant meaning even in subanalyses, and only critical changes need to be made from one run to the next. In particular, the input MATRIX can simply stay the same in all runs.

If the last row of the MATRIX section is compared to the matrix at the beginning of this chapter, it will be apparent that the variance of V6 has been divided by 100 while the covariance of V6 with each other variable has been divided by 10. Such a change reflects a rescaling of the raw scores of the original variable by multiplying by .1; that is, Final V6 = .1 times Original V6. In Windows, such rescaling can be done directly in the data file. Here we used the rule that if x = kX, then var(x) = var(kX) = k^2var(X) and cov(xY) = cov(kXY) = kcov(XY); here, k = .1. This change was made to make the variances of the input variables more similar to each other. In the original data, the variance of V6 was roughly 45 times larger than the variances of the other variables. In the modified matrix, the ratio of smallest to largest variance is more on the order of 1:3. It is known that the efficiency and accuracy with which programs like EQS can optimize a nonlinear function of many parameters, or solve a simple linear regression problem, depend on the "condition number" of the input matrix, which in turn is strongly affected by differences in scale of the input variables. If you have convergence troubles, you should consider whether variance differences are creating the problem. If nothing works, you could consider specifying ANALYSIS=ZSCORES; which forces the program to analyze a correlation matrix, which, of course, has equal unit variances. Once the model is acceptable, you should go back to the covariance matrix as noted next.

It is good practice, in general, to scale input variables so that they have similar variances. The scaling constants should, however, be numbers that do not depend on the particular sample of data under study. They should not be random variables such as would be chosen if one standardizes all variables to unit variance in the sample (transforms a covariance matrix into a correlation matrix). Such scaling "constants" would vary with the sample, which would complicate the statistical theory. The rather simple idea of moving a decimal place for some input variables (e.g., changing the format statement F3.0 to F3.1, indicating a field of 3 digits should be interpreted as having one digit to the right of the decimal, when reading raw score data) is a convenient way to deal with this problem. Of course, it must be remembered that the resulting parameter estimates and standard errors are based on the particular scales chosen. It will be obvious, of course, that rescaling V6 has no effect on the problem under consideration, which is the regression of V4 on V2 and V5. Thus the rescaling does not need to be done in this case, but it was done because the rescaled matrix will be used in subsequent analyses without further comment. Of course there are situations when you specifically want to use ANALYSIS=CORRELATION (stated in /SPEC) which forces EQS to analyze the correlation matrix using an appropriate statistical theory. See Chapter 5 for details.

EQS generated the LS regression solution to this problem as follows: V4 = .457*V2 - .192*V5 + E4. It would thus appear that the main effect on powerlessness in 1971 (V4) is by the same variable measured in 1967 (V2), even though education has a slightly negative impact. As expected, the variances of the two independent Vs are estimated at the sample values, as is their covariance. The variance of the residual is estimated at 6.991.

When variables have a known and interpretable metric, and in some other circumstances (see e.g., Greenland, Schesselman, & Criqui, 1986), regression weights can be interpreted for nonstandardized variables. This may be quite difficult in the social sciences, where no meaningful scale for the variables may exist. In order to alleviate this problem, EQS produces a completely *standardized solution*, one in which each and every variable in a model has been standardized to unit variance. In this case the standardized solution gives the ordinary standardized beta weights typically encountered in regression: V4 = .443*V2 - .189*V5 + .837E4. Note that the residual E4 is also standardized. The proportion of variance E4 accounts for in V4 is $.837^2$ = .70, and hence the squared multiple correlation between the predictors and V4 is 1 - .7 = .30. This is printed as R-SQUARED in the standardized solution output. Evidently powerlessness is not very stable, while it is marginally negatively affected by education. The program also provides the correlations among the standardized independent variables; the V5, V2 correlation is -.410.

The above example used the covariance matrix of all six variables as input to EQS. This is not necessary when only a subset of three of the six variables is being analyzed. Thus, an alternative program input would be the following:

```
/SPE                                          ! Can be abbreviated
 CAS=932; VAR=3; ME=LS;                       ! Only three variables
/EQU                                          ! Abbreviated
 V2 = 1*V1 - 1*V3 + E2;                       ! Note change in variable numbers
/VAR                                          ! Abbreviated
 V1 = 9*; V3 = 9*; E2 = 2*;                   ! Again note change in numbers
/COV                                          ! Abbreviated
 V3,V1 = -4*;                                 ! Again note change in numbers
/MAT                                          ! Abbreviated
  9.364                                       ! 3-by-3 matrix selected from 6-by-6 matrix
  5.028   9.986
 -3.889  -3.625  9.610
/END
```

Since EQS assumes that the input variables are numbered in sequence, and three variables are in the matrix, it assumes that the variables are numbered V1, V2, and V3. Thus the EQUations, VARiances, and COVariance sections must use a notation that corresponds to this sequence. If the equation had specified the model V4 = 1*V2 + 1*V5 + E4; as previously, the program would have searched for variables V4 and V5, not found them, and would have printed an error message and quit.

Input covariance or correlation matrices, whether they are in the /MATRIX section or on an external file, must be complete and have no missing elements or special missing data codes (such as 9's or other arbitrary numbers or symbols). Raw data may have missing values when special procedures (described in Ch. 12) are used to handle missing data.

Another job setup for the regression problem is the following: a correlation matrix rather than a covariance matrix is used as input, and standard deviations are provided to transform the correlations into covariances.

```
/SPE to /COV sections exactly as above.  Then:
/MAT
  1.00                                        ! Correlation input
   .52   1.00
  -.41   -.37   1.00
/STANDARD DEVIATIONS                          ! Required
  3.06   3.16   3.10                          ! No semicolon
/END
```

EQS will create the covariance matrix from the correlations and standard deviations without any further instruction, and the analysis will be done on the covariance matrix. It follows that the parameter guesses provided in the input should be of a magnitude consistent with the scaling implied by the covariances, and not that of the correlations. The output estimates will be in the covariance metric. The use of an input correlation matrix with standard deviations, as above, integrates well with some statistical packages that easily provide such information but do not readily provide covariance matrices. If you feel that you can ignore differences in variance between variables, you can, of course, do the analysis on correlations.

A regression equation in the context of a "causal" model is called a *structural equation*, and the parameters, structural parameters. Structural parameters presumably represent relatively invariant parameters of a causal process, and are considered to have more theoretical meaning than ordinary predictive regression weights, especially when the regression equation is embedded in a series of simultaneous equations designed to implement a substantive theory. The variables used in the equations must, of course, adequately represent critical substantive concepts. It is not necessarily easy to assure that variables measure a domain well (e.g., Little, Linderberger, & Nesselroade, 1999); see Saucier and Goldberg (2002) and Saucier (2002) for the case of the "Big 5" personality factors. In some domains, there may be so many items or variables that it may be necessary to use aggregate variables such as item parcels. While this can work well (Bandalos, 2002; Kishton & Widaman, 1994; Yuan, Bentler, & Kano, 1997), it can raise hard-to-resolve issues (Bandalos & Finney, 2001; Hall, Snell, & Foust, 1999; Little, Cunningham, Shahar,

& Widaman, 2002) that may require evaluating the parcels' unidimensionality (Kim & Hagtvet, 2003). Also the model design must be appropriate to the theoretical specification and should include relevant causal variables if at all possible (James, Mulaik, & Brett, 1982; Pratt & Schlaifer, 1984). The theoretical meaning of the model parameters is also based on the invariance characteristics of the solution made possible by the design. Thus, for example, ordinary regression coefficients can be arbitrarily affected by the amount of measurement error in the variables, but similar coefficients for the regression among latent variables would not be so affected (see e.g., Bollen, 1989b, Ch. 5; Shipley, 2000, Ch. 5). Multiple equation models, which are necessary to implement latent variable modeling, are introduced next.

A Two-Equation Path Model

Structural equation models are virtually always multiple equation models, with multiple dependent as well as independent variables. The simplest extension of the regression model is the multivariate regression model, in which several dependent variables are regressed on several predictor variables. In the ordinary multivariate regression model, there are again as many parameters as data points, so that as a structural model, it can always be fit exactly for any set of data. The next example is a trivial variant of such a model in which there is one over-identifying restriction, leading to a model that is testable. In particular, a consideration of the anomie and powerlessness variables measured on the two occasions leads to the standard cross-lagged panel design:

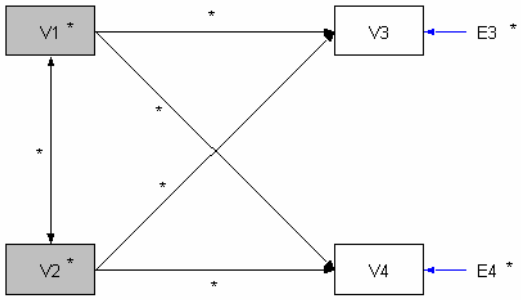


Figure 2.3

As before, measured variables are in rectangles, and independent measured variables are shaded. The arrows from the residuals E3 and E4, having no further designation, are assumed to represent fixed, known values of 1.0 (not shown in the diagram for simplicity). In addition, however, the diagram includes the nonstandard but useful feature of having a star (asterisk) in every location where a free parameter exists in the model. A star next to a variable indicates that the variable has a variance that is to be estimated. Thus, it can be seen quickly that there are nine parameters in the model, which, for 4(5)/2 = 10 data points leaves one df. The df is associated with the question of whether the covariation between variables V3 and V4 can be totally accounted for by variables V1 and V2. Allowing the errors E3 and E4 to covary would lead to a saturated model with as many parameters as data points, i.e., a model with zero df.

The cross-lagged panel design is often also used to test more restrictive models regarding the relative influence of the crossed variables on each other. In this instance, where anomie and powerlessness are each measured on two occasions, one might be interested in whether initial feelings of anomie (V1) more strongly affect feelings of powerlessness at a later time point (V4), than initial feelings of powerlessness (V2) affect feelings of anomie (V3). Such questions involve null hypotheses regarding whether the V1 → V4 path is equal to the V2 → V3 path, or whether one or the other might be statistically different from zero. These hypotheses will not be evaluated in this section. They are best evaluated in a design in which the key variables are latent rather than observed. Such designs have been used in studies on the effects of attitude on behavior vs. the effect of behavior on attitude. Some attitude-related models are given by Bentler and Speckart (1979, 1981), Bryan, Fisher, and Fisher (2002), Jackson and Aiken (2000), O'Callaghan, Callan, and Baglioni (1999), Perugini and Bagozzi (2001), Sideridis, Kaissidis, and Padeliadu (1998), Speckart and Bentler (1982), and Stacy, Bentler, and Flay (1994).

The path diagram clearly indicates how many equations exist in this model. In such a diagram, dependent variables can be unambiguously recognized because they have unidirectional arrows aiming at them. The above diagram has two variables that have unidirectional arrows aiming at them: V3 and V4. Thus, the model consists of two equations since there are as many equations in a model as dependent variables. (A single dependent variable is not allowed to have more than one equation. If such a situation occurs, the model can be rewritten so as to avoid this problem. An exception occurs in multilevel models, see Ch. 11). An analysis of the path diagram also indicates, for each dependent variable, how many variables (predictor plus residual) determine the dependent variable. This number is, of course, the number of terms on the right-hand side of an equation, when equations are written with the dependent variable on the left.

In this example, the diagram shows that there are three unidirectional arrows aiming at V3; thus, the equation for V3 must have three terms in it. Similarly, three unidirectional arrows aim at V4, so the equation for V4 has three terms in it as well. The particular variables involved in the equation are simply determined by tracing back from the dependent variable to the originating variables, e.g., the equation for V3 must involve V1, V2, and E3. Furthermore, the free and fixed coefficients in the equation can also be read from the diagram, since an asterisk denotes a free parameter and a fixed parameter has no star. If any such equation had to be represented symbolically, one could substitute a Greek letter for each free parameter. As was seen above, the input to EQS does not use such a symbol but utilizes a guess as to the parameter value. The program output would provide the improved, optimal value.

Any complete job setup also requires knowledge of the independent variables in the model. In a diagram, every variable that does not have a unidirectional arrow aiming at it is thus an independent variable; these are the shaded variables and the residual Es. In Figure 2.3, these are V1, V2, E3, and E4. Thus these variables must have variances that are parameters, and only these independent variables are allowed to covary. In this case, there is only one covariance because, in general, residuals are taken to be uncorrelated with predictor variables. The job setup in EQS, using variable labels instead of Vs, is as follows:

```
/TITLE
 PATH ANALYSIS MODEL
/SPECIFICATIONS
 CAS=932; VAR=6; ME=ML;                          ! ML = maximum likelihood
/LABEL                                           ! Optional
 V1 = ANOMIE67; V2 = POWRLS67;                   ! At most 8 characters per label
 V3 = ANOMIE71; V4 = POWRLS71;
/EQUATIONS
 ANOMIE71 = 1*ANOMIE67 + 1*POWRLS67 + E3;        ! Equations with labeled vars ok
 POWRLS71 = 1*ANOMIE67 + 1*POWRLS67 + E4;        ! No labels possible for Es
/VARIANCES
 ANOMIE67 = 10*; POWRLS67 = 10*;
 E3 – E4 = 2*;                                   ! Dash (-) or TO sets several in sequence
/COVARIANCES
 ANOMIE67,POWRLS67 = 7*;
/MATRIX
 (6 × 6 matrix, above)
/END
```

It will be up to the reader to learn to use a style of model setup that is convenient and meaningful, whether using variable labels or symbolic Vs. Vs are certainly easier to type, but their meaning could be mistaken. Substantively, this model follows the usual EQS setup, but now there are two equations. Any model ever done with EQS similarly will have equations (/EQU) for dependent variables, and /VAR and /COV sections for the independent variables. However, it is possible to have the program generate these sections using a /MODEL command. This gives a third way to specify a model.

A **/MODEL** *paragraph can be used as a substitute for the usual /EQU, /VAR, and /COV specifications.*

```
/MODEL                                          ! MODEL can replace EQUATIONS, VARIANCES and
 (V3,V4) ON (V1,V2) = 1*;                       ! COVARIANCES.  Here, V3 and V4 are dependent,
 VAR (V1,V2) = 10*;                             ! V1 and V2 are independent.  EQS automatically
 VAR (E3,E4) =  2*;                             ! creates error variables E3 and E4, which are
 COV (V2,V1) =  7*;                             ! independent.  See MODEL in Chapter 3.
```

This generates the setup shown above (with Vs, not labels). The statement (V3,V4) ON (V1,V2) = 1*; generates an equation for V3 with its two predictors V1 and V2, and similarly, an equation for V4 with the predictors V1 and V2. V3 and V4 are regressed *on* V1 and V2. The start values are 1.0 and free. The remaining statements specify for /VAR the variances of V1, V2, E3, and E4 (free with certain start values), and the covariance of V2,V1 for /COV.

A model with multiple Vs and multiple equations is often called a *path model*, although nowadays since path diagrams allow generic types of models, path models may include latent variables. Such a model is also called a *simultaneous equation model*, especially when there are reciprocal predictions in a model, e.g., V3 → V4 and also V4 → V3. Such a *nonrecursive model* is hard to estimate and test.

The output from EQS contains a test statistic that can be used to evaluate the adequacy of the model. The ML chi-square is 341.9, which is very large compared to 1 df; the probability value for the chi-square statistic is very small. That is, the model must be rejected. In fact, the model reproduces the observed covariances of all variables perfectly, except that the V4,V3 covariance is not well reproduced: it has a residual $(s_{43} - \hat{\sigma}_{43})$ between data and model of 4.282. The statistical test verifies that this residual is significantly different from zero. In this instance, with only 1 df, a test of the residual provides a test of the model. Typical models have many residuals $(s_{ij} - \hat{\sigma}_{ij})$, but only one model χ^2 test. In the future, EQS also will provide confidence intervals on specific residuals (Ogasawara, 2001a).

Since the covariance of V3 and V4 is not explained well by the model, one possibility to improve the model would be to add a correlation between the corresponding error terms E3 and E4. This recursive model would remain identified since the error correlation is not between a cause and its direct effect (see Brito & Pearl, 2002). But then the model would have no degrees of freedom, i.e., it would not longer be testable. So this is not a good idea.

Corresponding to each estimated free parameter in the estimating equation, the program reports its estimated standard error (S.E.) and prints that value immediately below the estimate. Dividing the estimate by S.E. makes a simple univariate test of whether the estimate is consistent with a population coefficient of zero; the resulting critical ratio is printed below the S.E. An @-sign indicates a *z*-test that is significant at the .05 probability level (critical ratio larger that 1.96 in absolute value). In this instance,

```
ANOMIE71=V3  =    .455*V1    +  .206*V2    + 1.000 E3
                  .037          .041
                12.407@       4.988@

POWRLS71=V4  =    .158*V1    +  .420*V2    + 1.000 E4
                  .034          .038
                 4.658@       11.042@
```

All the parameters above appear to be significantly different from zero. However, these tests should be considered suggestive rather than conclusive because they are associated with a model that does not fit the data (also, see the *caution* below). As in ordinary regression, the values of the coefficients may be difficult to interpret when the variables have different and relatively arbitrary scales, and the standardized solution may be of interest:

```
  STANDARDIZED SOLUTION:                                          R-SQUARED

ANOMIE71=V3  =    .443*V1    +  .178*V2    +  .818 E3               .331
POWRLS71=V4  =    .172*V1    +  .407*V2    +  .844 E4               .287
```

It appears as if anomie is slightly more stable than powerlessness (.443 vs. .407). The cross-lagged effects are small and about equal in magnitude (.172 and .178). A diagram of the standardized solution is:

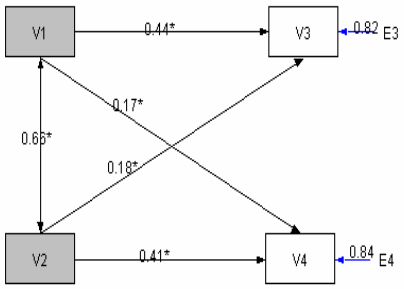


Figure 2.4

In EQS 6 for Windows, the standardized solution is inserted directly into the diagram as shown above. See the relevant section of the *EQS 6 for Windows User's Guide*.

Key features of this model, of course, extend to more complex path analysis models. A recent illustration involving the influences of alcohol and illicit drug use on each other across three time periods is given by Bentler, Newcomb and Zimmerman (2002). Another example is the reciprocal effects model of Marsh, Hau, and Kong (2002), addressing the effects of academic self-concept and achievement on each other across five time periods.

A Factor Analysis Model

Although a major application of structural modeling involves latent variables (LVs), the previous examples have not used any latent variables. Actually, exactly what a "latent variable" might be is not universally agreed upon. Bollen (2002) discusses four different definitions of the concept of latent variables – local independence, expected value, sample realization, and the viewpoint we espouse, due to Bentler (1982), that LVs are not deterministic functions of observed variables. In some approaches, residuals are latent variables, but in the approach used here that is not necessarily true. While residuals in variables of the regression and path models are unmeasured variables, in principle, once the coefficients in the equations are known (e.g., in the population), these residuals are simply linear combinations of observed variables. Hence, by our approach they are not latent variables. *In LV models, the latent variables cannot be expressed as a linear combination of measured variables.* The reason for this is that *the dimensionality of the space of latent variables exceeds the dimensionality of the space of measured variables.* Note that in the regression example there were three measured variables and three independent variables. In the path analysis example there were four measured variables and four independent variables. In true LV models, there will be more independent variables than measured variables.

Factor analysis is one of the oldest latent variable models. In its early form it was used as an exploratory data analysis technique, and it remains important in finding the LV dimensionality of data (e.g., Bentler & Yuan, 1996, 1998; Velicer & Fava, 1998; Yuan & Bentler, 2000b; Yuan, Fung, & Reise, in press). More recently, it has been developed as a hypothesis-testing tool, primarily due to the work of Jöreskog (1969). See, e.g., Hoyle (2000) and Russell (2002) for an overview, and Hayashi and Bentler (2000ab), Hayashi and Yuan (in press), Krijnen (2002), Shapiro and ten Berge (2002), and Yuan, Marshall, and Bentler (2002) for recent developments. In the exploratory factor analysis approach, the analysis is concerned with finding the smallest number of common factors to account for the correlations or covariances between variables. In the confirmatory approach, the analysis is concerned with implementing a theorist's hypothesis about how a domain of variables may be structured, and testing the adequacy of the hypothesis using statistical means. For maximum meaningfulness, such hypotheses should be generated from a model building enterprise; see, e.g., Dishion and Patterson (1999) on model development in developmental psychopathology. In practice, of course, structural theories are not perfectly clear-cut, and a variety of alternative model specifications may be evaluated. Although the structural modeling approach to factor analysis includes exploratory factor analysis, in practice specialized programs are superior in that domain because they are able to handle larger data sets and also include features such as orthogonal and oblique rotations that are typically not included in structural modeling programs. EQS for Windows contains an exploratory factor analysis front end based on the

equal unique variance model for standardized variables described in Hayashi and Bentler (2000a), and includes several rotation methods. EQS permits taking the output of such an analysis to automatically set up a confirmatory factor model.

The sociologists who provided the data for the stability of alienation example proposed that anomie and powerlessness may both be indicators of the same underlying construct, say, "alienation." The implication of such a concept is that the observed variables are correlated only to the extent that they share this underlying construct. Diagrammatically, considering only the 1967 measurement point, this hypothesis can be represented as follows:

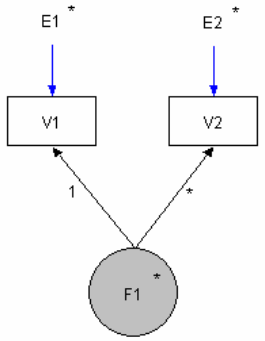


Figure 2.5

In this diagram, there are only two measured variables, V1 and V2. Notice from the direction of the arrows that both of *these variables are dependent variables, i.e., they are a linear combination of the latent variables, and not the reverse.* The variable F1, "alienation," is the common factor latent variable that generates any covariance between V1 and V2, while E1 and E2 are so-called unique factor variables. Factors such as F1 are sometimes called *constructs* and are usually shown in circles in a path diagram. The shaded variable F1 and the residual variables E1 and E2 are the independent variables in the model: there are three uncorrelated independent unmeasured variables, but fewer measured variables (only two), as is true by definition of a true latent variable model (Bentler, 1982). Although principal components analysis and factor analysis become equivalent when the number of indicators of a factor gets very large (Bentler & Kano, 1990; Ogasawara, 2000), with the usual few indicators per factor these are not equivalent models and the component loadings can be misleading (Sato & Ito, 2003). In fact, principal components are linear components of measured variables, while factors are not. See also Yuan and Bentler (1994).

An interesting issue is whether factors can ever be considered as being generated by measured variables, that is, whether the arrows in Fig. 2.5 could be reversed, to go from V1 and V2, to F1. The basic answer is "no". Factors need indicators (variables they point to, in a diagram), otherwise they do not exist. However, there are some larger model structures where variables indeed can predict factors. For example, adding a variable, say V20, that predicts F1 (V20 → F1) in Fig. 2.5 would be perfectly all right. A discussion of some of the issues can be found in Bollen and Lennox (1991), Bollen and Ting (2000), Edwards and Bagozzi (2000), Kenny (1979), and MacCallum and Browne (1993).

In the diagram each of the latent variables has its scale fixed by fixing a path from the variable to another variable. As before, the error variables E1 and E2 are taken to have unit paths aiming at the measured variables, and their variances are taken to be free parameters to be estimated (thus the star next to the variable). The common factor F1 also has its scale fixed by such a device; one path emanating from F1 is fixed, the other is free to be estimated. It is arbitrary which choice is taken. When the path is fixed, the variance must be a free parameter as noted previously (hence the star next to F1). Thus the diagram implies that there are four free parameters. But a quick count as to the data involved, namely the two variances of V1 and V2, and their covariance, indicates that there are three data points. Thus the model has more parameters than data, and cannot be uniquely estimated. It is *under-identified*.

The identification problem in this small factor model with one common factor could be resolved in several ways. Constraints could be placed on the parameters, so that there would be fewer than four effective parameters. For example, the residual variances of E1 and E2 could be set equal to each other, so that in this part of the model only one free parameter, rather than two, would be involved. A much better general solution is to have more indicators of

the factor. For example, one could entertain the model in Figure 2.6, where as usual, the independent factor F1 is shaded. The measured variables V1-V4 are not shaded, i.e., they are dependent variables.

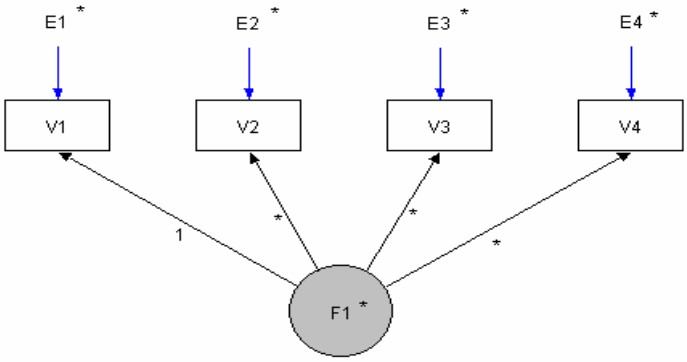


Figure 2.6

The single common factor F1 is presumed to generate variables V1-V4, and hence is the source of their correlations or covariances. In this model there are 4(5)/2 = 10 data points and only eight parameters, and it can be shown that the model is identified. In this instance, however, such a model may not make sense, since variables V1 and V2 represent data obtained in 1967, while V3 and V4 represent data obtained in 1971. While anomie and powerlessness may indeed represent indicators of alienation, the above model forces the construct of alienation (F1) to be identical in 1971 and 1967. Perhaps alienation is not this stable -- perfectly stable -- and a more realistic model is one with two factors where the original model for the 1967 data (Figure 2.5) is essentially replicated for the 1971 data, with a single factor underlying each measurement occasion, and the factors are allowed to covary. This idea is shown in Figure 2.7.

Although it appears in Figure 2.7 that factors F1 and F2 have only two indicators, and hence that the combined model should be under-identified because parts of the model are under-identified, the covariance between factors F1 and F2 acts like another indicator of each factor and serves to identify the model. This concept may not be obvious, but it might be noted that the covariance F2,F1 could be replaced by a unidirectional arrow from F1 to F2. Then F2 would act like V1 and V2 in identifying F1. However, a model such as is proposed above remains very fragile, because if the covariance F2,F1 is in fact precisely zero (which is one of the values that such a covariance can take) the model again would be under-identified; that is, the left and right parts of Figure 2.7 would suffer from the under-

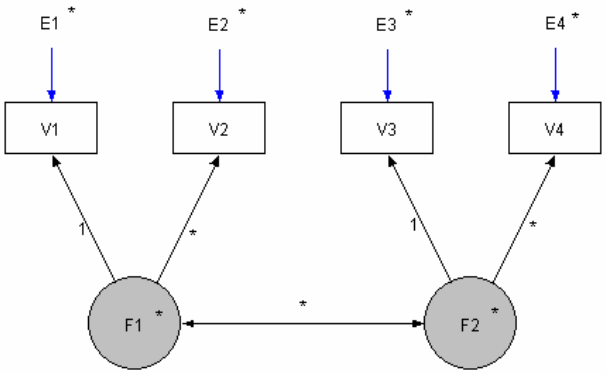


Figure 2.7

identification problem noted in Figure 2.5. This type of situation can arise in practice, when the researcher specifies a model such as Figure 2.7 and expects it to be identified because of the covariance term F2,F1. But when estimated from the data, the covariance may turn out to be very close to zero, leading to a situation called *empirical under-identification* (Kenny, 1979; Rindskopf, 1984a). In such a case EQS will indicate that *a parameter is linearly*

dependent on others, i.e., that there is a problem in obtaining unique estimates for all the parameters in the model. A good discussion of causes of linear dependencies and other anomalies is given by Chen, Bollen, Paxton, Curran, and Kirby (2001).

It should be clear by now that the two-factor model can be re-parameterized by using an alternative convention for fixing the scale of the unmeasured variables. In particular, factors F1 and F2 could have their variances fixed at 1.0 and the fixed path from each factor to its measured variable indicator could be set free to be estimated. In that case the covariance F2,F1 can be interpreted as a correlation. Thus the model of Figure 2.7 is equivalent to the following:

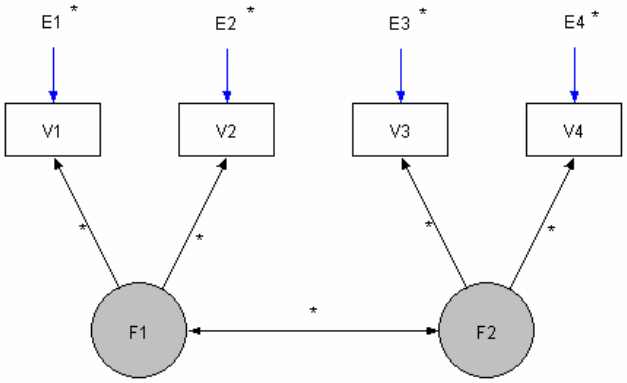


Figure 2.8

There is no star next to F1 and F2 because the variances of F1 and F2 are fixed at 1.0. This is the standard representation for a confirmatory factor analysis model, in which the common factors are usually standardized to have unit variance. This model can be estimated for the data by the following EQS setup, using the /MODEL command:

```
/TITLE
   FACTOR ANALYSIS MODEL
/SPECIFICATIONS
   CAS=932; VAR=6; ME=GLS;                    ! GLS = generalized LS
/MODEL
 (V1,V2) ON F1 = 2*;
 (V3,V4) ON F2 = 2*;
 VAR F1,F2=1;                                 ! Fixed at 1.0 (no star)
 VAR E1-E4=3*;                                ! E1-E4 specifies all 4
 COV (F2,F1)=.3*;                             ! Dangerous near 0*
/MATRIX
 (6 × 6 matrix, above)
/END
```

The program obtained a chi-square statistic T_{GLS} = 55.612 with 1 df. The value of 55.6 evaluated in a χ^2_1 table yields a probability less than .00001, indicating that if the model were correct in the population then sample data of the sort obtained in the study would be extremely unlikely to be observed. At this point, there is the option of concluding (a) the model is correct, but the sample is a very unusual one or (b) the model is incorrect. In the absence of a compelling reason to reject the sample, it is more prudent to reject the model. Actually, as you will see in Ch. 5, alternative (a) can be modified to (a') the model is correct, and the sample is appropriate, but one or more assumptions associated with the statistical method are violated, and hence the test statistic cannot be trusted. This typically occurs when nonnormal data are used with a normal theory statistical method such as ML or GLS. Here we will assume that using a normal theory method is correct – though without the raw data, we cannot test the multivariate normality of the data.

In Figure 2.8, there is no *direct* effect of F1 on V3 and V4, nor is there a direct effect of F2 on V1 and V2. These paths have been fixed to zero. This does not necessarily imply, however, that – under the model -- F1 will be

uncorrelated with V3 or V4, or that F2 will be correlated 0.0 with V1 and V2. When the factors F1 and F2 are correlated, typically these "structure coefficients" will be nonzero, and this should be kept in mind in interpreting the meaning of F1 and F2 (Bentler & Yuan, 2000). A detailed discussion of this issue is given by Graham, Guthrie, and Thompson (2003). These correlations are not normally printed out as a program default, but can be obtained with the /PRINT command along with CORRELATION=YES, which instructs EQS to print the entire matrix of model-reproduced correlations of Vs and Fs. See Chapter 3 for the /PRINT command.

It would be desirable to modify the model in some reasonable fashion to determine if an alternative model might not fit the data. Unfortunately, it is impossible to add parameters to the model without running out of degrees of freedom. Thus, for example, it would not be possible simply to add paths missing in the model, e.g., by allowing a complex rather than unifactorial factor structure. However, it is possible to add parameters to the model if, at the same time, constraints are added to the model to reduce the number of effective parameters.

If the same variables are measured on two or more occasions, the possibility of *correlated errors* should be considered. It was noted that the "errors" in a factor analytic model are really "unique factors," that is, they may contain specific but reliable variance that is not to be shared with other variables. This specificity may be similar in the same variable on other occasions, and thus the residuals may be correlated across time. In particular, E1 and E3, as well as E2 and E4, may covary. The following line can be added to the /MODEL section:

```
(E1,E2) PCOV (E3,E4) = .2*;      ! (E1,E3) and (E2,E4) are free parameters, starting at 0.2
```

The PCOV command pairs the variables on left and right, in sequence, here E1 with E3, and E2 with E4. Here, adding these two parameters yields -1 degrees of freedom; this model is under-identified and cannot be tested. Therefore, it is important to reflect further on the design of the study. In particular, since it is hoped that the same factor of alienation underlies the measures on both occasions of measurement, and the same variables are used as indicators of the factor, it may be hypothesized that the factor structure is actually the same on both occasions. The most restricted form of such factorial invariance would be to say that the relevant factor loadings are the same, that the relevant error variances are the same, and, perhaps the strongest hypothesis of all, that the covariances of errors are equal in magnitude (this is the strongest statement because the former hypotheses involve the same measured variable, but the latter involves different variables). These hypotheses would be specified as follows in EQS:

```
/CONSTRAINTS                              ! New program section
 (V1,F1) = (V3,F2);                       ! Two factor loadings equal
 (V2,F1) = (V4,F2);                       ! Other loadings equal
 (E1,E1) = (E3,E3);                       ! Error variances equal
 (E2,E2) = (E4,E4);                       ! Other error variances equal
 (E3,E1) = (E4,E2);                       ! Error covariances equal
```

The program uses a convention that allows each parameter in any model to be uniquely identified by the *double-label* of the pairs of variables involved in the specification. The parameter in any equation can be simply found by the pair (dependent variable, predictor variable). For example, in the equation V1 = 2*F1 + E1, the parameter 2* is clearly identified by (V1,F1) because V1 defines the equation, and F1 the relevant part of the right-hand side; similarly, (V1,E1) would define the fixed 1.0 implicit in the equation. Any parameter in the variances section can be identified by the variable involved, and considering it as the double-labeled diagonal element of a matrix. That is, for the E1 variable, (E1,E1) indicates its variance as a parameter. Any parameter in the covariance section is already defined by its standard double-label name, e.g., (E3,E1) is the covariance between E3 and E1. The /CONSTRAINT section can impose constraints only on free parameters. After all, if one wants to specify that the parameter (E2,E2) = 3, say, this could be done directly in the variance section and there would be no need to add a /CONSTRAINT section. Finally, it may be noted that the parameters are placed in parentheses in this section.

The estimated free parameters, standard errors, and critical ratios are given in Fig. 2.9. This restricted model has 11 parameters but 5 constraints. The degrees of freedom are 10 - 11 + 5 = 4. T_{GLS} = 2.958, giving a χ^2 probability value of .56. Evidently, the model fits the data very well. It will be noted that parameter estimates that are supposed to be equal, in fact *are* equal. Of course equality-constrained estimates have the same standard error as well.

Equations	Variances	Covariances
V1 = 2.937*F1 + 1.000 E1 .090 32.815@ V2 = 2.446*F1 + 1.000 E2 .081 30.262@ V3 = 2.937*F2 + 1.000 E3 .090 32.815@ V4 = 2.446*F2 + 1.000 E4 .081 30.262@	E1 = 3.520* .267 13.178@ E2 = 3.644* .206 17.715@ E3 = 3.520* .267 13.178@ E4 = 3.644* .206 17.715@	F2,F1 = .685* .026 26.574@ E3,E1 = .901* .122 7.415@ E4,E2 = .901* .122 7.415@

Figure 2.9

The factor correlation estimated by EQS is .685, with a very small S.E. of .026, and a critical ratio of 26.574. Thus, it is unlikely that the true factor correlation between F1 and F2 is 1.0. A model re-specification can be done and the model re-estimated under the restriction. The entire model is kept intact, except that in the /COV command of the MODEL section, .3* is replaced by 1.0, so that the factor correlation is fixed at 1.0 rather than free to be estimated. The associated GLS χ^2 = 162.292 based on 5 df, which clearly does not fit the data. Thus, this very restricted model, having two factors that are perfectly correlated, must be rejected.

When two models are *nested*, that is, when one model is a special case of the other, *chi-square difference tests* can be used to evaluate the structural importance of the parametric constraints that differentiate them. In the simplest and most typical application, two models would differ in that one model would contain extra parameters beyond those provided by the other model; all other parameters would be the same. In such a case, the chi-square difference test evaluates whether the added parameters, considered simultaneously, are necessary to the model. In the particular comparison given above, the two models are nested because the value of 1.0 for the factor intercorrelation is one special value that the parameter can take when it is simply a free parameter. Thus a chi-square difference test would evaluate whether the restricted value is a significantly poorer estimate for that parameter than the optimal freely estimated value. In this case this test of the restriction that differentiates the two models is given by 162.292 - 2.958 = 159.334, with degrees of freedom also given by the difference 5 - 4 = 1. Clearly, the hypothesis that (F2,F1) = 1.0 must be rejected. As noted above, chi-square difference tests are particularly useful when sets of parametric restrictions are tested simultaneously.

Caution: Although the z-test is distributed under the null hypothesis in large samples as a unit normal variable (mean zero, variance one), in small samples and especially under alternative hypotheses, its variance may not be 1.0 (Ogasawara, 2002) so that the assumed critical values may be off. Also, recent research (Gonzalez & Griffin, 2001) indicates that the *z*-test on a particular parameter (especially, latent variable relations such as correlations or regressions among factors) may not be as trustworthy as the χ^2 difference test, because the *z*-test value can depend on how the latent variables in a model are identified (fixed variance vs. fixed scale). This problem does not occur with the difference test. *If the significance level based on* z *is marginal, you should follow up with a* χ^2 *difference test.* This precaution is not needed in large samples.

Parameters that differentiate two nested models also can be evaluated for significance by two other procedures called *LM-tests* and *W-tests*. These tests are described in Chapter 6.

Since the model comparison yielded the conclusion that the (F2,F1) correlation is not 1.0, it is apparent that alienation is not perfectly stable across the four years between measurement occasions. The estimated correlation of .685 suggests that $.685^2$ = .469 is the proportion of the variance of F2 that is predictable from F1. However, this evaluation of the stability of alienation has not simultaneously considered the effects of SES. These effects are considered in the next section, in which a complete model for the data is evaluated.

The path diagram produced by the program for the model with correlated errors is given in Figure 2.10. This partial screen capture shows menu options (on top) and diagramming tools (left side) in the EQS for Windows *Diagrammer* that let you customize the diagram in various ways, add a statistical summary of the model, and so on.

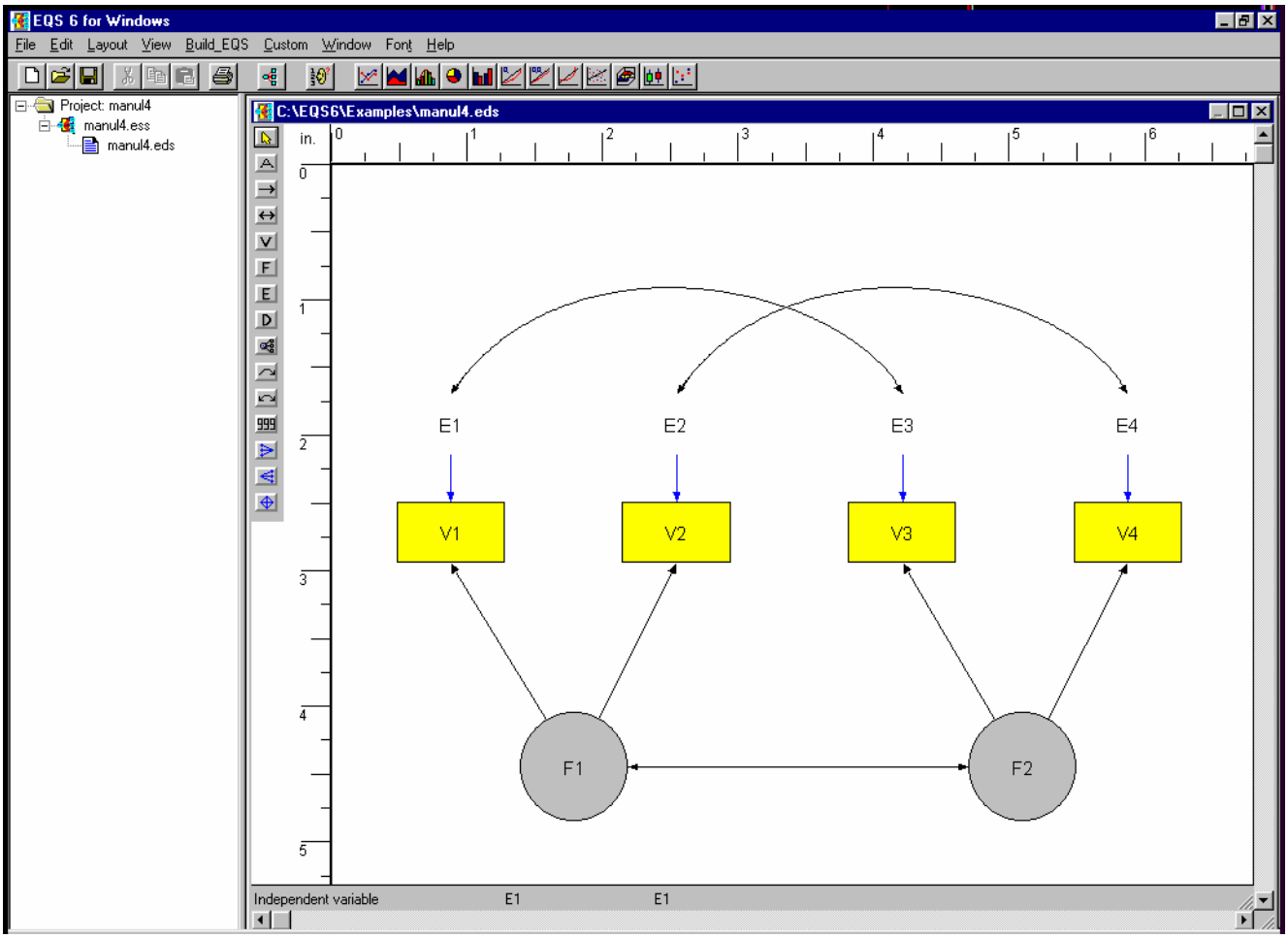


Figure 2.10

A Complete Latent Variable Model

Although factor analysis uses the concept of latent variables, it does not allow a regression structure on the latent variables. Factors are either correlated or not. The importance of allowing a simultaneous regression among factors was noted by Jöreskog (1973), Keesling, and Wiley (Keesling, 1972; Wiley, 1973) when they joined the *simultaneous equation* and *factor analytic* models to yield a single model. This is often called the LISREL model after the name of the computer program, but it is, in fact, the **f**actor **a**nalytic **s**imultaneous **e**quation **m**odel (FASEM, Bentler, 1986b). The concept is straightforward: allow latent variables to be connected by unidirectional arrows rather than two-way arrows, with the directionality reflecting a hypothesis about the process or causal structure of the variables.

It is easy to see how the previous factor models would be modified to encompass a regression structure on the factors. Since F1 and F2 represent the construct of alienation at time points 1967 and 1971, it would make sense to replace the covariance between F1 and F2 by the structural equation F2 = *F1 + D2, where * is a free parameter and

D2 is a disturbance in the factor. Then, F2 would be determined by F1 plus some residual, and * represents the extent of stability of the construct. If D2 were zero, F2 would be a constant times F1, and hence the construct would be totally stable. If * were zero, F2 would be an entirely new construct, unrelated to the prior F1 construct. Intermediate values of * would represent varying degrees of stability.

In the introductory section of this chapter it was noted that the stability of alienation might, however, be influenced by levels of education and SEI in the respondents. Whenever a theorist has reason to suspect that a structural relation might be impacted by other variables, it is important to try to include those variables in a simultaneous model of a phenomenon. In fact, biased estimates of the stability of alienation might be obtained if such stability were not evaluated in the context of socioeconomic status (SES). Since SES is a variable on which the respondents presumably have not varied substantially before 1967 (e.g., educational attainment expressed as years of school completed was probably the same before the attitudes were measured), the SES variables were considered to influence alienation in 1967 as well as in 1971. Furthermore, the education and occupational status measures V5 and V6 could reasonably be interpreted as indicators of SES, leading to the overall model in Figure 2.11.

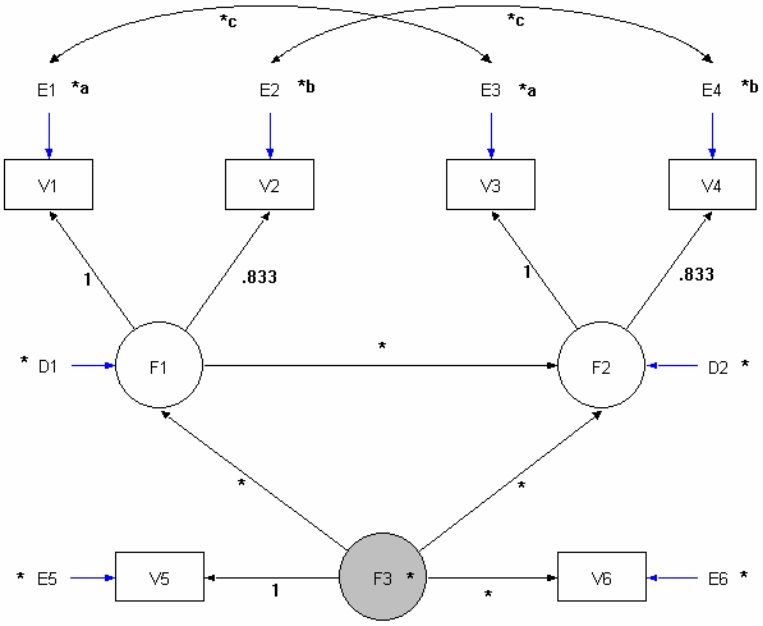


Figure 2.11
Note: Parameters marked “a”, “b”, “c” are equal

As can be seen, the upper part of the model is intended to be similar to the factor analytic models considered above, but the factor correlation is replaced by a unidirectional arrow representing the hypothesized across-time regression relation. In addition, the measured variables V5 and V6 are taken as indicators of the SES factor F3, which is hypothesized to affect alienation in 1967 (F1) as well as alienation in 1971 (F2). *Whenever a variable such as F1 or F2 changes its status from an independent variable to a dependent variable, a residual is added on the assumption that the variable cannot be perfectly predicted.* The disturbances are given in the diagram by D1 and D2.

As the diagram shows, F3 has its scale fixed by the unit path to V5; thus, the variance of the factor is taken to be a free parameter to be estimated. The factors F1 and F2, now dependent variables, cannot have their variances fixed, so the basic choice for fixing such factors is taken: let one path from the factor to a variable be fixed at 1.0. While there can be problems with this approach in complicated modeling situations (Steiger, 2002), it usually works correctly. Each of the error variables has its scale fixed by a unit path from the error variable to the corresponding measured variable. The disturbances D1 and D2 similarly have a path fixed at 1.0, so their variances become parameters to be estimated.

The model of Figure 2.11 shares many features of standard latent variable models. However, it also is intended to maintain most of the characteristics of the highly restricted factor analytic model considered in the previous section. In particular, the model constrains the error variances and covariances as before. These restrictions are shown in the diagram by the letters next to the stars; free parameters with the same letter are constrained to be equal. The model contains one additional restriction that will not be obvious: the paths F1 → V2 and F2 → V4 are fixed rather than free, and the parameters are set at the value .833. The very restricted hypothesis that these paths should equal the given value in this combined model was developed from the previous factor analytic results. In the factor analytic model these two paths were constrained equal, but they were free to be estimated. In that model, the estimated value of the F1 → V2 path was observed to be 2.446, while the F1 → V1 path was 2.937. Thus the F1 → V2 path was .833 as large as the value of the F1 → V1 path. In Figure 2.11, the F1 → V1 path is 1.0, and hence .833 times that value is .833. Thus, the model that is being tested includes the hypothesis that the ratio of factor loadings for the indicators of the F1 and F2 factors found in the four-variable factor analysis problem also holds in the larger problem including variables V5 and V6 and the associated factor and regression structure.

Although data-based model re-specifications are widely practiced in structural modeling, the legitimacy of imposing a constraint in one model based on the outcome of an analysis of a previous model on the same data raises some interesting statistical questions that cannot be pursued here (see e.g., Chou & Bentler, 2002; MacCallum, 1986; MacCallum, Roznowski, & Necowitz, 1992; Silvia & MacCallum, 1988). What is important here is the idea that any structural model requires a careful analysis of the parameters of the model, as well as their status as fixed, free, or constrained parameters. If the rationale for a constraint or fixed value is appropriate, it can be implemented in EQS. *Typically, however, one would not want to fix factor loadings at a nonzero value as was illustrated above,* especially if the variables' scales or variances did not have a well-justified rationale. If the scales of the measured variables are in any way arbitrary, one should allow factor loadings to be freely estimated parameters, except for those factor loadings that are fixed for identification purposes. See Cudeck (1989) for a discussion. In any case, the reader is urged to ponder how one might re-specify the current model so that the paths {F1 → V1} = {F2 → V3} and {F1 → V2} = {F2 → V4} are both free, precisely as in the restricted factor analytic model considered previously. The model of Figure 2.11 is specified as follows:

```
/SPE
 CAS=932; VAR=6; ME=ML;                          ! Maximum likelihood
/EQU
 V1 =     F1       + E1;
 V2 = .833F1       + E2;
 V3 =           F2 + E3;
 V4 =       .833F2 + E4;
 V5 =               F3 + E5;
 V6 =            .5*F3 + E6;
 F1 =           -.5*F3 +    D1;
 F2 =  .5*F1    -.5*F3 +    D2;
/VAR
 F3 = 6*;
 E1 – E6 = 3*; D1 – D2 = 4*;
/COV
 E1,E3 = .2*; E2,E4 = .2*;
/CON
 (E1,E1) = (E3,E3);                              ! "a" in Figure 2.11
 (E2,E2) = (E4,E4);                              ! "b" in Figure 2.11
 (E3,E1) = (E4,E2);                              ! "c" in Figure 2.11
/MAT
 (6 × 6 matrix, above)
/END
```

Figure 2.12

Above, spaces are introduced to make the model structure clearer. Spaces are not needed. But with the layout, it is immediately clear that there are 3 factors, that each observed V is influenced by only one F, and that it is possible to think of V5, V6, F1, and F2 as "indicators" of F3. The first set of 6 equations are the *measurement model*, relating observed to latent variables; while the equations for F1 and F2 are the *construct model*, or the *simultaneous equation model*, relating the latent variables to their predictors. A problem with this terminology is that, for example, one could add a V (say, V7 if we had it) as a predictor of V1-V6, and the equations no longer specify only a

measurement model; and one could add a V (say, V8) as a predictor of the factors F1 and F2, and the equations would no longer just relate constructs to each other.

As before, /EQU, /VAR, and /COV can be replaced with a /MODEL command. This is written as follows:

```
/MOD
 (V1) ON F1 = 1; (V2) ON F1 = .833;            ! .833 is fixed - no star
 (V3) ON F2 = 1; (V4) ON F2 = .833;
 (V5) ON F3 = 1; (V6) ON F3 = .5*;
 (F2) ON F1 = .5*;
 (F1,F2) ON F3 = -.5*;
 VAR D1,D2 = 4*;
 VAR E1-E6 = 3*;
 VAR F3    = 6*;
 (E1,E2) PCOV (E3,E4) = .2*;
```

It should be appreciated by now that the basic approach to setting up a model in EQS is the same regardless of whether path models, factor models, or more generic models are entertained. The basic requirement of any specification is information on the equations of the model and on the variances and covariances of independent variables. The definition of independent and dependent variables should be clear by now. Notice, for example, that F1 is on the right-hand side of some *ON*-commands but also on the left-hand side of one. Since F1 is once on the left, it is a dependent variable, and hence its variance is not a parameter to be specified. F3, in contrast, is never on the left-hand side and is thus an independent variable that must have a variance specification.

EQS adjusts the initial user-provided parameter estimates iteratively in the process of minimizing function Q in (1.5). The above input leads to the following iterative sequence, copied from the EQS output:

ITERATION	PARAMETER ABS CHANGE	ALPHA	FUNCTION
1	0.533285	1.00000	0.01598
2	0.040242	1.00000	0.01449
3	0.002323	1.00000	0.01448
4	0.000134	1.00000	0.01448

The program took four iterations to converge to an optimal set of parameter estimates. The second column gives the average change in parameters from iteration to iteration. The average is based on the absolute differences between individual parameters from iteration to iteration. It can be seen that by the fourth iteration, the parameters were changing only in the fourth decimal place. The function value is decreasing systematically as it should if the process is converging well. In all methods of estimation except ML, the printed function is the GLS function Q in (1.5) as specialized appropriately. In the case of ML estimation considered here, the function is a constant times the likelihood ratio statistic used for model evaluation. In almost all cases, the printed function times (N - 1), where N is the sample size, yields a test statistic that in large samples can be interpreted as a χ^2 statistic. Thus, $T_{ML} = 931 \times .01448 = 13.485$, which, based on 9 df yields a χ^2 probability value of .14186. This value is larger than the standard .05 cutoff probability used for model rejection. The model can be considered to fit the data. (The program also prints T_{RLS} based on Q, which is 13.280 and yields the same conclusion.)

Five of the eight equations of the model involve only fixed parameters, so their final values are the same as the initial values. The remaining equations, variance and covariance estimates, S.E.s, and critical ratios are:

Equations	Variances	Covariances
V6 = .537*F3 + 1.000 E6 .043 12.379@ F1 = -.630*F3 + 1.000 D1 .056 -11.181@ F2 = .593*F1 - .241*F3 + 1.000 D2 .047 .055 12.679@ -4.388@	F3 = 6.616* .639 10.352@ E1 = E3 = 3.608* .201 17.957@ E2 = E4 = 3.595* .164 21.856@ E5 = 2.994* .499 6.004@ E6 = 2.596* .183 14.176@ D1 = 5.670* .423 13.405@ D2 = 4.515* .335 13.464@	E3,E1 = E4,E2 = .906* .122 7.445@

The @-signs indicate that all parameters are significantly different from zero by the *z*-test on the critical ratio. SES (F3) more strongly affects F1 than F2, i.e., the contemporaneous effect is stronger than the lagged effect. These direct effects are more interpretable in the standardized solution

```
STANDARDIZED SOLUTION:                                          R-SQUARED

V1   =V1  =   .835 F1    + .551 E1                                .697
V2   =V2  =   .785 F1    + .620 E2                                .616
V3   =V3  =   .845 F2    + .535 E3                                .714
V4   =V4  =   .797 F2    + .604 E4                                .635
V5   =V5  =   .830 F3    + .558 E5                                .688
V6   =V6  =   .651*F3    + .759 E6                                .424
F1   =F1  =  -.563*F3    + .827 D1                                .316
F2   =F2  =   .569*F1    - .206*F3    + .708 D2                   .499,
```

with correlations E3,E1 = .251* and E4,E2 = .252*. The direct effect of SES on alienation remains most strongly negative contemporaneously (-.563 vs. -.206). As seen in the structural coefficient of .569, alienation remains quite stable even when SES is controlled. However, since R-squared = .499, only half of the variance in 1971 alienation is accounted for by 1967 alienation and SES. Substantial instability exists in this attitudinal construct. As compared to the factor model, only a trivial increase in variance explained in F2 was achieved by taking SES into account. Finally, it might be noted that *parameters that are constrained to be equal in the statistical metric do not necessarily remain equal in the standardized metric*. This can be seen, for example, in the standardized coefficients for the error variables E1 and E3. This argues for the usefulness of standard errors for standardized solution parameters. These will shortly be available in EQS (Jamshidian & Bentler, 2000).

In the above example, a specific rationale was given for proposing that the equations V2 = .833F1 + E2 and V4 = .833F2 + E4 should have a coefficient of .833 that is a fixed parameter. This rationale was based on the equations output in Fig. 2.9 from the factor analytic model, regarding the loadings of V1 and V2 on F1, giving the ratio of 2.446/2.937. It can be argued that creating a specific fixed parametric constraint on the basis of fallible data is an unnecessarily restrictive procedure in general. An approach in the spirit of the constraint that is liable to be more broadly applicable would be to use the results of one analysis to suggest bounds on parameters rather than specific values. In particular, the standard errors of the factor analytic solution can also be used to provide something akin to a *stochastic constraint* rather than fixed constraint. In Figure 2.9, the value of 2.446 is associated with an estimated standard error of .081. Under the usual assumptions, 2.446 ± .081 is quite likely (with $p \cong .68$) to contain the true parameter (or, one could use bounds of 2 or 3 S.E.s, depending on the certainty one might desire). Thus, the fixed

value of 2.446 could be replaced by lower and upper bounds of 2.365 and 2.527. In that case, it may be desirable to bound the (V2,F1) and (V4,F2) parameters between the values 2.365/2.937 = .805 and 2.527/2.937 = .860 rather than fixing the parameters at .833 in the new, larger model. Such an effect can be implemented quite easily in EQS using the /INEQUALITIES section that permits the direct implementation of inequalities that are otherwise somewhat difficult to specify (McDonald, 1980; Rindskopf, 1983, 1984b). The input of Figure 2.12 would be modified as follows: First, ".833" is replaced by ".833*" in the first two ON commands, and the following section is added:

```
/INEQUALITIES
 (V2,F1) GE .805, LE .860;                    ! GE means greater than or equal
 (V4,F2) GE .805, LE .860;                    ! LE means less than or  equal
```

This model specification has two more free parameters than the previous one, so that there are 7 df rather than 9 df. EQS yielded $\chi^2 = 12.656$, p = .081, implying a statistically acceptable model. The equations were estimated as

```
  V2   =V2  =     .860*F1    +  1.000 E2
                  .027
                31.559@
  V4   =V4  =     .855*F2    +  1.000 E4
                  .033
                26.033@
```

The (V2,F1) parameter was actually held to the boundary. The (V4,F2) parameter was inside the bounds and thus the value is not restricted. The rest of the solution is virtually identical to that previously described, so it is not presented.

A Second-Order Factor Analysis Model

The literature discusses structural models in which factors are further decomposed into other factors (e.g., Bentler, 1976; Bentler & Weeks, 1980; Bollen, 1989b; Byrne, 1994; Duncan, Duncan, Biglan, & Ary, 1998; Hancock, Kuo, & Lawrence, 2001; Rindskopf & Rose, 1988; Sayer & Cumsille, 2001; Wothke & Browne, 1990). In the ordinary factor model, there are a number of factors that are arbitrarily correlated, as in the model of Figure 2.8 where there are two factors F1 and F2; these factors are independent variables in the model. Such factors, which are only one unidirectional arrow away from measured variable indicators, are usually called *first-order* factors. It may be desired to further analyze the intercorrelations among these first-order factors so as to yield *higher-order* factors. The number of higher-order factors, and the levels of order that one could use, depend on the number of first-order factors. With two first-order factors, as in Figure 2.8, at most one higher-order factor could be obtained, and, even then, only with appropriate restrictions. (See the discussion surrounding Figure 2.5.) This higher-order factor is thus a second-order factor. It is two levels away from the measured variables. The diagram below presents a second-order model for the variables in Figure 2.8.

The top part of the figure is a measurement model expressing the variables V1-V4 in terms of factors F1 and F2, plus residual errors. In contrast to Figure 2.8, however, factors F1 and F2 are no longer correlated, and, as dependent variables, they are now drawn unshaded and have residual variables aiming at them. The correlation is no longer a parameter of the model since F1 and F2 are now dependent variables, and such variables cannot have variances or covariance as parameters. Any correlation or covariance between F1 and F2 is explained by the second-order factor F3, which generates F1 and F2 (with imprecision, as indicated by the disturbances D1 and D2). The bottom part of the figure is really of the form of Figure 2.5, except that measured variables to be explained have been replaced by first-order factors to be explained. In order to identify the model, the two second-order factor loadings are constrained to be equal, as indicated by the letter "a". The variance of factor F3, an independent variable, is fixed at 1 for reasons of identification.

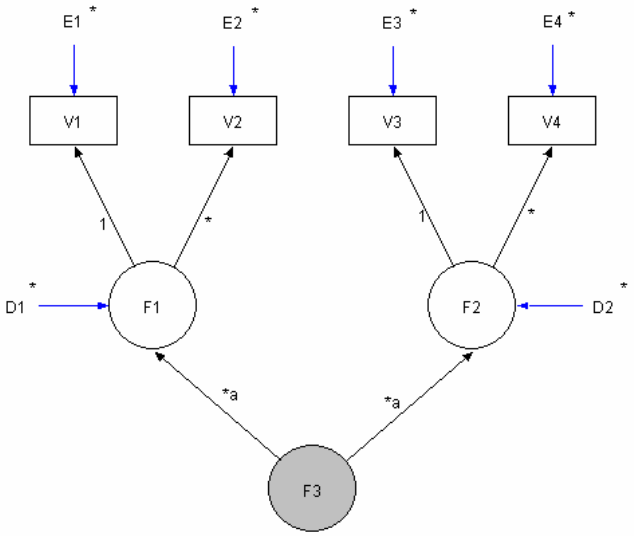


Figure 2.13
Note: Parameters Marked “a" Are Equal

The EQS setup for this model is as follows:

```
/TITLE
   A SECOND-ORDER FACTOR ANALYSIS MODEL
/SPECIFICATIONS
   CAS=932; VAR=6; ME=GLS;
/EQUATIONS
 V1 =   F1 + E1;                          ! F1 and F2 are 1st-order factors
 V2 = 2*F1 + E2;
 V3 =   F2 + E3;
 V4 = 2*F2 + E4;
 F1 = 1*F3 + D1;                          ! F3 is the 2nd-order factor
 F2 = 1*F3 + D2;                          ! F1 and F2 are now dependent
/VARIANCES
 F3 = 1;                                  ! fixed
 D1-D2 = 1*;                              ! free
 E1-E4 = 3*;                              ! free
/CONSTRAINTS
 (F1,F3) = (F2,F3);                       ! Necessary!
/MATRIX
 (6 × 6 matrix, above)
/END
```

In the factor model of Figure 2.8, F1 and F2 are never on the left side of any equation, while in this model they are. The equations for F1 and F2 specify how these factors are generated by a second-order factor F3 plus a residual. F1 and F2 serve as indicators of F3, in the same way that V's usually serve as indicators of F's. If there had been more than two indicators of F3, the constraint of equal loadings for F1 and F2 on F3 would not have been needed. A second-order factor is interpreted by its indicators in the same way that a first-order factor is interpreted via its indicators. In this example, F3 might be considered to be that part of “alienation” that is stable across the measurement span 1967-1971. D1 and D2 represent the part of alienation that is specific to a given time period. If V1 through V4 in Figure 2.13 were first-order factors rather than variables (with their variable indicators not shown), then F1 and F2 would be interpreted as second-order factors and F3 would be a third-order factor. Quite complex hierarchies of factors can be constructed with appropriate data.

The model of Figure 2.13 contains exactly the same number of degrees of freedom as the model of Figure 2.8. The second-order parameterization did not gain degrees of freedom as it usually would with a greater number of indicators. In this case, the two parameterizations are exactly equivalent, and the fit of the model is also the same. Thus, the GLS χ^2 for the above model is 55.612 with 1 df. A residual covariance between errors, along with some further constraints, would lead to a model that fits the data. Space limitations preclude giving more details.

It will be apparent by now that *the residual variables E and D are essentially always independent variables*, see, e.g., Figs. 2.11 and 2.13. Thus *EQS has been written so that E and D variables cannot be specified as dependent variables in an equation.* If a residual is intended to be predicted by other variables, that residual should be called an F variable.

The previous examples have made it clear that essentially any linear structural model can be specified by a series of equations with known and unknown coefficient parameters. In covariance structure models, the variances and covariances of the independent variables complete the specification. Thus, a single, unified approach to model specification based on the Bentler-Weeks (1979, 1980) concepts of dependent and independent variables is used in EQS, irrespective of the particular type of model being studied: whether measured-variable, latent-variable, higher-order, or any other model. The simple principles described above always remain applicable, and no special cases are needed. Given an adequate rationale for a model, and appropriate data against which to evaluate the model, the main difficulty in practice will be in deciding whether the model is identified, so that it can be estimated and tested. A helpful device in this regard is to visually search a complex diagram for parts in which identification problems are understood, e.g., to see the bottom part of Figure 2.13 as a factor model with two indicators.

Models in Practice

Models are often motivated by the attempt to capture, in the equation structure, some causal process that explains the data, especially the interrelations among the variables. Whether any particular model truly reflects a causal process is very hard to determine. In general, this cannot be done on the basis of empirical results, but hinges on placing a proposed model into a larger conceptual framework (see, e.g., Bollen, 1989b; Cliff, 1983; Gollob & Reichardt, 1987; James, Mulaik, & Brett, 1982; MacCallum, 1995; Reichardt & Gollob, 1986; Sobel, 1994, 1995; Wegener & Fabrigar, 2000; West, Biesanz, & Pitt, 2000). At a minimum, key variables of relevance to the structural system must not have been omitted, error of measurement must have appropriately been taken into account by the use of latent variables, and conditions and times of measurement must have been correctly specified to permit obtaining the hypothesized effects. These considerations have led to the use of some rather standard designs, often involving latent variable models with repeatedly measured constructs, with each construct having multiple indicators. Control for initial levels of the variables is thus done at the latent variable level, avoiding bias due to ignoring the error of measurement in the predictors. An example of such a design is given by the typical use of the standard FASEM model developed by Jöreskog, Keesling, and Wiley (Jöreskog, 1973; Wiley, 1973), and illustrated in Figure 2.11. Such a model has a measurement model and a latent variable simultaneous equation model. It is quite probable, however, that the use of such a design will not guarantee that a "causal" interpretation may be appropriate. In fact, the standard design may overlook some very important effects. When logically necessary effects are omitted from a model, it will generally be misspecified, and the resulting parameter estimates will be biased.

The standard types of models illustrated above do not exhaust the kinds of models that can be studied in practice. Not all variants of interesting models can be shown here, e.g., circumplex models (Fabrigar, Visser, & Browne, 1997; Remington, Fabrigar, & Visser, 2000), multitrait-multimethod models (e.g., Barbaranelli & Caprara, 2000; Byrne & Goffin, 1993; Coenders & Saris, 2000; Corral-Verdugo & Figueredo, 1999; Cudeck, 1988; Dumenci, 2000; Eid, 2000; Kenny & Kashy, 1992; Marsh & Grayson, 1995; Scherpenzeel & Saris, 1997; Wothke, 1996; Wothke & Browne, 1990), multimode models (Bentler & Lee, 1979; Bentler, Poon, & Lee, 1988; Oort, 1999, 2001), simplex models (Hays, Widaman, DiMatteo, & Stacy, 1987; Marsh, 1993; Raykov, 1998a; Sivo & Willson, 2000) or trait-state models (Kenny & Zautra, 1995, 2001). The key point is that you should not be "straitjacketed" into thinking about models using the few standard designs that tend to be used in practice. First, nonstandard models allow variants that might not be obvious. Second, creative reconceptualization of a model often makes it possible to implement a model in EQS that, at first glance, might not work (e.g., Bentler, Poon, & Lee, 1988).

Nonstandard Models

In a FASEM model, various types of effects of interest, for example, direct paths or arrows between V variables, paths from Vs to Fs, paths from residuals such as E variables to F factors, and so on are not permitted. While a standard model such as this answers an interesting question about general latent variable effects -- the consequences of general constructs on each other -- it does ignore potentially important specific effects. To illustrate the flexibility of EQS in setting up non FASEM models, and to illustrate how conceptual thinking should drive the design of a model, we consider a design that permits evaluation of specific effects as well as general effects using Bentler's (1990c) hypothetical example involving the drugs and health domains. An important problem associated with the isolation of unique consequences of the use of specific drugs is that some of the effects may not best be conceptualized as effects of latent drug use variables. Such latent variables are usually rather general constructs that contain common variance among several indicators, in this case, of drug use. A specific effect, however, would be associated with the unique action of a drug (possibly mediated by various social and personal processes). Of course, the unique consequence of taking a given drug cannot be identified unless one also simultaneously identifies common actions stemming from use of a group of drugs. A series of *nonstandard* models evaluating the consequences of general and specific drug use in adolescence on psychological and social functioning in young adults was given by Newcomb and Bentler (1988) and, in a related context, by Newcomb (1994). Spirtes, Scheines, and Glymour (1990) provide other examples of nonstandard models.

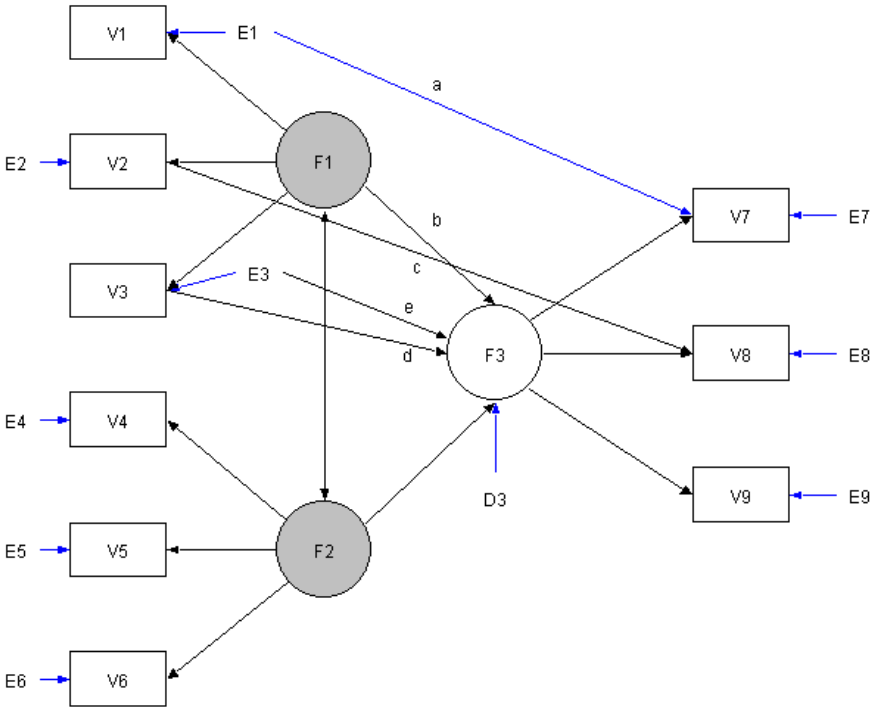


Figure 2.14

An example of how drug consequences might be detected in a structural model is shown in Figure 2.14 above. This represents a hypothetical example involving nine measured variables (V1-V9). Three latent common factors F1-F3 are assumed to generate these variables, along with errors in variables E1-E9. With all the paths shown, the model is not identified. Imagine first that the paths marked "a", "c", "d", and "e" are not in the diagram. Then this model would be a standard FASEM model that might represent, for example, data obtained at two time points, with V1-V6 measured initially, and V7-V9 measured some years subsequently. Consider next that F2 might represent initial general health status, and F3 might represent final health status. If F1 represents early general drug use, path "b"

would represent the effect of early drug use on subsequent health, holding constant initial levels of health (because the F2 to F3 path, representing health stability, is also in the model). D3 is the latent variable regression residual. Such a model, without the paths "a", "c", "d", and "e", would represent a standard FASEM model for drug use consequences. Of course other important control variables would also be included in the model, but these are not shown in the figure to conserve space. Also not shown are the correlated errors between identically measured variables across time (e.g., E4-E7), or comparable paths (e.g., E4 → V7), and the designation of free or fixed parameters.

If V1-V3 represent alcohol, marijuana, and hard drug use, respectively, F1 would be interpretable as a general drug use factor. With V4-V6 and V7-V9 representing, say, indices of hepatic (liver), pulmonary (lung), and cardio-vascular functioning at the two time points, path "b" represents the effect of general drug use on general health. However, it is possible and, on the basis of prior knowledge, perhaps probable, that a cause-effect (drug use) → (poor health) sequence is much more specific. The figure shows four possible examples of specific effects that might be added to the general model first discussed, though in any application only one or another might be tested. Path "a" would represent the effect of alcohol use on hepatic functioning, in the context of the general model. Such a path might represent the idea that alcohol use has a particular effect on liver functioning. Since the path originates at E1, the effect shown in "a" is based on that part of alcohol use that is not in common with use of the other substances, i.e., it is a drug-specific effect. Path "c" would represent the effect of marijuana use on pulmonary functioning. This effect is specific, but more general than the previous effect: in this case general drug use (F1) also effects pulmonary functioning indirectly via V2 and path "c". (It already does so in the original model, via F3 to V8). Two very different types of effects are shown in paths "d" and "e". Both of these paths represent the effects of a specific drug, say, hard drug use, on general health (rather than on a specific aspect of health as noted above). The difference between these effects is that "d" also permits F1 to have an indirect effect on F3, while "e" does not permit such an indirect effect. Thus there are many potential nonstandard effects, and substantive knowledge must be used to choose the particular effects to evaluate in any given situation.

It is apparent from this example that the evaluation of drug use consequences may need to involve very careful theorizing and specialized model evaluation. It would not be desirable to utilize only standard designs to evaluate such effects, since these effects may be subtle and localized as illustrated above. If a model with specific effects such as is shown in Figure 2.14 represents the true nature of a causal process, then focusing only on the FASEM structure of these variables would mis-specify the true model. Although the illustration dealt with the field of health, the idea that the structural model being investigated should mirror, as well as possible, the true causal process involved in the variables, is applicable to all fields. While standard designs typically will help to focus on important model features, e.g., general influences in the context of control for errors of measurement, nonstandard designs should be considered whenever substantive theory so dictates. At times the substantive questions studied with nonstandard effects could, with the availability of a different set of measured variables, be studied with a standard model. For example, the specific part of an error-in-variable residual in one model can, with a different set of variables and measurement structure, become a common factor.

The typical application of nonstandard models is much less dramatic than considered above. In particular, a FASEM model may be desired in which several latent factors influence other latent factors, but it is recognized that one of the factors has only one indicator. Even though *a single measured variable cannot create a factor* unless external information is brought to bear on the problem, the literature shows many examples of a diagram and model setup in which a factor does appear to have been created from the single indicator: the diagram shows both a factor and its indicator, with a path such as F → V. However, careful study will show that the F and V actually are identical, and that a trick was used merely to allow the LISREL program to run. Such tricks are not necessary in EQS. In fact, they are not desirable because they obscure the model actually being tested. A simple example is shown next.

Nonstandard Model: Example

Factors and variables can have the same status in a model. This is typically not the case in standard models. Bagozzi (1980) reported a study on performance and satisfaction in an industrial sales force. Job satisfaction as a latent variable was predicted by an achievement motivation latent variable, a self-esteem latent variable, and two

measured variables of performance and verbal IQ. A path diagram for this model is given in Figure 2.15. An interesting feature of this model is that the measured and latent variables are used as both predictors and criteria. In the left part of the figure, factors F2 and F3 have the same status as variable V8; they are arbitrarily correlated independent variables. In the right side of the figure, V1 is a mediator of the effects of F3 and V8 on F1. From a substantive point of view, factors of need for achievement (F2) and self-esteem (F3), as well as verbal intelligence (V8), are hypothesized to explain the factor of job satisfaction (F1). The effects of self-esteem and verbal IQ on satisfaction are mediated by performance (V1).

Figure 2.15 could be compared to the rather different diagrams for the same model given by Bagozzi (1980) and Dillon and Goldstein (1984), based on a LISREL representation. Since the basic LISREL or FASEM model does not allow Fs and Vs to correlate, dummy variables were used to create "latent variable" counterparts to measured variables V1 and V8, and the corresponding latent variables were then correlated and regressed. In fact, these so-called "latent" variables are in fact nothing other than measured variables. In Figure 2.15 there is no confusion that the V variables are just measured variables. See Bentler and Satorra (2000) for related illustration showing that phantom variables are often not necessary to achieve a modeling objective.

Immediately below Figure 2.15, we give a condensed printout of the EQS output for this problem. Some comments have been added to the output to clarify the results. Material to the right of the exclamation mark "!" signifies a comment, added to the output file after the EQS run.

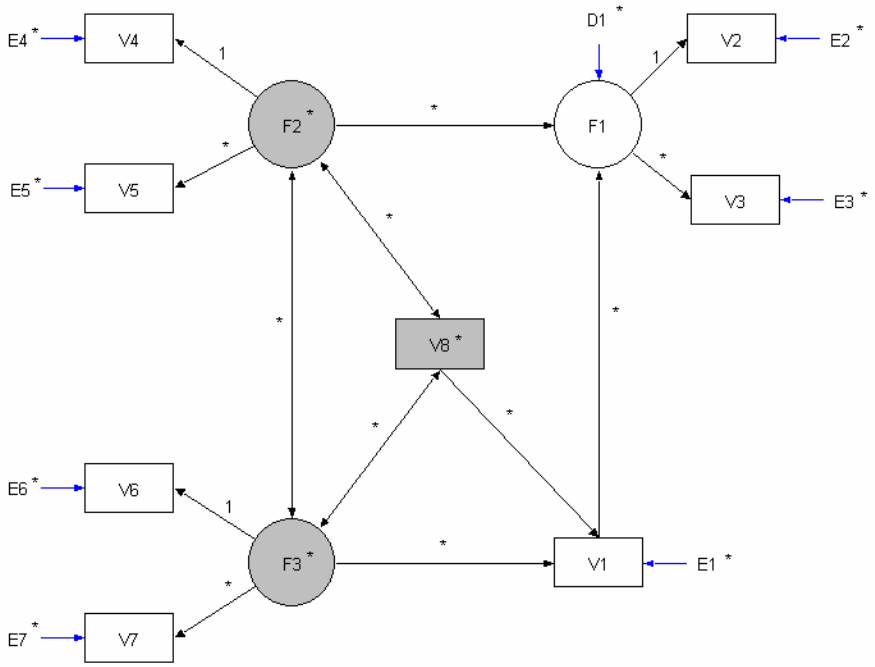


Figure 2.15

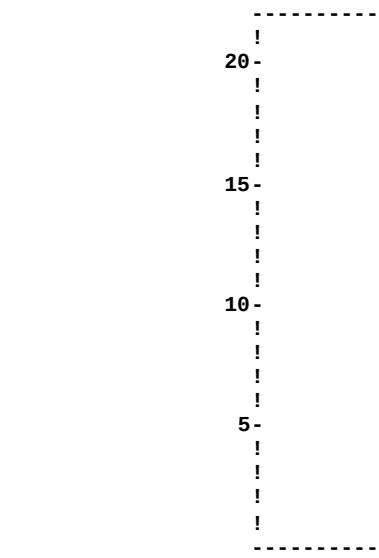
```
PROGRAM CONTROL INFORMATION

     1  /TITLE
     2   BAGOZZI (1980) EXAMPLE STUDIED BY DILLON & GOLDSTEIN (1984)
     3   PERFORMANCE AND JOB SATISFACTION IN AN INDUSTRIAL SALES FORCE
     4   CORRELATION MATRIX & STANDARD DEVIATIONS
     5   SCALE OF V1 HAS BEEN MODIFIED BY MOVING DECIMAL PLACE
     6  /SPECIFICATIONS
     7   CASES = 122; VARIABLES = 8; MATRIX=CORRELATION; ANALYSIS=COVARIANCE;
     8  /LABELS
     9   V1 = PERFORM;  V2 = SATISFC1; V3 = SATISFC2; V4 = ACHMOTV1;
    10   V5 = ACHMOTV2; V6 = SLFESTM1; V7 = SLFESTM2; V8 = VERBALIQ;
    11   F1 = JOBSATSF; F2 = N-ACHIEV; F3 = ESTEEM;
    12  /MODEL
    13   (V2,V3)    ON F1 = 1*;  ! EQS FORCES F1 TO BE IDENTIFIED BY SETTING (V2,F1) = 1
    14   (V4,V5,F1) ON F2 = 1*;  ! SIMILARLY, (V4,F2) = 1.0
    15   (V1,V7)    ON F3 = 1*;  ! (V6,F3) MUST BE FIXED AT 1.0 EXPLICITLY.  IF WE TRIED
    16   (V6)       ON F3 = 1 ;  ! (V6,V1,F7) ON F3 = 1*;
    17   (V1)       ON V8 = .5*; ! (V1,F3) WOULD BE FIXED AND (V6,F3) AND (V7,F3) FREE
    18   (F1)       ON V1 = .5*;
    19   VAR V8    = 10*;
    20   VAR E1-E7 = 5*;
    21   VAR F2,F3,D1 = 1*;
    22   COV (F2,F3,V8) ;
    23  /MATRIX
    24   1.000
    25    .418 1.000
    26    .394  .627 1.000
    27    .129  .202  .266 1.000
    28    .189  .284  .208  .365 1.000
    29    .544  .281  .324  .201  .161 1.000
    30    .507  .225  .314  .172  .174  .546 1.000
    31   -.357 -.156 -.038 -.199 -.277 -.294 -.174 1.000 ! NO SEMICOLON TO END DATA
    32  /STANDARD DEVIATIONS
    33   2.09  3.43  2.81  1.95  2.08  2.16  2.06  3.65  ! NO SEMICOLON TO END DATA
    34  /print                                           ! NOTE THAT UPPER AND LOWER CASE
    35    effect=yes; covariance=yes; correlation=yes;   ! MAY BE USED INTERCHANGEABLY.
    36  /LMTEST                                          ! JUST THE DEFAULT TESTS ARE
    37  /WTEST                                           ! REQUESTED.
    38  /END

MODEL SECTION GENERATED THESE SECTIONS: ! CHECK THAT THE GENERATED MODEL IS WHAT WE WANT

 /VARIANCES
  V8=10*;
  E1-E7=5*;
  F2,F3,D1=1*;

 /EQUATIONS
  V1=.5*V8+1*F3+E1;                    ! Vs AND Fs PREDICT Vs, SAME WAY
  V2=1F1+E2;                           ! MORE TYPICAL F PREDICTING V
  V3=1*F1+E3;
  V4=1F2+E4;
  V5=1*F2+E5;
  V6=1F3+E6;
  V7=1*F3+E7;
  F1=.5*V1+1*F2+D1;                    ! Vs AND Fs PREDICT Fs, SAME WAY

 /COVARIANCES
  F3,F2=*;
  V8,F2=*;V8,F3=*;                     ! NO PROBLEM HAVING Vs AND Fs CORRELATE
```

```
 COVARIANCE  MATRIX TO BE ANALYZED:   8 VARIABLES (SELECTED FROM   8 VARIABLES)
  BASED ON   122 CASES.
                       PERFORM    SATISFC1   SATISFC2   ACHMOTV1   ACHMOTV2   SLFESTM1   SLFESTM2
                         V  1       V  2       V  3       V  4       V  5       V  6       V  7
     PERFORM  V  1       4.368
     SATISFC1 V  2       2.997     11.765
     SATISFC2 V  3       2.314      6.043      7.896
     ACHMOTV1 V  4       0.526      1.351      1.458      3.802
     ACHMOTV2 V  5       0.822      2.026      1.216      1.480      4.326
     SLFESTM1 V  6       2.456      2.082      1.967      0.847      0.723      4.666
     SLFESTM2 V  7       2.183      1.590      1.818      0.691      0.746      2.429      4.244
     VERBALIQ V  8      -2.723     -1.953     -0.390     -1.416     -2.103     -2.318     -1.308

                       VERBALIQ
                         V  8
     VERBALIQ V  8      13.322

  MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

  PARAMETER ESTIMATES APPEAR IN ORDER,
  NO SPECIAL PROBLEMS WERE ENCOUNTERED DURING OPTIMIZATION.
! IMPORTANT MESSAGE.  IF IT DOES NOT APPEAR, THERE WILL BE MESSAGES ON LINEAR DEPENDENCIES, ETC.


RESIDUAL COVARIANCE MATRIX  (S-SIGMA) :

                       PERFORM    SATISFC1   SATISFC2   ACHMOTV1   ACHMOTV2   SLFESTM1   SLFESTM2
                         V  1       V  2       V  3       V  4       V  5       V  6       V  7
     PERFORM  V  1       0.000
     SATISFC1 V  2      -0.049     -0.039
     SATISFC2 V  3      -0.011     -0.030     -0.023
     ACHMOTV1 V  4      -0.165     -0.137      0.322      0.000
     ACHMOTV2 V  5      -0.030      0.191     -0.185     -0.054      0.000
     SLFESTM1 V  6      -0.008      0.090      0.446      0.148     -0.137      0.000
     SLFESTM2 V  7       0.029     -0.151      0.489      0.081     -0.007     -0.014      0.000
     VERBALIQ V  8       0.000      0.826      1.731      0.007     -0.348     -0.331      0.428

                       VERBALIQ
                         V  8
     VERBALIQ V  8       0.000

             AVERAGE ABSOLUTE  COVARIANCE  RESIDUALS  =        0.1808 ! LARGE? HARD TO TELL IN
 AVERAGE OFF-DIAGONAL ABSOLUTE  COVARIANCE  RESIDUALS  =        0.2303 ! COVARIANCE METRIC

  STANDARDIZED RESIDUAL MATRIX:

                       PERFORM    SATISFC1   SATISFC2   ACHMOTV1   ACHMOTV2   SLFESTM1   SLFESTM2
                         V  1       V  2       V  3       V  4       V  5       V  6       V  7
     PERFORM  V  1       0.000
     SATISFC1 V  2      -0.007     -0.003
     SATISFC2 V  3      -0.002     -0.003     -0.003
     ACHMOTV1 V  4      -0.041     -0.021      0.059      0.000
     ACHMOTV2 V  5      -0.007      0.027     -0.032     -0.013      0.000
     SLFESTM1 V  6      -0.002      0.012      0.074      0.035     -0.031      0.000
     SLFESTM2 V  7       0.007     -0.021      0.085      0.020     -0.002     -0.003      0.000
     VERBALIQ V  8       0.000      0.066      0.169      0.001     -0.046     -0.042      0.057

                       VERBALIQ
                         V  8
     VERBALIQ V  8       0.000

             AVERAGE ABSOLUTE STANDARDIZED RESIDUALS  =        0.0247 ! MUCH EASIER TO INTERPRET
 AVERAGE OFF-DIAGONAL ABSOLUTE STANDARDIZED RESIDUALS  =        0.0315 ! IN CORRELATION METRIC
```

LARGEST STANDARDIZED RESIDUALS: ! MUCH EASIER TO ANALYZE THAN THE ABOVE MATRIX

NO.	PARAMETER	ESTIMATE	NO.	PARAMETER	ESTIMATE
1	V8, V3	0.169	11	V5, V3	-0.032
2	V7, V3	0.085	12	V6, V5	-0.031
3	V6, V3	0.074	13	V5, V2	0.027
4	V8, V2	0.066	14	V7, V2	-0.021
5	V4, V3	0.059	15	V4, V2	-0.021
6	V8, V7	0.057	16	V7, V4	0.020
7	V8, V5	-0.046	17	V5, V4	-0.013
8	V8, V6	-0.042	18	V6, V2	0.012
9	V4, V1	-0.041	19	V5, V1	-0.007
10	V6, V4	0.035	20	V2, V1	-0.007

DISTRIBUTION OF STANDARDIZED RESIDUALS



! DISTRIBUTION IS CLOSE TO SYMMETRIC, AND
! CENTERED ON ZERO

	RANGE			FREQ	PERCENT
1	-0.5	-	--	0	0.00%
2	-0.4	-	-0.5	0	0.00%
3	-0.3	-	-0.4	0	0.00%
4	-0.2	-	-0.3	0	0.00%
5	-0.1	-	-0.2	0	0.00%
6	0.0	-	-0.1	20	55.56%
7	0.1	-	0.0	15	41.67%
8	0.2	-	0.1	1	2.78%
9	0.3	-	0.2	0	0.00%
A	0.4	-	0.3	0	0.00%
B	0.5	-	0.4	0	0.00%
C	++	-	0.5	0	0.00%
	TOTAL			36	100.00%

EACH "*" REPRESENTS 1 RESIDUAL(S)

MODEL COVARIANCE MATRIX FOR MEASURED AND LATENT VARIABLES

	PERFORM V 1	SATISFC1 V 2	SATISFC2 V 3	ACHMOTV1 V 4	ACHMOTV2 V 5	SLFESTM1 V 6	SLFESTM2 V 7	VERBALIQ V 8	JOBSATSF F 1
PERFORM V 1	4.368								
SATISFC1 V 2	3.046	11.804							
SATISFC2 V 3	2.325	6.073	7.919						
ACHMOTV1 V 4	0.691	1.488	1.136	3.802					
ACHMOTV2 V 5	0.852	1.835	1.400	1.535	4.326				
SLFESTM1 V 6	2.464	1.992	1.520	0.698	0.861	4.666			
SLFESTM2 V 7	2.153	1.740	1.328	0.610	0.752	2.443	4.244		
VERBALIQ V 8	-2.723	-2.779	-2.121	-1.424	-1.755	-1.987	-1.736	13.322	
JOBSATSF F 1	3.046	7.957	6.073	1.488	1.835	1.992	1.740	-2.779	7.957
N-ACHIEV F 2	0.691	1.488	1.136	1.245	1.535	0.698	0.610	-1.424	1.488
ESTEEM F 3	2.464	1.992	1.520	0.698	0.861	2.796	2.443	-1.987	1.992

	N-ACHIEV F 2	ESTEEM F 3
N-ACHIEV F 2	1.245	
ESTEEM F 3	0.698	2.796

! THIS IS THE COVARIANCE MATRIX REPRODUCED FROM THE PARAMETERS
! OF THE MODEL. IT ALSO CONTAINS COVARIANCES OF V AND F
! VARIABLES, AND F-F COVARIANCES.

```
MODEL CORRELATION MATRIX FOR MEASURED AND LATENT VARIABLES

                PERFORM  SATISFC1 SATISFC2 ACHMOTV1 ACHMOTV2 SLFESTM1 SLFESTM2 VERBALIQ JOBSATSF
                  V  1     V  2     V  3     V  4     V  5     V  6     V  7     V  8     F  1
PERFORM  V  1    1.000
SATISFC1 V  2    0.424    1.000
SATISFC2 V  3    0.395    0.628    1.000
ACHMOTV1 V  4    0.170    0.222    0.207    1.000
ACHMOTV2 V  5    0.196    0.257    0.239    0.378    1.000
SLFESTM1 V  6    0.546    0.268    0.250    0.166    0.192    1.000
SLFESTM2 V  7    0.500    0.246    0.229    0.152    0.176    0.549    1.000
VERBALIQ V  8   -0.357   -0.222   -0.206   -0.200   -0.231   -0.252   -0.231    1.000
JOBSATSF F  1    0.517    0.821    0.765    0.271    0.313    0.327    0.300   -0.270    1.000
N-ACHIEV F  2    0.296    0.388    0.362    0.572    0.661    0.290    0.265   -0.350    0.473
ESTEEM   F  3    0.705    0.347    0.323    0.214    0.248    0.774    0.709   -0.326    0.422

                N-ACHIEV ESTEEM  ! THIS REPRODUCED CORRELATION MATRIX OF V AND F VARIABLES
                  F  2     F  3  ! IS EASIER TO INTERPRET THAN THE COVARIANCE MATRIX
N-ACHIEV F  2    1.000
ESTEEM   F  3    0.374    1.000

GOODNESS OF FIT SUMMARY FOR METHOD = ML

INDEPENDENCE MODEL CHI-SQUARE =        256.499 ON    28 DEGREES OF FREEDOM

INDEPENDENCE AIC =   200.49879   INDEPENDENCE CAIC =  93.98620 ! COMPUTED FOR THOSE WHO LIKE
       MODEL AIC =   -14.60016          MODEL CAIC = -71.66048 ! AKAIKE INFORMATION CRITERION.

CHI-SQUARE =       15.400 BASED ON    15 DEGREES OF FREEDOM      ! CHI-SQUARE AND PROBABILITY
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS     0.42302    ! INDICATE A GOOD FIT
THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS  15.418.

FIT INDICES
-----------
BENTLER-BONETT NORMED    FIT INDEX=       0.940  !FIT INDICES ARE VERY HIGH, INDICATING A VERY
BENTLER-BONETT NONNORMED FIT INDEX=       0.997  ! GOOD MODEL FIT
COMPARATIVE FIT INDEX (CFI)       =       0.998  ! PREFERRED INDEX, SEE CHAPTER 14

RELIABILITY COEFFICIENTS
------------------------
CRONBACH'S ALPHA                                          =     0.554
GREATEST LOWER BOUND RELIABILITY                          =     0.780  !NEW IN EQS
BENTLER'S DIMENSION-FREE LOWER BOUND RELIABILITY           =     0.780  !SEE CH. 14
SHAPIRO'S LOWER BOUND RELIABILITY FOR A WEIGHTED COMPOSITE  =     0.869

                    ITERATIVE SUMMARY

                 PARAMETER
ITERATION        ABS CHANGE         ALPHA              FUNCTION
    1             1.398404        1.00000             1.14871 !CONVERGENCE WAS QUICK AND
    2             0.476569        1.00000             0.70931 ! SMOOTH: NO STEP HALVINGS
    3             0.435321        1.00000             0.23518 ! (alpha always 1.0)
    4             0.178585        1.00000             0.13277
    5             0.041384        1.00000             0.12742
    6             0.015907        1.00000             0.12729
    7             0.004635        1.00000             0.12727
    8             0.001867        1.00000             0.12727
    9             0.000674        1.00000             0.12727
```

```
MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
  STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

 PERFORM =V1  =   -.082*V8    +  .823*F3    + 1.000 E1
                   .047          .147
                  -1.754        5.590@               ! EXCEPT FOR (V1,V8), ALL PARAMETERS ARE
                                                     ! SIGNIFICANT AT THE 95% LEVEL
 SATISFC1=V2  =   1.000 F1    + 1.000 E2             ! THAT IS, P  < .05

 SATISFC2=V3  =    .763*F1    + 1.000 E3
                   .132
                  5.784@

 ACHMOTV1=V4  =   1.000 F2    + 1.000 E4

 ACHMOTV2=V5  =   1.233*F2    + 1.000 E5
                   .400
                  3.084@

 SLFESTM1=V6  =   1.000 F3    + 1.000 E6

 SLFESTM2=V7  =    .874*F3    + 1.000 E7
                   .142
                  6.148@

  CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
  STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

 JOBSATSF=F1  =    .557*V1    +  .886*F2    + 1.000 D1
                   .139          .372
                  4.010@        2.385@


VARIANCES OF INDEPENDENT VARIABLES            ! AS CAN BE SEEN, THESE ARE ORGANIZED BY COLUMNS
  ---------------------------------
  STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

              V                        F                        E                        D
             ---                      ---                      ---                      ---
 V8  -VERBALIQ   13.323*I F2  -N-ACHIEV    1.245*I E1  -PERFORM     2.117*I D1  -JOBSATSF  4.941*
                  1.713 I                   .534 I                   .382 I                1.382
                  7.778@I                  2.329@I                  5.547@I                3.576@
                        I                        I                        I
                        I F3  -ESTEEM      2.796*I E2  -SATISFC1    3.848*I
                        I                   .665 I                  1.301 I
                        I                  4.202@I                  2.957@I
                        I                        I                        I
                        I                        I E3  -SATISFC2    3.283*I
                        I                        I                    .818 I
                        I                        I                   4.012@I
                        I                        I                        I
                        I                        I E4  -ACHMOTV1    2.558*I
                        I                        I                    .513 I
                        I                        I                   4.988@I
                        I                        I                        I
                        I                        I E5  -ACHMOTV2    2.434*I
                        I                        I                    .668 I
                        I                        I                   3.643@I
                        I                        I                        I
                        I                        I E6  -SLFESTM1    1.870*I
                        I                        I                    .445 I
                        I                        I                   4.198@I
                        I                        I                        I
                        I                        I E7  -SLFESTM2    2.109*I
                        I                        I                    .396 I
                        I                        I                   5.330@I
```

```
COVARIANCES AMONG INDEPENDENT VARIABLES          ! AGAIN, ORGANIZED BY COLUMNS
  -------------------------------------
  STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.
                 V                                F
                ---                              ---
 F2  -N-ACHIEV           -1.424*I F3  -ESTEEM              .698*I
 V8  -VERBALIQ             .576 I F2  -N-ACHIEV             .303 I
                         -2.470@I                         2.305@I
                                I                               I
 F3  -ESTEEM             -1.987*I                               I
 V8  -VERBALIQ             .693 I                               I
                         -2.867@I                                I

  DECOMPOSITION OF EFFECTS WITH NONSTANDARDIZED VALUES        !THESE ARE NOT EQUATIONS
  STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

  PARAMETER TOTAL EFFECTS
  -----------------------
PERFORM =V1  =    -.082*V8    +  .823*F3    + 1.000 E1

SATISFC1=V2  =     .557 V1    + 1.000 F1    -  .045 V8    +  .886 F2
                   .139                       .028          .372
                  4.010@                     -1.602         2.385@

                +  .459 F3    +  .557 E1    + 1.000 E2    + 1.000 D1
                   .140          .139
                  3.287@        4.010@

SATISFC2=V3  =     .425 V1    +  .763*F1    -  .035 V8    +  .676 F2
                   .111          .132          .022          .288
                  3.840@        5.784@       -1.591         2.348@

                +  .350 F3    +  .425 E1    + 1.000 E3    +  .763 D1
                   .110          .111                        .132
                  3.192@        3.840@                      5.784@


 ACHMOTV1=V4  =   1.000 F2    + 1.000 E4

 ACHMOTV2=V5  =   1.233*F2    + 1.000 E5

 SLFESTM1=V6  =   1.000 F3    + 1.000 E6

 SLFESTM2=V7  =    .874*F3    + 1.000 E7

 JOBSATSF=F1  =    .557*V1    -  .045 V8    +  .886*F2    +  .459 F3
                   .139          .028                        .140
                  4.010@       -1.602                       3.287@

                +  .557 E1    + 1.000 D1
                   .139
                  4.010@

  DECOMPOSITION OF EFFECTS WITH NONSTANDARDIZED VALUES
  STATISTICS SIGNIFICANT AT THE 5% LEVEL ARE MARKED WITH @.

  PARAMETER INDIRECT EFFECTS    !SOME INDIRECT EFFECTS ARE NOT SIGNIFICANT AND PROBABLY
  --------------------------    !SHOULD NOT BE INTERPRETED

 SATISFC1=V2  =    .557 V1    -  .045 V8    +  .886 F2    +  .459 F3
                   .139          .028          .372          .140
                  4.010@       -1.602         2.385@        3.287@

                +  .557 E1    + 1.000 D1
                   .139
                  4.010@

 SATISFC2=V3  =    .425 V1    -  .035 V8    +  .676 F2    +  .350 F3
                   .111          .022          .288          .110
                  3.840@       -1.591         2.348@        3.192@
```

```
              +  .425 E1    +  .763 D1
                 .111          .132
               3.840@        5.784@

JOBSATSF=F1  =   -.045 V8    +  .459 F3    +  .557 E1
                  .028          .140          .139
                -1.602        3.287@        4.010@


 DECOMPOSITION OF EFFECTS WITH STANDARDIZED VALUES  !STANDARDIZED EFFECTS ARE OFTEN EASIER TO
                                                    !INTERPRET THAN NON-STANDARDIZED EFFECTS
 PARAMETER TOTAL EFFECTS
 -----------------------
PERFORM =V1  =   -.143*V8    +  .659*F3    +  .696 E1

SATISFC1=V2  =    .339 V1    +  .821 F1    -  .048 V8    +  .288 F2
               +  .223 F3    +  .236 E1    +  .571 E2    +  .647 D1

SATISFC2=V3  =    .316 V1    +  .765*F1    -  .045 V8    +  .268 F2
               +  .208 F3    +  .220 E1    +  .644 E3    +  .603 D1

ACHMOTV1=V4  =    .572 F2    +  .820 E4

ACHMOTV2=V5  =    .661*F2    +  .750 E5

SLFESTM1=V6  =    .774 F3    +  .633 E6

SLFESTM2=V7  =    .709*F3    +  .705 E7

JOBSATSF=F1  =    .413*V1    -  .059 V8    +  .351*F2    +  .272 F3
               +  .287 E1    +  .788 D1

 DECOMPOSITION OF EFFECTS WITH STANDARDIZED VALUES

 PARAMETER INDIRECT EFFECTS
 --------------------------

SATISFC1=V2  =    .339 V1    -  .048 V8    +  .288 F2    +  .223 F3
               +  .236 E1    +  .647 D1

SATISFC2=V3  =    .316 V1    -  .045 V8    +  .268 F2    +  .208 F3
               +  .220 E1    +  .603 D1

JOBSATSF=F1  =   -.059 V8    +  .272 F3    +  .287 E1

 STANDARDIZED SOLUTION:                                          R-SQUARED

PERFORM =V1  =   -.143*V8    +  .659*F3    +  .696 E1                .515
SATISFC1=V2  =    .821 F1    +  .571 E2                             .674
SATISFC2=V3  =    .765*F1    +  .644 E3                             .585
ACHMOTV1=V4  =    .572 F2    +  .820 E4                             .327
ACHMOTV2=V5  =    .661*F2    +  .750 E5                             .437
SLFESTM1=V6  =    .774 F3    +  .633 E6                             .599
SLFESTM2=V7  =    .709*F3    +  .705 E7                             .503
JOBSATSF=F1  =    .413*V1    +  .351*F2    +  .788 D1               .379

! IN THE STANDARDIZED SOLUTION ALL VARIABLES HAVE BEEN TRANSFORMED TO HAVE UNIT VARIANCE.
! THIS IS A WRIGHT-LIKE PATH ANALYSIS SOLUTION, OFTEN EASIER TO INTERPRET THAN THE
! NONSTANDARDIZED SOLUTION.

CORRELATIONS AMONG INDEPENDENT VARIABLES  ! USEFUL TO CHECK FOR OUT OF RANGE VALUES
  -------------------------------------

                V                              F
               ---                            ---
F2  -N-ACHIEV           -.350*I F3  -ESTEEM              .374*I
V8  -VERBALIQ                I F2  -N-ACHIEV                  I
                             I                               I
F3  -ESTEEM             -.326*I                              I
V8  -VERBALIQ                I                               I
```

```
 ---------------------------------------------------------------------------
                          E N D    O F    M E T H O D
 ---------------------------------------------------------------------------
! THE ABOVE MARKS THE END OF A STANDARD OUTPUT.  IF OTHER METHODS WERE REQUESTED
! THEY WOULD PRINT NEXT. WTEST AND LMTEST COME AT THE END OF THE OUTPUT.

  WALD TEST (FOR DROPPING PARAMETERS)
  MULTIVARIATE WALD TEST BY SIMULTANEOUS PROCESS  ! DEFAULT TEST

         CUMULATIVE MULTIVARIATE STATISTICS              UNIVARIATE INCREMENT
         ----------------------------------              --------------------

  STEP  PARAMETER   CHI-SQUARE  D.F.  PROBABILITY       CHI-SQUARE  PROBABILITY
  ---- -----------  ----------  ----  -----------       ----------  -----------
    1     V1,V8        3.075      1      0.079             3.075       0.079

! THE TEST FOUND ONLY ONE PARAMETER THAT COULD BE DROPPED, SAME AS THE Z-TEST ABOVE

LAGRANGE MULTIPLIER TEST (FOR ADDING PARAMETERS)
       ORDERED UNIVARIATE TEST STATISTICS:
                                                 HANCOCK              STANDAR-
                                    CHI-           15 DF   PARAMETER   DIZED
   NO    CODE     PARAMETER       SQUARE     PROB.  PROB.    CHANGE    CHANGE
   --   ------    ---------       ------     -----  --------  --------  --------
    1    2  11       V3,V8        4.991     0.025     0.992     0.136     0.013 ! SIGNIFICANT?
    2    2  15       F1,V8        3.749     0.053     0.998     0.159     0.015 ! SEE BELOW
    3    2  12       V3,F3        2.461     0.117     1.000     0.356     0.076
    4    2  11       V5,V8        1.879     0.170     1.000    -0.109    -0.014
    5    2  11       V6,V8        1.662     0.197     1.000    -0.077    -0.010
    6    2  11       V7,V8        1.662     0.197     1.000     0.068     0.009
    7    2  12       V2,F3        0.843     0.359     1.000    -0.251    -0.044 ! LARGER PARAM
    8    2  20       V6,F1        0.409     0.522     1.000     0.050     0.008 ! CHANGE, BUT NS
    9    2  16       F1,F3        0.347     0.556     1.000     0.180     0.038
   10    2  12       V5,F3        0.276     0.599     1.000    -0.110    -0.032
   11    2  12       V1,F2        0.246     0.620     1.000    -0.109    -0.047
   12    2  12       V6,F2        0.127     0.722     1.000     0.082     0.034
   13    2  20       V4,F1        0.106     0.745     1.000     0.041     0.007
   14    2  11       V2,V8        0.103     0.748     1.000    -0.024    -0.002
   15    2  20       V7,F1        0.059     0.808     1.000     0.019     0.003
   16    2  20       V5,F1        0.057     0.811     1.000     0.036     0.006
   17    2  12       V4,F3        0.057     0.812     1.000     0.041     0.013
   18    2  20       V1,F1        0.021     0.885     1.000    -0.017    -0.003
   19    2  12       V2,F2        0.019     0.890     1.000     0.068     0.018
   20    2  12       V3,F2        0.019     0.890     1.000    -0.052    -0.016
   21    2  12       V7,F2        0.006     0.936     1.000     0.017     0.008
   22    2  11       V4,V8        0.001     0.982     1.000     0.001     0.000
   23    2   0       V2,F1        0.000     1.000     1.000     0.000     0.000 ! EXACT 0.0 OFTEN
   24    2   0       V6,F3        0.000     1.000     1.000     0.000     0.000 ! IMPLIES FREEING
   25    2   0       V4,F2        0.000     1.000     1.000     0.000     0.000 ! WOULD GET MODEL
                                                                                ! NOT IDENTIFIED

  MULTIVARIATE LAGRANGE MULTIPLIER TEST BY SIMULTANEOUS PROCESS IN STAGE  1
! THIS IS THE DEFAULT PROCEDURE, MANY OTHER OPTIONS COULD HAVE BEEN CHOSEN

  PARAMETER SETS (SUBMATRICES) ACTIVE AT THIS STAGE ARE:
     PVV PFV PFF PDD GVV GVF GFV GFF BVF BFF
! NOTE THAT CORRELATED ERRORS, PEE, ARE NOT PART OF THE DEFAULT PROCEDURE

        CUMULATIVE MULTIVARIATE STATISTICS      UNIVARIATE INCREMENT
        ----------------------------------      ----------------------------
                                                                 HANCOCK'S
                                                                 SEQUENTIAL
 STEP   PARAMETER   CHI-SQUARE  D.F.   PROB.   CHI-SQUARE   PROB.  D.F.   PROB.
 ----  -----------  ----------  ----   -----   ----------   -----  ----   -----
    1       V3,V8        4.991     1   0.025        4.991   0.025    15   0.992


! ONLY ONE PARAMETER SIGNIFICANT, OTHERS DON'T ADD ANY USEFUL INFORMATION
! A DIRECT PATH FROM V8 TO V3 COULD BE ADDED, AND WOULD BE EXPECTED TO IMPROVE THE MODEL
```

```
! CHI-SQUARE BY ALMOST 5 POINTS.  BUT MODEL FIT IS VERY GOOD ALREADY, SO IS NOT NEEDED

  LAGRANGIAN MULTIPLIER TEST REQUIRED     6355 WORDS OF MEMORY.
  PROGRAM ALLOCATES   100000 WORDS.
! IF THE MEMORY IS INSUFFICIENT, EITHER ALLOCATE MORE MEMORY, OR TEST FEWER PARAMETERS
```

Both the univariate and multivariate *LM* tests propose adding a *direct effect* from V8 to V3, that is V8→V3. Using the standard probability values for a 1 df *LM* test, this is significant (p = .025). However, as noted in the first line of the ordered univariate test statistics, and the probability for the univariate increment, Hancock's (1999) conservative test procedure for exploratory model modification based on 15 df would yield nonsignificance (p = .992). See a discussion of Hancock's method in Chapter 6, which gives details on the /LMTEST and /WTEST options. If you accept Hancock's reasoning, you would not add this parameter to the model.

The model currently has an *indirect effect* of V8 on V3 (V8→V1→F1→V3 in Fig. 2.15). Even without the direct effect, V8 does affect V3, but only through *mediating* variables. The indirect effect here is not significant (with a small estimate of -.035). You should think of whether a direct effect is meaningful here, especially since it is not necessary for model fit. If it were conceptually meaningful, it could be added to the model, and the model could be rerun with the modification. This could be accomplished by changing the equation V3 = 1*F1+E3 to V3 = *V8+1*F1+E3. Direct effects are typically interpreted as causal effects and are associated with parameters, as shown in modified model setup. In contrast, indirect effects are computed from the parameters of the model, and are obtained in EQS in the output stream by a specification in /PRINT as shown above. A more detailed discussion of the input options for EQS, and the resulting output, is given in the next two chapters.

Ever since the distinction between *mediator* and *moderator* variables became clear, especially due to Baron and Kenny (1986), models involving mediating variables have become very popular (Finch, West & MacKinnon, 1997; Holmbeck, 2002; Hoyle & Kenny, 1999; Judd, Kenny, & McClelland, 2001; MacKinnon, 2000; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; MacKinnon, Krull, & Lockwood, 2000; Rascle & Irachabal, 2001; Shrout & Bolger, 2002). One reason is that they attempt to elucidate the process by which independent variables influence ultimate outcome variables. For examples of mediation, see Abrams, Viki, Masser, and Bohner (2003), Bryan, Fisher, and Fisher (2002), Goodyear, Newcomb, and Allison (2000), Holmbeck (1997), Linares, Heeren, Bronfman, Zuckerman, Augustyn, and Tronick (2001), Reynolds, Yaroch, Franklin, and Maloy (2002), Shapiro and Levendosky (1999), Singhapakdi, Vitell, and Franke (1999), and Wood, Read, Palfai, and Stevenson (2001). In spite of the popularity of mediational models, you should realize that they require strong and usually untestable assumptions about omitted variables. In a chain of 3 variables, say, V1→V2→V3, with or without the additional direct effect V1→ V3, there cannot be an unmeasured variable (a V, F, E, or D) that directly influences both V1 and V3. Similarly, there cannot be such a variable that directly influences V2 and V3. Nor can there be a correlation between V1 and E3 (the error in V3), or between E2 (the error in V2) and E3. As was shown by Pearl (2000, section 5.3.2) and Herting (2002), for example, such omitted variables or correlations can equally well account for hypothesized direct or indirect effects in mediational models. These two classes of models are observationally equivalent, i.e., they cannot be distinguished empirically. As a result, a credible mediational model should be accompanied by rational arguments as to why omitted variables and correlations are unlikely in the given application. For a more general critical view of effect decomposition in relation to causal interpretations, see Sobel (1990).

3. PROGRAM INPUT

This chapter provides detailed information on how to set up a job to run EQS. It presumes that you have a basic understanding of structural modeling, in particular, the basic concepts of the Bentler-Weeks approach and the estimators that might be appropriate to the model and data under consideration. In the absence of such knowledge, you should first review the relevant sections of Chapter 2. At some point you should also read the other chapters of this *Manual* relevant to the specific methods that you want to implement.

In all the computing environments to which EQS has been ported to date, EQS runs as a stand-alone batch program. In all of these environments, you must create a basic input file that contains all the key model information required to specify an EQS job. The information that is needed in this *model file* was illustrated in the last chapter, and is explained more fully in this chapter. In graphics-oriented environments such as MS Windows or the Macintosh, the model file can be created automatically from a diagram or other model setup tool. This input file must be given a name in accord with the computer's operating conventions, say, INPUT.FIL for PCs. Our recommendation is that the input file be called **.eqs*, where * is a shorthand for the meaning of the model (e.g., for a study of intelligence, iq.eqs). EQS is then run (executed) by using the computer's conventional name for the program, typically, by typing EQS (or EQSS). This command calls the program, which then prompts you for the name of the INPUT.FIL. You will also be prompted to provide the name of an output file, say, OUTPUT.FIL, into which the results of the run will be placed. If necessary, a file containing the raw data (scores for all cases on the variables) can also be specified; in most computing environments, this will be done in the INPUT.FIL as described below. Directories, subdirectories, and other path information required to specify where these files are located are allowed, using the computer's typical conventions.

In general, an input model file contains several sections or paragraphs of material. Each of these sections or paragraphs is preceded by a slash and a keyword (e.g., /TITLE), where the keyword can be abbreviated to the first three characters. The keywords can be started in any column of an input line, but *the line that contains the keyword should contain no other input information.* Specific statements of information regarding the model and data must be provided within each section. Such information is typically separated by a semicolon (;). In this chapter and the next we will maintain ";" to get you used to it; later, we will drop it in the *Manual* even though it is almost always needed in the model file. You are urged to develop a standard practice that makes it easy to review the program input, e.g., starting keywords in column 1, and beginning statements below the keyword, perhaps indented a few characters. Not all of the sections are needed in any particular job.

As a shortcut, a list of consecutive variables may be specified using dash (minus sign) or TO, e.g. V11-V25, F1 TO F33, E6-E9, or D5-D7. In addition, variable labels defined in the LABELS section, below, may be used instead of V- or F-labels, anywhere in the model instructions. For instance, if the /LABELS section includes V5=HEIGHT; and V9=WEIGHT;, then HEIGHT-WEIGHT is equivalent to V5-V9. Labels should not be used before TO, since that will cause an error if the label is not exactly eight characters long. For this reason, we recommend using dash instead of TO.

In-line comments can be placed in any line of the input file by use of the exclamation mark (!). The program will ignore any material to the right of the exclamation. This feature is helpful in documenting the ideas behind a run.

It is easy to be overwhelmed by the many possibilities in a program like EQS. If you are new to structural modeling, you should make use of the simplest job specification possible, e.g., as shown in Chapter 2. Advanced options can be a source of confusion if not selected judiciously, building upon prior experience. This chapter also describes some options, such as /SIMULATION, that will only rarely be invoked by the data analyst, but that may be of special interest to the methodologist.

The input to EQS must be a plain file that contains no invisible control characters that may be generated by a program but that are invisible on a video display. Ordinary word processors use special characters and formats that must be avoided, although usually they also provide a way to save an acceptable plain text (*.txt) or ascii (*.asc) file

that strips the file of such special characters. The program expects the input to be specified with no more than 80 columns of information in each input line.

Covariance or correlation matrix data to be used in a model may reside in a separate data file. Raw data is always assumed to reside in a separate file. These data files may contain as many columns per line as the computer involved will allow. Specific instructions on how to integrate such files into EQS are given below.

In addition to providing input information, in a minicomputer/mainframe environment, you may have to provide some computer hardware/software-dependent control statements relevant to implementing EQS on your computer system. These statements control such things as the available CPU time, memory, etc. Information on these matters should be obtained from your computer center. Such information is not needed in interactive or standalone environments, e.g., on a PC.

/TITLE (Optional)

This section is identified by specifying the keyword /TITLE in an input record. On the next line following the keyword /TITLE (abbreviated to /TIT is acceptable), one or more lines of job information can be entered. An example of a typical /TITLE input might be the following:

```
/TITLE
 CONFIRMATORY FACTOR ANALYSIS RUN.
 PROBLEM BASED ON A RANDOMLY SELECTED SUBSET OF
 OBSERVATIONS FROM THE 1972 COHORT.
```

While the /TITLE section is not needed to run EQS, it is a good idea to be liberal with title information since it is sometimes the major source of information to help you identify the purpose of the analysis at a later time. A familiar occurrence is that what was all too obvious at the time of job submission turns out to be totally mysterious a few months later. Liberal use of title information is especially urged when dozens of EQS runs are generated. In that case, carefully labeled title information along with a path diagram provides a quick review of the purpose of an analysis. EQS prints all /TITLE information once, and repeats the first line as a header on all output pages. You should also consider making a photocopy of any critical notes, path diagrams, etc. associated with the run, and attaching this material to the printed output. In addition, a few notes marked in large colored writing on the face of the output are usually very helpful. If you do not print your output, keeping a log of runs with notes and annotations is even more critical!

/SPECIFICATIONS

This section is identified by entering the keyword /SPECIFICATIONS (abbreviated /SPE) on a new line. It gives the number of cases, the number of input variables, and the method (or methods) of estimation desired, as well as a variety of information to guide the EQS run. The items that can be specified are shown in the table below. The information in the /SPEC section can be placed in any order, but each separate type of information must be delineated by the semicolon (;). Information on numbers of cases and variables must be provided, unless the data file is an ESS file.

CASES

"CASES" refers to the number of subjects, observations, entities, respondents, etc. that were used to generate the covariance matrix, or that are in the input data file (if raw data are to be read). This number should be the number of cases in the input file even if cases are to be deleted during computations (i.e., this number will not change with such deletions). With covariance matrix input, this number is required in order to be able to compute statistics such as chi-square values and standard errors. Typically, e.g., in Windows, data are on an *.ESS file, i.e., a system file created by EQS; this file internally knows information about cases and will supply it as needed. In other situations,

when no number is given, the default value (100 cases) is used. If the default is wrong, the statistics will be wrong as well. Example: CASES = 200;

If you do not know how many cases are on a raw data file, you may give a large number, larger than the actual number of cases. EQS keeps a running tally of the number of cases of raw data read. It will continue to read data until the end of file is found. The actual number of cases read will be used in the computations and statistics. A warning message will also be printed.

VARIABLES

This number represents the number of variables in the input raw data file to be read, or in the covariance matrix to be read. It may or may not represent the actual number of variables analyzed in the program, because EQS automatically uses the model specification to search out the subset of input variables needed to perform the analysis. That is, if the input contains ten variables, but the model requires only six, the relevant six variables will be selected automatically by the labeling V1, V4, etc. used in the /EQUATIONS and /VARIANCES section. The program assumes that variables are ordered in the input matrix as V1, V2, V3, … . Example: VAR = 15;. Abbreviations, defaults, and other options available under /SPEC are summarized in the following table.

/SPECIFICATIONS Keywords

KEYWORD	ABBREVIATION	DEFAULT	MEANING
CASES	CAS = n;	100	Number of cases
VARIABLES	VA = n;		Number of input variables
METHOD	ME = = LS; = GLS; = ML; = ELS; = EGLS; = ERLS; = HKGLS; = HKRLS; = AGLS; = xx,ROBUST;	ML	Estimation methods: LS=Least squares GLS=Generalized LS ML=Maximum likelihood ELS=Elliptical LS EGLS=Elliptical GLS ERLS=Elliptical reweighted LS Heterogeneous kurtosis GLS Heterogeneous kurtosis RLS Arbitrary distribution GLS Robust statistics on method xx
DATA_FILE	DA ='INP.FIL';		Name of external file where input data resides
MATRIX	MA = COV; = RAW; = COR; = MOM;	COV	Covariance matrix input Raw data input Correlation matrix input Moment matrix input
ANALYSIS	AN = COV; = COR; = MOM; = ZSCORES;	COV	Analyze covariance matrix Analyze correlation matrix Analyze means & covariances Analyze covariance matrix calculated from *z*-scores
GROUPS	GR = m;	1	No. of input samples (at most 100)
KAPPA	KA = x;	EQS computes	Elliptical kurtosis parameter to override default
DELETE	DE =a,b,...,z;		Case numbers to be deleted (at most 20)
MCHAR	MC = c;	*	Character which denotes a missing

KEYWORD	ABBREVIATION	DEFAULT	MEANING
			value
MISSING	MI = COMPLETE; = ML; = PAIR;	COMPLETE	Discard cases with missing values ML, using all cases Pairwise present computations
FIELDS	FI = n;		Max. entries in row of rearranged covariance matrix
FORMAT	FO = (….);	Free format	FORTRAN format for input matrix
UNIT	UN = n;	UN=9; UN=8;	Input unit for raw data Input unit for covariance matrix (on unit-oriented computers)
CATEGORY	CAT = V7,V8;		List of categorical variables (at most 200)
WEIGHT	WE = 'WT.NAM';		File name for AGLS weight matrix
WFORMAT	WF = (….);	Free format	Format for WEIGHT file
LOOP	LO = n;	1	Run same analysis on n equal-size samples
WT or WTN or WTE	WT = V5;		Name of case-weight variable Use WTN for normalized weights Use WTE for effective weights
SE	SE = FISHER; = OBSERVED;	FISHER	Calculate standard errors from Fisher or observed information matrix
OUT	OU = HTML;		Write .HTM file in addition to .LOG file
CROBUST	CRO = #,#;		2 numbers for case-robust weighting
HK	HK = AVE; = GEO;	GEO	Use average kurtosis for HK methods Use geometric mean kurtosis
MULTILEVEL	MU = HLM; = ML; = MUML;		Hierarchical linear model ML multilevel analysis Muthén's approx. ML analysis
CLUSTER	CL = V4;		Variable for grouping cases in multilevel run
XDATA	XD ='XTR.FIL';		File written by lower-level run when MU=HLM;
XVAR	XV = n;		Number of variables to read from XDATA file
VSORT	VS = V1;		Variable for sorting data from DATA_FILE, above, so that case-order matches XDATA file

METHOD

EQS estimates by any one of the methods listed above and described in more detail in Chapter 5. If no method is specified, normal theory maximum likelihood (ML) is used. By the use of compound statements such as ME = ML, AGLS; the program allows the running several estimation methods with one job submission. In actuality, several methods may be run in a single job submission even if only one method is specified. The elliptical and arbitrary distribution methods are always preceded by a normal theory run, which is not necessarily printed out. E methods are by default preceded by their normal theory counterpart (ELS by LS; EGLS by GLS; ERLS by ML), although you can override this by specifying ME = LS, ERLS; for example. AGLS is by default preceded by LS, but the results are not printed. If you specify any prior method, by indicating xx, AGLS; where xx is any valid normal or

elliptical theory method, the method will be printed. The program can yield up to seven computational methods in a single job submission. Examples are:

METHOD	COMMENTS
`ME = LS, GLS, ML;`	are always computed in this order even if the methods are permuted;
`ME = EGLS, LS;`	gives LS, GLS, EGLS in sequence;
`ME = AGLS;`	gives LS, AGLS; but LS is not printed;
`ME = ERLS, EGLS, ELS;`	gives LS, ELS, GLS, EGLS, ML, ERLS.

The fact that a number of methods can be run with a single specification does not mean that it is a good idea to do this as a routine matter. It is easy to get overwhelmed with output.

Normal theory estimates are always obtained by an iterative process that starts at the input parameter values provided by you, or the program's default values, even if two methods such as ME = LS, ML; are specified. Elliptical and arbitrary theory estimates are obtained by starting the iterative process with converged values based on the previous method. If you want a run of these methods with parameter estimates starting at the values provided in program input, you must do this with a statement in the /TECHNICAL section of the input.

ME=HKGLS; and ME=HKRLS; invoke the heterogeneous kurtosis theory of Kano, Berkane, and Bentler (1990), designed to handle large covariance structure models when variables may have differing kurtoses. Elliptical theory and normal theory methods are a special case. This method requires raw data, and involves parameters that relate 2^{nd} to 4^{th} order moments. By default, HK=GEO; is used, where these relations are based on the scale free methodology of Bentler, Berkane, and Kano (1991) designed to yield appropriate weight matrices under a wide range of kurtoses. These relations involve the geometric mean of certain marginal kurtosis parameters. To obtain the original relational parameters based on averages, use HK=AVE;. The old methodology performed quite well in Hu, Bentler, and Kano (1992), while the new methodology seems not to have been further studied. As usual, GLS means that the weight matrix is not updated iteratively, while RLS means that it is reweighted and updated, as in ML.

AGLS is the arbitrary distribution GLS method for continuous variables, or the asymptotically distribution free (ADF) or minimum distance (MD) methodology for covariances, as it is sometimes known. It is best used when sample size is quite large (see Yuan & Bentler, 1997). The corresponding arbitrary distribution methodology for the analysis of correlation matrices (Leung & Chan, 1998; Steiger & Hakstian, 1982) is obtained when both ME=AGLS; and ANAL=COR; are used together. This forces the correlation matrix to be analyzed using a correct distribution-free weight matrix. AGLS estimates cannot be computed without raw data input. Elliptical estimates also require the use of raw data unless a value of KAPPA is provided.

ME = xx, ROBUST; means that you want to accept the xx estimates, but want to correct the chi-square and standard errors. The option provides robust statistics for any method xx other than AGLS. Robust statistics include 1) Satorra-Bentler scaled test statistic that is designed to have a distribution that is more closely approximated by χ^2 than the usual test statistic; 2) three residual based statistics namely, T_{RES}, $T_{YB(RES)}$, and $T_{F(RES)}$; and 3) robust standard errors that are correct in large samples even if the distributional assumption regarding the variables is wrong. The Satorra and Bentler (1988, 1994) scaled test and the residual χ^2 and F tests, $T_{YB(RES)}$ and $T_{F(RES)}$ (Yuan & Bentler, 1998a; Bentler & Yuan, 1999b) are the most reliable test statistics at this time under distributional violation. The robust standard errors (Bentler & Dijkstra, 1985; Browne, 1984) are sometimes known as coming from the "sandwich" estimator of asymptotic variances; it is based on the triple product of certain matrices (Kauermann & Carroll, 2001). These statistics are extended in EQS to cover a wide variety of situations beyond those described in the original publications, e.g., to multiple samples, correlation structures, categorical variable methods, case-robust methods, and so on. Robust statistics can only be computed if raw score data are provided on the input variables. Robust statistics are computationally very demanding. However, they perform better than uncorrected statistics where the normal distribution assumption is false and better than AGLS in all but the largest samples.

DATAFILE

This keyword can be used if: 1) the data to be analyzed is in an external file, i.e., it is not the covariance or correlation matrix given in the /MATRIX section; and 2) your computer system is interactive, as most systems today are. DA = 'INPUT.FIL' uses single quotes to enclose the filename, which can be any name that is acceptable to the computer's operating system. The contents of the file can be raw data, or a correlation or covariance matrix. Missing data is not permitted for matrix input.

If the operating system is not interactive, it will have a procedure for specifying external file names. See UNIT, below. Please consult your computer center's staff for assistance.

MATRIX

When a covariance matrix is input you do not need to specify MA = COV;, since this will be assumed by default. The notation MA = COR; specifies a correlation matrix. Typically, when a correlation matrix is input, standard deviations must also be provided, unless you also use ANALYSIS=COR;. MA = RAW; must be given when raw case scores are input. The words MATRIX, COVARIANCE, and CORRELATION may be spelled out in full. MA = MOM; refers to input in which the raw product moment matrix ($S + \bar{X}\bar{X}'$) has been computed externally, and is read. If MA=MOM; is used with ANAL=MOM; (see below), the means must also be read in.

The matrix to be read may be in the input or in a separate data file, with no missing elements in COR or COV. MA = RAW; allows missing data. While a covariance or correlation matrix may be in either location, the raw data matrix is always assumed to reside in an external file. The name of the file is usually specified with the DATA_FILE statement, above. In Windows and Mac environments, the raw score matrix will be read in and become automatically available as an *.ess file. In some mainframe systems, external files are located with reference to a unit number and a file name, as described below.

ANALYSIS

In addition to specifying the type of matrix that is in the input file, or in the external data file, you can specify the type of matrix to be analyzed. The program will transform the input data into the correct form for analysis; e.g., a correlation matrix can be transformed into a covariance matrix if you also provide standard deviations. The default is ANAL = COV; i.e., a covariance matrix will be analyzed. If this default is desired, the ANALYSIS statement is not needed. The program will automatically transform available input data into a covariance matrix for analysis, without a special instruction. If you use this default, or specify ANAL=COV; when in reality your input matrix is a correlation matrix, some statistics may be questionable (see, e.g., Cudeck, 1989).

If you require the analysis of a correlation matrix that will assure correct statistics, you must specify ANAL=COR;. If you specify a normal theory method, such as ME=ML;, the program will use an extension of Jennrich's (1970) method for a correct correlation structure analysis. With raw data, you have the option of correcting the test statistics and standard errors with ME=ML, ROBUST; using versions of the Satorra-Bentler corrections, residual test statistics, and robust standard errors that are appropriate to correlation matrices. A similar sequence occurs if, at the same time, you use CAT=xx,yy; indicating that polychoric/polyserial correlations should be computed. Estimates are first computed by the chosen method (e.g., ME=ML;) and then corrections are made. The corrected statistics are, obviously, the ones to rely on.

ANAL = COR; also yields correct statistics when AGLS and/or CAT models are implemented. If ANAL=COR; *and* ME=AGLS; are used together, then a statistically correct arbitrary distribution (AGLS) methodology for the analysis of correlation matrices (de Leeuw, 1983; Leung & Chan, 1998; Steiger & Hakstian, 1982) is obtained. When ANAL=COR; *and* ME=AGLS; *and* CAT=xx,yy; are used together, the Lee-Poon-Bentler (1995) methodology for categorical variables is implemented.

A new option is ANALYSIS=ZSCORES;, which converts raw data to *z*-scores, computes the covariance matrix from those *z*-scores, and then analyzes the covariance matrix using any method that you specify. You should recognize that this covariance matrix is, in fact, a correlation matrix. This may be a useful approach if variables are scaled so differently that convergence has proven impossible in the covariance metric. Or, it simply may be

interesting to see what estimates or other statistics look like when based on correlational data. However, some statistics may not be correct as noted in the literature (e.g., Cudeck, 1989).

When ANAL = MOM; is specified, 1st and 2nd moments, i.e., means and covariances, are to be analyzed. This option is used with structured means models (see Chapter 8), in which equations contain the intercept V999, and raw score data and/or means for the variables are available.

GROUPS

This keyword can be omitted unless a multisample analysis (see Chapter 7) is undertaken. In multisample analysis, several datasets are analyzed simultaneously. GROUPS = m; informs EQS that data from m groups or samples are to be analyzed. GR = 2; is an example. A maximum of 100 groups can be analyzed together. GR = 1; is the default and need not be specified.

KAPPA

This coefficient is used with elliptical computations as described in Chapter 5. As a default, with raw data input, EQS computes the Mardia-based coefficient and uses it in the computations. However, if this does not fall within the appropriate range, EQS substitutes the mean scaled univariate kurtosis. If this also does not fall in the appropriate range, it is adjusted further. Thus, if these program defaults are acceptable, KAPPA does not need to be defined in the /SPE section. However, you have the option of overriding the program's default values by specifying any given value. Whatever the value you provide, it will override any values computed in the program. Thus, you could evaluate the alternate coefficients mentioned above, or additional coefficients that are computed and printed out when AGLS estimation is used, and decide to use one of these coefficients in subsequent runs. You must provide the actual numerical value of the coefficient that is to be used, e.g., KAPPA = .2;.

Specification of KAPPA also allows for the use of elliptical estimation when covariance matrix input (i.e., no raw data) is being used. Thus, if many elliptical runs with very large data sets will be done, kappa and the covariance matrix could be computed once and saved, to be used in subsequent runs.

DELETE

When raw data are read in, it is possible to eliminate the scores of certain cases from the computation using the DELETE statement. EQS reads input score vectors and labels them sequentially 1, 2, ... The DEL = 2, 17, 45; statement tells EQS to ignore the data from case numbers 2, 17, and 45. Obviously, each integer number given in the DEL statement cannot exceed the total number of cases in the input file. Up to 20 cases can be deleted with the DELETE statement. The numbers in the DEL statement can be in any order.

MCHAR

If there are missing scores in the raw data file, you must give a character in place of the missing datum. This character should unambiguously distinguish between actual scores and non-scores. The character can be any keyboard character that does not represent a possible numerical value of scores in the data file, and can be specified as MC=c; where "c" is that character used in place of a missing value. By default, MC=*; that is, "*" is used in place of missing values.

MISSING

There are three options for dealing with missing data, specified with the MISSING command. MI=COMPLETE is equivalent to list-wise deletion. Only cases with complete data are used in the analysis. This is the default procedure. When only a few cases have missing data, it is appropriate. A technically more adequate procedure is to use case-wise maximum likelihood, specified as MI=ML;. Based on the methodology of Jamshidian and Bentler (1999), this is an EM-type computation that provides optimal results when the data are multivariate normally distributed. When data are not normal, however, the associated test statistics and standard errors can be misleading. Yuan and Bentler (2000d) provided the technical development for corrections to the test statistic and standard errors that parallel the ME=ML,ROBUST; option with complete data. This is specified using the statements MI=ML; and ME=ML,ROBUST;. As a consequence, a Yuan-Bentler scaled test statistic analogous to the Satorra-Bentler scaled chi-square is computed, and robust standard errors are also provided. The residual-based statistics T_{RES}, $T_{YB(RES)}$, and $T_{F(RES)}$ also are computed. Finally, you have a third option. You can analyze the pairwise present covariance

matrix, i.e., the matrix that is computed from all available data. Historically, this method did not provide accurate statistics (see, e.g., Marsh, 1998), but a new robust methodology as well as a distribution-free methodology based on van Praag, Dijkstra, and Van Velzen (1985) are now available in EQS. You specify MI=PAIR; and then the method that you want to use, e.g., ME=ML, ROBUST; or ME=AGLS;. Chapter 5 gives more information. Since the AGLS method for complete data is a special case, it is not surprising that the AGLS method requires huge sample sizes (Gold, Bentler, & Kim, 2002).

FIELDS

When a covariance or correlation matrix is read in, EQS assumes that the matrix is in lower triangular form. Computer file output probably already is in this format. There are two exceptions to a lower triangular requirement. First, a full symmetric input matrix may be used if a format statement is provided; this is discussed further under FORMAT. Second, a FIELDS statement can be used that allows a rearranged lower triangular input matrix to be read in. This type of rearranged matrix is available as printed output from standard statistical packages. Thus FIELDS would usually be used when a matrix created by such packages is used in the input stream or via an external data file. FIELDS is not relevant to raw data input. Elements of a seven variable lower triangular input matrix are given, and computer-read, in the sequence

1						
2	3					
4	5	6				
7	8	9	10			
11	12	13	14	15		
16	17	18	19	20	21	
22	23	24	25	26	27	28

where elements 1, 3, 6, 10 ... are variances and the other elements are covariances. If such a matrix has very few fields of data in a given line, it is universally printed as shown. The above example has a maximum of seven fields, which is the number of separate data in the last row. This way of printing output becomes difficult with a large number of variables. For example, if the matrix represents 50 variables, the last row would contain 50 covariances (including one variance). However, if each datum required four columns of space, these 50 numbers could not be printed in one row. Either the numbers must be printed sequentially using several lines for a single variable, in which case the matrix will no longer visually appear to be lower triangular, or the lower triangular matrix format must itself be modified. FIELDS describes the modified matrix.

The standard modification to printing the lower triangular matrix is to cut off the matrix after a certain number of fields (usually 12 or so), and to move the entire section after that number of fields to a subsequent part of the output. This permits all covariances or correlations having a certain column (variable) designation to be printed sequentially. In the above example, suppose one sets FI = 3;. In that case, the program expects the input matrix to be in the form

1		
2	3	
4	5	6
7	8	9
11	12	13
16	17	18
22	23	24
10		
14	15	

19	20	21
25	26	27
28		

When taken from some statistical packages, the matrix may also contain column variable designations, but EQS assumes, as in the example, that this extraneous information has been removed from the file. The FIELDS statement thus identifies the maximum number of input data in one row of a covariance matrix that has been rearranged as shown above. It can be used with a full symmetric matrix, provided a format statement is used. FI should not be used with an unmodified lower triangular matrix. *If a FORMAT statement is given, the number of fields in FO (e.g., 18F1.0) must be equal to the number of fields in FI.* FIELDS also directs EQS to reorganize the matrix into an ordinary lower triangular matrix prior to the computations. Verification of the program's actions can be found by looking at the printed output section labeled Matrix To Be Analyzed.

FORMAT

The format statement is needed if 1) numbers in the file abut one another with no separating space, comma, or tab, 2) some columns in the file must be skipped, 3) a full covariance or correlation matrix is input, and a row of the matrix occupies more than one record, or 4) some numbers without decimal points need to be scaled automatically, e.g. 123 is to be interpreted as 1.23. Raw data are assumed to be in a case (row) by variable (column) matrix. The format represents the scores for one case of raw data to be read. The CASES statement informs the computer how many such case scores are to be read.

The FORMAT statement should be embedded in a pair of parentheses, e.g., FO = (6F3.0); where, as usual, it must end with a semicolon. The format statement can wrap around to a new line, but cannot exceed a total of 240 characters, i.e., three lines of format is the maximum possible. Formats may include any of the following edit descriptors: D, E, F, G, I, P, X, T, TR, TL, BN, BZ, slash, colon, comma, and parentheses. See a FORTRAN manual for details. Descriptors which are commonly used are X, F, and slash. X means skip (ignore) one column, 8X means skip 8 columns, etc. F means floating-point input. For example, 3F7.2 will read three fields of 7 columns each. In each field, if no decimal point is found, one will be placed between columns 5 and 6. Thus F7.2 transforms the number 1234567 to 12345.67. When a data value being read contains a decimal point, the decimal point in the data overrides the decimal point implied by the F-type format, e.g., 10.3 can be read using F4.0, F4.1, F4.2, or F4.3. The minus sign, if present, is part of the field width. Slash (/) instructs the program to skip the remainder of the record and begin the next record. It may be used when the raw scores of a case are recorded on more than one line. An example is FO = (3X,2F2.0,1X,5F3.2); giving instructions on how to read 23 columns of input information.

UNIT

The UNIT keyword is not needed in interactive systems like PCs. In older mainframe and batch systems like IBM OS/MVS and IBM VM/CMS systems, if raw data are to be read, the location of the input data file must be given in integer form. The unit number is a Fortran logical unit, which can be any number less than 99 (except 4, 6, 8, 11, 15, 20). If no value is specified, UN = 8; is assumed when MATRIX = COV; and UN = 9; is assumed when raw data (MA = RAW;) is to be read. Information on assigning unit numbers to files should be available at your computer center.

CATEGORY

The category keyword specifies the variables that are categorical, in which ordinary correlations are to be replaced with polychoric and polyserial correlations based on Lee, Poon, and Bentler (1995). It is assumed that any categorical variables to be modeled are categorized versions of variables that are truly continuous, as well as multivariate normally distributed. The correlation between the underlying variables yielding two categorical variables is known as the polychoric correlation. The correlation between the underlying variables that generate a categorical and a continuous variable is the polyserial correlation. First, these correlations are estimated without any concern for the structural model under consideration. Then this correlation matrix is then considered to be a function of more basic parameters that can be estimated via two different approaches. The first is the ME = ML, ROBUST; approach that accepts EQS's RLS-ML estimator for correlations based on continuous variables, and corrects the chi-square and test statistics using a large optimal weight matrix that is appropriate for categorical data.

This method is new in EQS 6 and is probably the better option unless sample size is very large. The second uses the relevant AGLS procedure, also using the optimal weights. Both require ANAL=CORR;, the correlation matrix.

There are several requirements to run a model with categorical data. First, raw data must be analyzed. Second, those variables that are categorical must not have too many categories. In practice, about 5-7 or so is a practical maximum, but more are allowed. However, with many categories, the variable may as well be treated as continuous. See e.g., Babakus, Ferguson, and Jöreskog (1987) or Olsson (1979b). Third, a reasonably large sample size is highly recommended. Fourth, you may have up to 200 categorical variables, but more than 20 may cause problems with ordinary sized samples. A reason for these requirements is that all categorical data methods require the cross-tabulation tables of the categorical variables. These tables will be very sparse if there are too few subjects or too many categories, and hence the computational procedures can easily break down.

To specify a categorical data model, in the /SPECIFICATION section, you must indicate which variables are categorical, using a statement such as

```
CATEGORY = V2,V5;
```

The abbreviation CAT is sufficient. The program will figure out the scoring of your variables. Also, specify that MATRIX = RAW; and point to the external data file using the DATA statement. The program will use ME=ML, ROBUST; and ANAL=CORR; as the default method when doing the analysis. Remember that the initial method, the RLS/ML correlation structure method, is fine for the estimates, but that only the corrected statistics should be interpreted. As noted above, you also can choose ME=AGLS; instead.

Note: There is one limitation to the current implementation. Consistent with the statistical theory, all measured variables in models with categorical variables must be *dependent* variables. However, we can trick the theory. If you want to use a measured variable as an independent variable, you can create a dummy factor to represent it. For example, if you want to include V7 in your model as an independent variable, create an equation like V7=F7; and use F7 in the model as if it were V7. Research will be needed to evaluate this procedure.

WEIGHT and WFORMAT

An external matrix can be read in, and inverted, to replace the default weight matrix used in AGLS estimation. Elements must be arranged in sequence, with rows and columns of the matrix corresponding to the *rearranged* covariance matrix, in which dependent variables are followed by independent variables. The read-in matrix will be inverted within EQS to yield the W matrix given in equation (1.5). *Note that the matrix to be read in for correlation structures has the same dimensions as that for covariance structures.* The diagonal elements corresponding to variances should equal 1.0, and all corresponding off diagonal elements should be zero. Rules for WFORMAT are the same as for FORMAT, above.

LOOP

To run the same analysis repeatedly on equal-sized subsets of a raw data file, use LOOP. For instance, if the data file has 600 cases, then LOOP=12; CASES=50; could be used to analyze each 50-case subset of the data.

WT, WTN, or WTE

To produce a case-weighted mean and covariance matrix, a variable containing a nonnegative number for each subject must be defined. The case weight variable may be specified by using WT = …; WTN =…; or WTE =…. WT uses the raw values of the specified variable in the data file. WTN scales or normalizes the raw weights so that their mean is 1.0 or their sum is sample size. WTE produces a sum of weights equaling the sum of squares of the weights; the sum is the "effective" sample size. WTN or WTE are usually the more meaningful specifications. The weight variable is handled as in BMDP and SAS. A good discussion of the use of weights in the context of sampling is given by Kaplan and Ferguson (1999) and Stapleton (2002). Cases whose weight is missing or nonpositive have no effect on the calculations. If MISSING=ML; is given, missing values are imputed for these cases, but not for the weight variable itself. If raw data and/or factor scores are saved (see the SAVE section), the imputed values are saved for these cases.

SE

Standard errors for many basic methods may be calculated using either the Fisher or observed information matrix. The former, the ordinary information matrix, is the default. The latter is based on second derivatives or Hessian of the fitting function that may provide more accurate standard errors in small samples. Computations are based on a numerical method that conceivably might fail. If all i,j elements of the Hessian matrix equal their j,i elements, the matrix is computed correctly and a printed symmetry index (called SYMMETRY INDEX FOR NUMERICAL SECOND DERIVATIVES) will be close to zero. Specify SE = OB; to get the observed information matrix and its symmetry index.

OUT

The designation LOG is the default output file extension. To create a version with HTML flags so that it can be read by a browser or included in a web page, specify OUT=HTM;. The name of the HTM file is the same as the name of the LOG file, except that its extension (the part of the name after the period) is HTM instead of LOG.

CASE-ROBUST

Ordinary means and covariances are influenced strongly by outliers and other influential cases. *Case-robust* estimates of the means and covariance matrix are insensitive to such influences on case scores. EQS uses a methodology reviewed by Campbell (1980) and developed for structural modeling in Yuan and Bentler (1998) to get case-robust means and covariance matrices. These are called ROBUST ESTIMATES OF MEANS AND COVARIANCE MATRIX BASED ON CASE ROBUST METHOD. The command is CROBUST = #, #;. If only CROBUST or CRO is given, then b1=2 and b2=1.25 are used as default. If only b1 is given, then b2=1.25. Both numbers must be positive. MATRIX = RAW; is required. Most of the resulting statistics, including a scaled chi-square, are based on the Yuan-Bentler reference. However, EQS also produces several residual based test statistics, adapting and generalizing the methodology of Yuan and Bentler (1998a) to the case-robust situation.

HK

This command stands for Heterogeneous Kurtosis and defines the parameter that relates 2^{nd} to 4^{th} order moments under theory developed by Yutaka Kano and his colleagues. When METHOD = HKRLS; or ME=HKGLS;, by default the HK=GEO; specification is used. This involves the relational parameter defined in Bentler, Berkane, and Kano (1991) involving the geometric mean of certain marginal kurtosis parameters. In order to obtain the HK correction described in the original article (Kano et al., 1990), involving the average of marginal kurtosis parameters, use HK = AVE;. If the average method fails, the alternative HK=GEO; specification can still yield a workable methodology.

MULTILEVEL

The keywords MULTILEVEL, CLUSTER, XDATA, XVAR, and VSORT are explained in Chapter 11 on multi-level analyses.

Examples

Examples of valid /SPECIFICATIONS section keywords are:

```
/SPECIFICATION
CAS = 348; VAR = 15; METHOD = ML;

/SPE
ME = LS; VAR = 17; CAS = 200; DATA = 'INPUT.FIL';

/SPECS
VARIABLES = 15; CASES = 257; MATRIX = COR; ANALYSIS = COV;

/SPEC
CASES = 60; VARIABLES = 6; FIELDS = 5;
MATRIX = COVARIANCE; METHOD = GLS;
```

```
/SPECIFICATIONS
VAR = 6; MA = RAW;

/SPEC
CAS = 112; VAR = 6; ME = ML,AGLS; UNIT = 9; MA = RA;
FO = (6X,3F5.1,10X,3F5.1); DEL = 24, 88;
```

/LABELS (Optional)

/LABELS (abbreviated /LAB) is a keyword that signals that identifying labels for the observed and/or latent variables are to be used. The use of V followed by a numerical value (such as V3) denotes that a label for the third measured variable follows. An F denotes a label for a factor or latent variable.

Labels may be one to eight characters in length and may be assigned only to F or V type variables (see **Types of Variables** below). Blanks will be suppressed in printing, so a blank should be replaced by an underscore or similar symbol. A label may not have the same name as a variable in the model, e.g. V3=E5; is disallowed if E5 is in the model. Of course, V3=V3; is allowed. As an example, the name V3=MATH; defines the label for the third variable of the variance-covariance or raw data matrix as MATH. Similarly, F2=VERB; attaches the label VERB to the second latent variable. Labels are not needed for variables that are not used in an analysis. Thus, if a covariance matrix of 20 variables is input in the MATRIX section, but only 13 variables are used, labels for the unused variables are superfluous. Fewer than 13 variables may be labeled if so desired.

As stated previously, the program automatically assigns V1 to the first variable, V2 to the second, and so forth. If the data are read from a *.ESS file, labels of the input variables are also read from the file. Latent variables will automatically be assigned the names F1, F2, F3, etc. Any or all of the labels may be overridden by labels specified in this section. The labels may be given in any order, but each must be followed by a semicolon. Examples of information for the /LABELS section follow.

```
/LABELS
V1 = SES1; V3 = SES2; V8 = VERB; V11 = PERFM;
F1 = SES; F2 = IQ;

/LAB
F1 = SEXRESPN; F2 = SEXEXPER;
V1 = ORG_FREQ; V2 = ORG_INTS; V3 = NUMBDATS;
V4 = NUMBPART; V5 = AGE;
```

/EQUATIONS (Optional)

The /EQUATIONS keyword (abbreviated /EQU) signals that detailed information about the particular model under examination follows next. One and only one equation is required for each dependent variable. The dependent variables may be either observed or latent, and parameters within the equations may be specified as either fixed or free. Possibly you may not need to specify /EQU yourself. Models that contain a large number of equations can be tedious to specify, and in many cases, a model can be defined quickly and easily by using the MODEL or RELIABILITY commands (which are described below). However, these sections always build an /EQUATIONS section, and hence you must know how it works.

Types of Variables

Equations can be written in a text-like format, as is described below in the section titled **EQU, VAR, COV with Labels**. However, the basic model specification in EQS uses four types of variables, and it is important to understand their meaning. The measured variables and latent variables have already been referred to by the

characters “V" and “F", respectively. Residuals of observed variables are denoted by “E", while residuals or disturbances of latent variables are denoted by “D". For reference, these codes are summarized here:

CODE	NAME	MEANING
V	Variable	Measured variable
F	Factor	Latent variable
E	Error	Residual of measured variable
D	Disturbance	Residual of latent variable

It is recommended that you label all the variables in your path diagram in accord with these conventions. Information necessary for the /EQUATIONS section can then easily be read off the diagram. Following traditional usage in such diagrams, the observed variables (denoted by rectangles) will be labeled by V's, while the latent variables (diagrammed as circles or ovals) are labeled by F's.

Each variable having a V, F, E, or D designation can be followed by a one-, two-, or three-digit integer. Thus, the range of variable numbers within each variable type is from 1 to 999. Examples are V3, F6, D278. Note that leading zeros are suppressed. We also recommend that your residual variable numbers match the corresponding V or F variable, that is, that the equation for V23 uses the residual E23, and the equation for F12 uses D12 as its residual.

Independent and Dependent Variables

Independent variables are defined in EQS as those measured variables or latent variables that are never structurally regressed on any other variables in the system. When working from a path diagram, independent variables are easily recognized; they are the variables that have no directional arrows aiming at them. Dependent variables are those that are structurally regressed on at least one other variable in a model. It is possible for a dependent variable to also be a predictor of other variables in a model. However, this does not mean it is an independent variable in the sense used by EQS. If a variable is ever structurally regressed on any other variable in the model, i.e., if it appears at least once on the left-hand side of one equation in the /EQU section, then it is a dependent variable, regardless of its relationships to other variables in the system. Dependent variables are also easily recognized in a path diagram; they are those variables with one or more unidirectional arrows aiming at them.

Only V and F variables can be dependent variables, i.e., on the left side of an equation. E and D variables must never be on the left side of an equation. As residuals, they are always independent variables. If your model requires that an E or D variable be a dependent variable, you can simply rename it as an F variable.

Equations Format

Each dependent variable in the model must be given one and only one equation. Thus there are exactly as many equations as there are dependent variables in a model. For each parameter within an equation, an asterisk or star (*) after a numerical value specifies that the parameter is to be estimated, with the number to the left of the asterisk as a starting value. If no number is given next to *, the program will insert the start value. Numerical values not followed by an asterisk specify that the parameter is to be fixed to the numerical value.
As an example, consider the following two equations:

```
/EQUATIONS
V2 = .6*F1 + E2;
F3 = .78*F1 + .64*F2 + D3;
```

The first equation characterizes the information for the measured variable V2. It states that V2 is based upon a regression on F1 plus a residual term E2. The parameter for V2 on F1 is to be estimated, and a starting value of .6 is to be used. If the equation had been V2 = .6F1 + E2; then the value .6 would be interpreted as a fixed value, never changed by the program. Of course, the first equation is exactly the same as V2 = .6*F1 + 1.0E2; where 1.0 is a

fixed parameter. It is possible to use the equation V2 = .6*F1 + 1.0*E2; where 1.0 is a free parameter, but this approach will almost always require fixing the variance of the E2 variable. The second equation describes the structural equation for the latent variable F3. It shows that F3 is a function of F1, F2, and the residual D3. Two parameters are to be estimated in this equation, with starting values of .78 and .64. Note that embedded blanks are permissible in the equations, as well as completely blank lines. Each equation must end with a semicolon.

A variable can appear only once in an equation. Thus, V1 = .2*F1 + .6*F1 + E1; is not permitted. If such a situation is encountered, EQS will ignore all but the last occurrence of the duplicated variable. This example would be interpreted as V1 = .6*F1 + E1. If a variable appears on both the left and right sides of an equation, EQS prints an error message and stops the run.

Equations can be of any length and can span several lines if necessary. However, the parameter estimate and the associated variable designation (e.g., .5*F2) cannot be split and appear on two lines. An equation can be wrapped around to the next line, but, as usual, the equation must end with a semicolon.

The sequence of equations is arbitrary, so they can be put in any order. You are urged, nonetheless, to group the measurement equations (equations for V variables) together, and similarly to group the construct equations (equations for F variables) together, as well as to sequence equations in ascending numerical order within each set. Equations organized in this way are more easily scanned for possible errors. Blanks can be inserted into equations so that they more visibly reflect the intended structure. Thus, V1 = .5*F1 + E1; and V2 = .3*F2 + E2; might be written as follows when embedded in a large number of equations

```
/EQU
V1 = .5*F1      + E1;
V2 =      .3*F2 + E2;
```

making the factor structure more visible. However, care must be taken to avoid unseen control characters such as those stemming from tabs. And, as noted above, whenever possible, the numerical designation for V's and E's should correspond in a given equation. Thus, while one could write V2 = .6*F1 + E13; where the number 13 is arbitrary, this practice is liable to lead to confusion. Similarly, it is suggested that the disturbances D be numbered in accord with the number of the dependent factor F.

Start Values

Start values are not needed for most jobs. That is, equations of the form V4 = *F1 + *F2 + E4; are acceptable, and the program will generate values for the * parameters that, in most cases, will allow the iterative procedure to find the optimal estimates. The typical default start value is 1.0*. However, it is probably good practice to provide start values because EQS's start values may be inadequate in some models and thus the program may not converge. More importantly, providing start values helps to clarify which coefficients are expected to be large positive or large negative. The final results then can be informative about one's expectations.

Because of the varying types of models that are run with EQS, it is difficult to give rules about the relative size of coefficients in equations. Actually, it is only necessary to have a reasonably good guess about the size and sign of a few key start values, since the relative size of others then tends not to matter much in the iterative calculations. Thus a few key factor loadings should always be on the large size. What is “large," of course, depends on the scale of the variables, how unmeasured variables are identified, and on the true model. (If necessary, a few runs with small models on parts of the data, or of a complete model, will quickly establish this information.) Coefficients for predictor and criterion variables of a similar designation (e.g., predicting a V from a V, or an F from an F) and conceptual nature (e.g., first-order factors) can usually be started at small values such as .1 or zero. Of course, if one knows that an effect, such as a factor loading, is strong, the start value should reflect this information. The better the start values, the quicker the iterative calculations will converge.

Selecting a Subset of Variables

The information in the /EQUATIONS section can be used to select a subset of variables out of the original input covariance or data matrix. EQS scans the equations, and then selects only those variables from the input covariance or data matrix that are specified in the /EQUATIONS or in the /VARIANCES section. For example, if an input covariance matrix has fifteen variables, but the equations contain only V2, V6, V8, V9, and V10, then only those five variables will be selected as data to be analyzed. The remaining variables will be ignored unless the variable is mentioned in the VARIANCES section, in which case it will also be selected. The covariance matrix to be analyzed will be constructed with V2 as the first variable, V6 as the second, and so forth.

/VARIANCES (Optional)

The /VARIANCES section (abbreviated /VAR) specifies fixed and free values for the variances of the independent variables. In EQS, E and D variables are always independent variables. *Dependent variables are not allowed to have variances as parameters, whether fixed or free.* All other variables, the independent variables, must have variances and in principle need to be specified. However, if by mistake you do not do so, the program will correct your error and add the specification internally.

As in the /EQUATIONS section, numerical values followed by "*" denote free parameters that are to be started at the numbers you provide. Numerical values not followed by "*" are considered fixed parameters. As always, each sentence or piece of information is terminated by a semicolon. A sequence of variances can be started at the same value by using the dash "-" convention, or equivalently, the "TO" convention. ("TO" but not "-" may break down in the special case LABEL1 TO LABEL2, if LABEL1 has fewer than 8 characters.) Some examples are:

```
/VARIANCES
E1-E6 = .67*; E7-E9 = .5*;
F1 TO F2 = 1.0; D3 = 3.4*;
```

Here the variances of the two latent variables F1 and F2 are fixed parameters with values of 1.0, while the remaining variances are free to be estimated. The first six E1, E2,..., E6 are started at .67, while the variances of the residuals E7, E8, and E9, are started at .5. The latent variable residual D3 has its variance started at 3.4. As usual, the better the start value, the quicker the program will converge to a solution.

Variables also can be listed in sequence with a single start value. When some variables are in a sequence of variables generated by the "-" convention, and then *subsequently* listed separately with particular start values, the subsequent start values *override* the start values given with "-". Thus,

```
/VAR
F2,F4,V8 = 1;
E14-E20 = 0.5*; E15,E17,E19 = 0;
```

fixes the variance of F2, F4, and V8 at 1.0, while the second line is equivalent to E14=.5*; E16=.5*; E18=.5*; E20=.5*; since the fixed 0 values override the sequence specified by "-". Blanks are ignored.

As noted in Chapter 8, V999 is a special independent variable (a constant) that has variance fixed at zero. This need not be specified; it will be assumed.

As always, a logical ordering of information will make viewing the input and understanding its purpose much easier. It is possible to write multiple sentences of information on a single line, as long as a sentence is not split. *Each line must end with a semicolon.* You are urged to group the variance information in a meaningful way across several lines whenever possible.

If a measured variable is specified in this section, it will also be selected in the creation of a subset covariance matrix to be analyzed. If a measured variable is not mentioned in either /EQUATIONS or /VARIANCES, it will not be used in the analysis.

Start Values

As usual, EQS will provide simple and typically adequate start values where none are given. Most variances are started at 1.0*, while E variables will have variances started at .9 times the variance of the associated V variables. Independent V's have their variances started at the input data values. Of course, user-specified values can be helpful in creating quicker convergence to a solution, though it can be difficult to get a feel for the appropriate size of a start value. In the case of variances, it is a good practice to make them quite large relative to the covariances that might be in a model. For example, the variance for an E term should be some substantial fraction of the variance of the corresponding measured variable. Overestimation of the variances of the independent variables only rarely creates computational difficulties. On the other hand, underestimation will frequently lead to singularities in the model matrix, which may create iterative problems with several types of estimators. Variances started near zero are usually problematic.

/COVARIANCES (Optional)

The /COVARIANCES keyword (abbreviated /COV) begins the section specifying the fixed and free covariances among independent variables. *Dependent variables cannot have covariances.* If a variable is involved in a covariance, its variance must also be specified in /VAR. A pair of variables whose covariance is to be estimated by EQS is specified by giving both variable names followed by a numerical value, and optionally, an asterisk or star if the parameter is to be estimated. If the "*" is deleted, the covariance will be fixed at the number. For example,

```
/COVARIANCES
F2,F1 = .3*;
```

specifies that the covariance between variables F2 and F1 is a free parameter, started at the .3 value. As always, fields in the /COVARIANCES section must be delimited by semicolon. Although the pair of variables involved in a covariance can be specified in any arbitrary sequence (e.g., the above example could be given as F1,F2 = .3*), you are urged to adopt a consistent way of specifying covariances, for example, by systematically listing the covariances as if taken from a lower or upper triangular matrix form. In the former case the subscript for the variable with the larger number is always written first e.g., F2,F1 = .4* and rows of the lower triangular matrix are listed sequentially.

Covariances that are not specified in this section are assumed to be fixed zeros, so there is no need to specify fixed zero covariances. However, fixed nonzero covariances among independent variables also have to be specified. And, of course, all free covariance parameters must be included and provided an asterisk for identification.

The "TO" or dash "-" convention can be used to generate all possible lower triangular pairs among a set of variables of the same type. For example,

```
/COV
E1-E4 = .2*;
```

is equivalent to

```
/COV
E2,E1 = .2*;
E3,E1 = .2*; E3,E2 = .2*;
E4,E1 = .2*; E4,E2 = .2*; E4,E3 = .2*;
```

It is apparent that /COV excludes the diagonal elements of a covariance matrix, which must be specified in the /VAR section. Furthermore, TO or dash cannot be used to connect different types of variables, e.g., E1 TO F3 = .3* does not make any sense to the program.

When the TO or dash convention is *followed* by specific covariances that are not consistent with the TO or dash statement, the specifically designated covariances *override* the statements generated by TO or dash. For example,

```
/COV
E1 TO E4 = .2*; E3,E2 = 0; E4,E3 = 0;
```

is equivalent to

```
/COV
E2,E1 = .2*;
E3,E1 = .2*;
E4,E1 = .2*; E4,E2 = .2*;
```

so that four rather than six free covariances are created.

As in the /VAR section, V999 is a special variable; it always has fixed zero covariances with other variables that need not be specified. Also, blanks in the input will be ignored, *and each line must end in a semicolon*. As before, a good visual layout for the specifications will make them easy to understand and review.

Start Values

Start values for covariances to be estimated should usually be small, compared to the size of the variances of the variables involved. In fact, unless you have some knowledge that gives reason to predict a relatively large positive or negative covariance between a pair of variables, a start value of 0* is usually safe, as well as effective. In contrast, large values frequently lead to iterative problems. The default start value of a covariance is 0*, except for V variables whose covariances are started at the input values. Thus a statement like E2,E1 = *; will have the initial parameter estimate taken at zero.

EQU, VAR, COV with Labels

Although the basic model setup involves the use of V, F, E, and D variables, this is not essential. It is possible to use labels for the V and F variables instead. Once the labels have been defined, /EQS, /VAR and /COV sections, as well as /CON will allow the labels to be used directly. However, the E and D variables must be specified as usual. The idea is easily illustrated with an example.

```
/LABELS
V1 = PERFORM;  V2 = SATISFC1; V3 = SATISFC2; V4 = ACHMOTV1;
V5 = ACHMOTV2; V6 = SLFESTM1; V7 = SLFESTM2; V8 = VERBALIQ;
F1 = JOBSATSF; F2 = N-ACHIEV; F3 = ESTEEM;
/EQUATIONS
PERFORM  =                      1*ESTEEM + .5*VERBALIQ + E1;
SATISFC1 =   JOBSATSF                                  + E2;
SATISFC2 = 1*JOBSATSF                                  + E3;
ACHMOTV1 =             N-ACHIEV                        + E4;
ACHMOTV2 =           1*N-ACHIEV                        + E5;
SLFESTM1 =                       ESTEEM                + E6;
SLFESTM2 =                      1*ESTEEM               + E7;
JOBSATSF =           1*N-ACHIEV          + .5*PERFORM  + D1;
/VARIANCES
VERBALIQ = 10*;
N-ACHIEV TO ESTEEM = 1*;
E1 TO E7 = 5*;
D1 = 1*;
```

```
/COVARIANCES
N-ACHIEV, ESTEEM   = *;
N-ACHIEV, VERBALIQ = *;
ESTEEM,   VERBALIQ = *;
```

/MODEL (Optional)

The MODEL section can be used *instead of* EQUATIONS, VARIANCES, and COVARIANCES, above. (If used in addition to one or more of these 3 sections, the program will stop.) In large models, MODEL is much simpler and easier to use than those sections, unless there are many different start values or fixed values for parameters. There are seven different types of commands in the MODEL section:

(variable list)	ON	(variable list)	= #*;
	VAR	(variable list)	= #*;
	COV	(submatrices of PHI)	= #*;
	COV	(variable list)	= #*;
(variable list)	COV	(variable list)	= #*;
(variable list)	PCOV	(variable list)	= #*;
	EQM;		

Here, # is any number, and (variable list) denotes any list of variables and/or lists generated by TO or dash, separated by commas, inside parentheses. For example, (V5, F3-F8) is one variable list, and (F2, V2-V9, F1) is another. The parentheses may be omitted from the list *following* a keyword, above, but not from the list preceding ON, COV, or PCOV. The asterisk may be omitted from any command, above, to indicate a fixed parameter. The number may be omitted, to let EQS generate start values. If the three characters = #* are all omitted from the VAR command, that is equivalent to =1.0;. If they are omitted from any other command, that is equivalent to =*;.

The ON statement is used to generate sets of equations. For each variable in the left-hand list, one equation will be generated, having that variable as the dependent variable. Each variable in the right-hand list will appear on the right side of the equation, as a predictor. If a variable appears on the left side of more than one ON statement, its equations will be combined automatically into one equation, having all its predictors on the right. See the example, below. A set of ON statements generates all nonzero coefficients in measurement and construct equations, i.e., everything in the EQUATIONS section.

VAR is used for each fixed variance, or to give an initial estimate of a free variance. Variances not mentioned in VAR commands are free, with an initial estimate calculated by EQS.

COV and PCOV are used for nonzero covariances. As seen in the table, there are three variants of COV, which differ in whether variables are specified to the left or right of the command, and the type of variable designation. The first COV command accepts to its right a list of submatrices of PHI. The full list of possibilities is (VV, VF, FF, VE, FE, EE, VD, FD, ED, DD). See **Types of Variables** in the EQUATIONS section, above. For instance COV (VF) = 1*; means that each covariance between an independent variable and an independent factor is a free parameter, with a start value of 1. The second COV command generates all pairs (X,Y) such that Y precedes X in the variable list given to its right. The third COV command generates all pairs such that X is in the left list, Y is in the right list, and X is not equal to Y. The PCOV (i.e. paired COV) command generates all pairs (X,Y) such that X is the nth element of the left list and Y is the nth element of the right list, for n = 1, 2, … . The two lists must have the same number of elements. This is good for pairwise correlations.

The EQM command generates a RETEST file, including EQUATIONS, VARIANCES, and COVARIANCES sections, with final parameter estimates. See RETEST in the PRINT section, below. If used, EQM should be the first command in the MODEL section. This will make the LOG file easier to read. The VARIANCES, EQUATIONS, and COVARIANCES generated by the MODEL section are printed on the LOG file. For example,

```
/MODEL
       (V2, V3)     ON F1;
       (V4, V5, F1) ON F2;
       (V1, V7)     ON F3;
       (V6)         ON F3 = 1;
       (V1)         ON V8;
       (F1)         ON V1;
       VAR V8, E1-E7 = 6*;
       COV F2, F3, V8;
```

generates these sections:

```
/VARIANCES
       V8, E1-E7 = 6*;
/EQUATIONS
       V1 = *V8 + *F3 + E1;
       V2 = 1F1       + E2;
       V3 = *F1       + E3;
       V4 = 1F2       + E4;
       V5 = *F2       + E5;
       V6 = 1F3       + E6;
       V7 = *F3       + E7;
       F1 = *V1 + *F2 + D1;
/COVARIANCES
       F3, F2 = *;
       V8, F2 = *; V8, F3 = *;
```

Note that the coefficients (V2,F1) and (V4,F2) are fixed at 1. EQS fixes the first coefficient of a factor at 1 if all coefficients of that factor are free, and that factor does not have fixed variance. This is necessary to avoid an underidentified model. Also note that error and disturbance residuals are added automatically.

/RELIABILITY (Optional)

The RELIABILITY section can be used instead of EQUATIONS and VARIANCES when the model has exactly one independent factor. This section has one keyword, SCALE. For example,

```
/RELIABILITY
       SCALE = V1 TO V4;
```

generates these sections:

```
/EQUATIONS
       V1 =  F1 + E1;
       V2 = *F2 + E2;
       V3 = *F3 + E3;
       V4 = *F4 + E4;
/VARIANCES
       F1 = *;
```

In addition to estimating the model parameters, EQS calculates the internal consistency reliability coefficient, rho, based on this one-factor model. (See Raykov, 1997, Eq. 5).

/CONSTRAINTS (Optional)

The parametric constraints available in EQS are quite general in that individual parameters may be fixed to specific numerical values, may be freely estimated, may be constrained equal to other free parameters, may be constrained with other parameters to satisfy a general linear equality, or may be constrained to lie between an upper and/or lower bound. In the /CONSTRAINTS section (abbreviated /CON), two types of equality constraints may be imposed on estimated parameters, namely simple equality and general linear equality constraints. Inequalities are discussed below.

Simple equality constraints are specified when two or more parameters are to be estimated with the same numerical value. For example, one may want $\theta_i = \theta_j$, where θ_i and θ_j are some parameters. Such a constraint can be specified as just given. However, as a lead-in to the more complicated general linear constraint, the program also gives you the option of specifying the constraint such that a constant appears on the right of the equation. In this example, we have $\theta_i - \theta_j = 0$, where the constant is zero. If several parameters are to be set equal, this can be done directly by a statement of the form $\theta_i = \theta_j = \theta_k$. However, as an introduction to the general linear constraints, you can also specify three equalities with two equations. That is, if $\theta_i = \theta_j = \theta_k$, two equations are needed, for example: $\theta_i - \theta_j = 0$, and $\theta_i - \theta_k = 0$, or else $\theta_i - \theta_j = 0$ and $\theta_j - \theta_k = 0$. Note that with m equalities, m - 1 equations are needed. That leaves m - (m - 1) = 1 free parameter to be estimated. It does not matter which m - 1 equations are used, as long as they are not redundant (linearly dependent). Thus $\theta_i - \theta_j = 0$ and $\theta_i - \theta_j = 0$ as two constraints would not be correct.

General linear equality constraints represent the constraint hypotheses that are found in most linear models. They allow the weighting of parameters in arbitrary fashion, as long as the weights *w* are known. Thus, one may believe that, or be interested in testing whether, one parameter is a linear combination of others. Thus, if $\theta_i = .5\theta_j + .3\theta_k + 1$, one would need to specify the constraint equation in a form with the constant on the right-hand side, as $\theta_i - .5\theta_j - .3\theta_k = 1$. In the general case, $w_i\theta_i + w_j\theta_j + ... + w_k\theta_k = c$. As in the case of simple linear equalities, it is necessary to assume that the constraint equations are not redundant. If the constraint equations are linearly dependent, the program will print an error message and will adjust the degrees of freedom. At a minimum, there should be fewer constraint equations than free parameters involved in these equations.

It is important to note that only free parameters can be involved in constraint equations. If a fixed parameter needs to be included in such an equation, you must do the arithmetic that will absorb the fixed parameter into the constant on the right-hand side of the equation. Thus, if $w_i\theta_i + w_j\theta_j + ... + w_k\theta_k = c$, but θ_k is a fixed or known number, the equation must be specified in the form $w_i\theta_i + w_j\theta_j = \bar{c}$, where $\bar{c} = c - w_k\theta_k$ is the new constant. The program will not do this arithmetic for you.

Both types of constraints can be imposed on any of the free variances, covariances, and measurement or construct equation parameters. Such parameters must be associated with asterisks (*) to denote a value to be estimated. If a parameter defined in the /CONSTRAINTS section is not a free parameter, an error message will be printed and the program will terminate. The error must be corrected, and the job resubmitted.

Double Label Convention

To implement equality and general linear constraints, a unique designation must be available for each parameter of a model. Since EQS does not use Greek symbols such as θ_i, a “double-label" convention based on the model specification language is used instead. Each parameter is defined by a {left parenthesis, variable name, comma, variable name, right parenthesis}, such as (V1,F1), so two variable labels uniquely define a parameter. Since all parameters are either coefficients in equations, or variances or covariances, there are three cases to consider.

Equations. Regression coefficients appearing in equations are identified by the pair (dependent variable, predictor variable). Thus in the equation V1 = 2*F1 + E1, the parameter 2* is identified as (V1,F1) since V1 defines the unique equation involved, and F1 the relevant part of the right-hand side. Note that (V1,E1) defines the fixed 1.0 parameter; however, since only free parameters can be involved in constraints, the parameter (V1,E1) could not be used in a constraint specification. In the longer equation F1 = .6*F2 + .1*F3 - 2*F4 + D1;, the parameter -2* is given by (F1,F4). The ordering of the labels is critical; (F4,F1) does not define a parameter in the equation. Each of these parameters must be given by (F1,XX), where XX is any of the three predictor variables F2, or F3, or F4.

Variances. The variance of an independent variable is uniquely defined by repeating the variable name. Thus, if V6 = .3*; was specified in the /VAR section, the parameter .3* would be given by (V6,V6). The mnemonic is, of course, the row and column designation for the diagonal element of a matrix. If /VAR contained F1 TO F4 = .8*; V1 = 1.0; E5 = .3*; etc., then (F2,F2) would define the relevant .8* free parameter, (V1,V1) could not be used in a constraint since it is a fixed, not free, parameter, etc.

Covariances. The covariances among independent variables are specified as in the /COV section, except that in the /CON section they are surrounded by parentheses. Thus, with /COV specifying E6,E3 = .3*; V1,F5 = .2*; etc., these parameters are double-label identified as (E6,E3) and (V1,F5). To avoid confusion, the reader is urged to use the same double-labels in the /COV section as in the /CON section, e.g., the E6,E3 = .3* parameter should not be labeled (E3,E6).

Simple Equality Constraints

After using the keyword for the /CON section, any free parameter can be set equal to one or a group of parameters, whether a regression coefficient, variance, or covariance, by a simple statement. Examples are:

	COMMENTS
/CONSTRAINTS	Required
(E1,E1) = (E2,E2);	Simplest case
(V5,F1) = (V3,F2) = (V9,F3) = (V12,F1);	Cannot exceed 80 cols!

A maximum of ten equalities can be specified in a single statement due to the 80 column limitation. Each equality statement must be terminated with a semicolon on each line. More than one statement per line is not permissible, i.e., *exactly one ";" per line must be given*. Of course, the parameters specified in an equality must 1) exist in the model, 2) be free, and 3) start at the same value. The latter point requires you to assure that the start values provided for the free parameters meet the constraints to be imposed. If the start values do not meet a constraint, the program may not be able to impose the constraint during optimization. Note that no weighting constants can be used when simple equalities are specified. Thus, (E1,E1) = 2(E2,E2); is an illegal specification.

General Linear Constraints

As noted above, m equalities require m - 1 linear constraint equations. For example,

	COMMENTS
/CON	
3(V4,V4) - (V8,V8) = 0;	No "*" permitted
.2(E1,E1) + .8(E2,E2) = 1;	One constraint per line

The first constraint sets the free parameter (V4,V4) at 1/3 the value of the free (V8,V8) parameter. The second says that the weighted sum of the (E1,E1) and (E2,E2) parameters equals 1.0.

Constraints must be used with a great deal of care in order to avoid problems. An example of an *inappropriate* job set up is the following.

```
/EQUATIONS
V4 =  .68*F3 + E4;
V5 = 1.0  F3 + E5;
V6 =  .8* F4 + E6;

/CONSTRAINT
(V6,F4) - (V5,F3) - .2(V4,F3) = 0;
```

Because the parameter (V5,F3) is a fixed parameter that takes on the value 1.0, the correct CONSTRAINT specification should be

```
/CON
(V6,F4) - .2(V4,F3) = 1.0;
```

The above example, however, is not strictly correct due to one more requirement of the program. *The starting values of the free parameters should meet any constraints that are to be imposed.* In this instance, the constraint is not met by the given estimated parameters, because .8 - .2(.68) ≠ 1.0. The constraint would be met if the /EQUATIONS were respecified as

```
V4 = -1.0*F3 + E4; V6 = .8*F4 + E6;
```

because now .8 - .2(-1.0) = 1.0, as required. Note that such a radically different estimate of the (V4,F3) parameter, with its sign reversed, may not make sense substantively; alternatively, the constraint may not make sense with the original, well-estimated, parameters. Imposing constraints on an otherwise sensible and well-fitting model may make a model fit badly, thus suggesting that a constraint is not reasonable. *Constraints should only be considered in the context of an otherwise plausible model.* Estimation difficulties can easily occur with implausible models, or implausible constraints. The appropriateness of a constraint can be evaluated directly by the /LMTEST. See Chapter 6.

Cross-group Constraints

The method of specifying constraints described above is applicable to models involving data from several samples, but the description is limited to the within-sample constraints only. That is, the /CON section is placed within the input stream of the model for a given sample, and each of the parameters involved in the constraint is a parameter in the model for that sample. When the data contains only one group, this is, of course, the only possibility, but when multisample analysis is undertaken, with GROUPS in /SPEC being specified as 2 or more, there are other possibilities. Multisample covariance structure analysis is described in Chapter 7. In multisample analysis, there are within-sample constraints, but also cross-sample or cross-group constraints.

When GROUPS = 2; or greater, there are as many input segments (with sections from /SPEC to /END) as there are groups in the analysis. A /CONSTRAINTS section can be placed within each of these input segments, as usual, to describe the within-sample constraints. However, the /CONSTRAINTS section of the *last* group is modified to permit specifying, in addition, cross-group constraints.

Cross-group equality constraints must be specified in the last group, using a special format. The format is essentially the same as in the one-group case, except that a group number is now added as the first specification inside parentheses. In a 3-group example with only one parameter constrained to be equal across the three groups, one would have, symbolically: (1,P1) = (2,P1) = (3,P1); where 1, 2, and 3 are the group numbers and P1 is any free parameter, designated in practice by a specific double-label name such as F2,F1. Thus an actual example would be

```
/CONSTRAINTS
(1,E1,E1)= (2,E1,E1);
(1,V3,F1)= (2,V3,F1);
(1,F2,F1)= (2,F2,F1);
```

where there are two groups and three parameters constrained to be equal. In these examples, the same type of parameter such as an error variance (E1,E1) is being constrained to be equal across the two groups. However, any free parameter from one group can be specified to be equal to any other free parameter in any other designated group. As usual, all the parameters listed must be free parameters in the model, and their start values must be equal.

General linear constraints across groups involving constant multipliers or additive constants can also be used. These constraints must be written in the form .5(1,V1,F1) + 3(2,V1,F1) = 1; and the start values must meet the constraint. The appropriateness of constraints can be evaluated by the /LMTEST as described in Chapter 6.

The SET keyword now allows most cross-group constraints to be specified very simply. The CONSTRAINTS section in the last group may include

```
SET = LIST; or SET = ALL;
```

LIST is a list of submatrices of the pattern matrices, as in the SET keyword in the LMTEST section. SET = ALL; means all submatrices of the pattern matrices. For each submatrix in the SET command, each free parameter in that submatrix is constrained to be equal, across groups, to each identically-named parameter that is in the same submatrix. For example, if SET = BVF;, and (1,V1,F2), (2,V1,F2), and (3,V1,F2) are free parameters in the beta matrix, then EQS generates two constraints:

```
(1,V1,F2) - (2,V1,F2) = 0;
(2,V1,F2) - (3,V1,F2) = 0;
```

The list of submatrices may be specified with one SET keyword, or with several. SET=BVF,BFF; is equivalent to SET=BVF; SET=BFF;

A parameter may be excluded from the cross-group constraints by a forced-free statement, e.g. if

```
(1,V1,F2) = *;
```

then the second constraint, above, will be generated, but not the first. Each parameter excluded from the cross-group constraints must be mentioned in a separate equation.

If SET is used, the generated constraints are printed in the LOG file, and if a RETEST file is created, the constraints from the last group are written on the file.

/INEQUALITIES (Optional)

Because of the frequent occurrence of improper solutions and the uninterpretability of some associated parameters such as negative variance estimates (Boomsma, 1985; Dillon, Kumar, & Mulani, 1987; Gerbing & Anderson, 1987; Rindskopf, 1983, 1984a,b; Sato, 1987; Chen, Bollen, Paxton, Curran, & Kirby, 2001) and suppressor effects in which coefficients have unexpected signs and sizes (Maasen & Bakker, 2001), some care in evaluating a solution is in order. Certain undesirable phenomena are handled automatically by EQS, while you may need to intervene to avoid others. The program automatically constrains variance estimates to be nonnegative, and correlations between variables having fixed variances as lying between -1 and +1. However, EQS cannot anticipate that a coefficient should be, say, positive, and you may want to impose such a constraint on the solution. The /INEQUALITIES keyword (abbreviated /INE) permits an estimated free parameter to be constrained between an upper and/or lower bound. Fixed parameters cannot be constrained this way; they already are, by their nature, maximally constrained. Possible corrections to test statistics due to population boundary values of parameters (Shapiro, 1985b; see also Dijkstra, 1992) are not incorporated in the program; population values are rarely known.

Only one parameter at a time can be specified as constrained. This is done by specifying a constraint on a single line of the input stream using one of two conventions. The basic convention uses capitalized FORTRAN-like greater than (GT), greater than or equal to (GE), less than (LT), and less than or equal to (LE) symbols. An alternative convention, described further below, uses a more mathematical notation. For example, the inequality (V3,F1) GT 0, LE 1; specifies that the free parameter (V3,F1) is to be greater than zero and less than or equal to one. The inequality (V8,F4) GT 50; specifies that the parameter (V8,F4) is to be greater than fifty. The inequality can also specify only an upper bound, for example (E6,E4) LT 100; inequality constraints on variances must be denoted by a double identifier as in (V10,V10) LT 100;. As before, every parameter involved in an inequality is inside parentheses.

EQS does not allow negative variance estimates. *The program automatically imposes an inequality constraint of nonnegativity on each variance,* except in correlation structure analysis (ANAL=CORR;), where this constraint is not imposed. Since most programs do not impose this logical constraint, the results obtained under such a constraint

may not match results obtained without the constraint. If a constraint is necessary to statistically-based estimation methods such as maximum likelihood; a solution with a negative variance estimate is not an ML solution. You may, of course, override this feature by specifying that each variance be greater than or equal to a large negative number. This constraint replaces the current implicit constraint that a variance be GE 0. Thus (E4,E4) GT -999999; allows large negative values.

The program automatically imposes an inequality constraint on a covariance such that the corresponding correlation is in the ±1 range; this is done if and only if the relevant variances of the variables are set to fixed numbers. If F1 and F2 are factors whose variances have been specified to equal fixed 1.0, the covariance F2,F1 actually represents a correlation. This correlation is automatically forced to lie in the range between -1 and +1. If the variances of F1 and F2 are fixed at 4 and 9, the free parameter covariance F2,F1 is forced to lie in the range ±6, since the end points of the range correspond to correlations of 1. This program feature can be overridden by specifying an appropriate very small lower bound and very large upper bound for the covariance in question. Note that since these inequalities are not automatically imposed in most covariance structure programs, results from EQS may not match those from similar programs. If the variances of variables are free to be estimated, EQS will not impose any inequality on the associated covariances. Thus, after standardization as provided in the program output, the researcher may find in pathological cases that the correlation that is associated with a covariance is out of the ±1 range. Considering the standard error associated with such a covariance estimate, this often implies that the variables are perfectly associated and the researcher may wish to respecify the model based on this knowledge. In principle, the stronger condition that the covariance matrix of independent variables be nonnegative definite should be imposed. A method for accomplishing this was developed by Bentler and Jamshidian (1994), and will appear in a future release of EQS.

The program cannot effectively distinguish between GE and GT, or between LE and LT. In practice, GE and LE are used in the program. So, if you say (V3,V2) GT 0; we may obtain a value of 0.0. If you want a value strictly greater than 0, you will have to decide how much larger to specify. Thus, (V3,V2) GT 0.0001; will insure the value is greater than 0. But the program implements (V3,V2) GE 0.0001;.

When no /INEQUALITY constraint is imposed on a parameter, the program acts as if any number that can be carried in the computer is an appropriate estimate. Such an estimate may be very small or very large. This effect is achieved in practice by imposing inequalities such as GE -9999999. and LE 9999999. When an /INEQUALITY is imposed, *you must be sure that the start value for the parameter is inside the permissible range*. Thus, variances must be estimated initially as nonnegative and covariances must be such that the corresponding correlations are in the +1 to -1 range when the defaults noted above are used.

It should be noted that when a parameter is estimated at a specific boundary (e.g., at 0 for a variance), that parameter might not have a standard error associated with the estimate.

An alternative convention can be used to specify /INEQUALITIES. In this convention, an inequality sign (> or <) replaces the GT (GE) or LT (LE) symbol. Furthermore, the statement itself requires the inequality to be written in a standard mathematical form. To illustrate,

	Equivalent Convention
/INEQUALITIES	
-1.0 < (V4,V3) < 1.0;	(V4,V3) GT -1.0, LT 1.0;
1.0 > (V5,V4) > -1.0;	(V5,V4) LT 1.0, GT -1.0;
(E1,E1) < .5;	(E1,E1) LT .5;
.2 < (D2,D2);	(D2,D2) GT .2;

Thus while the original FORTRAN-like convention requires every inequality to start with the double-label name of the free parameter involved, statements involving the inequality sign embed the parameter inside the appropriately directed signs. As noted above, the program cannot distinguish between > and ≥ , or < and ≤, i.e., the program assumes that the meaning of ">" is really " ≥ ", and acts accordingly. If you are concerned with creating a strict inequality, an appropriately modified constant will have to be used in the statement. For example, .201 < (D2,D2);

will assure that (D2,D2) is not equal to .2. However, it must be recognized that the computational accuracy of the estimates sets a limit on such discriminations.

Inequalities for Convergence Control

/INEQUALITIES can sometimes effectively be used in difficult estimation problems, as a means of obtaining convergence when nothing else seems to work. This is especially true if it appears as if one estimated parameter is going outside of a reasonable range. Imposing a constraint that forces the parameter to stay in an appropriate range during the iterations may make the process work, yielding a final solution that is interpretable and statistically valid. The trick is to impose an /INE that will help the program converge, but that is not needed at the end. At convergence, ideally, the parameters are not held to their boundaries. If a parameter is held to its boundary at the solution, then the trick did more than help convergence; it obtained a restricted solution. If such a solution is not desired, wider bounds may be needed. There will be cases, of course, when the bounds will always be necessary. For example, inequalities should not be expected to solve the problem of underidentified parameters.

/MATRIX (Optional)

This section is used for covariance and correlation matrix input only, i.e., it is not applicable to raw data input. The keyword /MATRIX (abbreviated /MAT) signals to the program that the input covariance (correlation) matrix follows in the input file, beginning on the next record. Do not use a /MAT section if the matrix resides in a separate data file. As a default, EQS uses a "free-field" input format, which requires that elements of the matrix be separated by at least one blank space. It is assumed that no extraneous information, e.g., labels or missing data codes, exists.

Three different formats can be used for this (covariance or correlation) matrix, whether the matrix follows in the input or resides in a separate data file. The formats are lower triangular form, rearranged lower triangular form, and full symmetric form. When a small problem is to be analyzed, the lower triangular form is usually the format of choice. When a large problem is to be analyzed, and the matrix is imported from one of the standard statistical packages, the rearranged lower triangular form is frequently used. In certain circumstances, a full symmetric matrix may be available and it can be used without modification.

An example of the /MATRIX section in lower triangular free format form is

```
/MATRIX
```

2.0		
.67	1.8	
.32	-.09	1.4

It will be noted that this is only one of four sections of the input where *no semicolon is used to mark the end of the input sentence*. The simple rule is that the input consists of data only. Also, /MEANS, /STANDARD DEVIATIONS, and /KURTOSIS do not end their data segments with a ";".

EQS reads a lower triangular matrix as follows: in the first line, it expects one datum, and it stops reading after that is encountered. In the second line, it expects two data, and it stops thereafter. It continues in this fashion until more data would need to be read than can fit in a single line: at this point, it expects the data to "wrap" into the next line and it continues to read until it has read just enough to finish that variable's input. Consequently, it is not possible to list all the data as more or less continuous input, in as few lines as possible. In the example above, it would not be possible to place 2.0 .67 1.8 .32 -.09 1.4 into a single line of input. For p variables, at least p lines of input are needed, but in a big job, more than p will be needed because of the wrapping feature.

If a FIELDS statement is given in the /SPECIFICATION section, then the input matrix must be in the form discussed in detail in the FIELDS documentation. If the above data were to be used with a FIELDS = 1; specification (in practice this would never be done), the /MAT section would have to be followed on the next six

lines with the numbers 2.0 .67 .32 1.8 -.09 1.4, one number per line with no punctuation. Note that FIELDS = 1; causes the matrix to be read column-by-column, instead of row-by-row.

When a FORTRAN format statement is provided in the /SPEC section, the /MAT section will also accept a full, symmetric input covariance matrix. For example, with the format FO = (3F3.1); specified in /SPEC, the following input of 3 rows with 9 columns per row

```
/MATRIX
```

18	6	-1
6	12	4
-1	4	9

is equivalent to the free format stream

```
/MAT
```

1.8		
.6	1.2	
-.1	.4	.9

Two strategies are generally used with regard to the input covariance matrix. The covariance matrix may contain only the variables planned for a particular analysis. The variables that are used for an analysis may occur in any order in the input matrix; no special ordering is needed. Alternatively, a large matrix may be used, from which those variables needed in an analysis are then automatically selected. The latter, more general approach is useful for those problems that require different combinations of variables from among a large set. The automatic selection procedure has been discussed previously. However, if a matrix contains many variables that will never be modeled, it is best to reduce the matrix to the set to be modeled.

/STANDARD DEVIATIONS (Optional)

If the input matrix is a correlation matrix rather than a covariance matrix, unless you specify ANAL=CORR;, you should transform the correlations into covariances prior to analysis. In order to do this, you need to provide the standard deviations of the variables. This is done by providing the keyword /STA or /STANDARD DEVIATIONS, and, on the subsequent line, listing the standard deviations. There must be exactly as many standard deviations as there are variables in the correlation matrix to be read in. The standard deviations are expected to be in the same sequence as the variables in the input matrix.

The numbers are read in free format. Thus, there must be a blank space between two different numbers. More than one line can be used if necessary. *No semicolon is used to end this data input section.* Examples are:

```
/STANDARD DEVIATIONS
1.345  6.336  1.22  9.4

/STA
.87  4.334  .969  3.01  2.78  3.4
1.116  .45
```

Note that blanks are permissible (between numbers, but not within one number), and wrapping to the next line is permitted (however, a number cannot be split across two lines). Since the program expects to read a certain quantity of numbers as given by the number of variables in the input matrix, the first example must have VAR=4; and the second must have eight variables.

If the correlation matrix is read from an external file, and standard deviations are on the file immediately following the matrix (or immediately following the means), they may also be read from the file. Simply use the section /STA on the model file. EQS will read the standard deviations using the same format (or free format) as was used for the matrix. If means precede standard deviations on the file, /MEA must precede /STA on the model file.

/MEANS (Optional)

Means are not relevant to typical covariance or correlation structure models, and hence they usually are not needed in the input stream. However, when structured means models are analyzed, as described in Chapter 8, the model will contain the intercept V999, and variable or factor intercepts and means become parameters of the model. As a consequence, the means of the input variables become data to be analyzed, and they must be available. If raw data are read in, of course, the program computes the variable means, and they need not be provided separately. However, if a correlation or covariance matrix is used as input, the /MEANS section must be used. This information is used to create the covariance/mean partitioned matrix that will be analyzed.

There must be exactly as many means given under /MEANS as there are input variables in the matrix to be read in. The means are expected to be in the same sequence as the variables in the input matrix. The numbers are read in free format. Thus there must be a blank space between two different numbers. More than one line can be used if necessary. As data input, *no semicolon is used to end this input section*. Some examples are:

```
/MEANS
1.012 1.123 .564 2.10

/MEA
.87 4.334 .969 3.01 2.78 3.4
1.116 .45
```

In the first example, there must be four variables in the input covariance matrix, since four numbers are given. The second example shows wrapping onto a second line, with means given for 8 variables. Note that no ";" is used to end the means information.

If the covariance or correlation matrix is read from an external file, and means are on the file immediately following the matrix (or immediately after the standard deviations), they may also be read from the file. Simply use the section /MEA on the model file. EQS will read the means using the same format (or free format) as was used for the matrix. If standard deviations precede means on the file, /STA must precede /MEA on the model file.

/KURTOSIS (Optional)

This command provides marginal kurtoses of variables for the Heterogeneous Kurtosis (HK) methodology. Typically HK methods will be used with raw data, and the required marginal kurtoses $\eta_i = \sqrt{\sigma_{iiii} / 3\sigma_{ii}^2}$ will be computed from the data automatically and printed along with other univariate statistics as KURTOSIS (ETA). Hence, normally, there is nothing you need to provide. However, if you want to use the HK method but do not have raw data, or if you want to override the data-computed kurtosis values, you can provide these analogously to means or standard deviations. You use the keyword (or its abbreviation) along with actual kurtosis values, and *do not use a semicolon to end the input*:

```
/KUR
15.046 -.183 -.868  1.321  1.289
```

The kurtosis values you provide must represent the $g_{2(i)}$ coefficients given in eq. (4.2). Internally, EQS will compute $\eta_i = \sqrt{(g_{2(i)} + 3)/3}$ (with positive square roots), the HK values that are multiplied or averaged depending on whether HK=GEO; or HK=AVE; is specified.

/LMTEST (Optional)

The Lagrange Multiplier or /LMtest procedure is designed to test hypotheses on the statistical necessity of restrictions that exist in a model. The first type of restriction tested is whether the equality constraints that may have

been imposed in a model are appropriate. If some of the constraints are inappropriate, i.e., not consistent with the data, the overall fit of the model might be improved substantially by releasing the constraint in a future EQS run. The second type of restriction tested is whether fixed parameters, such as “missing" paths or covariances that are set to zero in the model, in fact are nonzero in the population, and hence would be better treated as free parameters and estimated in a future run. Univariate and multivariate LM statistics are produced to permit evaluation of the statistical necessity of these restrictions. The program also produces a parameter change statistic to estimate the value that would be obtained if a fixed parameter were freed in a future run.

The /LMTEST procedure has many options that are described fully in Chapter 6. The basic default test procedure can be invoked by simply placing:

```
/LMTEST
```

or /LMT on a separate line in the input file. This will produce tests on equality constraints and tests on certain standard types of fixed parameters. The univariate statistics that are printed out are modification indexes that can be consulted for suggestions as to which fixed parameters might be set free. The corresponding multivariate test should also be consulted, since it takes into account correlations among the many univariate tests. However, the results should not be accepted blindly. In this process, *substantive theory is very important*, because the /LMtest can suggest model changes that are theoretically meaningless, for example, effects that operate backward in time.

The options that are available in the /LMTEST include keywords such as APRIORI, HAPRIORI, PROCESS, SET, PVAL, and NOFREE, as well as BLOCK, and LAG. These options provide customized application of the test. The *a priori* application of the LM test to particular fixed parameters or the given equality constraints is the most defensible statistically, since the probability statements given by the test may be distorted when the test is blindly applied as it can be in the default mode. If particular subsets of matrices are to be scanned for fixed parameters to free, for example, “correlated errors,” the SET keyword needs to be implemented. Further information on these options is given in Chapter 6.

/WTEST (Optional)

The parameters denoted by “*" are estimated optimally in EQS, and standard errors are provided so that a *z*-test can be constructed to evaluate whether a given parameter might be zero in the population. When sets of parameters are to be evaluated, another procedure is needed. The /WTEST or Wald test is designed to determine whether sets of parameters that were treated as free in the model could in fact be simultaneously set to zero without substantial loss in model fit. The test is implemented by typing

```
/WTEST
```

or /WTE on a separate line in the model input file. This will produce a default test based on taking the least significant parameter (smallest *z*-test value), and adding other parameters in such a way that the overall multivariate test yields a set of free parameters that, with high probability, can simultaneously be dropped from the model in future EQS runs without a significant degradation in model fit.

The /WTEST procedure also contains a number of keywords (such as APRIORI, HAPRIORI, PVAL, NOFIX) that permit customized application. These are described in Chapter 6. The most important keyword provides a way of specifying specific parameters to be tested in an *a priori* way. The probability statements associated with W tests are most likely to be accurate when *a priori* tests are implemented. Gonzalez and Griffin (2001) point out that W tests require quite large samples to be trustworthy. χ^2 difference tests should be used to verify W tests in small samples.

/TECHNICAL (Optional)

This optional section is identified by entering the keyword /TECHNICAL (abbreviated to /TEC) in an input record. The /TECHNICAL section allows you to specify the maximum number of iterations for normal, elliptical, HK, and arbitrary distribution theory computations, as well as the convergence criterion and tolerance. This section also provides an option for using arbitrary start values for elliptical and arbitrary distribution theory computations. The keywords of the /TECHNICAL section must be separated by a semicolon.

The defaults of the keywords and their meanings are as follows, where n refers to an integer number you specify, which must be in the range 0-500, and x refers to a decimal number. For use of n = 0, see below.

KEYWORD	ABBREVIATION	DEFAULT	MEANING
ITR	IT = n;	30	Max. no. of iterations for normal theory (LS, GLS, ML), at most 500
EITR	EI = n;	30	Max. no. of iterations for elliptical theory (ELS, EGLS, ERLS), at most 500
HITR	HI = n;	30	No. of iterations for heterogeneous kurtosis theory (HKGLS, HKRLS), at most 500
AITR	AI = n;	30	Max. no. of iterations for arbitrary distribution theory, at most 500
CONVERGENCE	CO = x;	.001	Convergence criterion, based on mean of absolute differences of parameter estimates between two iterations
START	ST = ELL; ST = AGLS;		Arbitrary start values for elliptical methods (sets EITR = 30) Arbitrary start values for distribution-free methods (sets AITR = 30)
TOLERANCE	TO = x;	.000001	Tolerance to control accuracy in solving linear equations

Examples of valid /TECHNICAL section keywords are:

```
/TECHNICAL
ITR  = 50; AITR = 10;EITR = 20;
CONV =.0001; START = ELL,AGLS;

/TEC
START = ELL;    EITR = 50;
```

Iterations always converge faster if good start values are available. Elliptical theory usually starts with estimates based on normal theory. Typically, AGLS starts its estimation with the results based on LS. However, with ST=AGLS; it will use the given start values instead.

The convergence criterion can be loosened or tightened in accord with the needs of the researcher. Small values of criteria will typically require more iterations. The precision of your computer limits the maximum accuracy possible.

Tolerance is a technical concept associated with pivoting operations in solving linear equations. It determines the point at which one variable can be considered to be linearly dependent on a set of other variables. See Frane (1977) or the BMDP Statistical Software Manual (Dixon, 1988) for further information. In general, you would not adjust the default value. In some special situations, a difficult convergence problem may be cleaned up by adjusting TOL.

For example, when the program informs you that many parameters have been found to be linearly dependent on other parameters during the iterations, and the problem does not stem from identification difficulties, sometimes increasing TOL to, say, .001, may create a cleanly converged solution and sometimes also a dramatic reduction in the chi-square statistic.

Zero Iterations

You can do zero ITR, EITR, HITR and/or AITR iterations. In this case, the program produces estimates based on initial starting values without updating parameter estimates, i.e., there is no iterative process. This can be helpful in various circumstances, e.g., in power calculations, for obtaining statistics at particular parameter values, in checking the adequacy of an a priori set of estimates, in comparing results to those obtained by others, and so on.

/PRINT (Optional)

The /PRINT segment controls a variety of printed information that can help to make sense of a model and the quality of the estimates. PRINT keywords include several format-related commands, and some that involve additional computations.

KEYWORD	ABBREVIATION	DEFAULT	MEANING
DIGITS	DI = n;	3	Number of digits printed to right of decimal, 1 to 7
LINESIZE	LI = m;	80 or 132	Width of output, 80 or 132 characters
TABLE	TA=EQUATION; =MATRIX; =COMPACT;	EQUATION	Print equations as equations Print equations in matrix form Print equations in compact form
COVARIANCE	COV = YES;	NO	Model covariance matrix of all V and F variables
CORRELATION	COR = YES;	NO	Model correlation matrix of all V and F variables
PARAMETER	PA = YES;	NO	Correlation matrix of parameter estimates
FIT	FI = ALL;		Print fit indices
EFFECT	EF = YES;	NO	Effect decomposition
FMEANS	FM = BGLS;		Print factor means and the modified test for a potential structured means model
RETEST	RE='fil.nam';		Generate RETEST file
LMTEST	LM = YES;		Put LMTEST information on RETEST file
WTEST	WT = YES;		Put WTEST information on RETEST file
WPAR	WP = NO; = YES;		Put Wald test parameter names on RETEST file. Put Wald test parameter names and estimates on RETEST file

DIGITS

The number of digits that are printed after the decimal point in an EQS output file can be modified. The default is DIG = 3; but any number from 1 to 7 can be obtained. Printing with greater precision would imply specifying, say,

DIG = 5; in the /PRINT section. Unless your sample size is huge, using more digits will imply greater accuracy than can be justified. In small to intermediate samples, 2 digits would really suffice.

LINE SIZE

The width of the output LOG file is 80 characters for interactive runs, 132 for batch runs. This may be changed by LI = n; where n is either 80 or 132.

TABLE

This command controls printing of output in standard, matrix, or compact form. To print coefficients of the measurement and construct equations, and the variance/covariance matrix of independent variables, in compact form, specify TABLE = COMPACT; each coefficient will be printed in one line with its standard error, test statistic, etc. To print coefficients in matrix form, specify TABLE = MATRIX; all the coefficients of one predictor variable will be printed in a column. If there are many variables, you may have to cut and paste in order to see the whole matrix. The default is TABLE = EQUATION; which produces the usual output.

COVARIANCE/CORRELATION

The final solution given in the standard EQS output consists of optimal parameter estimates, standard errors, residuals between data and model, and so on. It is sometimes useful also to examine the *model matrices* based on the optimal parameter estimates. An important matrix is the reproduced covariance matrix $\hat{\Sigma}$ based on the final parameter estimates $\hat{\theta}$. In models with structured means, the reproduced mean vector $\hat{\mu}$ is also part of the partitioned covariance/mean matrix. This output is provided for the fitted or implied moments based on measured, or V, variables, and is automatically augmented by the reproduced or fitted moments of the factor, or F, variables. Thus the statement COVARIANCE = YES; informs EQS to print the reproduced covariance/mean matrix of all V and F variables. A standardized version of this matrix is available with CORRELATION = YES;. This is the reproduced covariance matrix of (V,F) variables, transformed into a correlation matrix. One can also write COV = YES; and COR = YES; in shorthand notation.

PARAMETER

The parameter estimates obtained in a model covary. For example $\hat{\theta}_i$ and $\hat{\theta}_j$ covary (of course, the parameters themselves, θ_i and θ_j, are fixed numbers that do not covary). In extreme cases, the covariance could be so high that the corresponding correlation could approach 1.0, in which case the parameter estimates become linearly dependent and there is an empirical, if not also theoretical, identification problem. Of course, a parameter estimate may be linearly dependent on others if it can be perfectly predicted (in a multiple correlation sense) from a set of other estimates, even though no single correlation seems to be approaching 1.0. In any case, it may be of interest to evaluate these relationships among estimates. In the usual output, only the standard error estimates for parameter estimates are printed. The entire covariance matrix of the estimator $\hat{\theta}$ is not printed. But it is possible to request that the standardized covariance matrix, i.e., the *correlation matrix of parameter estimates*, be printed. This is done with the statement PARAMETER = YES; or PAR = YES; the default is NO, and it need not be specified.

The printed matrix has as many rows and columns as there are free parameters to be estimated in the problem. In large problems, this matrix takes up a substantial amount of space in an output file or on printed paper. As a consequence, this print option should be used with caution.

FIT

All versions of EQS now compute and can print a wider range of *fit indices*. By default, the Independence Model Chi-square, Independence AIC, Model AIC, Independence CAIC, Model CAIC, Chi-square, Bentler-Bonett normed (NFI) and nonnormed (NNFI) fit indices and Bentler's comparative fit index (CFI) are printed. However, there is a great deal of literature on the virtues and problems of various fit indices (e.g., Hu & Bentler, 1998, 1999), so if your favorite index is not produced by default, you must request it. This is done by writing

```
/PRINT
 FIT = ALL;
```

As a result, you get (in addition to the above indices) Bollen's (IFI) Fit Index, McDonald's (MFI) Fit Index, LISREL's (GFI) Fit Index, LISREL's (AGFI) Fit Index, Root Mean Squared Residual (RMR), Standardized RMR, Root Mean Square Error of Approximation (RMSEA), and the 90% confidence interval of RMSEA. When METHOD = xx,ROBUST; EQS prints robust versions of IFI, MFI, RMSEA, and its 90% confidence interval. See Chapter 14 for more information.

EFFECT

The parameters of a structural model are the coefficients in the equations and the variances and covariances of independent variables, and, in a structured means model, the means of the independent variables and the intercepts of the dependent variables. These parameters are listed in the input stream with a "*", and so they are estimated directly in EQS. As a consequence, standard errors are also obtained and the significance of these effects can be determined. While these parameters of the model are also the "causal" parameters of a model, if it permits a causal interpretation, nonetheless they may not be the only aspects of the model of interest. In particular, *indirect effects* and *total effects* are also important interpretively (see, e.g., Alwin & Hauser, 1975; Bollen, 1987a; Fox, 1980, 1985; Graff & Schmidt, 1982; Greene, 1987; MacKinnon, Krull, & Lockwood, 2000; Sobel, 1987).

Suppose that the variables F1-F4 are connected by paths as follows: F1 $\rightarrow$ F2 $\rightarrow$ F3 $\rightarrow$ F4. (Residual variables are not shown.) Then F1 has a direct effect on F2, F2 has a direct effect on F3, and F3 has a direct effect on F4. These direct effects are coefficients in equations. There is no direct effect of F1 on F4; there is no path, and consequently the equation structure would mirror this fact, and there would be no "*" parameter for an F1$\rightarrow$F4 coefficient. However, F1 does have an *indirect* effect on F4, namely, through the variables F2 and F3. A measure of this indirect effect is given by the product of the coefficients represented by the arrows. If there are many sequences by which a variable like F1 can influence F4, the total indirect effect is a number indicating the size of this effect. Total indirect effects summarize how one variable influences another regardless of the particular paths chosen to trace from one variable to the other. In the example, if there were another variable F5, with F1 $\rightarrow$ F5 $\rightarrow$ F4 in addition to the above paths, F1 can influence F4 through F5 as well. While such specific indirect effects are sometimes of interest, EQS computes only the total indirect effects. Total effects are defined as the sum of direct and indirect effects. These are also computed and printed in EQS. To obtain these effects, one must put EFFECT = YES; in the PRINT section. The default is NO, and need not be specified.

The typical *effect decomposition* provides effects defined for variables in their natural metric, i.e., the metric implicit in the equations and the input variables. In many circumstances the size of these unstandardized effects are difficult to interpret. Hence, EQS also computes an effect decomposition for the standardized solution, i.e. the solution in which all variables have been rescaled to have unit variance.

Since indirect and total effects are sample statistics, they have sampling variability. One way to judge an unstandardized effect is to determine whether it is significantly different from zero, using the standard error estimate to form a normal *z*-test. The work of Sobel (1982, 1986, 1987) on indirect effects was extended and modified to be applicable to the classes of models, estimators, and constraints available in EQS. Similar standard errors and *z*-tests were developed for total effects. In addition, EQS provides robust standard errors and *z*-tests for effect estimators based on robust estimates of sampling variability.

FMEANS

Typical models are covariance structure models, and *factor means* are not relevant. However, it is possible that the means of the observed variables also can be structured in terms of regression coefficients and factor means of latent variables. Since factor means are not available in covariance structure analysis, when FMEANS = BGLS; is given in a latent variable covariance structure model, and observed means are available, factor means are estimated and the variable means are reproduced from the augmented model parameters. To evaluate whether a structured means model might make sense, the Bentler-Yuan (2000) modified test for a potential structured means model is calculated and printed.

RETEST

When specified, RETEST takes the final parameter estimates from a completed EQS run and inserts these into a new file that can be submitted, with only minor modifications, for another EQS run. Specifically, new /EQUATION, /VARIANCE, and /COVARIANCE sections are created that contain the optimal parameter estimates from the just-completed run. These sections can be used in a new run. If a previous run took many iterations to converge, and each iteration takes a lot of computing time, the next run that can be made with this feature will almost certainly converge much more rapidly than the previous run, even if a reasonable number of model adjustments (e.g. dropping parameters or adding parameters) are made. On big jobs, the savings in computational time can be substantial. With today's computers, this is a less critical issue than it was in the past.

RETEST is implemented by a statement in the PRINT section as follows:

```
/PRINT
RETEST = 'output.eqs'; lmtest=yes; wtest=yes;
```

where "output.eqs" is the name of a new file that is to be used in the future as an input file to the EQS run. The next two statements are discussed below. When such a specification is given, EQS does the following: (1) Puts the file that was used as the input file for the current run at the beginning section of output.eqs. (2) Puts new /EQU, /VAR, and /COV sections, based on final optimal parameter estimates, into output.eqs, following the current input file. The newly created output.eqs must then be edited to delete parts of the file that are obsolete, and to update the model in the desired way. For example, the /TITLE may need to be changed. The /SPEC section may be perfectly all right, or may need to be modified. The old /EQU, /VAR, and /COV sections and the /END commands, and some intermediate lines created where the new output joins the old output file, must typically be completely removed. This is usually easily done with the block-delete feature of an editor. Other sections, such as /LMTEST, may also need to be modified to be appropriate to the current run. *If RETEST was used in the previous run, the file name that was used must be updated, or the old file will be written over*. As usual, the file to be submitted for an EQS run must end with /END.

Automatic Model Modification in RETEST File. The Windows, Unix and Mac versions of EQS permit RETEST to be implemented in such a way that model modifications suggested in the previous run are automatically included in the new model file created by RETEST. If the LM test and W test have yielded meaningful information in the previous run, then incorporating their test results into the new model will be meaningful. However, if the LM test and W test have yielded some nonsense information, then the model setup created by RETEST must be edited to eliminate this nonsense. *It is your responsibility to assure that this advanced option is used with great care.*

Automatic model modification will make sense when the parameters tested by the LM test and/or the W test are theoretically meaningful, and their inclusion or deletion from the model can be well-justified from a substantive point of view. In that case, results based on the LM test will suggest parameters to add to a model that are both theoretically meaningful and statistically necessary, and it is entirely appropriate that such parameters be added to the model. Similarly, results based on the W test will suggest parameters that, although theoretically interesting, can be trimmed from the model due to lack of their statistical necessity. If these tests are thus performed on meaningful parameters, carrying forward the results to the new model automatically can save a substantial amount of model setup time.

If LMTEST=YES; all of the parameters that were found to be statistically significant in the multivariate LM test will be added in the appropriate sections of the model setup created by RETEST. That is, variances, covariance, or predictor terms in equations may be added automatically. *The newly added parameters can be recognized in the file because only these parameters will have no start value.* RETEST uses the previous run's final estimates as start values in the new file; but newly added parameters will not have such start values. They will appear with only a "*". Thus all of the newly added parameters can be easily scanned. *They must be scanned for meaningfulness*. If the LM test is used primarily for exploration of potential missing parameters, where many of the parameters may be uninterpretable (e.g., a coefficient that represents a causal path going backward in time), these uninterpretable parameters also will get incorporated into the new model setup if they were statistically significant. Obviously, it is imperative that such nonsense parameters be edited out of the newly-created file.

The W test results will get carried forward in a similar manner with the command WTEST = YES;. Parameters found to be nonsignificant by the W test will be eliminated from the new model setup, in the following sense. *The newly dropped parameters do appear in the file, but they can be located in the model setup as follows: 1) there is no "*" next to the parameter, and 2) the fixed start value of the parameter is 0.* Parameters that have a fixed zero weight, of course, do not contribute to the model. If desired, they can be edited out of the file, but this is not necessary. In the absence of WTEST =YES; the assumption is that "NO" is intended.

A danger in automatically dropping parameters is that variances of residual variables (Es, Ds) may be nonsignificant and hence may be suggested to be dropped. In general, it does not make sense to remove such parameters, even if the variances vanish. For example, variances may not vanish in other, similar models. It is suggested that the NOFIX option in the W test be used to avoid dropping of residual variances or any other parameters that are considered crucial to be left free. If a parameter is mistakenly suggested to be dropped, all that needs to be done is that a "*", perhaps with a start value, be added to the parameter in the new file.

Optional /WTEST Setup in RETEST File. There are times when it is desirable to have a complete listing of the free parameters in a model run. This can be accomplished with the WPAR print feature. *This feature must be used with the RETEST option.* WPAR refers to Wald (test) parameters. The parameters are listed in the output file created under RETEST using the double label convention of the EQS program. If desired, the value of the final parameter estimates, as computed in the run, can also be obtained. When both the parameter names and the estimates are required, the option is specified as WPAR = YES;. This produces the double-label listing of the parameters, as well as the associated estimated values, in the last section of the designated RETEST file. Each parameter is printed after the headings /WTEST and APRIORI in the illustrative format (V1,F1): 0.8925, giving the double label followed by the estimate. If the final estimates are not desired, /PRINT can be activated with WPAR = NO;. Then, if RETEST is active, the double-label parameter names are printed in the file created by RETEST, and the optimal estimates are omitted.

At least two purposes can be envisioned for the use of the WPAR option. First, the resulting listing and estimates can be moved from the file into a word processing program for use in a report or a table. Second, a complete /WTEST setup is created that can be used as is, if desired, or modified in some way to provide a desired W test in the next run.

/SIMULATION (Optional)

A data generation procedure has been built into EQS to permit simulation studies, which can be used to create artificial data to evaluate the empirical performance of various statistics under controlled conditions. Overviews of simulation methodology are given by Lewis and Orav (1989), Mooney (1997), Thompson (1999), and especially Paxton, Curran, Bollen, Kirby, and Chen (2001). A simulation in EQS requires (1) specification of the population, (2) a method of sampling, (3) an estimation run in every sample, (4) specification of the total number of samples, and (5) whether the output raw data files will be saved and what file names will be used. In addition, results from the EQS estimation in each sample may be saved in a file for further analysis using the appropriate /OUTPUT section options.

In EQS, the population is specified in the usual job file; the model in the input file, or the input matrix, defines the population. However, variables in the population may be normal or non-normal continuous variables, or even categorical transformations of normal variables, providing that category thresholds are given. EQS performs two types of sampling: simple random sampling from the population (the default) or resampling (input raw data are taken as the population). Characteristics of each sample, and its EQS estimation, are controlled by the usual statements in the /SPECIFICATIONS section (e.g., estimation method, number of cases). In general, the sample statistics obtained from each sample will vary probabilistically from the population parameters, and a simulation study seeks to describe the sampling results. Most details of the simulation are specified in the /SIMULATION section, using keywords to create the particular study desired.

KEYWORD	ABBREVIATION	DEFAULT	MEANING
POPULATION	PO = MODEL; = MATRIX;	MODEL	How simulated data is generated
REPLICATIONS	RE = n;	1	Number of replications (samples)
SEED	SE = 87654321;	123456789	Random number starting seed
DATA_PREFIX	DA = 'MYFIL';	SIM	Up to 5 characters of data file name
BOOTSTRAP	BO = n;		Number of bootstraps to do
MBB	MB = n;		Number of model-based bootstraps to do
CROBUST	CROB = #, #;	2, 1.25	Case-robust bootstrap
JACKKNIFE	JA;		Perform jackknifing
CONTAMINATION	CO = #, #;		Probability and scale factor
TRANSFORMATION	TR = …;		Univariate transformations
CATEGORIZATION	CA = list: thresholds;		Categorization of data
SAVE	SA = SEPARATE; =CONCATENATE;		Save data from each replication separately Save all replications on one file
FORMAT	FO=(…);	(6(1X,F12.7))	Format for saving data
MISSING	MI = #;	0	Proportion of missing values

POPULATION

This keyword states how the population model is to be generated. There are two choices. If POP=MOD; is specified, the model will be generated from the linear relationship of the variables defined in equations, and the means, variances and covariances of independent variables. The numbers next to * are taken as population parameters, and the fixed parameters are also taken as fixed population quantities. Consequently, random numbers are generated for all the independent variables (e.g., F, E, D and possibly V) before transforming them to measured Vs as input for EQS. POP = MOD; is the default. If POP=MAT; is specified, the matrix provided in the input file, typically a covariance matrix (specified in the /MATRIX section) but possibly also means, serves as the population matrix. In such a case any particular structural equation model may or may not be consistent with the matrix. The latter feature allows the study of misspecified models.

REPLICATIONS

A simulation requires repeated sampling from the specified population, in which new sample data is obtained and analyzed in each replication.

```
REPLICATIONS = n;
```

states that n samples or replications from the population are to be obtained. The maximum number of replications permitted is 999. The default is 1. In principle, a large number of replications is recommended for a simulation, but in practice a smaller number (say, 100) is often used due to limitations of computer resources.

SEED

EQS uses the pseudo random number generator of Deng and Lin (2000). It requires a seed, which is 123456789 by default. It may be changed by

```
SEED = n;
```

where n is an integer number. A large integer is recommended, but it cannot exceed 2147483647.

DATA_PREFIX

This keyword tells EQS where the data file of output from each sample analysis will be stored, if SAVE is used, below. The format for this keyword is

```
DATA_PREFIX = 'character string';
```

The first five characters of the character string will be used as the prefix of the output file names. The program will then append the replication number to that prefix to compose the file names. To illustrate, if REP=3; and DATA='EQS';, then three output files will be generated, namely, EQS001.DAT, EQS002.DAT, and EQS003.DAT. With a large number of replications, these files will take up a lot of computer storage space.

Resampling

In the standard simulation setup, sampling from the population is done with ordinary simple random sampling. Once the population is set up, EQS samples repeatedly from that population. However, three other options are available. These options are based on sampling an existing data file to create a new sample, resampling again from this same data file, and so on. For theory and application, see, e.g., Efron (1982, 2000), Boomsma (1986), Nevitt and Hancock (2001), Yuan and Marshall (2000), Yung and Bentler (1994, 1996a) or Zhu, Fang, Bhatti, and Bentler (1995). The input data file that will be resampled has to be specified in the DATA keyword of the /SPECIFICATION section. Sampling from this file is done in accord with three options: bootstrap, model-based bootstrap, and jackknife. Several options in /SIMULATION will automatically be turned off when the BOOTSTRAP, MBB, or JACKKNIFE keyword is elected. In particular, since the population is the data file given in /SPEC, POPULATION (see above) will be unavailable. In addition, CONTAMINATION, TRANSFORMATION, and CATEGORIZATION options (see below) are not usable.

BOOTSTRAP

This keyword activates the bootstrap sampling method. Bootstrap sampling independently and repeatedly selects data from the existing data file until the given number of cases has been selected.

```
BOOTSTRAP = n;
```

is the statement needed to invoke the bootstrap, where n is an integer number. In practice, one often uses BOOT; EQS takes n to be the number of cases in the data file being sampled.

MBB

This keyword activates the model-based bootstrap sampling method, in which the data are transformed so that the resulting data are consistent with a model. See Beran and Srivastava (1985) and Bollen and Stine (1993).

```
MBB = n;
```

is the statement that invokes this method, where n is an integer number. In practice, one often uses MBB; EQS takes n to be the number of cases in the data file being sampled.

CROBUST

A case-robust version of the regular and model-based bootstrap based on Yuan and Hayashi (2003) is implemented when this keyword accompanies either the bootstrap or MBB commands. It is implemented with the command

```
CROBUST = 2, 1.25;
```

where any two tuning constants can be specified. The default constants are shown.

JACKKNIFE

Jackknife simulation is a type of resampling that analyzes the existing raw data file, but excludes one observation on each replication. It excludes observation 1 on replication one, observation 2 on replication two, and so on. Thus, the number of replications must be less than or equal to the sample size. If REP is not given, REP will default to the number of CASES. The method is implemented by simply stating the jackknife keyword, abbreviated as JACK. An example is

```
/SIMULATION
REPLICATIONS = 20; JACKNIFE;
```

The generated jackknife data that is analyzed on a given replication will not be saved, that is, the SAVE keyword of the /SIMULATION section is not applicable. Furthermore, the DELETE keyword of the /SPEC section is not valid.

CONTAMINATION

The contaminated normal distribution is a member of the elliptical class of symmetric distributions governed by two parameters in addition to the standard parameters of a multivariate normal distribution. See e.g., Kano (1994). In essence, a contaminated normal distribution is made up of a large normal population with parameters μ and Σ, and a small normal population having the same μ but a covariance matrix that is $k\Sigma$, i.e., bigger than Σ by a positive multiple k. The smaller population is typically called the contaminated population, since the variances of the variables are increased by the factor k relative to the larger parent population. The scale factor k is one of the parameters of the distribution, and the second parameter can be thought of as the probability of selection from the smaller population. The contaminated normal distribution is often used in robustness studies, and it serves as a good model for studying the effects of outliers in multivariate analysis. Berkane and Bentler (1988) show how to estimate the parameters of this distribution, and study its relevance to the deletion of outliers in multivariate analysis. In the /SIMULATION section, this distribution is used as the parent population when the keyword

```
CONTAMINATION  = percentile, factor;
```

is invoked. Here, the two arguments are the parameters of the distribution that need specification. The first gives the probability of selection from the contaminated population. It is a number between 0.0 and 1.0, usually 0.1 or less. The second argument gives the scale factor k that is used to create the contaminated population. Typically, the scale factor is a number in the range (1.0,10.0). An example: CON=.05,3;

TRANSFORMATION

Simulated data will be multivariate normal unless the variables are transformed. All transformations are univariate. There are three major types of transformations, all mutually exclusive. These are as follows:

- Transformation to specified skew and/or kurtosis
- Transformation to contaminated normal
- Nonlinear transformations
    - EXP or Power function applied to a normal (default) or uniform
    - LOG function applied to uniform
    - Transformation to chi or chi-square variate

Nonlinear transformations are performed on uncorrelated variables, which are then standardized before the specified covariance structure is imposed. This insures that the simulated data are drawn from a population that has the covariance structure specified by the user. CATEGORIZATION (see below) will then be applied if necessary. A simple adjustment by the mean is also possible, if the user has provided M, the mean.

Which variables can be transformed depends on how POPULATION is specified in /SIM. When POPULATION = MATRIX;, only the measured variables (Vs) can be transformed. When POPULATION = MODEL;, only the independent variables (Vs, Fs, Es, and Ds) can be transformed. Dependent variables are obtained by the reduced form of the population structural model.

The following keywords are used to specify the transformations desired:

U	uniform distribution on the interval (0,1)
S	followed by the desired value of skew (do not use with C)
K	followed by the desired value of kurtosis (do not use with C)
M	followed by the desired value of the mean
C	followed by the value of the contamination factor
EXP	exponentiation
LOG	logarithm (default to uniform distribution)
**	followed by an integer, to denote power function
X1(n)	transformation to a chi-one variate with n degrees of freedom
X2(n)	transformation to a chi-square variate with n degrees of freedom

Transformations to create the desired values of skew and kurtosis are based on the procedure described by Vale and Maurelli (1983), using formulae based on Fleishman (1978). Recent extensions are Mattson (1997), Headrick (2002), Headrick and Sawilowsky (1999, 2000), and Reinartz, Echambadi, and Chin (2002). All transformations are implemented with the keyword TRANS and designations:

```
TRANSFORMATION = variable or variable list: univariate transformation formula & (or ;)
```

where the information to the right of the equal sign depends on the desired result for a particular variable. Transformations for a set of variables are created by repeating the material to the right of the equal sign, provided that the "&" sign is used to denote that more information is coming on the next line. As usual, the final transformation is terminated by the semicolon ";".

As an illustration of /SIMULATION with a nonnormal population, consider a six variable, two factor model, in which certain characteristics are desired. The model with its TRANSFORMATION can be specified as:

```
/SIMULATION
 SEED  = 123456789;
 REPLICATIONS = 50;
 DATA = 'EQS';
 POPULATION = MODEL;
 TRANSFORMATION = e1 to e3 : exp &
e4 : m2.5, s-0.2, k1.0 &
e5 : m2.0, s-0.2, k1.0 &
e6 : m1.5, s-0.2, k1.0 &
f1 to f2 : log;
```

CATEGORIZATION

Data generated by the simulation procedure is continuous unless the categorization option is chosen. Categorization (abbreviated CAT), if any, is carried out only after TRANSFORMATION has been implemented. Variables can be categorized in accord with the instruction:

```
CATEGORIZATION  = variable or variable list : category thresholds & (or ;)
```

where & refers to continuation, on subsequent lines, of information to the right of the equal sign. The semicolon ";" completes the specification. Dash or TO can be used to specify variables in a sequence. The number of thresholds must be at least 1, but no more than 15, and thresholds must be ordered. The number of categories created is the number of thresholds plus one. An example is:

```
/SIMULATION
       CATEGORIZATION = E1-E6 : -2.5,-1.5,-0.5,0.5,1.5,2.5;
```

The effect is to categorize each of the six error variables E1 to E6 using the thresholds stated. With 6 thresholds, there will be 7 scores, namely, 1, 2, ..., 7. For example, the continuous scores -.22 and .3 for an error variable would, with the above instruction, create the same category score 4 for both these numbers.

SAVE

This keyword tells the program whether to save the generated raw data file. If SAVE is not specified, the data file will not be saved. If, however, the SAVE keyword is provided, the program must be told whether to save the data from each replication separately, or to combine the data, end to end, in a single file. The SAVE keyword is implemented as

```
SAVE = SEPARATE; or CONCATENATE;
```

where either SEP or CON, but not both, can be specified. The data will be saved in the file whose name was provided with the DATA_PREFIX keyword. If DATA_PREFIX is not given, the default is SIM.

FORMAT

You can set the format for saving data, above. The default is (6(1X,F12.7)). If given, your format must obey the same rules as FORMAT in the SPECIFICATION section.

MISSING

You can force a proportion of generated data to be missing, by setting MISSING = #; here # is a number between zero and one. EQS will randomly force N elements of the generated dataset to be missing, where N is the nearest integer to NV*NC*#, and NV and NC are the number of variables and number of cases, respectively. The result is a missing completely at random data set. Using MISSING forces MISSING = ML; in the SPECIFICATION section.

Simulation in Multisample Analyses

Simulation may be performed when GROUP = n; is used in the SPECIFICATION section. The following keywords may be used in the first group only: REPLICATIONS, SAVE, SEED, and DATA_PREFIX. If FORMAT appears in more than one group, the last one will be used for all groups. BOOTSTRAP, MBB, and JACKKNIFE carry over from one group to the next, unless overridden. SAVE = SEPARATE creates one file for each replication, containing all groups. SAVE = CONCATENATE creates one file containing all groups for replication 1, followed by all groups for replication 2, etc.

/OUTPUT (Optional)

This section is designed to create EQS technical output in compact form and with greater precision. In the absence of an /OUTPUT section, all ordinary EQS output will be placed end to end in the LOG file specified in the job submission procedure. The size of this file may be large even with a small number of replications of a SIMULATION run, so /OUT is recommended for simulations. If /OUTPUT is requested, minimal output is put into the standard output file (an echo of the model, a guide to technical output that will be produced, and format of the technical output), while the specifically requested output is stored in an external file that can be easily accessed by other programs, for further analysis. The keywords that can be specified in /OUTPUT are DATA, LISTING, and a variety of specific information as given below.

KEYWORD	ABBREVIATION	DEFAULT	MEANING
DATA	DA = 'XY.RST';	EQSOUT.DAT	File where technical output will be stored
LISTING	LI;	Short LOG file	Regular LOG file produced
ALL	AL;		Turns on all options below
PARAMETERS	PA;		Parameter estimates
STANDARD or SE	ST;		Standard errors of parameters
GRADIENTS	GR;		Derivatives of minimized function

			w.r.t. parameters
COVARIANCE	CO;		Sample covariance matrix
SIGMA	SI;		Model covariance matrix
INVERSE	IN;		Inverted information matrix
DERIVATIVES	DE;		Derivatives of model covariance matrix w.r.t. parameters
WEIGHT	WE;		AGLS weight matrix
SS	SS;		Standardized solution
RSQUARE	RS;		Bentler-Raykov r-squares
FMEANS	FM		Factor means
WTEST	WT;		Wald test results
LMTEST	LM;		Lagrange multiplier test results

DATA

The file name where technical output will be stored is given as

```
DATA  = 'file name';
```

where file name is any appropriate name in the computer system. If no file name is given, EQSOUT.DAT will be used as a default. No format is needed; the output file will be in free format.

LISTING

This option controls only the regular EQS output, and not the technical output requested in /OUTPUT. When this keyword is not given, the regular output will be turned off, and only an abbreviated output will be produced. When this keyword is given, the regular EQS output will also be produced. The format is simply to state the name LISTING. An example that includes some specific output items (see below) is:

```
/OUTPUT
Listing; parameter estimates; standard errors;
```

Output Information For All Samples Combined

Some output information is produced automatically, and thus need not be specified. In particular, if /OUT is specified without any other keywords, EQS will generate some basic model statistics. These statistics will be printed on the first nine lines of output associated with each replication. In a multisample analysis, these nine lines are printed once, for all samples combined. The contents of the printout and the key nine lines are:

Tail probabilities not applicable to this analysis are written as -1.
Other statistics which are not applicable are written as -9.

Summary section contains--

Line 1 beginning: ANALYSIS, with the method, replication number and loop number

Line 2 containing these 11 elements of model statistics in format (11I8):
Estimation method (LS,GLS,ML,ELS,EGLS,ERLS,AGLS,HKGLS,HKRLS +10 if ROBUST), or +20 if ROBUST with MISSING=ML)
Condition code (0 for normal condition)
Convergence (0 for model converged)
Number of iterations for convergence
Degrees of freedom
Number of constraints
Denominator degrees of freedom for F-tests
Degrees of freedom for potential structured means model test

D.F. for GLS test of homogeneity of means
D.F. for GLS test of homogeneity of covariance matrices
D.F. for GLS combined test of homogeneity of means/covariances

Line 3 containing these 10 elements of model statistics in format (10F12.5):
Tail probability for model chi-square
Tail probability for residual-based test statistic
Tail probability for Yuan-Bentler residual-based test statistic
Tail probability for Yuan-Bentler AGLS F-statistic
Tail probability for Yuan-Bentler residual-based F-statistic
Tail probability for Satorra-Bentler scaled chi-square
Tail probability for Yuan-Bentler scaled chi-square
Tail probability for scaled chi-square (Yuan-Bentler)
Tail probability for GLS test of homogeneity of means
Tail probability for GLS test of homogeneity of covariance matrices

Line 4 containing these 3 elements of model statistics in format (10F12.5):
Tail probability for GLS combined test of homogeneity
Tail probability for potential structured means model
Tail probability for Yuan-Bentler corrected AGLS test statistic

Line 5 containing these 10 elements of model statistics in format (10G12.5):
Independence model chi-square
Model chi-square
LISREL GFI fit index
LISREL AGFI fit index
Bollen (IFI) fit index
McDonald (MFI) fit index
Bentler-Bonett normed fit index
Bentler-Bonett non-normed fit index
Comparative fit index (CFI)
Root mean-square residual (RMR)

Line 6 containing these 10 elements of model statistics in format (10G12.5):
Standardized RMR
Root mean-square error of approximation (RMSEA)
Confidence interval for RMSEA (lower bound)
Confidence interval for RMSEA (upper bound)
Residual-based test statistic
Yuan-Bentler residual-based test statistic
Yuan-Bentler AGLS F-statistic
Yuan-Bentler residual-based F-statistic
AGLS fit index
AGLS adjusted fit index

Line 7 containing these 10 elements of model statistics in format (10G12.5):
AGLS corrected comparative fit index
Cronbach's alpha
Greatest lower bound reliability
Bentler's dimension-free lower bound reliability
Shapiro's lower bound reliability for a weighted composite
Reliability coefficient rho
Maximal internal consistency reliability
Robust independence model chi-square
Robust Satorra-Bentler scaled chi-square
Robust Bollen (IFI) fit index

Line 8 containing these 10 elements of model statistics in format (10G12.5):
Robust McDonald (MFI) fit index
Robust Bentler-Bonett normed fit index
Robust Bentler-Bonett non-normed fit index
Robust comparative fit index
Robust root mean-square error of approximation (RMSEA)
Confidence interval for robust RMSEA (lower bound)
Confidence interval for robust RMSEA (upper bound)
Yuan-Bentler scaled chi-square (case robust weighting)
Scaled chi-square (Yuan-Bentler)
Scaled(Yuan-Bentler) independence model chi-square

Line 9 containing these 6 elements of model statistics in format (10G12.5):
Bentler-Yuan modified test for potential structured means model
Minimized model function value
Yuan-Bentler corrected AGLS test statistic
Chi-square for GLS test of homogeneity of means
Chi-square for GLS test of homogeneity of covariance matrices
Chi-square for GLS combined test of homogeneity of means/covariances

The last three statistics are for missing data analysis.

Output Information For Each Sample

Most of the remaining output information is optional. However, the first line (record) is always present; it contains the sample size, number of measured variables (Vs), and number of factors (Fs). Next is optional output, PARAMETERS through FMEANS, in the order given in the table, above. Last are two standardization arrays, which are always written. The LOG file indicates how many records each array occupies. In a multisample analysis, the above information (sample size through standardization vectors) is written for group 1, followed by information for group 2, etc. WTEST and LMTEST results are at the end, for all groups combined.

All the above information is evaluated at the final solution of the particular method of estimation applied. In this technical printout, *the input data and model covariance (or, covariance/mean) matrices have been rearranged* from their normal sequential input form so that dependent variables are given first, in ascending order, and independent variables subsequently, in ascending order. The residuals and weight matrices are printed in accordance with this sequence of variables. Parameters are ordered sequentially as indicated in the LOG file, with free parameters in Φ, γ, and β given in sequence. Symmetric matrices are printed in full symmetric form.

A short description of the optional technical output is as follows:

PArameter estimates provides the vector $\hat{\theta}$ at the solution.

STandard errors give the estimated sampling variability of $\hat{\theta}$, based on the covariance matrix of parameter estimates as defined in "INverted information matrix", below. When ME = xx, Robust; is used, two sets of standard errors are provided. Robust standard errors follow those of ME = xx.

GRadient elements are $\partial Q / \partial \hat{\theta}$ for the particular function chosen. For ML, the elements are $\partial F / \partial \hat{\theta}$.

COvariance matrix is the sample matrix used in the analysis, rearranged with dependent variables first, followed by independent variables. In structured means models, this is the augmented covariance/mean matrix. It describes the sample data after all transformations (if any) have been applied.

SIgma is the model covariance matrix rearranged with dependent variables first, followed by independent variables. In structured means models, this is the augmented covariance/mean model matrix.

INverted information matrix is n times the covariance matrix of parameter estimates. In the case of LS and ELS estimation, it is the sandwich matrix with a normal optimal matrix. When ME= xx, Robust; is used, two full matrices are printed: first, the usual inverse information, then the sandwich with a distribution-free optimal matrix.

DErivatives are the elements of $\partial\sigma/\partial\hat{\theta}$ for the model and free parameters chosen. Each row of the matrix represents a parameter, and, for each parameter, the derivatives are given in the sequence of the rearranged lower triangular model matrix.

WEight matrix refers to the AGLS weight matrix *W*, which is the inverse of the optimal matrix for the theory being invoked. The rows and columns of this matrix correspond to the ordering of the rearranged covariance matrix.

SS is the standardized solution.

RSquare is the Bentler-Raykov corrected R-square (Bentler & Raykov, 2000).

FMeans are the estimated factor means (Bentler & Yuan, 2000). See FMEANS in the PRINT section.

WTest refers to the multivariate statistic obtained for an APRIORI or HAPRIORI test. In LS and ELS estimation, and ROBUST estimation, the appropriate sandwich covariance matrix is used.

LMtest refers to the multivariate statistic obtained for an APRIORI or HAPRIORI test.

/SAVE

A file of type ASCII or ESS may be saved, containing raw data (and/or factor scores), or matrices. The keywords are:

KEYWORD	ABBREVIATION	DEFAULT	MEANING
FILE	FI = 'file.nam';		Name of the file, at most 60 characters
CONTENT	CO = RAW; = NO; = MATRIX; = PARAMETERS;	RAW	Raw data Factor scores without raw data Sample covariance matrix Parameter estimates
FSCORE	FS = GLS; = REG;	None	GLS factor scores Regression factor scores
TYPE	TY = ESS; = TEXT;	ESS	ESS file is written Text (ASCII) file is written
DELIMITER	DEL = SPACE; = COMMA; = TAB;	SPACE	Character which separates data values on a TEXT file
DECIMAL	DEC = n;	5	Number of digits to the right of the decimal point on a TEXT file, at least 3 and at most 8
VARIABLES	VA = m; = ALL;	6	Number of variables per line on a TEXT file
LABELS	LA = YES; = NO;	YES	Variable labels are written at the top of a TEXT file, or not
MISSING	MI = c;	*	Missing value character on TEXT file

FILE must be specified. In a multisample analysis, specify all relevant keywords in the first sample, and only FILE in each subsequent sample. The files must be distinct; otherwise one will overwrite another.

When CONTENT=MATRIX; and a 4^{th} moment weight matrix is present, that matrix is written along with the covariance matrix. CONTENT = PARAMETERS; applies only to hierarchical analyses. It forces TYPE = ESS;, and causes a file to be written, containing parameter estimates for each group. The variable names on the ESS file are the eight-character parameter names. GLS factor scores are described in Bentler and Yuan (1997), and specialize to the Bartlett (1937) scores in standard situations.

Numbers on a TEXT file are written in free format, using the delimiter character between numbers. Each number is written in E-format, with one digit to the left of the decimal point and n digits to the right, e.g. 0.123456 is written as 1.23456E-01. Numbers in the range (1,10) are written without the E+00.

VARIABLES = ALL; should not be used if many variables are saved, since the file may be difficult to read.

The MISSING character is only used in writing missing values of the weight variable, if any. A case is not written if MISSING=COMPLETE; is given in the SPECIFICATION section and any model variable is missing, or if MISSING=ML; and all model variables are missing. If MISSING=ML; and some model variables (but not all) are missing, imputed values of missing variables are written on the SAVE file.

/END

This keyword signals the end of the input.

4. PROGRAM OUTPUT

A run of EQS produces several major sections of output information in a titled and paginated output file, using the usual 80- or 132-column print format. This chapter provides a brief description of the major output sections. You can use it mainly for reference purposes. Here we discuss the output of the typical single sample analysis found in any modeling run. Some information is given on output that is printed only under specialized circumstances, e.g., when analyzing incomplete or categorical data. Details on such specialized topics will be found in the relevant chapters. The output from a multisample analysis is basically the same within each sample, but modified slightly and extended, as is described at the end of the chapter. Listings of actual EQS runs are given elsewhere in this *Manual*.

The problem of how to transform a SEM study and its results into a meaningful report is discussed in such sources as Boomsma (2000), Hoyle and Panter (1995), and McDonald and Ho (2002). A complete report may well include information about your data beyond that provided by a SEM analysis, e.g., graphical descriptions of your data (Wainer & Velleman, 2001). If you are running EQS in a Windows or Mac environment, you can do data visualization and will have additional options for analysis as well as the associated output possibilities. These are discussed in the *User's Guide* (Bentler & Wu, 2002).

Computer Memory Problems

The amount of memory available to EQS is under your control as well as the operating system. If you do not specify how much memory EQS should use, a default allocation is made. On PCs, the default is 2000000 integer words. If the memory allocation is insufficient, EQS produces some messages that may permit an unsuccessful run to be resubmitted with modified computer requirements and then run successfully.

EQS checks memory dynamically in three stages, and it is possible for the program to abort at any of these stages without completing the computations required for the problem. If the allocation is insufficient at the first or second stage, a message will be printed indicating how many integer words were allocated, with a rough estimate of how many are needed for that stage, or a message `"MUCH MORE MEMORY WILL BE NEEDED…"`. In general, the 3rd stage needs more memory than the first two, so the estimate should be multiplied by roughly 2. If you see the latter message, multiply the memory allocation by 3, at least. If the program successfully passes the first two stages of dynamic allocation, a message (`3RD STAGE OF COMPUTATION...`) will be printed after the Bentler-Weeks structural representation has been computed and printed. If the amount required at this stage is less than the amount allocated, the program will not be hindered by any allocation problems in further computations. If the amount required is larger than the amount allocated, the program will again abort. The following message will also be printed: `INSUFFICIENT MEMORY, PROGRAM TERMINATED`. In this case, the number of required words printed is accurate, and can be used as a basis for resubmitting the job. However, the LMTEST calculations, which are a 4th stage, may need even more memory than the 3rd stage.

Program Control Information

The first output page, after any computer system-generated message, echoes back a numbered sequence of the program input. At the end of the input echo, EQS prints `xx RECORDS OF INPUT MODEL FILE WERE READ`. The number of records or lines of input read should correspond with the actual input file. This input feedback is given to help locate any problems in the input, and also to provide a summary of the proposed analysis.

If a /MODEL section is present, EQS prints `MODEL SECTION GENERATED THESE SECTIONS`: followed by the VARIANCES, EQUATIONS, and COVARIANCES sections that are translations of the MODEL section. You

should check these translations. If they are not what you intended, change the MODEL section, and rerun. Similarly, a /RELIABILITY section produces EQUATIONS and VARIANCES. Use of the SET keyword in the CONSTRAINT section of a multisample run produces a list of the additional constraints generated. Again, you should check that they are correct.

If the DELETE option was requested, the program also prints out `CASE NUMBERS DELETED FROM RAW DATA ARE:` and lists the cases that were deleted during computations. This information can be used to verify the accuracy of the specification.

If GROUPS is greater than one, the program prints out the input information from the first group first, then from the second group, and so on.

Error Messages

Minor errors. A variety of syntax-checking error messages will be printed if the input contains specifications that the program finds problematical. If the errors are relatively minor, in the sense that the program can make an educated guess as to the intended specification, a warning message will be printed and the job will proceed. Nevertheless, you should check that the program's corrective actions are appropriate.

If the errors are minor, or there are no errors, the program will continue with the computational phase. For example, if you specify a certain number of CASES, but fewer actually exist in the data file, the program will print `***WARNING*** USER SPECIFIED 100 INPUT CASES, BUT ONLY 80 CASES WERE FOUND IN DATA FILE. THE CASE NUMBER IN DATA FILE WILL BE USED TO COMPUTE CHI-SQUARES.` (Of course, the specific numbers printed depend on the situation.)

Serious errors. If the program cannot logically correct the input errors, they are considered fatal, and the program will not proceed to the computational sections. You will need to study these messages to locate the problem, and correct the errors in the input. The corrected input file will have to be resubmitted. It is possible that the error message refers to EQS commands generated by a MODEL section, RELIABILITY section, or use of the SET keyword in the CONSTRAINT section. You may either correct the original control language, or replace it by the generated language, and correct that.

Simulation Definitions

This section of output appears when a /SIMULATION is being run. The specifications of the simulation are summarized on this page of output, giving such information as the number of replications, how the sample was generated, the sample size, the normality or non-normality of the data, contamination (if any), seed, and instructions on saving the data file. The univariate characteristics of the simulated data are summarized, and the population matrix used to generate the data is shown.

Categorical Variable Analysis

When CAT =… is given in the /SPECIFICATION section, EQS prints the number of categorical variables and the number of categories of each as follows.

```
YOUR MODEL HAS SPECIFIED CATEGORICAL VARIABLES
        TOTAL NUMBER OF VARIABLES ARE         8
        NUMBER OF CONTINUOUS VARIABLES ARE    6
        NUMBER OF DISCRETE   VARIABLES ARE    2
        INFORMATION ON DISCRETE VARIABLES
           V7  WITH    3 CATEGORIES
           V8  WITH    3 CATEGORIES
```

Computations, such as thresholds and their standard errors, that are needed to compute the polyserial and polychoric correlations are then printed. The individual standard errors for these correlations may be useful as an adjunct to modeling results.

Missing Data Pattern Analysis

When there are missing data, EQS analyzes the input matrix and prints out characteristics of the missing data file. In this example, 66 of 73 subjects have some missing scores on the 6 variables. There are 29 different patterns of missing data. We show only 3 different patterns:

```
NUMBER OF CASES USED                    =   73
NUMBER OF CASES WITH POSITIVE WEIGHT    =   73
NUMBER OF CASES WITH MISSING DATA       =   66
NUMBER OF MISSING PATTERNS IN THE DATA  =   29

IN THE SUMMARY OF MISSING PATTERNS, M REPRESENTS A MISSING VALUE

                          VARIABLES
    #        #       %
 MISSING    CASES   CASES  123456
 -------    -----   -----
       0        7    9.59
       4        1    1.37   MM MM
       3        1    1.37   MM M
```

Seven subjects exhibit the first pattern, namely scores with no missing data at all. Only one subject exhibits the next pattern, with data on variables 1 and 4, but no data on variables 2, 3, 5, and 6. The third pattern again is exhibited by one case, and this person has no data on variables 2, 3, and 5. Subsequently, EQS prints out the pairwise present covariance matrix, and the sample sizes for these computations.

Sample Statistics

When raw data are read in, a number of sample statistics are computed and printed. These can be used descriptively in the usual manner. However, they also serve an important diagnostic purpose, namely, to permit evaluation of the adequacy of the input data and the appropriateness of the statistical assumptions underlying the estimation methods to be used in the analysis.

In certain circumstances, these statistics may not be meaningful. Typically, either a warning message is given, but sometimes the statistics are thus not reported. For example, ordinary univariate statistics are not too meaningful for categorical variables, or in a multilevel run, and EQS so informs you.

Univariate Statistics

Each variable's mean, standard deviation and coefficients of skewness and kurtosis are printed, based on all N data scores available for that variable when there is missing data. The mean and standard deviation are the usual statistics. Skewness is defined as

$$g_{1(i)} = N^{1/2}\Sigma_1^N (z_{it} - \bar{z}_i)^3 / [\Sigma_1^N (z_{it} - \bar{z}_i)^2]^{3/2}. \quad (4.1)$$

Under normality with N>200 or so, its standard error is approximately $(6/N)^{1/2}$. Kurtosis is given as

$$g_{2(i)} = N\Sigma_1^N (z_{it} - \bar{z}_i)^4 / [\Sigma_1^N (z_{it} - \bar{z}_i)^2]^2 - 3. \quad (4.2)$$

A useful discussion of kurtosis is given by DeCarlo (1997). When a structured means model is requested with ANALYSIS = MOMENT;, and the intercept V999 is included in the /EQUATIONS section, V999 is treated as just another variable in this section. As a constant, its mean is 1.0, and its standard deviation, skewness and kurtosis are all zero. Of course, V999 is a program-created variable for which actual input data is not provided.

Multivariate Kurtosis

Two variants of Mardia's (1970, 1974) coefficient are printed. The deviation from expectation version of Mardia's coefficient is given for *p* variables by

$$g_{2,p} = N^{-1}\Sigma_1^N [(z_t - \bar{z})' S^{-1} (z_t - \bar{z})]^2 - p(p+2) \quad (4.3)$$

where z_t and $\bar{z}$ are individual case and mean vectors. The more practical formula to use is

$$g_{2,p} / \{8p(p+2)/N\}^{1/2}, \quad (4.4)$$

the normalized estimate. This is distributed, in very large samples from a multivariate normal population, as a unit normal variate, so that large values indicate significant positive kurtosis and large negative values indicate significant negative kurtosis. In practice, values larger than 3 provide evidence of nontrivial positive kurtosis, though modeling statistics may not be affected until values are 5, 6, or beyond. When sample size is small, the significance of Mardia's coefficient is tested using the Monte Carlo method developed by Bonett, Woodward, and Randall (2002).

With missing data, we compute Yuan, Fouladi and Lambert's (2000) extension of Mardia's coefficient as

$$g_{2,p} = N^{-1}\Sigma_1^N [(z_t - \bar{z}_t)' S_t^{-1} (z_t - \bar{z}_t)]^2 - N^{-1}\Sigma_1^N p_t(p_t+2), \quad (4.5)$$

where, for the given pattern of missing data with p_t complete variables, S_t is a subset of the saturated model maximum likelihood covariance matrix estimate. The corresponding normalized estimate is

$$g_{2,p} / \{8\Sigma_1^N p_t(p_t+2)/N^2\}^{1/2}, \quad (4.6)$$

Again, this is referred to a normal distribution with mean zero and variance one.

Elliptical Theory Kurtosis Estimates

A number of estimates of the elliptical kurtosis coefficient κ are computed and printed. The Mardia-based kappa estimate, and the mean scaled univariate kurtosis, which work for complete or incomplete data, are always printed:

$$\hat{\kappa}_1 = g_{2,p} / N^{-1}\Sigma_1^N p_t(p_t+2) \quad (4.7)$$

$$\hat{\kappa}_2 = (3p)^{-1}\Sigma_1^p q_{2(i)} \quad (4.8)$$

If $\hat{\kappa}_2$ is inadmissible, it is replaced by $\hat{\kappa}_3 = (3p)^{-1}\Sigma_1^p g^*_{2(i)}$, where $g^*_{2(i)} = \max\{g_{2(i)}, -6/(p+2)\}$. Two additional estimates, the multivariate mean kappa and multivariate least squares kappa, $\hat{\kappa}_4$ and $\hat{\kappa}_5$, are also computed with AGLS estimation. See Chapter 5, eqs. (5.32) and (5.33).

The Mardia-based kappa (4.7) is a generalization of that recommended by Bentler and Berkane (1985, 1986), Browne (1982, 1984), Steiger and Browne (1984), and Shapiro and Browne (1987) for complete data. It is used as a

default in the elliptical theory computations. If this coefficient is smaller than a permissible bound it is replaced by another coefficient. The first attempted substitute is (4.8). If this also is out of the permissible range, EQS will use $\hat{\kappa}_3$ in the computations. $\hat{\kappa}_4$ and $\hat{\kappa}_5$ are not printed with missing data. The program prints out the name and value of the coefficient that is actually used.

Cases with Large Impact on Normalized Multivariate Kurtosis

The five case numbers and the corresponding estimates of the case contributions to the multivariate kurtosis normalized estimate are printed. The normalized estimate, equation (4.4) or (4.6), is the mean of such estimates across all cases in the sample. Consequently, these estimates can be compared to the normalized estimate and a judgment can be made as to whether any of these five cases are very deviant from the entire sample. If one or two cases are extremely deviant, i.e., have extremely large estimates, their raw scores and other data about these cases might be studied in order to evaluate problems such as whether any gross errors in recording may have occurred or whether, for some reason, the cases may have been mistakenly included in the sample. If not, it may be worthwhile to consider using the case-robust methodology in EQS.

Matrix To Be Analyzed

Covariance Matrix

The lower triangular section of the covariance matrix, including the diagonal, will be printed. If user-defined labels were specified, they will be printed along the rows and columns of the matrix. If labels were not specified, default labels will be printed. If the model specification is such that you have selected only part of the data to be analyzed, then just the selected portion will be printed. The number of observations is also printed.

The matrix to be analyzed will differ from the input matrix if ANALYSIS and MATRIX of /SPECIFICATIONS do not match, for example, if a correlation matrix with standard deviations was used in the input and analysis is to be on the covariance matrix. The matrix to be used in the analysis will be printed, e.g., covariances will be computed from correlations and standard deviations as input, and the covariance matrix will be printed. In addition, if the input covariance matrix is in rearranged format, it will be scrambled into the lower triangular form used by EQS during the computations. In such a case, it is valuable to check the matrix to be sure that the rearrangement was accomplished correctly. Only the variables selected for analysis will be printed.

Covariance/Mean Matrix

When ANALYSIS = MOMENT; is specified and the intercept V999 is included in the equations, and appropriate input data exists to create the matrix, a covariance/mean matrix will be analyzed and is printed. This matrix contains the covariance matrix of the selected variables in the upper triangle, and the means of these variables in the final row of the matrix. The row designation of this final row is V999, and the final entry in the row is the mean of this constant, which is 1.0.

Correlation Matrix

When ANALYSIS=CORRELATION; is specified, the correlation matrix rather than the covariance matrix is printed. This may be the ordinary product moment matrix for continuous variables, or, if CAT=xx,yy; is given, the matrix or polychoric and polyserial correlations.

Missing Data Saturated Model Matrix

If you have missing data, information is given about unstructured means and covariances under the following heading:

```
MAXIMUM LIKELIHOOD ESTIMATES OF MEANS AND COVARIANCE MATRIX (S)
BASED ON THE SATURATED (UNSTRUCTURED) MODEL
```

This is printed for information only. It is primarily used to calculate test statistics.

Missing Data Imputed Sample Means & Covariances

With missing data (not otherwise), the Jamshidian-Bentler (1999) ML method computes intermediate unstructured sample means and covariances during the estimation of a structured SEM model. These are not the saturated model estimates. These are used like the usual sample means and covariances with complete data to judge discrepancies between data and model. This is printed after the Bentler-Weeks Structural Representation (below) with the heading:

```
IMPUTED ESTIMATES OF MEANS AND SAMPLE COVARIANCE MATRIX (S)
BASED ON THE STRUCTURED MODEL
```

Bentler-Weeks Structural Representation

The program input is decoded to generate a matrix specification consistent with the Bentler-Weeks designation of dependent and independent variables. The number of dependent variables is printed, and index numbers are given for all measured and latent dependent variables, with all V variables listed first, then F variables. Then the number of independent variables is printed along with the index numbers for the variables, in the sequence V, F, E, and D. The number of free parameters and the number of fixed nonzero parameters are also printed. This information can be used, if necessary, to determine whether the model was specified as intended.

Technical Information For /OUTPUT

This segment of output will not appear in a standard EQS run. It will appear only if the DATA statement of /OUTPUT gives a file name where the technical information is to be stored, or if MULTILEVEL=HLM; in the SPECIFICATION section. Titled `FOLLOWING TECHNICAL INFORMATION HAS BEEN STORED IN XX.YY`, the section gives a summary of what information can be found in the file. It gives the parameter names that are associated with parameter estimates and standard errors, using the usual double label convention. It includes a description of the model statistics on the file, how many standard error estimates, etc. It also provides the format of the output file, and the number of lines per set of information, so that the output file can be easily decoded for further analysis.

Estimation Solution

A line of information on the estimation method and the distributional theory used in the analysis is printed next. This printed information will head each page of output associated with the given method. When this information changes, a new solution method is being reported. Immediately following the initial mention of the solution method, some critical information on the solution is printed, as discussed next.

Parameter Condition Codes

If estimates for all parameters seem to be technically acceptable, the program prints out the message:

```
PARAMETER ESTIMATES APPEAR IN ORDER.  NO SPECIAL PROBLEMS WERE ENCOUNTERED DURING OPTIMIZATION.
```

This is the ideal case, and this message should be located, prior to evaluating the meaning of any results. If any of the parameter estimates have become linearly dependent on other parameters, or if an estimate is held at a boundary, the program will print out this information in the following format.

```
PARAMETER    CONDITION CODE

V5,F1        LINEARLY DEPENDENT ON OTHER PARAMETERS
E6,E6        CONSTRAINED AT UPPER BOUND
E7,E7        CONSTRAINED AT LOWER BOUND
```

The parameter designation is based on the standard double-label designation of the parameter discussed in the /CONSTRAINTS section of Chapter 3. The condition code for a given parameter will represent one of the following:

1) *Linearly dependent on other parameters*. This code indicates that the covariance matrix of parameter estimates or its equivalent is singular, with the given estimated parameter being a linear combination of other estimated parameters. Typically this will occur because the parameter is underidentified in an equation, but it can also represent the effects of empirical underidentification, due to the data (and not the model per se).

2) *Constrained at upper bound*. This code indicates that the parameter estimate is not inside the boundaries that you specified, rather, the parameter is being held at the upper boundary.

3) *Constrained at lower bound*. This code indicates that the parameter estimate is not inside the specified boundaries. The parameter is being held at the lower boundary specified for the problem.

The constraint of a parameter at an upper or lower boundary may be a cause for celebration or a reason for distress. If the bound is desired, the solution may be totally acceptable. If the bound is not desired, it implies a possible problem. If a parameter is held to a boundary, it may be that releasing the boundary constraint would lead to a substantial improvement in fit. However, when the boundary is natural, such as nonnegativity for a variance estimate, it may not make sense to release the constraint. A useful recent discussion of improper solutions and what to do about them, if anything, is given by Chen, Bollen, Paxton, Curran, and Kirby (2001).

Linear dependence among parameters is a potentially serious problem, because the solution probably cannot be fully trusted. When the source is an identification problem, typically not fixing the scale of a factor, you should make the correction and resubmit the job. However, in a properly identified model it may reflect computational problems stemming from the data, the start values, the default technical parameters, etc. For example, modifying (often, increasing) TOLERANCE may alleviate the problem and lead to a substantially improved solution. You are urged to trace down the source of the problem, experimenting if necessary with various attempts at eliminating the difficulty.

```
VARIANCE OF PARAMETER ESTIMATE IS SET TO ZERO
```

indicates that the statistical variability of the parameter estimate cannot be accurately computed. Thus the standard error of the estimate will not, in later sections, be printed. Furthermore, the correlation of the estimate with other estimates is indeterminate. Such messages typically accompany solutions having computational difficulties, in which a diagonal element of the covariance matrix of estimates is zero or even negative.

It should be noted that even though some error messages may point to a problem in some part of the output, it is still possible that other parts can be relied upon. For example, if a solution converged cleanly, it is quite likely that the parameter estimates are correct even if there is a problem with the standard error of an estimate.

If linear constraints among parameters were specified in the /CON paragraph, and the program was able to produce estimates that meet the constraints, the program will print

```
ALL EQUALITY CONSTRAINTS WERE CORRECTLY IMPOSED.
```

If this message is not printed, the program could not impose all constraints. One possible reason for such a result is that the start values you provided did not meet the constraints. This problem is simple to fix. Other reasons are more difficult to trace, but are typically associated with general difficulties in finding a legitimate, converged solution. When the estimation method has not been able to converge to a minimum of the function, the difficulty of imposing an equality constraint may be either a cause of the problem, a consequence of it, or both. Then the main problem becomes one of tracing the origin of the convergence problem, e.g., due to poor start values, parameter underidentification, etc.

If the constraint equations in /CON appear to be linearly dependent or redundant, the program will correct the degrees of freedom to reflect the actual number of independent constraints involved. EQS will print the message: `NOTE: 2` (or some other number) `CONSTRAINTS APPEAR TO BE LINEARLY DEPENDENT. THE DEGREES OF FREEDOM HAVE BEEN ADJUSTED.`

Correlations of Parameter Estimates

If requested in the /PRINT section, the correlation matrix summarizing correlations among estimates of free parameters is printed. The number of rows and columns of this matrix represents the number of free parameters in the model. The matrix gives the large-sample correlation matrix of the parameter estimates obtained by standardizing the estimated covariance matrix of parameter estimates. The form of this covariance matrix is given by various formulae in Chapter 5. Often (e.g., ML) it is the standardized inverse of the information matrix. In many other cases, it is a triple product type "sandwich" estimator that corrects the information matrix estimate for distributional violation or other feature of the data.

This was at one time considered a valuable matrix to study, but now it is often ignored. It is a difficult matrix to study, because it is so large. It may be worthwhile to print and scan this matrix if `PARAMETER CONDITION CODES` indicate that there was some problem in parameter estimation that cannot be corrected with alternative start values, model respecifications, and so on. Most likely there will be problems associated with this matrix, for example, some of the estimated correlations may be outside the range ±1. Correlations greater than 1.0 in absolute value are easy to spot because the program prints the symbol # in each row that has such a value. The parameters having such a problem are identified by the double-label descriptions that are associated with each row and column of this matrix. These parameters are prime candidates to be studied for possible model respecification.

If a parameter estimate has no standard error or variance, there will be a nonpositive diagonal element in the covariance matrix of estimates. This would make it impossible to standardize the row or column to yield a correlation matrix in a meaningful way. In that case the corresponding row and column of correlation matrix of estimates will be zeroed out, indicating a problem with the parameter estimate. The parameter may lie on the boundary of the permissible parameter space, i.e., equal to its lower or upper bound, there may be linear dependencies among parameter estimates, or the estimate may be so poor that if it is changed the value of the function may hardly be affected.

If a parameter estimate is linearly dependent on others and some of its correlations are outside the range -1 to +1, one should suspect that the computer program failed to reach an appropriate optimum point. More generally, such solutions should be considered suspect, even if other error indicators seem all right. Currently, the only solution to such a problem in EQS is to try other runs using quite different start values and different technical constants (e.g., TOL, CON). This type of problem can also occur when a parameter is held on the lower or upper boundary, in which case the boundary constraint may be imposed inappropriately or unnecessarily. If the boundary is reasonably imposed, these standardized values should be in the range ±1 (or at least approximately so; very minor deviations will not matter much in practice).

If the parameter is involved in a linear equality constraint, as specified via the /CON paragraph, its correlation with the relevant other parameters that are involved in the constraint will be 1.0. Such a result is necessary and not a cause for concern. It also may be worthwhile to scan this matrix to determine whether there are any extremely high correlations among unconstrained parameters, indicating near dependence of the parameter estimates. Such a situation might suggest potential problems in future runs even if no problems were encountered in a given run.

Residual Covariance Matrix

The residual covariance matrix $(S - \hat{\Sigma})$ is printed on a new page. This is the discrepancy between the matrix to be modeled, usually the sample covariance matrix, and the model covariance matrix. Its values should be small and evenly distributed among variables if the model is a good representation of the data. Large residuals associated with specific variables indicate that the structural model explaining the variable is probably inadequate. The program also computes and prints two averages (ignoring signs) of the residuals. One average is based on all the elements of the lower-triangular residual matrix, while the other average ignores the diagonal elements. Usually, the off-diagonal elements are more critical to the goodness-of-fit chi-square statistics. In the future, EQS may provide indications of the significance level of individual residuals, based on formulae in Bentler and Dijkstra (1985).

Residual Covariance/Mean Matrix

If the analysis involves the constant V999, the matrix that is analyzed is the augmented covariance/mean matrix. In that case, the residual matrix that is printed contains the residual covariances in the upper triangle, and the residual means in the bottom row. The residual means are the vector $(\bar{z} - \hat{\mu})'$, with 0 as the last element for the residual of V999. It is possible for a model to fit well in means, but not covariances, and vice versa. A well-fitting model will have all residuals small.

Standardized Residual Matrix

Standardization is performed on the residual covariance matrix so that the elements are in a more interpretable range. The residual matrix $(S - \hat{\Sigma})$ is pre- and post-multiplied by the diagonal matrix of inverses of standard deviations of the variables, so that the resulting residual matrix can be interpreted in the metric of correlations among the input variables. The standardized residual matrix contains elements $r_{ij} - \hat{\sigma}_{ij} / s_i s_j$, where r_{ij} is typically the observed correlation between two variables, $\hat{\sigma}_{ij}$ is the reproduced model covariance, and s_i and s_j are the standard deviations of the measured variables V_i and V_j. Note: Bollen (1989) defines a related correlational residual as $r_{ij} - \hat{\sigma}_{ij} / \hat{\sigma}_{ii}\hat{\sigma}_{jj}$. In addition to printing this matrix, the averages of the absolute standardized residuals are printed. One average includes the diagonal elements in its computations; the other does not.

When means and covariances are analyzed, the standardized residual matrix that is printed will include residuals for means as well as covariances. The standardized mean residuals are $(\bar{z}_i - \hat{\mu}_i) / s_i$.

You will see problems with a model more easily in the standardized residual matrix than in the residual covariance matrix. Large values of standardized residuals point to the variables that are not being well explained by the model. Alternatively, *a model with tiny standardized residuals fits the data very well, regardless of what other types of fit indices may imply.* In the future, EQS may provide indications of the significance level of individual standardized residuals, based on Ogasawara (2001a).

Largest Standardized Residuals

The elements from the standardized residual matrix are ordered from large to small in absolute value, and the largest twenty of these are printed along with a designation of which pairs of variables are involved. Even though with some estimation methods, under some models, variance elements such as V5,V5 may be large, large covariance residuals tend to be more influential in lack of model fit. If V8,V2 is the largest standardized residual, and the model does not fit the data, both variables V2 and V8 must be studied as to why they are not explained well by the model. In general, it will be necessary to modify the structural antecedents or correlates of these variables, rather than their consequences. If these variables are dependent variables, residuals associated with them are antecedents to the variables, and at times it may be desirable to let them correlate.

Distribution of Standardized Residuals

A frequency distribution of the standardized residuals is printed next. Ideally, the distribution is symmetric and centered around zero. The legend for the figure describes how many residual elements are represented by a single asterisk in the figure, and gives information on the specific frequencies and percentages that fall within a given range used in the figure. Note that the diagram is labeled 1, 2, ..., 9, A, B, C. Each of these numbers or letters refers to a given range of numbers that describes the size of the residuals. "1" corresponds to residuals smaller than -.5; "2" to those in the range -.4 to -.5; "A" to those in the range .3 to .4; etc.

Model Covariance Matrix for Vs and Fs

If COVARIANCE = YES; was specified in /PRINT, at this point the program prints the model-implied, or reproduced, covariance matrix of the V and F variables. Under the title `MODEL COVARIANCE MATRIX FOR MEASURED AND LATENT VARIABLES`, the covariances of sets of V-V, F-V, and F-F variables are printed. The segment of the matrix concerned with V-V variables, of course, contains $\hat{\Sigma}$, the reproduced covariance matrix. The covariances of F-V and F-F variables may be of interest in interpretation, since these elements are not necessarily parameters of the model unless the pairs involved are independent variables.

When a structured means model is undertaken, with ANAL = MOM; this matrix will also contain the intercept V999. Entries in the row and column corresponding to V999 give the reproduced means $\hat{\mu}$ of the V and F variables based on the model. These entries may be intercepts or total effects, as well.

Model Correlation Matrix for Vs and Fs

If CORRELATION = YES; was specified in the /PRINT segment of input, the program will print the standardized version of the matrix described in the previous paragraph. This is the `MODEL CORRELATION MATRIX FOR MEASURED AND LATENT VARIABLES.` The reproduced model correlation matrix is often easier to interpret than the corresponding covariance matrix because of the standardized metric of the variables. The standardization also will affect the means that are part of a structured means model with the constant V999.

SAVE File Summary

If the SAVE section is present, EQS prints information about the file, including the name of the file, the type (ESS or TEXT), the number of variables, and the content. If the file is a TEXT file, EQS prints the number of lines

occupied by each type of information on the file (variable labels, cases of data, matrices). If an error occurred while writing the file, that will also be printed. An error is extremely unlikely, but possible, e.g. trying to write a 2-megabyte file on a 1.44-megabyte diskette.

Goodness of Fit Summary

Summary information on goodness of fit is printed next. Careful attention must be paid to the heading, since there may be several sections on fit, each one referring to a particular method being run. For example, it may say `GOODNESS OF FIT SUMMARY FOR METHOD = ML`. Then similar information may be given for another method, e.g., ROBUST.

Independence Model

The χ^2 associated with the model of independent or uncorrelated variables is presented first. This model is estimated and tested because, in small samples, it may fit as well as the model actually under consideration, and, if it does, serious questions can be raised about the explanatory meaningfulness of the hypothesized model. In large samples, the independence model is also a good baseline model against which other models can be evaluated for the gain in explanation that they achieve.

AIC

The values of Akaike's (1987) information criterion (AIC) are printed next, first for the independence model and then for the model of interest. Bozdogan's (1987) consistent version of this statistic (CAIC) is also presented for these two models. These criteria, developed for ML estimation, are applied in EQS to all the estimation methods available. It has been argued that, when selecting a model from a large number of models, one should take into account both the statistical goodness of fit and the number of parameters that have to be estimated to achieve that degree of fit. AIC and CAIC are measures designed to balance these two aspects of model fit. The model that produces the minimum AIC or CAIC may be considered, in the absence of other substantive criteria, to be a potentially useful model. Hence if you only have one given model, these indices may not be meaningful.

Let d_i be the degrees of freedom of the independence model and d_k be the degrees of freedom of the model of interest. Then EQS prints out the following:

Independence AIC = Independence model χ^2 - $2d_i$
Model AIC = Model χ^2 - $2d_k$.

Independence CAIC = Independence model χ^2 - $(\ln N + 1)d_i$
Model CAIC = Model χ^2 - $(\ln N + 1)d_k$.

Chi-Square Statistics

Next, EQS prints out the χ^2 goodness of fit test for the model of interest. A statistic T is used to evaluate the null hypothesis $\sigma = \sigma(\theta)$, i.e., whether the model holds in the population. With a sample size N, the model test statistic T is usually computed as $T = (N-1)\hat{Q}$, where Q is some variant of (1.5), the function being minimized, evaluated at the computed function minimum (see **Iterative Summary** below). Typically, the test statistic T is simply written as `CHI-SQUARE.` Sometimes a different title is used, e.g., `LIKELIHOOD RATIO CHI-SQUARE` for missing data, where a different function is optimized. The specific statistics are given in Chapter 5.

The given statistic T and tabled values of the $\chi^2_{(df)}$ distribution are used to determine the probability of obtaining a χ^2 value as large or larger than the value actually obtained, given that the model is correct. This is printed as the probability value for the χ^2 statistic. When the null hypothesis is true, the model should fit the data well and this probability should exceed a standard cut-off in the χ^2 distribution (such as .05 or .01). Thus, in a well fitting model, the probability will be large. In a poorly fitting model, the probability will be below the standard cut-off. Of course, this assumes that you are using an appropriate chi-square test.

In various situations, a variety of additional test statistics are printed. When ME=AGLS; you get the additional statistics

```
YUAN-BENTLER CORRECTED AGLS TEST STATISTIC
YUAN-BENTLER AGLS F-STATISTIC
```

and the associated p values. These are more accurate in small samples than the basic chi-square. See Chapter 5, or Yuan and Bentler (1997a, 1999a). When ME=ML,ROBUST; several distribution-free tests based on the distribution of the residuals $(S - \hat{\Sigma})$ are given. These include the test developed by Browne (1982, 1984), which works well primarily in very large samples, and its extensions to smaller samples by Yuan and Bentler (1998a) and Bentler and Yuan (1999b). Also, extensions of these tests to a variety of other modeling situations have been made in EQS 6, and hence these statistics will be computed and printed under various other circumstances (e.g., correlation structure models) as well. You will see the following

```
RESIDUAL-BASED TEST STATISTIC
YUAN-BENTLER RESIDUAL-BASED TEST STATISTIC
YUAN-BENTLER RESIDUAL-BASED F-STATISTIC
```

along with the associated p-values. The first two are referred to χ^2 distributions, while the latter is an F test.

Fit Indices

The usual test statistic $T = (N-1)\hat{Q}$ is directly proportional to sample size. That is, for some nonzero degree of discrepancy between model and data, measured by $\hat{Q}$, there will always be some sample size N where T becomes so large that rejection of the model is implied by the χ^2 test. Hence, fit indices that are less sensitive to sample size have been developed and recommended to assist in model evaluation. EQS prints out a variety of indices, depending on your specification and print request. You should select one or two to use, since these indices are very highly correlated and hence redundant (see, e.g., Hu & Bentler, 1998; Ogasawara, 2001). Hu and Bentler (1999) recommend the use of a two-index presentation strategy, selecting from among the least redundant ones. See Chapter 14 for a discussion and detailed definitions.

The following indices generally vary between 0-1, and should approach 1.0 for a well fitting model (actually, NNFI, IFI, and MFI can exceed 1).

```
BENTLER-BONETT NORMED FIT INDEX
BENTLER-BONETT NON-NORMED FIT INDEX
COMPARATIVE FIT INDEX (CFI)
BOLLEN   (IFI) FIT INDEX
MCDONALD (MFI) FIT INDEX
LISREL    GFI  FIT INDEX
LISREL   AGFI  FIT INDEX
```

Among these, the CFI index (Bentler, 1990a) generally performs well. Several fit indices should approach zero for a well-fitting model:

```
ROOT MEAN-SQUARE RESIDUAL (RMR)
STANDARDIZED RMR
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)
```

The standardized RMR is very similar to the average absolute standardized residual printed out with the standardized residual matrix. If the SRMR value is very small, the model fits the data well regardless of what other measures of fit may imply. This is a good index to pick up model misspecification (see Hu & Bentler, 1998). The RMR has a scale that depends on the variances of the variables and hence may vary in size for the same degree of misfit. An interesting feature of the RMSEA is its confidence interval, printed out as `90% CONFIDENCE INTERVAL OF RMSEA`. Confidence intervals and related results for a wider variety of indices are given by Ogasawara (2001ab). Some will be incorporated into EQS.

With AGLS estimation, only a subset of the above indices is printed. In addition, the following are given:

```
AGLS CORRECTED COMPARATIVE FIT INDEX
AGLS FIT INDEX
AGLS ADJUSTED FIT INDEX
```

The latter two were developed by Bentler (1983a), discussed by Tanaka and Huba (1985), and are sometimes also called GFI and AGFI indices (Jöreskog & Sörbom, 1994).

Reliability Coefficients

As discussed in Chapter 14, EQS computes a variety of reliability coefficients to describe the internal consistency of a hypothetical composite given by the sum of the scores on the variables being analyzed. These coefficients are printed as follows.

```
CRONBACH'S ALPHA
RELIABILITY COEFFIENT RHO
GREATEST LOWER BOUND RELIABILITY
BENTLER'S DIMENSION-FREE LOWER BOUND RELIABILITY
```

Cronbach's alpha (1951) is well-known. Developed for EQS, rho is based on the latent variable model being run. This could be any type of model with additive error variances. When this is a factor model, this gives Bentler's (1968, eq. 12) or equivalently Heise and Bohrnsted's (1970, eq. 32) omega for an equally weighted composite. With the /RELIABILITY command, it is based on a one-factor model (Raykov, 1997ab), and then rho is the same as McDonald's (1999, eq. 6.20b) coefficient omega. [Note: McDonald's omega (l.c. ω) is a special case of Heise and Bohrnsted's omega (cap Ω)]. The greatest lower bound (Jackson & Agunwamba, 1977; Bentler & Woodward, 1980) and dimension free lower-bound coefficients (Bentler, 1972) are based on a factor model with an arbitrary number of factors that underlie the variables. The former requires the estimated unique variances to be nonnegative, while the latter does not.

There are also two reliability coefficients for weighted sums, i.e., the variables are weighted differentially to achieve a higher reliability. The first, based on Shapiro (1982a), is based on no specific assumption about the number of factors and is usually computed. The coefficient is printed as

```
SHAPIRO'S LOWER BOUND RELIABILITY FOR A WEIGHTED COMPOSITE
```

followed by the title `WEIGHTS THAT ACHIEVE SHAPIRO'S LOWER BOUND`, and then, the variable weights. When a 1-factor model is run, as with the /RELIABILITY command, EQS also prints out the

```
MAXIMAL INTERNAL CONSISTENCY RELIABILITY
```

based on a 1-factor model. This was given in Bentler (1968) and has recently become popular, e.g., under the concept of construct reliability (Hancock & Mueller, 2001). Then, EQS prints the message `MAXIMAL RELIABILITY CAN BE OBTAINED BY WEIGHTING THE VARIABLES AS FOLLOWS`, and gives the weights.

Goodness of Fit for ME=ROBUST

When ME = xx, ROBUST; has been specified, a separate goodness of fit summary for METHOD = ROBUST is computed and printed as well. A key feature of this is the

```
SATORRA-BENTLER SCALED CHI-SQUARE
```

that corrects the prior xx method model chi-square in accord with principles laid out by Satorra and Bentler (1988, 1994). In EQS this concept is implemented for a wider variety of statistics than originally described. The correction applies to many more classes of correlations, initial estimators, and tests. This section also gives results on the independence model, various AIC indices, and fit indices based specifically on scaled statistics. As noted previously, you also get several residual based test statistics: `RESIDUAL-BASED TEST STATISTIC, YUAN-BENTLER RESIDUAL-BASED TEST STATISTIC, YUAN-BENTLER RESIDUAL-BASED F-STATISTIC`, and the latter two, along with the Satorra-Bentler, may be the most trustworthy under a wide a variety of conditions.

Goodness of Fit for MISSING=ML

Some special fit information is provided with missing data analyzed by ME=ML,ROBUST; The basic model test statistics are the following.

```
LIKELIHOOD RATIO CHI SQUARE
RESIDUAL-BASED TEST STATISTIC
YUAN-BENTLER RESIDUAL-BASED TEST STATISTIC
YUAN-BENTLER RESIDUAL-BASED F-STATISTIC
```

The first is the ML χ^2 test that is appropriate if the data are normal, while the others should be used if the data are not normally distributed. The latter three are new extensions in EQS 6 of the Browne (1982, 1984) and Yuan and Bentler (1998a) distribution-free residual based tests to the incomplete data situation. We use the same nomenclature for these tests in a variety of contexts, though in principle other names would be more precise. Again the latter two are the best candidates for successful performance in small samples, though no empirical information about such performance is available at this time. In addition, EQS provides several tests to determine whether the various patterns of missing data could be samples from a common population with a single mean vector and covariance matrix. These are given as follows.

```
GLS TEST OF HOMOGENEITY OF MEANS
GLS TEST OF HOMOGENEITY OF COVARIANCE MATRICES
GLS COMBINED TEST OF HOMOGENEITY OF MEANS/COVARIANCES
```

Based on Kim and Bentler (1999) and updated to reflect the most recent version of this paper, these are χ^2 tests under normality. If the null hypothesis of homogeneous means and covariances cannot be rejected, you can act as if the data are MCAR and use all patterns of responses to develop a single structural model. Distribution-free tests of homogeneity will be implemented in EQS in due course. Then there is a section that gives two test statistics developed by Yuan and Bentler (2000d) for correcting the ML statistics for nonnormality.

```
GOODNESS OF FIT SUMMARY FOR YUAN-BENTLER CORRECTION

  SCALED CHI-SQUARE (YUAN-BENTLER)     =      10.535 BASED ON       8 DEGREES OF FREEDOM
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS      .22944

  YUAN-BENTLER RESIDUAL-BASED AGLS CHI-SQUARE      =     2.135
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS      .97660
```

Goodness of Fit for MISSING=PAIR

In this case the pairwise present correlation matrix is analyzed. In principle, this can run AGLS and also ML, ROBUST, but we recommend that you use the robust methodology unless sample size is very large. The AGLS methodology is based on van Praag, Dijkstra, and Van Velzen (1985) as an extension of the Browne-Chamberlain methodology to missing data. The robust methodology was developed in EQS 6. This gives a variety of new test statistics. These include new variants of the residual based test statistics: `RESIDUAL-BASED TEST STATISTIC, YUAN-BENTLER RESIDUAL-BASED TEST STATISTIC, YUAN-BENTLER RESIDUAL-BASED F-STATISTIC` discussed previously, along with the Satorra-Bentler extension which is printed out as follows:

```
SCALED (SATORRA-BENTLER TYPE) CHI-SQUARE =      6.7271 ON     8  D.F.
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS      .56634
```

Iterative Summary

The iterative summary contains information obtained from each iteration. The `PARAMETER ABS CHANGE` is the mean of the absolute values of the vector of parameter estimates from iteration to iteration. `ALPHA` is the value of a step size parameter used during the iterations. `FUNCTION` depends on the current value of $\hat{\Sigma}$, and represents the value of the function (usually, *Q*) being minimized in that iteration. This depends on the data vector *s* being modeled, which might be a covariance matrix, a pairwise present covariance matrix, a robust covariance matrix, an ordinary correlation matrix, or a polychoric/polyserial correlation matrix. Then it also will depend on the weight matrix *W* used in the function definition, which will depend on the distribution assumed. Typically the basic χ^2 statistic is given by n = (N – 1) times the final function value, but as was noted above, a wide variety of other test statistics are also computed. They use different formulae. See Chapter 5.

The iterative summary should be scanned prior to evaluating the meaning of any result. The iterative process fails if the average of absolute values of elements of the parameter change vector is larger than about .001 (or the current convergence criterion specified in /TECH). Then the output cannot be considered to have any optimal properties. It may be necessary to restart the program, possibly with the currently best parameter estimates, to obtain further iterations to a better function value. Note that the program uses a default maximum of 30 iterations, so if this number is reached the program will very likely not have converged. A message will be printed warning you not to trust the output if convergence was not achieved. If subsequent estimation methods were requested in the input, these will be aborted since you must evaluate the seriousness of the problem. If the problem is not theoretical (e.g., underidentification), it may be practical (e.g., poor start values), and modifications of program input, including technical constants, may improve performance. If a very large number of iterations has been performed without convergence, it is typically unlikely that simple resubmission with a request for more iterations will clean up the difficulty.

Alphas other than 1.0 indicate that the program did step-size adjustment. If done frequently, this indicates that optimization problems may have been occurring. Such difficulties do not matter if, at the final iteration, the average parameter change is very small and the function is at a minimum.

In the absence of a converged solution, the iterative summary may provide a hint as to problems that the program encountered. In general, the function value should decrease across iterations, and the same effect should be noted for the parameter changes. If this did not happen, an iterative difficulty was encountered. One basis for failure is a model covariance matrix $\hat{\Sigma}$ that is not positive definite. In this case the message `IN ITERATION #n, MATRIX W_CFUNCT MAY NOT BE POSITIVE DEFINITE` will be printed after the Bentler-Weeks structural representation. Iterative difficulties do not matter if a proper solution meeting the convergence criterion is obtained in the end. However, it may suggest that future EQS runs based on similar models should use better start values.

It is still possible to do linearized estimation in EQS, where the parameters are updated only once with one iteration. This is no longer a default. In this methodology, the absolute parameter change from the starting estimates may be

quite substantial. This is usually not a cause for alarm, since the statistical theory guarantees that in large samples, under the model, linearized estimation will perform equivalently to fully iterated estimation. However, if you do not trust the estimates, it may be worthwhile to obtain a fully iterated estimator.

Measurement Equations

Unless TABLE = MATRIX; or TABLE = COMPACT; is used in the PRINT section (for which, see below), the input equations are printed in two sections. The measurement equations that relate measured dependent variables (Vs) to other variables are printed first, while the construct equations that have dependent latent variables (Fs) are printed in a separate section. Within each section the equations are printed sequentially, and thus may also be reordered as compared to their input sequence. The equations of course now contain the updated parameter estimates rather than the starting values.

Each measurement equation with optimal estimates for free parameters is printed using a three-line format. The equation is printed on the first line., e.g., F4 = .4*F3 - .143*F1 + D4. The standard error is printed on the second line immediately below each estimate. The test statistic (estimate divided by standard error) is printed below the standard error. This same format is repeated for all measurement equations.

Within the context of an appropriate model, the test statistics are univariate large-sample normal *z*-tests of the null hypothesis that a given parameter is zero in the population. Values exceeding ±1.96 (corresponding to significance at the 5% level) are marked with the symbol @, indicating that the null hypothesis can be rejected, i.e., that the structural coefficient is not zero. The standard errors should be scanned to see whether they are abnormally small. If a standard error approaches zero, the test statistic cannot be defined. This may reflect an estimation problem rather than excellent precision: it may be that a parameter is almost linearly dependent on others. It may be best to rescan the messages preceding the residual covariance matrix for possible error flags. This same evaluation should be made in subsequent sections.

The coefficients of the measurement equations can be evaluated for significance using the z test, as just stated. In all cases, however, the researcher should note that conclusions drawn from a series of univariate tests on parameters might not yield the same conclusion drawn from a simultaneous test on the significance of a set of parameters. A multivariate test of hypotheses on several coefficients, or for that matter, any other parameters, including construct coefficients, variances, and covariances, would be highly desirable. This can be done in three ways: by the χ^2 difference test, the Wald test, and the Lagrange Multiplier test. See Chapter 6 for details. A simple Wald test can be conducted automatically by specifying /WTEST in the input stream. This test would be concerned with locating parameters that can be considered zero simultaneously. The χ^2 difference test can be done as follows. One would create a restricted model in which all of the free parameters to be tested are simultaneously set to zero. The fit of this more restricted model can be compared to the previous, more highly parameterized model by taking the difference in χ^2 values obtained from the two runs, as well as by calculating the corresponding difference in df. This test would evaluate the hypothesis that all of the model-differentiating parameters can simultaneously be set to zero. EQS will run this test automatically when you use COMPARE=YES; with APRIORI or HAPRIORI in the /WTEST section; see chapter 6. The /LMTEST can also be requested when the more restricted model is run. In some situations, several parameters may be marginally unnecessary to a model by the univariate z ratios, but the multivariate test may indicate that not all can be dropped at the same time because doing so would create a significant decrement in fit. In that case, perhaps one but not another parameter can be dropped simultaneously with a set of parameters due to the differing correlations of their estimates.

Robust Statistics

When the job is run with METHOD = xx, ROBUST; robust standard errors and test statistics are also computed. See Chapter 5. They are printed in parentheses immediately below the test statistic based on the xx method. Again, the standard error estimate is printed first, and the test statistic below that. Test statistics significant at the 5% level are marked with the symbol @. Preliminary indications are that when the distributional assumption underlying the

xx method is not met, the robust statistics are more trustworthy than the ordinary statistics. Robust standard errors are, of course, computed for free parameters in construct equations, variances, and covariances as well as measurement equations.

Construct Equations

Each equation for an F-variable or factor is printed using the same format as for the measurement equations. First, the equation is printed with its final estimate. Then, on the next line just below the estimate, its standard error is printed. Finally, on the line below the standard error, the test statistic is printed. This format is repeated for all F-type or construct equations. If requested, robust standard errors and test statistics are printed inside parentheses. Test statistics significant at the 5% level are marked with the symbol @.

Equations in Matrix Form

When TABLE = MATRIX; is used in the PRINT section, the equations are printed in one section. For each variable that appears on the right side of an equation, all of its coefficients are printed in one column. As described above, standard errors, test statistics (and robust versions, if requested) are printed. Test statistics significant at the 5% level are marked with the symbol @. If there are many variables, it may be necessary to paste several pages together in order to see the full matrix of coefficients. The printout of the equations in this form is usually 25-75% longer than the separate printouts, above.

Equations in Compact Form

When TABLE = COMPACT; is used in the PRINT section, the equations are printed in one section. Each coefficient in an equation is printed in one line, with its standardized coefficient, R-squared, standard error, and test statistic (and robust versions, if requested). Test statistics significant at the 5% level are marked with the symbol @. The printout of the equations in this form is usually much shorter than the separate printouts, above.

Variances of Independent Variables

Unless TABLE = MATRIX; or TABLE = COMPACT; is used in the PRINT section (for which, see below), the variances of V, F, E, D variables are printed in four separate columns. The general format within each column is the same. The variable name is given first, followed by the label. On the same line at the right side, one finds the parameter estimate of a free parameter, or its fixed value. A standard error for the free parameter is given just below the estimate. The test statistic for the estimate is printed on the next line, immediately below the standard error. Test statistics significant at the 5% level are marked with the symbol @. This format is repeated for all independent variables in these respective categories. An arbitrarily large test statistic is replaced by a row of 9's.

Variances should always be scanned to determine whether any unusual numbers are being reported. In particular, it is necessary to evaluate whether there are any negative variance estimates, or zero estimates. Generally, negative estimates should not occur unless you overrode the program's automatic nonnegativity constraint on these parameters; however, they may occur with categorical variable models where this constraint may not be imposed. Zero estimates are likely to be constrained estimates, most likely on the boundary set by the program. If so, there may be no standard error for the estimate. A discussion of some of the causes of improper estimates, issues involved, and solutions is given by Chen, Bollen, Paxton, Curran, and Kirby (2001).

Robust standard errors and test statistics are given below the usual ones, when they are requested.

Covariances among Independent Variables

The covariances are presented in a table that immediately follows the variances of independent variables. The same format and column headings are used as in the VARIANCES output. The labels for the pair of variables involved in a covariance are printed below each other.

Variance-Covariance Matrix of Independent Variables

When TABLE = MATRIX; is used in the PRINT section, the variances and covariances, above, are printed in the form of a lower triangular matrix. Again, standard errors, test statistics (and robust versions, if requested) are printed, and test statistics significant at the 5% level are marked with the symbol @. If there are many variables, it may be necessary to paste several pages together in order to see the full matrix.

Variances in Compact Form

When TABLE = COMPACT; is used in the PRINT section, each variance is printed in one line, with its standard error and test statistic (and robust versions, if requested). Test statistics significant at the 5% level are marked with the symbol @. The printout of the variances in this form is usually much shorter than the separate printouts, above.

Covariances in Compact Form

When TABLE = COMPACT; is used in the PRINT section, each covariance is printed in one line, with its standard error and test statistic (and robust versions, if requested) and the corresponding correlation. Test statistics significant at the 5% level are marked with the symbol @. The printout of the covariances in this form is usually much shorter than the separate printouts, above.

Effect Decomposition

If the /PRINT section contains the statement EFFECT = YES; the program prints the effect decomposition at this point. If TABLE = COMPACT; is used, the direct and indirect effects are printed in a form identical to that used for **Equations In Compact Form**, above. Otherwise, the effect decomposition is printed in two segments. The `DECOMPOSITION OF EFFECTS WITH NONSTANDARDIZED VALUES` is first. Within this segment, total effects and (total) indirect effects are given, in sequence. Effects are printed in an equation-like format, with the affected consequent variable on the left of an equal sign, and the variables influencing this variable on the right-hand side. Each effect is associated with a standard error that is printed directly below the effect estimate. The z statistic for the hypothesis that the effect is zero in the population is printed immediately below. Some total effects may be direct effects, and some direct effects will be associated with free parameters so that an asterisk "*" will appear. Indirect effects are printed in the same equation-like format.

The second segment of printout is the `DECOMPOSITION OF EFFECTS WITH STANDARDIZED VALUES.` The format for this printout is similar to that for the nonstandardized values, that is, total effects first, followed by indirect effects. These standardized effects are those that obtain when all V, F, E, and D *model* variables are transformed to unit variance, as in the standardized solution (see next section). Because the effect sizes in this solution are not arbitrary, they may in many instances be easier to interpret than the nonstandardized values. However, no standard errors or test statistics are provided for the indirect effects in this solution. Statistical significance is a quality reserved for the nonstandardized effects in the current printout.

Standardized Solution

If TABLE = COMPACT; is used, the standardized solution is printed in the table of **Equations in Compact Form**, above. Otherwise, each measurement and construct equation is printed, giving the standardized solution. This is a completely standardized path analysis type of solution, except that observed variables are not standardized: all V, F, E, and D *model* variables are rescaled to have unit variance. Standardization is done for all variables in the linear structural equation system, including errors and disturbances. Consequently, all coefficients in the equations have a similar interpretation, and the magnitude of these standardized coefficients may be easier to interpret than the magnitudes of the coefficients obtained from the covariance or raw data metric. Furthermore, parameters in this metric are invariant to arbitrary identification choices. It should be noted explicitly that the standardized solution produced by EQS is not the same solution as provided by Jöreskog and Sörbom (1994), who do not standardize measured variables, errors in variables, or disturbances in equations.

The proportion of variance accounted for in each dependent variable, the squared multiple correlation (see, e.g., Aiken, West, & Pitts, 2003, p. 485), is printed in the column labeled R-SQUARED on the right side of the page. If any R-squared cannot be calculated, or differs by more than .01 from the corresponding Bentler-Raykov corrected R-squared (Bentler & Raykov, 2000), the corrected R-squared values are printed below. The latter are especially needed for nonrecursive models.

There are situations where standard errors are desired on parameters in the standardized metric. You can compute them using the bootstrap. EQS also will provide standard errors based on methods developed by Jamshidian and Bentler (2000). When available, they will be printed out in the usual format.

One feature of the standardized solution is that previously fixed parameters will take on new values. Another feature is that certain parametric constraints that were met in the problem may no longer be met in the solution. A consequence of these changes is that the interpretation of a standardized solution may, at times, be problematic.

Correlations among Independent Variables

If TABLE = COMPACT; is used, the correlations are printed in the table of **Covariances Among Independent Variables**, above. Otherwise, the correlations accompany the standardized solution when it is computed. Since the variances of the independent variables in a standardized solution must be 1.0, these are not printed. However, covariances will now be correlations. These should be scanned to see whether they are in the necessary ±1 range. If not, the solution has a problem that may imply theoretical or empirical underidentification. The problem is not necessarily a computational one, in the sense that other programs may also yield the same estimates. However, an out-of-range correlation is a theoretical problem that may imply an inadequacy in the model. Assuming that theoretical identification is not a problem, it is possible that a constrained solution could be attempted. Sometimes, one way to achieve the constraint is to reparameterize the model so that variables' variances are fixed (rather than free), and the related regression coefficients are estimated. The program then can force the relevant covariance estimate to be such that the corresponding correlation lies in the range ±1. This is done automatically by the program, but the feature can be overridden in the /INE section. However, this approach cannot solve parameter underidentification.

End of Method

This phrase delineates the end of output from a given estimation problem with a given method. If more than one estimation method is being computed, output for the subsequent method follows this message, using the output format described in the previous pages. If an /LMTEST or /WTEST was requested, it follows the last segment of output from the last method of estimation requested.

Wald Test

If a /WTEST was requested, it is printed at this point. The Wald test is a test for evaluating whether any free parameters can be restricted without significant loss of information. See Chapter 6 for more information on the purposes of the test, its method of specification, and its printed output.

Lagrange Multiplier Test

If an /LMTEST was requested, it is computed for the last method requested and printed at this point. The LM test may not run if too many parameters were requested to be evaluated, and the computer did not have enough memory to do the computations. The program prints out the message: `LAGRANGE MULTIPLIER TEST REQUIRES xx WORDS OF MEMORY. PROGRAM ALLOCATES yy WORDS`. If xx is less than or equal to yy, EQS can do the computations and will report a result. If xx is greater than yy, this segment of the program will be aborted. You will have to increase the memory allocation, or re-specify the LM test by reducing the potential fixed parameters that need to be evaluated, and then resubmit the job to be run.

The specific output from the LM test, and the interpretation thereof, is discussed in depth in Chapter 6, and will not be repeated here.

Summary of Replications

When DATA=XX.YY; is given in the OUTPUT section, and either a SIMULATION section is used, or LOOP=… is given in the SPECIFICATION section, summaries are printed, giving mean, standard deviation, skewness, kurtosis, and upper and lower 5% values for these elements of the file XX.YY: parameter estimates, standard errors, robust standard errors, sample covariance matrix, model covariance matrix, standardized solution, R-squares, and model statistics. For all those except model statistics, a separate printout appears for each group in a multisample analysis. For each estimation method, summaries are printed for all replications, for successes, and for failures. A success is a model that converged with no condition code. Quantiles are calculated by interpolation. In a lower-level hierarchical run (MULTILEVEL =HLM; in the SPECIFICATION section) a summary of parameter estimates over all hierarchical groups is printed.

Multisample Analysis

In multisample analysis, models from several groups are analyzed at the same time, typically, subject to cross-sample constraints. The output from a multiple group analysis is very similar to that described above for a single group. However, some additional features are present that bear discussion.

The program echoes back each group's input file under `PROGRAM CONTROL INFORMATION`. This permits evaluating whether the job was correctly submitted. As usual, errors, if found, will be identified, with minor errors corrected and fatal errors leading to program termination. Subsequently, the **Matrix To Be Analyzed** and **Bentler-Weeks** representation for Group 1 is given, followed by the same information for Group 2 and all subsequent groups. A complete section of output is then produced for Group 1, consisting of the information discussed above. Similar information is then printed for Group 2, followed by Group 3, and continuing to the final group being analyzed.

The crucial `GOODNESS OF FIT SUMMARY` is produced only once, at the end of all groups' output. As usual, this gives one or more χ^2 statistics, the relevant fit indices, as well as the `ITERATIVE SUMMARY` that should be consulted to determine whether the estimation procedure converged to an appropriate solution.

The goodness of fit indices are computed as in a one-group setup, though the statistics needed for their computation have been redefined in an obvious way. For the Bentler-Bonett NFI and NNFI indices, and Bentler' CFI index, a null or baseline model is defined and the current model is evaluated against this baseline. The baseline model in the multisample analysis is the model of within-group independence or uncorrelated variables (with means freely estimated, if appropriate), with no cross-group constraints. The χ^2 statistic for this multisample independence model is reported, in the usual format. The χ^2 for the model of interest is based on the entire parametric specification across all groups, taking into account all within-group and cross-group constraints. These statistics, along with their degrees of freedom, provide the components needed for calculation of NFI, NNFI, and CFI, as well as the AIC and CAIC indices, using the formulas provided above. In distribution-free estimation, the FIT and ADJUSTED FIT indices are computed in the usual way, using information from all groups. The RMSEA index is not, as of now, based on Steiger's (1998) redefinition.

The LM test, if requested, is printed at the end of the output using the standard format. In multisample analysis, this test is available for evaluating cross-group equality constraints, as well as the more standard within-group constraints. Results from this test are printed in the usual way, except that a group identifier is present for each free parameter. This test can be invaluable in evaluating whether cross-group restrictions are consistent with the data.

Execution Summary

The final printout from the program is the execution summary, which indicates the time the job began, the time it was completed, and the elapsed time (on a PC) or the CPU time (on a mainframe). This information can be used for planning or reporting purposes.

Computer system and compiler messages, if any, may be printed after this point, but they are not under the control of the EQS program.

5. STATISTICAL METHODS

A number of illustrative models were presented in Chapter 2 as a way of introducing both structural modeling and the EQS program. Chapters 3 and 4 provided specific instructions on program input and output. This chapter returns to the didactic material on structural modeling initiated in Chapter 2, emphasizing the wide range of choice in statistical methods available to you.

In the "old days," it has been said, life was simpler. So, of course, was structural modeling. If you could conceptualize a model in a form appropriate to your SEM program, the statistics would be taken care of automatically by its basic defaults. It is now known that basic defaults often do not work well. As a result, you need to know important choice points that will help you use options to get the most accurate statistical conclusions possible. In this chapter we discuss some of the fundamental statistics available to you in EQS 6. This is a far larger set than was available in EQS 5 and is substantially larger than is available in competing programs. EQS uniquely gives you access to recently published, in press, and even unpublished methods. In this chapter we provide an overview of these alternatives, focusing on test statistics for evaluating models, and standard errors for evaluating parameters. Additional options and statistics, such as Lagrange Multiplier, missing data, or multilevel statistics, also exist in EQS. These are discussed in later chapters.

Manuals for a variety of statistical packages, including virtually all SEM programs, make it hard or impossible to know just what statistics are provided. Perhaps this is done so as not to scare you away. We do not want to scare you either. But you should know that equations exist to define precisely what EQS computes. We do not want EQS to be a "black box" that cannot be penetrated by even the most technically skilled. Hence, in its **Technical Details**, this chapter gives a number of equations – many in matrix form – that define various computations and output from a modeling run. You do not need to know these equations, but you should know they exist. Hopefully, the more practical and less technical discussion on **Assumptions and Use** will be clear enough to convey the purpose of a method.

Basic Modeling Statistics

In this section we discuss the basic options available to you when you use METHOD=xx, where xx is some method. [Note: for simplicity, from here on we drop the ";" that is required in EQS input statements.] As noted in the /SPECIFICATION discussion of Chapter 3, there are 9 of these basic methods for the covariance structure analysis of a sample of continuous variables. However, these 9 methods are grouped into four classes of methods that depend on the assumptions that can be made about the distribution of your variables: normal, symmetric with tails that deviate a certain extent from normal, symmetric but with tails that deviate differentially from normal, and distributed in any arbitrary way. Subsequent parts of this chapter deal with corrections to these methods for more accurate performance under violation of assumptions, as well as with the analysis of correlations and case-weighted data.

Are you a beginner in this field and don't want to get overwhelmed with statistical complexities as you get started? OK, for now, just do the following: Set METHOD=ML,ROBUST and skip most of the material below. Use the robust output. But you should return here after you get some experience, because this method also may not be optimal for you. You need to know that there are alternatives. We start our discussion with a very general approach, though, as you will see, it often is not the most practical one.

Arbitrary Distribution Theory

Assumptions and Use

The simplest theory to consider is one in which your p continuous variables can have any distribution, that is, an arbitrary distribution. We simply want good estimates of model parameters, a chi-square test of the appropriateness of the model, and standard errors for the parameter estimates, without making strong assumptions about the data. In EQS this leads to ME=AGLS, the arbitrary distribution generalized least squares method that, for data vector s, model σ, parameter vector θ, and weight matrix W, minimizes the generalized least squares (GLS) function

$$Q = (s - \sigma(\theta))' W (s - \sigma(\theta)) \tag{1.5}$$

to get the optimal parameter estimator $\hat{\theta}$, a chi-square test

$$T_{\text{AGLS}} = n\hat{Q}, \tag{5.1}$$

based on Q evaluated at $\hat{\theta}$, and standard error estimates. The function Q is printed out during the iterations. In covariance structure analysis, s is the vector of elements in the lower triangle of S, the unbiased sample covariance matrix, and σ is the vector of elements in the lower triangle of Σ, the population model covariance matrix. In SEM, $\Sigma = \Sigma(\theta)$, and hence $\sigma = \sigma(\theta)$. The χ^2 test has d = (p* - q + r) degrees of freedom, for p* sample covariances, q free parameters, and r nondependent constraints (possibly, r = 0, since there may be no constraints). Here and elsewhere, we will use n = (N-1) when S is the unbiased sample covariance matrix; N can replace n with the biased sample covariance matrix.

A lot hinges on the proper or optimal choice of W, because T_{AGLS} may not be χ^2 distributed with a poor choice of W. This is discussed below. Even though we present the AGLS method first, it is not the oldest in the SEM field; the normal theory ML method (e.g., Jöreskog, 1969, 1971ab) as discussed below has a longer history. Nonetheless, the statistical development of estimation under minimization of Q with minimal distributional assumptions is by now almost two decades old. In its general forms, it was provided by such authors as Browne (1982, 1984), Chamberlain (1982), Dijkstra (1981), and Shapiro (1983) for covariance structures, by Bentler, Lee, and Weng (1987) for multiple population covariance structure analysis under arbitrary distributions, and by Bentler (1983a, 1983b), Bentler and Dijkstra (1985), Shapiro (1986) and Satorra (1989, 2001, 2002, 2003). The results of the latter authors are sufficiently general to include structured means and multiple population models. For example, the data vector s can contain means as well as covariances from more than one sample, and σ remains the corresponding model vector. Our language is closest to the Bentler and Dijkstra approach.

In covariance structures, the AGLS method is known as the generalized least squares (Dijkstra, 1981), minimum distance (Chamberlain, 1982; Satorra, 2001) or the asymptotically distribution free (Browne, 1982, 1984), or ADF, method. Distribution free means the same thing as arbitrary distribution, but what does "asymptotic" mean? It means that sample size N gets arbitrarily large, i.e., approaches infinity. In practice, we may only have a small to intermediate sized sample. What happens then? This can be analyzed. In (1.5), s is a vector of nonredundant elements (lower triangle) of S, the sample covariance matrix. So if S is p×p, it has p*=p(p+1)/2 distinct elements that make up the vector s. This means that the weight matrix is p*×p*. So if we work with p=30 variables, W is a 465×465 matrix, with 465(466)/2=108,345 elements. Or if we have p=35, W has 198,765 elements, i.e., five more variables almost doubles the number of elements we need to estimate. Now if our sample size is small, say N=200, we are clearly estimating a lot of weight matrix elements based on very few cases. This is not liable to work well. In practice things are even more difficult, since we obtain W by inverting a matrix made up of elements based on 4$^{\text{th}}$-order moments (see eq. 5.7 below) that can be quite unstable in small samples. Inverting this huge matrix based on few cases can fail. One can use a generalized inverse (e.g., Satorra, 1989), but getting a stable estimate may be hard. An obvious solution might be to take $W=I$. Then (1.5) is just a sum of squares of residuals, which is easy to understand and a reasonable quantity to try to minimize. Unfortunately (5.1) is then not χ^2 distributed.

All of these remarks are meant to point out that a beautiful theory is liable to have problems in implementation, and this is what empirical research has found. Unless sample size is really huge, the ADF test statistic (5.1) based on an

optimal weight matrix *W* yields distorted conclusions about the adequacy of a model (e.g., Chou, Bentler, & Satorra, 1991; Curran, West, & Finch, 1996; Hoogland, 1999; Hu, Bentler, & Kano, 1992; Muthén & Kaplan, 1992; Nevitt, 2000; Yuan & Bentler, 1998a). For example, in the simulation study of Hu, Bentler, and Kano (1992), at the smallest sample sizes the ADF test statistic virtually always rejected the true model, and sometimes 5000 cases were needed to yield nominal rejection rates. More generally, the AGLS method may not converge, i.e., it may not be possible to minimize (1.5) at all, or if it converges, it may yield problematic estimates, inaccurate test statistics (5.1), and questionable standard error estimates. "The ADF method is not recommended" (Powell & Schafer, 2001, p. 127).

There are, however, some promising adjustments to the classical theory. Yung and Bentler (1994), for example, proposed some computationally intensive modifications to the ADF test statistic. These improved its performance, but are not fully adequate either. A simple, and substantially improved, version of the T_{AGLS} statistic was developed by Yuan and Bentler (1997d, eq. 17 and CADF). They proposed

$$T_{\text{YB}} = T_{\text{AGLS}} /(1 + T_{\text{AGLS}}/\text{n}) \quad . \tag{5.2}$$

In EQS, this is printed as the `YUAN-BENTLER CORRECTED AGLS TEST STATISTIC`. As sample size gets large, T_{YB} becomes equal to the standard test (5.1), but at smaller sample sizes it more or less eliminates the typical inflation of the classical test. At the smallest sample sizes, it may overcorrect a bit, but generally it performs very well (Fouladi, 2000). A second statistic developed by Yuan and Bentler (1999a, eq. 3.1) is given by

$$T_{\text{F(AGLS)}} = \{(\text{N} - \text{d})/ (\text{nd})\}T_{\text{AGLS}} \, . \tag{5.3}$$

This is referred to an *F* distribution with d and (N – d) degrees of freedom. Called the `YUAN-BENTLER AGLS F-STATISTIC` in EQS, it is probably even better than (5.2) at the smallest sample sizes. But you should remember, there always will be some sample size that is too small and where AGLS cannot be resurrected. We recommend using robust statistics or residual-based statistics (see below) for those situations.

Additional developments also are incorporated into EQS. The AGLS standard error estimator has been improved in two ways, as discussed below. The first is obtained automatically below the standard results as the `YUAN-BENTLER CORRECTED AGLS STATISTICS`. The second is obtained with the command SE=OBS. The AGLS method also has been extended to incomplete data using the theory of van Praag, Dijkstra, and Van Velzen (1985). This will be discussed in Chapter 12 on missing data.

Technical Details

As just noted, EQS minimizes the GLS function $Q = (s - \sigma(\theta))'W(s - \sigma(\theta))$ for some choice of weight matrix *W*. Ideally, one would use some type of optimal GLS weights. Optimal weights are those that allow T_{AGLS}=n$\hat{Q}$, based on parameter estimates $\hat{\theta}$ obtained at the minimum $\hat{Q}$, to be interpreted as an asymptotic χ^2_d variate. A matrix V^{-1} gives these optimal weights. In practice, a consistent estimator

$$W = \hat{V}^{-1}, \tag{5.4}$$

is used. The matrix *V* is defined by the asymptotic distribution of the residual

$$\sqrt{n}(s - \sigma) \xrightarrow{D} \mathcal{N}(0, V). \tag{5.5}$$

That is, as n goes to infinity, under the null hypothesis, the discrepancy between data and model is asymptotically normal with covariance matrix *V*. (In many presentations, *V* is called *Γ*, but this is a standard symbol in SEM equations.) In covariance structures, without any further assumptions, the typical elements of *V* are given by

$$v_{ij,kl} = \sigma_{ijkl} - \sigma_{ij}\sigma_{kl} \, , \tag{5.6}$$

where the multivariate product moment for four variables *i,j,k,* and *l* is defined as

$$\sigma_{ijkl} = E(z_i - \mu_i)(z_j - \mu_j)(z_k - \mu_k)(z_l - \mu_l) \, , \tag{5.7}$$

and σ_{ij} is the usual sample covariance. There are other ways to compute *V* (Mooijaart & Bentler, 1985). Below we will see that under specialized distributions, such as normal, (5.7) can be simplified and hence *V* will have a more

specialized form. In EQS, when ME=AGLS, the optimal choice of weights (5.4) is used, based on sample estimates of (5.6) and (5.7), that is, with s_{ij} in place of σ_{ij}, $\bar{x}_i$ in place of μ_i, and the sample average replacing the expectation operator *E*. This results in a slightly biased estimator, but using an unbiased estimator of (5.6)-(5.7) does not improve estimation and testing (Chan, Yung, & Bentler, 1995); neither does simple iterative updating (Tanaka & Bentler, 1985). There can be a substantial bias in (5.4) even if the estimator of (5.6)-(5.7) is unbiased, see Yuan and Bentler (1997c) and eq. (5.11) below. However, a matrix of the form (5.4) assures that T_{AGLS} in (5.1) is χ^2 distributed in large samples. Further, as noted, in order to deal more effectively with nonasymptotic samples, the alternative tests (5.2) and (5.3) are also computed.

The AGLS estimator $\hat{\theta}$ is a minimum χ^2 estimator (Berkson, 1980; Ferguson, 1958). It is asymptotically equivalent to a minimum discrepancy function estimator (Browne, 1984; Shapiro, 1986), see Shapiro (1985a). Whether the test statistics are central or non-central χ^2 variates will depend on whether the modeling null hypothesis is true or not. The noncentral χ^2 also may not be appropriate in some circumstances (see e.g., Shapiro, 1983; Bentler, 1994; Yuan & Marshall, 2000; Curran, Bollen, Paxton, Kirby, & Chen, 2002).

When there are no constraints, the asymptotic covariance matrix of the AGLS estimator is given by the inverse of a matrix similar to the "information matrix" for ML estimation

$$\Delta = (\dot{\sigma}' W \dot{\sigma})^{-1}, \tag{5.8}$$

with $W = V^{-1}$, and where the matrix of partial derivatives of the model with respect to the parameters is given by $\dot{\sigma} = \partial\sigma / \partial\theta'$. In practice, this has to be adjusted for equality constraints $c(\theta) = 0$ in the model. This gives the typical AGLS covariance matrix of the parameter estimator as

$$M = \Delta - \Delta\dot{c}'(\dot{c}\Delta\dot{c}')^{-1}\dot{c}\Delta, \tag{5.9}$$

where $\dot{c} = \partial c / \partial\theta'$. In practice, an estimator $\hat{M}$ of *M* is used, with all matrices in (5.9) evaluated at $\hat{\theta}$, and the sample size divisor n is added, so that the standard error estimate for the i[th] parameter under AGLS is obtained as

$$S.E.(\hat{\theta}_i) = \sqrt{\hat{m}_{ii} / n}. \tag{5.10}$$

This is printed out immediately below the parameter estimate. Methods for testing with active inequality constraints in the population (Shapiro, 1985b) are not yet implemented in EQS, but at times it is appropriate to treat such parameters as fixed zeros (Dijkstra, 1992).

In EQS we provide two additional standard error estimators that we believe are more appropriate to use. Empirical research has shown a discrepancy between the empirical standard errors and the theoretical ones based on (5.10) (e.g., Chou & Bentler, 1995; Henly, 1993). For example, Curran (1994) found the standard errors based on (5.10) were significantly negatively biased in normal samples with Ns of 100 and 200, and also that they "cannot be trusted even under conditions of multivariate non-normality at samples of N = 1000 or less" (p. 198). Similarly, Hoogland (1999, p. 142) noted, "It is not recommended to use the ADF standard error estimator, unless the sample size is extremely large" relative to p*. Yuan and Bentler (1997c) noted that, in ordinary multivariate normal theory, it is known that even though *S* may be an unbiased estimator of Σ, S^{-1} is not unbiased for Σ^{-1}. An improved, unbiased estimator is given by $\{(n-p-1)/n\}S^{-1}$. Although we are not in the standard normal data situation, a consistent estimator of (5.8) and (5.9) that involves an inverse is needed, and hence they proposed that a similar correction would be useful here. Hence, EQS computes the Yuan-Bentler (1997c, eq. 2.6) corrected covariance matrix as

$$M_c = \{n/(n - p^* - 1)\}\, M. \tag{5.11}$$

In practice, the standard error estimate for the *i*[th] parameter, taking into account sample size, comes from the *i*[th] diagonal position of an estimator of (5.11) as

$$S.E.(\hat{\theta}_i) = \sqrt{\hat{m}_{c(ii)} / n}, \tag{5.12}$$

where $\hat{M}_c$ is an estimator of (5.11), evaluated at $\hat{\theta}$. These are printed out as `YUAN-BENTLER CORRECTED AGLS STATISTICS`, and are given in parentheses below the standard results. In small samples, especially with large p*, (5.12) is larger than (5.10); however, as $n \to \infty$, they become equal. Finally, it should be noted that the optional EQS printout `CORRELATIONS OF PARAMETER ESTIMATES` comes from $\hat{M}_c$ transformed into a correlation matrix.

In another approach to obtaining more adequate standard errors, EQS 6 for the first time also provides standard errors based on the inverse of an "observed" information-type or Hessian matrix, say M_H, instead of from the inverse of the "expected" information-type matrix as in (5.8)-(5.9). This is obtained with the command SE=OBS. Computation of M_H is based on an estimate of the second derivatives of (1.5), using the method of Jamshidian and Jennrich (2000) as extended here to AGLS methodology. M_H then replaces M in all the relevant formulas above. Standard errors based on observed information have been considered as optimal in the context of ML estimation (Dolan & Molenaar, 1991). They are likely to be more accurate in small samples, but fact little is known about their performance in practice since they have not been available in general SEM programs.

It should be noted that the theory implemented here is far more general than that of covariance structure analysis. The methodology allows modeling not just means and covariances, but also higher order moments such as skew and kurtosis, see e.g. Bentler (1983a, 1995, Ch. 10) or Mooijaart (1985a). The development of the GLS approach is a variation on statistics developed for nonlinear least squares (Jennrich, 1969). It is not restrictive (Shapiro, 1985a). Other approaches, such as estimating equations (e.g., Yuan & Chan, 2002; Yuan & Jennrich, 1998, 2000), could have been used. Now we turn to various specializations of the general theory, starting with the most restricted statistical approach, SEM based on multivariate normal distributions of variables.

Normal Theory

Assumptions and Use

If your data are continuous, or can be treated as continuous (e.g., you have Likert scales with a large number of scale points), and they are multivariate normally distributed, you should use one of the normal theory methods: LS, GLS, and ML. They are the simplest to compute and give reliable results. The maximum likelihood or ML method is the default method in EQS because of its optimal statistical qualities, namely, when sample size is large, the estimates are the most precise possible, and the standard errors are the smallest possible. This has been the classical choice ever since Jöreskog (1969) emphasized its importance in the statistics of confirmatory factor analysis. An advantage of ML is that it can be applied even when sample size is quite small, perhaps only slightly larger than the number of variables in the analysis. In this circumstance, you may also want to try the new standard errors based on the observed information matrix rather than the Fisher information matrix (with SE=OBS), since these should be more accurate in small samples. The GLS method has similar characteristics to ML, and in some circumstances is even better in small samples. However, it has some features that we do not like quite as much, e.g., in some models when ML will perfectly reproduce variables' variances, GLS may not do so.

You can evaluate normality in many different ways (e.g., Fang, Yuan, & Bentler, 1992; Liang & Bentler, 1999b). Kurtosis is a key issue. You can look at individual variables in this regard (see eq. 4.2 in Ch. 4), but Mardia's normalized multivariate kurtosis (eq. 4.4 or 4.6) is a simpler summary. It should be roughly in the +3 to –3 range, though somewhat larger values are probably not too worrisome. In small samples, we provide the Bonett, Woodward, and Randall (2002) significance test. EQS does not explicitly evaluate distributional symmetry, but if you have very skewed individual variables (see eq. 4.1), your data are clearly not multivariate normal. However, kurtosis and not skew is the key issue for covariance structure analysis (Yuan & Bentler, 1999c, 2000a). Skew matters more when you are modeling the means. (In practice, though, skew and kurtosis are correlated.) You also should not have outlier cases. You can detect these in many ways, including EQS's case contributions to Mardia's normalized kurtosis.

Normal theory estimates seem to be quite good, even under violation of normality. However, the test statistics and standard errors break down under violation of normality (Boomsma & Hoogland, 2001). For example, Hu, Bentler, and Kano (1992), Curran, West, and Finch (1996), and Nevitt (2000) showed that the normal theory χ^2 could substantially break down in nonnormal samples. More generally, in a meta-analysis of 25 studies, Powell and

Schafer (2001) found ML and GLS χ^2 statistics to be highly sensitive to nonnormality. Based on his review of 34 studies, Hoogland (1999) concluded that large positive kurtoses lead to poor performance of these statistics. He also concluded that the standard error estimates are too small with large kurtoses.

Certain statistics based on normal theory, such as the ML model χ^2 test, can at times be valid for nonnormal data. The basis for this is known as asymptotic robustness theory. To be relevant, it requires some specific conditions on the latent variables, such as the independence of errors and independence between factors and errors (e.g., Amemiya & Anderson, 1990; Anderson & Amemiya, 1988; Browne, 1987; Browne & Shapiro, 1988; Mooijaart & Bentler, 1991; Satorra, 1992, 1993, 2001, 2002; Satorra & Bentler, 1990, 1991; Shapiro, 1987; Yuan & Bentler, 1999c, 2000a). The assumption of independence is much stronger than simple uncorrelatedness that is typically used in developing SEM models. Since it cannot be effectively evaluated in practice, we recommend that you use robust corrections when data are not multivariate normal.

Technical Details

In maximum likelihood, parameter estimates $\hat{\theta}$ are chosen for the model $\Sigma = \Sigma(\theta)$ to minimize the normal theory ML function

$$F = \ln|\Sigma| + tr(S\Sigma^{-1}) - \ln|S| - \mathrm{p}\ . \tag{5.13}$$

We suppress writing $\Sigma(\theta)$ in our optimization function, here and below, and simply write Σ when this is clear. When ME=ML is requested, EQS minimizes and prints out F during the iterations, and calculates the ML based χ^2 statistic (with the usual d degrees of freedom) as

$$T_{\mathrm{ML}} = \mathrm{n}\,\hat{F}\,, \tag{5.14}$$

where $\hat{F}$ is (5.13) at the minimum. This test, which actually compares the current model to the saturated model, is sometimes called the likelihood ratio test. However, we also use a more general theory due to Browne (1974). To do this, we want to minimize Q in (1.5) under a special condition associated with normal distributions. Under normal theory assumptions, the optimal weight matrix $W = V^{-1}$ given by (5.4) based on (5.6) is not needed. Instead, we find a way to simplify. We consider a class of normal theory weights

$$W_N = .5D_p'(W_2 \otimes W_2)D_p\,, \tag{5.15}$$

where D_p is the $\mathrm{p}^2 \times \mathrm{p}^*$ [=p(p+1)/2] duplication matrix (Magnus & Neudecker, 1999). Under (5.15), Q simplifies to

$$Q_N = tr[(S-\Sigma)W_2]^2/2 \tag{5.16}$$

where again $\Sigma=\Sigma(\theta)$. This is substantially easier to minimize because W_2 is a much smaller p×p matrix. This function, given by Jennrich (1970) and developed by Browne (1974), is a weighted sum of squares of residuals between the sample covariance matrix S and Σ and is printed out during the iterations. Three main versions of this method are available:

$$\begin{array}{ll} W_2 = I & \text{for least squares (ME=LS;)} \\ W_2 = S^{-1} & \text{for generalized least squares (ME=GLS;)} \\ W_2 = \hat{\Sigma}^{-1} & \text{for reweighted least squares (RLS) (ME=ML;)} \end{array} \tag{5.17}$$

From (5.15) and (5.17) it is obvious that there are no 4th order moments involved in the minimization such as are used in (5.6), and hence this method should be much more stable in small samples than AGLS. In a sense, LS is distribution-free, since it only involves the sum of squared residual between data (S) and model (Σ). But the LS statistics that are computed by EQS are justified by multivariate normality of the variables. With GLS, the weight matrix is fixed, while with RLS the weight matrix $\hat{\Sigma}^{-1}$ is updated each iteration to reflect the current estimate of the model. EQS prints out the RLS chi-square as a variant of Browne's standard GLS method. With $\hat{Q}_N$ being (5.16) at the minimum, the test statistics are

$$T_{\mathrm{GLS}} = \mathrm{n}\,\hat{Q}_{N(\mathrm{GLS})} \text{ and } T_{\mathrm{RLS}} = \mathrm{n}\,\hat{Q}_{N(\mathrm{RLS})} \tag{5.18}$$

when based on GLS and RLS, respectively. These are large sample χ^2_d variates. If the model and distribution assumptions are correct, T_{GLS} and T_{RLS} will give about the same results in large samples, and about the same result as T_{ML}. When ML is run, EQS also prints `THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS` and gives T_{RLS}. Furthermore, Lee and Jennrich (1979) showed that RLS estimates based on updated weight matrices are simultaneously ML under normality. That is, the estimates from RLS are identical to the estimates that minimize (5.13). In fact, this is how EQS computes the ML solution.

The standard errors of ML and GLS $\hat{\theta}$ are obtained from the inverse of the information matrix, adjusted for equality constraints $c(\theta)$=0 in the model. We will use a similar notation as in (5.9), namely

$$M_N = \Delta_N - \Delta_N \dot{c}'(\dot{c}\Delta_N \dot{c}')^{-1}\dot{c}\Delta_N, \tag{5.19}$$

where now $\Delta_N = (\dot{\sigma}' W_N \dot{\sigma})^{-1}$ is the inverse of the information matrix for W_N. Thus M_N is a normal theory counterpart to M. And as before, having an estimate of (5.19), the square roots of diag($\hat{M}_N$)/n give the estimated standard errors. These standard errors are correct if the normal distribution assumption is correct, but not necessarily otherwise. As noted previously, you also can specify SE=OBS and get the standard errors based on the actual second derivatives, the Hessian or observed information, instead of the Fisher information matrix.

Statistics for LS are a bit more complicated. The chi-square test is <u>not</u> given by n $\hat{Q}_N$ in this case. We make use of information about the LS estimate, evaluated in various ML equations. Let $\hat{F}$, $\hat{M}_N$ in (5.19), and $\hat{g}_{\mathrm{ML}} = \partial\hat{F}/\partial\hat{\theta}$, the gradient of the ML function, be evaluated at the LS estimator. Then,

$$T_{\mathrm{LS}} = \mathrm{n}(\hat{F} - \hat{g}_{\mathrm{ML}}'\hat{M}_N\hat{g}_{\mathrm{ML}}) \tag{5.20}$$

gives the chi-square test for LS (Bentler, 1995, eq. 10.44). Note that if there are no constraints and LS and ML estimates coincide, $\hat{F}$ will be the same as in ML, $\hat{g}_{\mathrm{ML}}$ will be null, and hence LS and ML chi square statistics will coincide as well. The standard errors are obtained from

$$\Omega_N = M_N \dot{\sigma}' W_N V_N W_N \dot{\sigma} M_N, \tag{5.21}$$

where M_N is based on (5.19), W_N is based on the LS version of (5.15), and V_N is the optimal covariance matrix V when the data are multivariate normal

$$V_N = 2D_p^+(\Sigma \otimes \Sigma)D_p^{+\prime}. \tag{5.22}$$

Using the Moore-Penrose inverse D_p^+ of D_p, this is the inverse of W_N in (5.15) with $W_2 = \Sigma^{-1}$. We will meet this type of triple product (*MBM*) sandwich estimator (Kauermann & Carroll, 2001) again in many other situations, with M and B redefined, of course, for those situations. It is generally a good estimator, though an alternative bootstrap estimator, not computed automatically in EQS, might be even better (Yuan & Hayashi, 2000). Placing consistent estimates into (5.21), the estimated standard errors are obtained as

$$S.E.(\hat{\theta}_i) = \sqrt{\hat{\omega}_{N(ii)}/\mathrm{n}}. \tag{5.23}$$

where $\omega_{N(ii)}$ is the ith diagonal of (5.21). Typically, these LS standard errors will not be smaller than those for ML, and, as with ML, they are liable to be incorrect if the normality assumption is false. When using SE=OBS, (5.21)-(5.23) are appropriately redefined. Note: Ω and ω as used here have nothing to do with reliability coefficients omega discussed in Chapter 14. Finally, it should be noted that all standard error estimates are sample-based estimates, and hence they have their own sampling variability. Ogasawara (2002) addresses this issue and gives estimators of this variability.

It should be noted that we have taken a very general approach to normal theory statistics for arbitrary covariance structure models. There are interesting and useful alternative methods for special models and situations such as factor analysis and factor scores, intraclass correlation, errors in variables, ipsative data, truncated data and so on (e.g., Amemiya, Fuller, & Pantula, 1987; Bentler & Yuan, 1997; Berkane & Bentler, 1993, 1997; Chan & Bentler, 1993, 1996; Hayashi & Bentler, 2000; Krijnen, Wansbeek, & ten Berge, 1996; McDonald & Krane, 1979; Tang &

Bentler, 1997). Related normal theory statistics for topics such as structured means, incomplete data, and multilevel modeling is covered elsewhere in this *Manual*.

There is another methodology, called partial least squares (PLS), that is sometimes used in structural modeling. It is different in philosophy and methodology from ML and LS. PLS is not yet fully developed and hence is not implemented in EQS. Some interesting background, especially a comparison of ML and PLS, is given by Jöreskog and Wold (1982) and Dijkstra (1983). More recent references are Chin (1998), Chin and Newsted (1999), Lohmöller (1989) and Tobias (1997).

Elliptical Theory

Assumptions and Use

If variables have heavier or lighter tails than the normal distribution, but remain symmetrically distributed, they may be elliptically distributed. To get some insight, we can define the kurtosis parameter kappa for variable *i* as

$$\kappa = \sigma_{iiii} / 3\sigma_{ii}^2 - 1 . \tag{5.24}$$

Here $\sigma_{iiii} = E(z-\mu)^4$ is a special case of (5.7). It is a fourth-order moment about the mean and σ_{ii} is the variance of variable *i*. In normal distributions, κ=0, i.e., normal distributions are a special case with no "excess" kurtosis. In heavier tailed distributions (as compared to the normal), κ is positive. In lighter tailed distributions, κ is negative. In elliptical distributions, all variables have the same κ parameter (5.24), i.e., they have homogeneous kurtoses. A formal test of this exists (Berkane & Bentler, 1987b), but it is not yet implemented in EQS, so you will have to evaluate this approximately by looking at the univariate kurtosis coefficients $g_{2(i)}$, since $g_{2(i)}$ /3 is an estimate of κ. In practice, because of its optimality properties (Bentler & Berkane, 1986; Berkane & Bentler, 1990), EQS uses the Mardia-based estimator $\hat{\kappa}_1$ in (4.7) as its default estimator of (5.24).

The important advantage of elliptical distributions is that only one extra parameter, κ, needs to be estimated as compared to normal theory methods and when data are actually normal, the class of elliptical methods reduce to their normal theory counterparts. This makes the method theoretically interesting as well as very practical from a computational viewpoint (Shapiro & Browne, 1987). A disadvantage of elliptical methods is that they may not be as robust to violations of distributional assumptions as are normal theory methods. Initially optimistic performance found in simulations (e.g., Harlow, 1985) has become less so (e.g., Boomsma & Hoogland, 2001; Hoogland, 1999). There seem to be better alternatives to elliptical methods today.

Technical Details

When the multivariate distribution of the variables *z* has a mean vector μ and a covariance matrix Σ, and can be described by a density of the form

$$\mathrm{k}_1 \,|\, \Sigma \,|^{-1/2}\; g\{\mathrm{k}_2 (z-\mu)' \Sigma^{-1} (z-\mu)\} \tag{5.25}$$

where k_1 and k_2 are constants and g is a nonnegative function, contours of constant probability are ellipsoids and the variables have a common kurtosis parameter given by (5.24). More generally (Bentler, 1983a) the expected values of the fourth order multivariate product moments are related to κ by

$$\sigma_{ijkl} = (\kappa+1)(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) \tag{5.26}$$

The multivariate normal distribution is a special case of the multivariate elliptical class of distributions, having κ = 0. The parameter κ is one of a series that characterizes elliptical distributions (Berkane & Bentler, 1986, 1987a). As shown by Bentler (1983a) and Browne (1984) (see also Tyler, 1982, 1983), and as can be developed by substituting (5.26) in (5.6), in covariance structure analysis with elliptical distributions the optimal inverse weight matrix *V* can be simplified to

$$V_E = D_p^+ [2(\kappa+1)(\Sigma \otimes \Sigma) + \kappa Vec(\Sigma) Vec(\Sigma)'] D_p^{+\prime} . \tag{5.27}$$

If κ = 0, this specializes to the normal theory matrix in (5.22). However, if we generalize a bit and let W be expressed as

$$W_E = D_p'[.5(\kappa+1)^{-1}(W_2 \otimes W_2) - \delta vec(W_2)vec(W_2)']D_p, \tag{5.28}$$

the general function Q in (1.5) can be simplified to

$$Q_E = .5(\kappa+1)^{-1}tr[(S-\Sigma)W_2]^2 - \delta[tr(S-\Sigma)W_2]^2, \tag{5.29}$$

where $\delta = \kappa/[4(\kappa+1)^2 + 2p\kappa(\kappa+1)]$ (Bentler, 1983a; Browne, 1984). If $\kappa = 0$, this is the normal theory function (5.16). Under elliptical theory, minimizing (5.29) with W_2 an estimator of Σ^{-1} yields optimal or efficient estimates and a chi-square distribution for the test statistic $n\hat{Q}_E$, since in this case $W_E = V_E^{-1}$. But this is not true if Q_E is minimized when the distribution is not elliptical or if W_2 is not a consistent estimator of Σ^{-1}.

EQS defines three special cases of Q_E in (5.29), based on three definitions of W_2:

$W_2 = I$	for elliptical least squares (ME=ELS;)	
$W_2 = S^{-1}$	for elliptical generalized least squares (ME=EGLS;)	(5.30)
$W_2 = \hat{\Sigma}^{-1}$	for elliptical reweighted least squares (ME=ERLS;).	

where ERLS is based on an iteratively updated estimate of Σ. EGLS and ERLS yield a χ^2 statistics via $n\hat{Q}_E$

$$T_{\text{EGLS}} = n\hat{Q}_{E(\text{EGLS})} \text{ and } T_{\text{ERLS}} = n\hat{Q}_{E(\text{ERLS})}. \tag{5.31}$$

The large sample covariance matrix of the associated estimator $\hat{\theta}$ is given by

$$M_E = \Delta_E - \Delta_E \dot{c}'(\dot{c}\Delta_E\dot{c}')^{-1}\dot{c}\Delta_E, \tag{5.32}$$

where M_E is the constraint-corrected inverse of the information matrix $\Delta_E = (\dot{\sigma}'W_E\dot{\sigma})^{-1}$ based on W_E as given in (5.28). As usual, the square roots of diag($\hat{M}_E$)/n give the estimated standard errors. To take into account extra small sample variability, you can use SE=OBS, and (5.32) is redefined appropriately.

As in the normal theory case, ELS requires a more complicated computation as

$$T_{\text{ELS}} = n(\hat{Q}_{E(\text{ERLS})} - \hat{g}'_{\text{ERLS}}\hat{M}_{\text{ERLS}}\hat{g}_{\text{ERLS}}), \tag{5.33}$$

with definitions parallel to (5.20) but based on ERLS computations. That is, the function, gradient, and constraint adjusted inverse information matrix M are based on ERLS formulae, but evaluated at the ELS estimator $\hat{\theta}$. The standard errors cannot be taken from (5.32), but are obtained from

$$\Omega_E = M_E\dot{\sigma}'W_EV_EW_E\dot{\sigma}M_E, \tag{5.34}$$

where M_E is based on (5.32), W_E is based on the ELS version of (5.28), and V_E is (5.27), the optimal elliptical covariance matrix V. Placing consistent estimates into (5.34), the estimated standard errors are obtained as

$$S.E.(\hat{\theta}_i) = \sqrt{\hat{\omega}_{E(ii)}/n}. \tag{5.35}$$

As usual, in EQS you can replace these expected information type standard errors by observed information type standard errors using SE=OBS.

Browne (1982, 1984) and Tyler (1983) have concentrated on models that are invariant with respect to a constant scaling factor. In such models functions $f = f(\Sigma)$ have the property that for all $\alpha > 0$, $f(\Sigma) = f(\alpha\Sigma)$ (Tyler, 1983), or, for any θ there exists a θ^* such that $\Sigma(\theta^*) = \alpha\Sigma(\theta)$ (Browne, 1982, 1984). In these models the normal theory ML estimator is asymptotically efficient, the second term in (5.29) drops out at $\hat{Q}_E$, and the ML test statistic T_{ML} based on (5.14) below can be corrected by the multiplier $(\kappa + 1)^{-1}$. EQS does not specifically print out this corrected ML chi-square statistic since its appropriate use depends on verifying the scale-invariance of the model (the correction

can be easily hand-computed). For a discussion of scale invariance, see e.g., Dijkstra (1990). However, when the model does possess this property, ERLS automatically yields Browne's estimator, covariance matrix, and the quadratic form test statistic. But EGLS and ERLS yield asymptotically efficient statistics, under the assumed distribution, regardless of whether scale invariance holds.

An obvious issue is how to estimate κ in (5.24). As noted in Chapter 4, several possibilities are computed in EQS. The Mardia-based kappa, generalized here to missing data as $\hat{\kappa}_1 = g_{2,p} / N^{-1}\Sigma_1^N p_t(p_t+2)$ in (4.7), is the default. Based on (5.24), the `MEAN SCALED UNIVARIATE KURTOSIS` $\hat{\kappa}_2 = (3p)^{-1}\Sigma_1^p g_{2(i)}$ given in (4.8) is a good option if $\hat{\kappa}_1$ is not admissible. If $\hat{\kappa}_2$ is also not admissible, then with $g^*_{2(i)} = \max\{g_{2(i)}, -6/(p+2)\}$, the average $\hat{\kappa}_3 = (3p)^{-1}\Sigma_1^p g^*_{2(i)}$ is used. In addition, EQS computes multivariate mean and multivariate least squares estimates. These depend on multivariate product moments that are only computed with AGLS estimation. Based on the relation (5.26), sample centered moments s_{ijkl} and sample covariances s_{ij} are used to generate a kurtosis estimate as $\hat{\kappa}_{ijkl} = s_{ijkl}/(s_{ij}s_{kl} + s_{ik}s_{jl} + s_{il}s_{jk}) - 1$. EQS computes the `MULTIVARIATE MEAN KAPPA` as their mean across the $a = p(p+1)(p+2)(p+3)/24$ different terms

$$\hat{\kappa}_4 = \Sigma_1^a \hat{\kappa}_{ijkl} / a \ . \tag{5.36}$$

Arranging s_{ijkl} into the vector s_4, and s_2 as the vector with elements $(s_{ij}s_{kl} + s_{ik}s_{jl} + s_{il}s_{jk})$,

$$\hat{\kappa}_5 = s_2{}'s_4 / s_2{}'s_2 - 1 \tag{5.37}$$

gives the `MULTIVARIATE L.S. KAPPA`. It is also based on (5.26), using sample estimates. There are other estimates. Kano (1992) provides a general estimator that holds more broadly than only in elliptical theory; it turns out to be the same as the scaling correction of Satorra and Bentler (1988, 1994).

Since its early introduction (e.g., Beran, 1979; Cambanis, Huang, & Simons, 1981; Muirhead & Waternaux, 1980), the theory and applications of elliptical and related distributions have continued to grow (Berkane, Kano, & Bentler, 1994; Berkane, Oden, & Bentler, 1997; Fang & Anderson, 1990; Fang, Kotz, & Ng, 1990; Fang, Li, & Zhu, 1998; Gupta & Varga, 1993; Kano, 1994; Kano, Berkane, & Bentler, 1993; Liang & Bentler, 1998; Purkayastha & Srivastava, 1995; Seo & Toyama, 1996; Steyn, 1996; Wakaki, 1997). It is possible that some of the recent developments will yield an improved methodology for SEM.

Heterogeneous Kurtosis Theory

Assumptions and Use

Heterogeneous kurtosis (HK) theory (Kano, Berkane, & Bentler, 1990) deals with SEM models for symmetric distributions. This method is more general than the elliptical, because different variables can have different kurtoses. Normal and elliptical theories are special cases of HK, so that if HK is used but data are normal or elliptical, there is no penalty, and the method specializes as needed. At this time, HK is the most general method that is implemented below a fully distribution free approach. But HK has a substantial advantage over AGLS and robust methods (see below) in that its computations are hardly more difficult than those of normal theory. Only a covariance matrix and marginal kurtoses are needed, and it makes use of a small p×p weight matrix. As a result, this method can handle large sets of variables that cannot be analyzed by methods requiring a general p*×p* weight matrix *W*. Furthermore, the method is fully scale invariant, so that for models that maintain their form when variables are rescaled, the χ^2 statistic remains identical and key parameters such as factor loadings are similarly rescaled. Most models are of this form.

Although this method was developed over a decade ago, and found to be promising in one study beyond that of its developers (Hu, Bentler, & Kano, 1992), it seems not to have been studied further. Hence the limitations of the method are really not well known. An obvious limitation is that like elliptical methods, HK does not take into account nonzero skew of variables. Hence with skewed variables the method may break down. On the other hand,

this may not be a serious problem in covariance structure analysis, where the statistics depend primarily on the kurtosis of variables. One abstract problem is that the classes of variable distributions that are covered by HK theory are not really understood.

To use this method, the marginal kurtoses of the various variables are estimated and then combined in a way to approximate the fourth-order moments of variables. EQS gives the choices AVE versus GEO to accomplish this. The *Average* approach uses the arithmetic mean of the two kurtoses of a pair of variables in weights used for estimating the distribution of covariances. The *Geometric Mean* method (Bentler, Berkane, & Kano, 1991) takes the geometric mean, a square root of their product instead. The GEO approach holds for a wider variety of nonnormal distributions, and should perform better, though there is no empirical data on this. For example, in the Hu et al. (1992) study with AVE, there was a tendency for HK to overcorrect the chi-square statistic. But because of the arithmetic mean ≥ geometric mean inequality, there should be less overcorrection using HK with GEO.

Technical Details

Kano, Berkane, and Bentler (1990) proposed a structure that would simplify the general distribution free version of the optimal weight matrix V associated with (5.5) and (5.6). In particular, they proposed

$$\sigma_{ijkl} = (a_{ij}a_{kl})\sigma_{ij}\sigma_{kl} + (a_{ik}a_{jl})\sigma_{ik}\sigma_{jl} + (a_{il}a_{jk})\sigma_{il}\sigma_{jk}, \tag{5.38}$$

where $a_{ij} = a_{ji}$, elements of a matrix A, are arbitrary parameters except for the restriction that V in (5.5) is positive definite. If we define the element-wise product C_2=A*Σ (that is, $c_{2ij} = a_{ij}\sigma_{ij}$), and the vector of its lower triangular elements as c_2, then using (5.38), we can obtain the HK specialized version of the optimal V as

$$V_{HK} = 2D_p^+(C_2 \otimes C_2)D_p^{+\prime} + c_2c_2' - \sigma\sigma'. \tag{5.39}$$

Comparing this to V_N, it is clear that the first term on the right is of the form (5.22), but with C_2 replacing Σ. It also follows immediately that if all a_{ij} = 1, $C_2 = \Sigma$ and $c_2 = \sigma$, then $V_{HK} = V_N$. A similar approach can show that the elliptical V_E in (5.27) is also a special case of (5.39). Next, with some current or fixed estimator $\hat{\Sigma}$

$$W_2 = \hat{C}_2^{-1}, \tag{5.40}$$

it follows that Q specializes to the quadratic form test statistic of Kano et al. (1990, eq. 11)

$$Q_{HK} = .5tr[(S-\Sigma)W_2]^2 - .25(b'W_{HK}^{-1}b), \tag{5.41}$$

where $b' = \{tr[(S-\Sigma)W_2], tr[(S-\Sigma)W_2\hat{\Sigma}W_2]\}$, $\Sigma = \Sigma(\theta)$ to be minimized, and

$$W_{HK} = \begin{pmatrix} 1+.5p & .5tr(W_2\hat{\Sigma}) \\ .5tr(W_2\hat{\Sigma}) & .5tr(W_2\hat{\Sigma}W_2\hat{\Sigma}) - 1 \end{pmatrix}. \tag{5.42}$$

For this method to yield efficient estimates, W_2 should be a consistent estimator of $(A*\Sigma)^{-1}$. Thus (5.40) is specialized in EQS to

$$\begin{aligned} W_2 &= (A*S)^{-1} && \text{for HK generalized least squares (ME=HKGLS;)} \\ W_2 &= (A*\hat{\Sigma})^{-1} && \text{for HK reweighted least squares (ME=HKRLS;).} \end{aligned} \tag{5.43}$$

In HKGLS, W_2 is not updated, and S replaces $\hat{\Sigma}$ in W_{HK}. In HKRLS, $\hat{\Sigma}$ is updated iteratively in W_2 and W_{HK}. Associated with (5.41) is the covariance matrix of $\hat{\theta}$ as

$$\Delta_{HK} = (\dot{\sigma}'V_{HK}^{-1}\dot{\sigma})^{-1} \tag{5.44}$$

when there are no constraints, or, more generally

$$M_{HK} = \Delta_{HK} - \Delta_{HK}\dot{c}'(\dot{c}\Delta_{HK}\dot{c}')^{-1}\dot{c}\Delta_{HK}. \tag{5.45}$$

The square roots of diag($\hat{M}_{HK}$)/n give the estimated standard errors based on the expected information. With SE=OBS, (5.44) is replaced by an estimated observed information matrix.

So far, nothing was actually said about how kurtosis enters the picture, nor even why the method is called heterogeneous kurtosis. This terminology is used because variables' marginal kurtoses, that may differ, are used to define the a_{ij} in (5.38). There are two main possibilities. Kano et al. (1990, eq. 12) propose the arithmetic average method, while Bentler et al. (1991, eq. 10) propose the geometric mean method. Specifically,

$$a_{ij} = .5(\eta_i + \eta_j) \qquad \text{for HK=AVE;}$$
$$a_{ij} = \sqrt{\eta_i \eta_j} \qquad \text{for HK=GEO;}. \tag{5.46}$$

In these, $\eta_i^2 = \sigma_{iiii} / 3\sigma_{ii}^2$ is a measure of excess kurtosis for variable *i*, and is equal to (κ+1) in (5.24). With elliptical theory, all variables are presumed to have the same kurtosis, but here variables may differ in kurtosis, i.e., this is the heterogeneous kurtosis method.

For models that are invariant under rescaling of variables, with HK=AVE, Kano et al. (Theorem 2) show that (5.41) specializes to the normal theory GLS/RLS form (5.16) with redefined weights W_2. Similarly, Bentler et al. (1991) show that if the model is invariant under a constant scaling factor, with HK=GEO, the function reduces to the GLS/RLS elliptical form (5.29) for a redefined multiplier δ.

Heterogeneous Skew and Kurtosis Theory

Although we have discussed elliptical and heterogeneous kurtosis distributions as models for the actual distribution of multivariate data, in the next few years additional distributions most likely will be applied to structural equation modeling. For example, Azzalini and Valle (1996) developed a multivariate skew-normal distribution that permits asymmetric marginals. More general distributions yet, allowing greater skew and kurtosis, were developed by Yuan and Bentler (1997b, 1999b, 1999c, 2000a). Because these distributions allow modeling the 4th moments of the data, analogously to (5.26) and (5.38), but in wider classes, as well as 3rd moments, when implemented they will provide new estimators, tests, and standard errors for SEM that specialize to certain ones given above. You can expect to see these in EQS in the future. To illustrate, in one of their distributions, the asymptotic covariance matrix of sample covariances is given by

$$V_{YB} = D_p^+[2(\kappa+1)(\Sigma \otimes \Sigma) + \kappa Vec(\Sigma)Vec(\Sigma)']D_p^{+\prime} + (\kappa+1)\sum_{j=1}^{m}(\kappa_j - 3)Vech(t_j t_j')Vech'(t_j t_j') \; . \tag{5.47}$$

Here, κ is a kurtosis parameter, κ_j is another kurtosis parameter of a generating variable, t_j is a column vector associated with a decomposition of Σ, and Vech(.) stacks the lower triangle of the subsequent matrix in a vector. This is a special case of (5.5), and hence provides a new specialization of *Q* in (1.5), with all its consequences. Its generality can be seen by noting that when all κ_j=3, V_{YB} specializes to the elliptical theory matrix V_E in (5.27). When furthermore κ=0, this specializes to the normal theory matrix V_N in (5.22). But in both cases, unlike in true elliptical and normal distributions, the variables here can have arbitrary skew, and hence Yuan and Bentler call the distributions pseudo-elliptical and pseudo-normal. Since skew does not affect V_{YB}, asymptotically it does not affect covariance structure statistics.

Corrections to Basic Statistics

Robust Test and Standard Errors

Assumptions and Use

An important disadvantage of all normal theory methods, including ML, is that they can yield distorted conclusions about model adequacy under violation of the requisite distributional assumption, here, multivariate normality (e.g., Bentler & Yuan, 1999b; Curran, West, & Finch, 1996; Fouladi, 2000; Hoogland, 1999; Hu, Bentler, & Kano, 1992;

Muthén & Kaplan, 1985, 1992; Nevitt, 2000; Powell & Schafer, 2001; Yuan & Bentler, 1998a). That is, when data are not normal, but ME=ML is implemented, the test statistic T_{ML} is not distributed as a chi-square variate, and the standard errors obtained are generally not correct. Similar consequences may be true of statistics (e.g., T_E and T_{HK}) based on methods that assume any other specific distributional form (e.g., elliptical, heterogeneous kurtosis) when these assumptions are not true, though it might be expected that the problem should be less severe the more general the basic methodology. Unfortunately, in many fields, data often are nonnormal and the distributions involved are hard to specify. In psychology, for example, Micceri (1989, p. 156) reported that "An investigation of the distributional characteristics of 440 large-sample achievement and psychometric measures found all to be significantly nonnormal at the alpha .01 significance level. Several classes of contamination were found...the underlying tenets of normality-assuming statistics appear fallacious for these commonly used types of data."

In EQS, the specification ME=xx,ROBUST for any legitimate method, provides three types of added computations:

1. The Satorra-Bentler scaled chi-square test $\bar{T}$ is computed;
2. Robust standard errors are computed; and
3. Three residual-based test statistics T_{RES}, $T_{YB(RES)}$, and $T_{F(RES)}$ are computed.

The Satorra-Bentler (1988, 1994) *scaled* test statistic $\bar{T}_{ML}$ has been found to correct the problematic behavior of T_{ML} under distributional violations. It can be obtained in EQS with ME=ML,ROBUST. In principle, $\bar{T}_{ML}$ is simple to compute, via $\bar{T}_{ML} = T_{ML}/k$, where k is a scaling constant that corrects T_{ML} so that the mean of the sampling distribution of T_{ML} will be closer to the expected mean d (the degrees of freedom) under the model. Note that if k is larger than 1.0, as it will be when the distributions have more outlying cases or heavy tails than expected from normal theory, $\bar{T}_{ML}$ will be smaller than T_{ML}, i.e., it will reduce the inflation of T_{ML}. Chou, Bentler, and Satorra (1991), Fouladi (2000), Hu, Bentler, and Kano (1992), Chou and Bentler (1995), Curran, West, and Finch (1996), Hoogland (1999), and Nevitt (2000) showed with simulation studies that $\bar{T}_{ML}$ performed very well under a wide variety of nonnormal conditions, as well as with normal data. See also Powell and Schafer (2001). Though special conditions can be constructed where the statistic breaks down (e.g., Bentler & Yuan, 1999b), it remains one of the best test statistics for evaluating models with ML estimation in practice. As noted, in EQS you can now use this procedure with a wide variety of estimators with ME=xx, ROBUST, where xx could be an elliptical or HK estimator. These new combinations have not been studied. In fact these types of corrections can be applied fruitfully in a wider range of statistical models than available in EQS (e.g., Bentler & Xie, 2000).

If you are evaluating a given model, the scaled statistic $\bar{T}$ is a good choice. However, if you are intending to compare a series of nested models (see Chapter 6) using χ^2 difference tests, this may not be the optimal choice of test. Usually, the difference between two test statistics, say $T_0 - T_1$, is itself distributed as a χ^2 variate, with df given by the difference in df of the two models. However, more complicated computations are needed when the base tests are scaled tests $\bar{T}$, since the distribution of $\bar{T}_0 - \bar{T}_1$ is not well described by χ^2. Satorra (2000) gives the computations in general form. Satorra and Bentler (2001) provide a simpler procedure based on the information available from the separate $\bar{T}_0$, $\bar{T}_1$ statistics; their method is described at the end of this chapter. These issues can be avoided by using the new *residual-based* test statistics provided in EQS, e.g., $T_{YB(RES)}$, for which the usual difference test procedures hold. In fact, as discussed in the next section, these tests seem to perform as well as, or better than, the scaled test.

In the future, the *adjusted* test statistic $\bar{\bar{T}}$ given by Satorra and Bentler (e.g., 1994) also will become available. Still other variants have recently been studied and may be added as well. Fouladi (1999) recommended using a Bartlett (1950) correction based on an exploratory factor model, and applying it to the Satorra-Bentler (1994) scaled and adjusted statistics. In a simulation, she found that the Bartlett-corrected adjusted statistic performed best. Fouladi (2000) also reported that the adjusted test outperformed the scaled test in power. Nevitt (2000) also found that a Bartlett-corrected statistic performed best in his study, which was more extensive than that of Fouladi. Nevitt's final conclusion was that a number of factor corrected Satorra-Bentler scaled (not adjusted) test statistic $\bar{T}$ is best used to evaluate model fit in small samples.

In addition to incorrect model tests, incorrect parameter tests also can result from use of normal theory methods when data are nonnormal. That is, uncorrected ML-based parameter tests and confidence intervals easily can be scientifically misleading. As noted earlier, the default normal theory standard errors are obtained from the inverse of the information matrix Δ_N, see (5.8), when there are no constraints, or from M_N in (5.19) more generally. "Standard errors" from these matrices are computed with most programs, but they have no special meaning if the data are not multivariate normal. What is needed is a type of triple product, or sandwich estimator of the form $\text{acov}(\theta_i) = M_N B M_N$, where B is a matrix that takes the nonnormality into account and, if the data are normal, $M_N B M_N = M_N$, i.e., the product specializes down to the usual simpler form. This has been known for some time (e.g., Bentler & Dijkstra, 1985; Browne, 1984; Shapiro, 1986). The correct standard errors, known as the "robust" standard errors, have been available in EQS and LINCS (Schoenberg & Arminger, 1988) for over a decade, but surprisingly, researchers, reviewers, and editors often do not insist on using these statistics in spite of their good performance (e.g., Chou & Bentler, 1995; Hoogland, 1999). They are, however, not perfect. No doubt they exhibit excess variance and below nominal confidence interval coverage as shown by Kauermann and Carroll (2001) for the case of regression and estimating equations. In EQS, we have generalized this procedure. EQS produces robust standard errors not only for normal theory methods, but also for elliptical and heterogeneous kurtosis methods.

Technical Details

Satorra and Bentler (1988, 1994) developed two major modifications of standard normal theory goodness-of-fit tests such as T_{ML} to yield distributional behavior that should more closely approximate a χ^2 variate. One of these, the scaled test statistic, is available in EQS, where it has been implemented to deal with various types of initial test statistics T, whether based on normal, elliptical, or heterogeneous kurtosis theory, and also extended to various correlation structure models. The concept behind the scaled statistic is very old, going back to Bartlett (1950). The idea is to scale the given test statistic with a simple multiplier to yield a new statistic whose mean is closer to that of the reference χ^2 distribution. In this context, Satorra and Bentler (1986) noted that the general distribution of any test T is in fact not χ^2, but rather is a mixture of χ^2 variates

$$T \overset{D}{\rightarrow} \sum_1^d \alpha_i \tau_i \, . \tag{5.48}$$

Here, α_i is one of the d=(p* - q + r) nonnull eigenvalues of the matrix UV, the τ_i are independent one degree of freedom χ^2 variates, V is the true asymptotic covariance matrix of the moment data given in (5.5), and U is a residual weight matrix under the model, the constraints, and weight matrix W used in the estimation

$$U = W - W\dot{\sigma} M \dot{\sigma}' W \, . \tag{5.49}$$

In an application, W may be W_N in (5.15), W_E in (5.28), or W_{HK} based on (5.40), and M is the associated putative asymptotic covariance matrix (if the distribution were chosen correctly) of the form (5.19), (5.32), or (5.45). Note that V carries information about the true data distribution, while U carries hypothesized model as well as hypothesized data distribution information. Bentler (1994) showed how to evaluate the test statistic T directly from the mixture distribution (5.48). This approach, which involves estimating the individual $\hat{\alpha}_i$, is not currently implemented in EQS. Its performance requires further study. Satorra and Bentler proposed using only the mean and variance of (5.48) to develop corrections to test statistics. The expected value or mean of the asymptotic distribution of T is given by trace(UV). An estimate of the average $\hat{\alpha}_i$ is given by

$$k = tr(\hat{U}\hat{V})/\text{d} \, , \tag{5.50}$$

where $\hat{U}$ is a consistent estimator of (5.49) that depends on the model and assumed distribution, and $\hat{V}$ is the distribution free estimator associated with (5.5). Then the Satorra-Bentler scaled test statistic is given by

$$\bar{T} = T/k \, . \tag{5.51}$$

This statistic is referred to a χ^2 distribution with d degrees of freedom. Of course, in practice, (5.51) may be $\bar{T}_{ML}$, $\bar{T}_{GLS}$, $\bar{T}_{HK}$, depending on which version of T is being scaled. The beauty of this is that it is a simple one-parameter correction. The mnemonic $\bar{T}$ reminds you that a mean correction is used.

When all the eigenvalues α_i are equal to one, T will be χ^2 distributed without correction. Also, estimated k then will approximate 1.0, the correction (5.50) will have no effect asymptotically, and the scaled statistic (5.51) will behave like T itself. If U is based on normal theory but the variables actually are elliptically distributed, the α_i are all equal to a constant. In this case (5.50) can be used to provide an estimate of κ, and (5.51) will be asymptotically χ^2 distributed. When the α_i have a high dispersion, the scaled statistic $\bar{T}$ will be only approximately χ^2 distributed. In fact, Yuan and Bentler (1998a) showed that if the coefficient of variation of the eigenvalues is too large, $\bar{T}$ is not the best available statistic. Satorra and Bentler provide a more accurate adjustment to the statistic T to cover this circumstance, but this adjustment is not currently available in EQS. Based on Bentler and Yuan (1999b) and Yuan and Bentler (1998a), the residual-based tests discussed below should be more promising in this situation.

As noted above, ME=xx,ROBUST corrects standard errors as well, using the triple product sandwich estimator

$$\Omega = M\dot{\sigma}'WVW\dot{\sigma}M \tag{5.52}$$

as given by Bentler and Dijkstra (1985), Browne (1984), Shapiro (1986) and others. In practice, the standard errors are obtained from the square roots of $diag(\hat{\Omega}/n)$. M and W are defined above for various distributions being assumed, and often represent the inverse of an information matrix adjusted for equality constraints $c(\theta)=0$ in the model and the associated weight matrix. V is the asymptotically distribution free covariance matrix in (5.5). Note that in contrast to asymptotic distribution free theory, here V is not inverted to yield W, and hence one might expect (5.52) to be better behaved in relatively smaller samples as compared to (5.9) in the ADF case. This seems to be what happens in practice. Finally, it should be noted that in EQS, M also can be based on the Hessian rather than the expected information matrix, giving a new robust sandwich estimator that has not been studied.

Residual-based Tests

Assumptions and Use

Several new distribution-free tests computed in EQS are not simply a multiple of the function being minimized, that is, they do not take the usual form $T = n\hat{Q}$. These tests, which are automatically computed with the specification ME=xx,ROBUST are based on the distribution of the residuals $(S - \hat{\Sigma})$. The original test T_{RES}, which seems not to have been incorporated into SEM packages previously, was developed by Browne (1982, 1984). It is available in EQS as the `RESIDUAL-BASED TEST STATISTIC`. A theoretical advantage of T_{RES} is that if sample size is large enough, its distribution is fully known. Unlike the Satorra-Bentler corrected test, which attempts to center the statistic so that its mean will be closer to that of a χ^2 variate, T_{RES} is precisely distributed as an asymptotic χ^2 variate. Unfortunately, this advantage turns out to be theoretical only. When sample size is not very large, T_{RES} breaks down as spectacularly as does the ordinary AGLS/ADF test (5.1).

Browne's residual test was extended to smaller samples by Yuan and Bentler (1998a) and Bentler and Yuan (1999b). One extension is a simple modification to yield the Yuan-Bentler version $T_{YB(RES)}$ of Browne's residual-based test. The modification is similar in form to that given in (5.2) for the AGLS test. This test also is referred to a χ^2 distribution. The modification helps the test perform remarkably better in small samples without any loss of its large sample properties since, in large samples, $T_{YB(RES)}$ becomes identical to T_{RES}. In the print-out, this test is called `YUAN-BENTLER RESIDUAL-BASED TEST STATISTIC`.

Another test based on residual calculations provided by EQS is the Yuan-Bentler (1998a) residual-based F-test, $T_{F(RES)}$. It is evaluated by reference to an F-distribution. This is a more radical modification of T_{RES}, similar in form to (5.3), designed to take sample size into account more adequately. Bentler and Yuan (1999b) studied the normal theory based statistic T_{ML}, the Satorra-Bentler scaled statistic $\bar{T}_{ML}$, T_{RES}, $T_{YB(RES)}$, as well as their residual F-test. They found $T_{F(RES)}$ to perform best among these at the smallest sample sizes. It also performed very well in Nevitt's (2000) study. To summarize, the new residual-based tests can perform better than the Satorra-Bentler corrected test in small samples, with the F-test perhaps being the best available test at this time. In EQS, $T_{F(RES)}$ is printed out as the `YUAN-BENTLER RESIDUAL-BASED F-STATISTIC`.

Technical Details

Consider an optimal counterpart to the residual weight matrix U in (5.49), that is, with V^{-1} replacing W and with M in (5.9) also based on V. This is given by

$$U_{\mathrm{V}} = V^{-1} - V^{-1}\dot{\sigma}M_{\mathrm{V}}\dot{\sigma}'V^{-1}. \tag{5.53}$$

In practice, this can be evaluated at any consistent estimator $\hat{\theta}$ (to yielding $\hat{\dot{\sigma}}$ and $\hat{M}_{\mathrm{V}}$), along with a distribution-free estimator of V, to obtain a consistent estimator $\hat{U}_{\mathrm{V}}$. Browne (1982, 1984) showed that the residual-based test statistic

$$T_{\mathrm{RES}} = \mathrm{n}(s\text{-}\sigma(\hat{\theta}))'\hat{U}_{\mathrm{V}}(s\text{-}\sigma(\hat{\theta})) \tag{5.54}$$

is an asymptotic χ^2_{d} variate that can be used to evaluate model fit. In fact, T_{RES} is an appropriate test statistic to evaluate a model based on any consistent estimator $\hat{\theta}$, including those obtained under an incorrect distributional assumption. Theoretically, this is an optimal solution, but, unfortunately, as noted above, T_{RES} is often substantially inflated. Although T_{RES} is based on a more forgiving estimate $\hat{\theta}$ that can be more easily computed than the ADF estimate, it has the same practical difficulty in model evaluation as T_{AGLS}. This test statistic cannot be trusted except at the very largest sample sizes.

Yuan and Bentler (1998a, eq. 7) and Bentler and Yuan (1999b, eq. 5) proposed to modify Browne's test with

$$T_{\mathrm{YB(RES)}} = T_{\mathrm{RES}}/(1 + \mathrm{N}T_{\mathrm{RES}}/\mathrm{n}^2). \tag{5.55}$$

Aside from the small discrepancy caused by the difference between N and n (=N-1), this is similar in form to T_{YB}, the correction given for the AGLS test in (5.2). Since the denominator in (5.55) is almost always greater than 1.0, $T_{\mathrm{YB(RES)}}$ will be smaller than T_{RES} and will reduce its inflation. However, as N goes to infinity, $T_{\mathrm{YB(RES)}}$ and T_{RES} become equivalent. Simulation results in the cited studies verify this theoretical advantage of (5.55), though Bentler and Yuan (1999) found that $T_{\mathrm{YB(RES)}}$ may overaccept models at small sample sizes in comparison to nominal levels.

Yuan and Bentler (1998a, p. 294) and Bentler and Yuan (1999b, eq. 6) also propose the test

$$T_{\mathrm{F(RES)}} = \{(\mathrm{N}-\mathrm{d})/(\mathrm{nd})\}T_{\mathrm{RES}}, \tag{5.56}$$

which is similar in form to (5.3), the F-test $T_{\mathrm{F(AGLS)}}$ for AGLS. The residual-based F-test is referred to an F-distribution with d and (N-d) degrees of freedom. Asymptotically, this test is equivalent to T_{RES}, but in realistic sized samples, $T_{\mathrm{F(RES)}}$ takes degrees of freedom of the model and sample size into account in a more appropriate way. Simulation work verifies that this statistic performs well under a wide range of distributional conditions and sample sizes. Although more work is needed, indications are that this may be the best single test currently available to evaluate models under general conditions.

Correlations with Continuous Variables

Up to this point, this chapter has been concerned with proper methods for fitting a model $\Sigma(\theta)$ to a sample covariance matrix S in such a way that the resulting statistical conclusions are not compromised by violation of assumptions. There are times, however, when you may wish to analyze or model a sample correlation matrix R instead of S. Of course, you could simply use covariance structure modeling methods in fitting R by $\Sigma(\theta)$. If the model is fully scale-invariant, i.e., the model structure is maintained when rescaling S to yield R, this may yield correct test statistics as well as parameter estimates for the scale-invariant parameters, but it will lead to incorrect standard errors for scale-dependent parameters. When the model is not fully scale invariant (e.g., it has some fixed nonzero factor loadings beyond those needed for identification), incorrect parameter estimates, standard errors, and tests may be obtained. This occurs because a sample-dependent scaling is used to transform S to R. Sample rather than known population standard deviations are used in this transformation, and this sample dependency is not taken into account when fitting $\Sigma(\theta)$ to R using covariance structure statistical methods. A good discussion of the issues can be found in Cudeck (1989), Lee (1985a), and Shapiro and Browne (1990).

The obvious thing to do is to reconceptualize the problem. One way is to continue to use your covariance sample data *S*, and to incorporate into your model the "random" rescaling that takes *S* into *R*. This can be accomplished by rewriting the model as $\Sigma(\theta)=DP(\theta)D$, where $P(\theta)$ become your substantive model, now considered as a model for the population correlation matrix, and D is a diagonal matrix of population standard deviations of the variables. Thus, P has diagonal elements of 1.0, and D is estimated even though it may be substantively uninteresting (i.e., it is a so-called nuisance parameter matrix). A way of doing this correctly was proposed by Bentler and Lee (1983). See also Jamshidian and Bentler (1993), Lee (1985b), Lee and Bentler (1980), and Lee and Poon (1985). This type of approach is not currently implemented in EQS.

Another approach is the natural one of simply fitting $P(\theta)$ to R. That is, we use *correlation structure analysis*. Standard deviations or variances are simply not used. They do not show up in the model, and are not used as data. To make this work, we have to use the sampling theory of sample correlations. This is what we do in EQS. In this section we discuss how this is done with ordinary Pearson correlations for continuous variables. In the subsequent section we discuss how this is done with polychoric and polyserial correlations, which are used when some of the variables are ordinal categorical variables.

Before you do a correlation structure analysis, you should be sure that it makes sense to ignore information given by the standard deviations or variances of the variables. This would not be a good idea if changes in variances represent intimate aspects of the process being studied. For example, in a panel study, variables associated with educational outcomes often increase in variance across time. Similarly, comparisons of groups (say, ethnic differences) typically reflect differences in means and variances, in addition to differences in correlations. Clearly you should be prepared to justify a focus on only the correlations when you use correlation structures.

Assumptions and Use

To do a correlation structure analysis, use the specification command ANALYSIS=CORRELATIONS. The effect is that the sample correlation matrix R is taken as the input to structural modeling. Other than this, you proceed as you usually would. In the current version of the program, the only restriction is that you cannot use equality constraints on error variances. This restriction will be released in the future.

As usual, you have to consider the distribution of your variables and choose an appropriate estimation method. The range of options is conceptually the same as in covariance structure analysis, though in its initial release, EQS 6 will only permit normal theory and AGLS methods. Elliptical and heterogeneous kurtosis methods will be added shortly. In general the performance of the various methods can be expected to be similar to that found with covariances, e.g., with similar robustness properties and susceptibility to distortion due to violation of assumptions, but no doubt surprising new results will be obtained.

Normal theory will be the default unless you specify the ROBUST option, and as in covariance structures, you can compute standard errors based on the observed rather than expected information. This is done with the SE=OBS command. You should be aware of a peculiarity of the ML designation for correlation structures. Although EQS allows you to specify "ML", in fact no one has managed to obtain strict maximum likelihood estimates when analyzing a correlation matrix. What you do get, as you always get with this specification, is the RLS or reweighted least squares methodology. That is, an optimal weight matrix is iteratively updated, depending on current parameter values, and the RLS version of a generalized least squares test statistic is computed. In many situations, for example, in covariance structure analysis (Lee & Jennrich, 1979), this type of reweighting does in fact lead to ML estimates, but no proof exists that this occurs in the correlation context. However, like the GLS estimates, the RLS estimates are optimal in large samples. So we can say that RLS is just as good as ML would be if we could compute it, but we cannot be sure here that it is identical.

We also give a distribution free approach to correlation structures. As in covariance structures, this is called AGLS, but of course a methodology appropriate to correlation structures is used. This is critical, since simply using the correlation matrix in covariance structure AGLS leads to unreliable inference. Of course, it is probable that you will need huge sample sizes to have this method perform adequately, just as you do in covariance structures. Leung and Chan (1997) verified this difficulty in their simulation study, although, in contrast, Fouladi (2000) found quite good performance in small models.

The use of a normal theory method, or some distribution-based method, followed by ROBUST corrections, is liable to be a superior option when data are not normal. As described previously under **Corrections To Basic Statistics**, you will obtain:

1. The Satorra-Bentler scaled chi-square test $\bar{T}$;
2. Robust standard errors; and
3. Three residual-based test statistics T_{RES}, $T_{\mathrm{YB(RES)}}$, and $T_{\mathrm{F(RES)}}$.

We use the covariance structure labels for these statistics, although they have been reconceptualized in EQS to be relevant to correlation structure specifications. In fact, these statistics are new in this context. Although to our knowledge there has been no empirical study of the performance of these statistics in correlation structure models, we have every reason to believe that covariance structure findings will generalize to correlations. Thus, we would expect $\bar{T}$ to perform generally well, T_{RES} to perform badly except at very large sample sizes, and $T_{\mathrm{F(RES)}}$ to be most reliable under a wide range of situations.

Technical Details

Since the diagonal elements of the sample correlation matrix R do not vary from sample to sample, the statistical theory is only concerned with the distribution of the $\tilde{p}$ = p(p-1)/2 lower triangular elements of R. Suppose that these elements are arranged into the $\tilde{p} \times 1$ vector r, and the corresponding elements of the model population covariance matrix $P=P(\theta)$ are placed into the vector $\rho = \rho(\theta)$ with elements ρ_{ij}. For AGLS, EQS uses the distribution-free approach to correlation structures given by de Leeuw (1983, eqs. 1-4). In this approach, our general function (1.5) to be minimized is maintained, but its terms are redefined. Correlations are used instead of covariances, and the weight matrix is suitably redefined. That is, we minimize

$$Q = (r - \rho(\theta))'\tilde{W}(r - \rho(\theta)) \tag{5.57}$$

for some appropriate $\tilde{p} \times \tilde{p}$ weight matrix $\tilde{W}$. Obviously this is identical to (1.5) with the specification that s is the vector of sample correlations and $\sigma(\theta)$ is a model for those correlations. We have simply rewritten the function here to be clear. The optimal choice of $\tilde{W}$ depends on the large sample distribution

$$\sqrt{n}(r - \rho) \xrightarrow{D} \mathcal{N}(0, \tilde{V}). \tag{5.58}$$

The covariance matrix $\tilde{V}$ gives the sampling distribution of r. The general form of this matrix has been known for a long time, and rediscovered several times (e.g., Browne & Shapiro, 1986; Hsu, 1949; Neudecker & Wesselman, 1990; Steiger & Hakstian, 1982). It is a $\tilde{p} \times \tilde{p}$ matrix with elements

$$\tilde{v}_{ij,kl} = \rho_{ijkl} + .25\rho_{ij}\rho_{kl}\{\rho_{iikk} + \rho_{jjkk} + \rho_{iill} + \rho_{jjll}\} - .5\rho_{ij}\{\rho_{iikl} + \rho_{jjkl}\} - .5\rho_{kl}\{\rho_{ijkk} + \rho_{ijll}\}. \tag{5.59}$$

Here, the fourth-order term

$$\rho_{ijkl} = \sigma_{ijkl}/(\sigma_{ii}\sigma_{jj}\sigma_{kk}\sigma_{ll})^{1/2} \tag{5.60}$$

is a standardized counterpart to (5.7), obtained by dividing the 4-th order moments by the standard deviations of the variables involved. Clearly, (5.58)-(5.60) are variants of (5.5)-(5.7) as described for covariance structures with AGLS estimation. As in (5.4), we choose $\tilde{W} = \tilde{V}^{-1}$, yielding an optimal estimator with minimal variance. In practice, (5.59)-(5.60) are estimated by substituting sample moments for population moments. Mooijaart (1985b) gave a simple alternative way to estimate $\tilde{V}$ as the ordinary sample covariance matrix of appropriately defined terms.

Optimal test statistics and standard errors are obtained precisely as in AGLS covariance structure estimation for the redefined terms (5.57)-(5.60). That is, we obtain T_{AGLS}, T_{YB}, $T_{\mathrm{F(AGLS)}}$ as test statistics as in (5.1)-(5.3). The degrees of freedom for T_{AGLS} and T_{YB} are d = ($\tilde{p}$ - q + r). In typical models, this is actually the same as in covariance

structures, since either (a) p variable error variances are not estimated, and are obtained subtractively from 1.0, or (b) the identical model is estimated but p constraints are added to assure that diag($\hat{P}$)=I. The df for the F-test are similarly the same.

Although T_{AGLS} may perform acceptably when the number of variables is small (Fouladi, 2000), on the basis of Leung and Chan (1997) we would expect T_{AGLS} to perform badly in medium to large models, except at very large sample sizes, and T_{YB} to perform in a superior way. The correlation-based $T_{F(AGLS)}$ has not been studied, but we would expect it to perform best. Standard errors are obtained as in (5.10) and (5.12), and we would expect (5.12) to be superior for the reasons discussed under covariance structures. Hessian based standard errors are available also. As in (5.57), for precision we could give new symbols and names to these statistics in the correlation structure context, and write a set of new equations. However, since they are conceptually identical to their covariance structure counterparts after correlational components (5.58)-(5.60) are appropriately substituted, *the same notation is used for statistics in covariance and correlation structures* where there is no ambiguity. More importantly, we have taken statistics that were originally derived in a covariance structure context, and generalized them so that they apply to correlation structures. And to make it easy to use these statistics across modeling contexts, we are using the same names wherever possible.

Next we describe correlation structures with normal data. This can be obtained using the normal theory special case of $\tilde{V}$ in the general function (5.57). Substituting the normal theory version of (5.26) in (5.59) and simplifying yields

$$\tilde{v}_{ij,kl(N)} = .5\rho_{ij}\rho_{kl}(\rho_{ik}^2 + \rho_{il}^2 + \rho_{jk}^2 + \rho_{jl}^2) + \rho_{ik}\rho_{jl} + \rho_{il}\rho_{jk} - \rho_{ij}(\rho_{jk}\rho_{jl} + \rho_{ik}\rho_{il}) - \rho_{kl}(\rho_{jk}\rho_{ik} + \rho_{jl}\rho_{il}) . \quad (5.61)$$

This form of the asymptotic covariance matrix is well known (Olkin & Siotani, 1964; Steiger & Hakstian, 1982). With this definition, we can obtain an optimal normal theory weight matrix by choosing $\tilde{W} = \tilde{V}^{-1}$. We could compute this numerically for any estimator, but Jennrich (1970, eq. 3.8) gave the symbolic form of the inverse of (5.61), thus providing a simple approach to optimal normal theory GLS estimation through use of the AGLS function (5.57). However, (5.57) requires the use of a large $\tilde{p} \times \tilde{p}$ weight matrix, which is not desirable. Jennrich (eq. 3.9) also showed how to simplify a quadratic function like (5.57) in the normal theory case so that the resulting χ^2 test statistic requires only small p×p matrices. Jennrich's quadratic form is general, and thus obviously applies to correlation structures as noted by de Leeuw (1983) and Shapiro and Browne (1990). We make use of Jennrich's function (7.1) and our (5.16) and extend it. Specifically, we optimize

$$Q_N = tr[(R-P)W_2]^2/2 - e'W_r e \quad (5.62)$$

where W_2 is a weight matrix that, as in covariance structures, we take to be

$$\begin{array}{ll} W_2 = I & \text{for least squares (ME=LS;)} \\ W_2 = R^{-1} & \text{for generalized least squares (ME=GLS;)} \\ W_2 = \hat{P}^{-1} & \text{for reweighted least squares (RLS) (ME=ML;),} \end{array} \quad (5.63)$$

e is the column vector of diagonal elements of the modified residual matrix

$$e = diag[(R-P)W_2] , \quad (5.64)$$

and W_r is a weight matrix obtained with the use of Hadamard (element-wise) multiplication,

$$W_r = (I + W_2 * W_2^{-1})^{-1} . \quad (5.65)$$

The function (5.62) is a variant of the GLS covariance structure counterpart (5.16). Jennrich recognized this, stating that the subtractive term "may be viewed as a correction employed when testing correlation matrices" (p. 908). In his (7.1), Jennrich took P to be a known fixed matrix, and $W_2 = P^{-1}$. We also take W_2 to be a fixed matrix, though possibly a non-optimal ($W_2 = I$) or an iteratively updated ($W_2 = \hat{P}^{-1}$) matrix. Shapiro and Browne (1990, eq. 7.1) proposed use of the inverse of an unspecified consistent estimator of $P(\theta)$, or $W_2 = P^{-1}$ considered as a matrix function of parameters θ (eq. 7.2).

The test statistics defined in (5.18), namely, $T_{GLS} = n\hat{Q}_{N(GLS)}$ and $T_{RLS} = n\hat{Q}_{N(RLS)}$ are χ^2_d variates when defined on (5.62) in this context. We also obtain standard errors as given in (5.19), or, with SE=OBS, using 2nd derivatives of the function in place of expected 2nd derivatives. *Again the statistics for covariance and correlation structures have the same labels, though the defining formulae may be the same or may vary slightly.* In EQS, the ANALYSIS and METHOD statements pick the correct function and the appropriate weights. *The specification "ME=ML" is used only for consistency of notation. ML estimates are not obtained – RLS estimates are obtained!* Finally, we also provide the LS method, which does not conform to Jennrich's specification, but for which we obtain test statistics and standard errors in the correlation structure analogue to (5.20) and (5.21).

As noted previously, correlation structures estimated under normal theory can be misleading if the variables do not have normal distributions (Fouladi & Steiger, 1999). In this situation, you should use the ROBUST option. Equations (5.48)-(5.52) that define the Satorra-Bentler scaled chi-square test $\bar{T}$ and robust sandwich covariance matrix Ω are computed in EQS for correlations, using suitably redefined correlation structure statistics and matrices. That is, $\bar{T}$ is based on a correlation structure version of the scaling correction *k*, which is obtained using a matrix $\tilde{U}$ analogous to *U* in (5.49), based on $\tilde{W}$ and, in obvious notation, $\tilde{M}$, as well as an estimate of $\tilde{V}$ in (5.59). Robust standard errors are obtained using the redefined matrices in definition of the sandwich covariance matrix Ω in (5.52). The estimates also may be based on the observed instead of expected information. The ROBUST option also prompts the computation of the three residual based statistics (5.53)-(5.56), namely, T_{RES}, $T_{YB(RES)}$, and $T_{F(RES)}$. They depend on the correlation structure version of U_V in (5.53), based on $\tilde{V}$, as well as $\tilde{M}_V$. These are new statistics for correlation structures, described here and provided in EQS for the first time.

Correlations with Categorical Variables

The previous section described how a correlation structure $P(\theta)$ could be fit to *R*, the matrix of ordinary Pearson correlations for continuous variables. If the variables are ordinal categorical variables, however, the matrix *R* may not be the best matrix to model. It may be better to model a matrix, say R_p, which contains sample polychoric and polyserial correlations as well as Pearson correlations, and to fit the model $P(\theta)$ to R_p. The null hypothesis is that the population counterpart to R_p depends on parameters, i.e., $P_p = P(\theta)$. This hypothesis is appropriate when the scores on some variables cannot be treated as continuous, but rather represent ordered numbers such as {1, 2, 3, 4, 5} that have no special meaning besides their ranking. Almost everyone has encountered Likert scales with 5-point, 7-point, or 9-point category scores. Often the categories are given verbal labels (e.g., "disagree strongly," "disagree," "neutral," "agree," "agree strongly"), and the subject is asked to respond by choosing one. Numerical scores used to represent these categories clearly should be rank-ordered to reflect increasing agreement (or disagreement). Other than rank-ordered, a given numerical score, say "2" for "disagree," may be considered as an arbitrary score, providing that it falls between the scores for "disagree strongly" and "neutral." Without use of optimal scaling (e.g., Michailidis & de Leeuw, 1998), or even a stronger measurement methodology (Savage & Ehrlich, 1992; but see Cliff, 1992), it probably is impossible to rule out {1, 1.3, 3, 4, 5} or any other set of increasing numbers as equally acceptable category scores. As a result, the ordinary correlations among such ordinal variables may change with – i.e., are not invariant to – an arbitrary ordinal transformation of category scores. Luckily, a choice of optimal category scores can be avoided by using only the rank order information in each variable to generate correlations. The R_p methodology is the most developed among such methods and is discussed in this section. Other approaches can be used to handle specialized situations (e.g., Bock, Gibbons, & Muraki, 1988; Ferrando & Lorenzo-Seva, 2001; Long, Feng, & Cliff, 2003; Maydeu-Olivares, 2001).

Ordinary correlations using ordinal variables with few categories, say 2-3, and especially with skewed responses, will tend to underestimate the "true" correlations, and functions of them such as factor loadings, as well as possibly erroneously implying lack of fit of a true SEM model (Olsson, 1979b). See MacCallum, Zhang, Preacher, and Rucker (2002) for a recent discussion. Typically, Likert response scales have 5, 7, or 9 scale points. When respondents distribute themselves well across all categories (especially, symmetrically around the center category), this may be large enough so that treating the ordinal scores as continuous will lead to only minor distortions. For example, the Pearson correlation between a variable with given category scores, and any other variable based on the

same data but using another ordinal scoring scheme, can be sufficiently large that it may hardly matter which scoring scheme is used. In the context of SEM, this implies, and it has been verified, that there are circumstances when ignoring the ordinal nature of the variables, and using ordinary correlations for continuous variables, will lead to minimal distortion (e.g., Bollen & Barb, 1981; Coenders, Satorra, & Saris, 1997; Collins, Cliff, McCormick, & Zatkin, 1986; Nandakumar, Yu, Li, & Stout, 1998). In particular, the sample size requirements with ordinary correlations are much lower than needed for ordinal methods.

Conceptually, it makes sense to treat ordinal data as continuous when the scores themselves can be interpreted as equal interval numbers, where distances between numbers have the same meaning at all points on the continuum. However, if there is a real question as to whether, say, the distance between 1 and 2 is the same as the distance between 4 and 5, then a method that simply uses the order information in the data seems more appropriate. Such a methodology, based on the idea of underlying latent variables, is described in this section. It is not the only one that might be used; for example, item response theory is another alternative (see Moustaki, 2001; Jöreskog & Moustaki, 2001). But the polychoric-polyserial correlation approach is the most well-developed approach to date.

Assumptions and Use

To use this methodology, you must use two commands. First, since this is a correlation structure analysis, you use the specification command ANALYSIS=CORRELATIONS. Second, you have to give the names of the categorical variables. For example, you state CATEGORICAL=V7,V8; and the program does the rest. Of course, as always you should state what statistical method you want to use. You have two classes of options: ME=xx,ROBUST and ME=AGLS. The former allows you to specify LS, GLS, and ML (RLS) to get *continuous variable* correlation structure estimates and tests based on our extension of Jennrich (1970) as described above. Note that this is a mis-specification, since the variables are actually not continuous. While the *estimates* will be good ones, you must follow up with the ROBUST option to give you the correct robust Satorra-Bentler and Yuan-Bentler tests and standard errors as described earlier. The AGLS alternative requires larger sample sizes.

Now that you know how to do it, let us discuss further when you should do so. A fundamental question in ordinal SEM is whether a linear model such as the Bentler-Weeks model (1.2) $\eta = \beta\eta + \gamma\xi$ is appropriate when some of the observed variables are categorical. Linear models imply the ability to do linear transformations on variables such as η and ξ, but linear transformations do not make sense on ordinal numbers. So, technically, linear models are not ideal for ordered categorical variables, especially dependent variables. Actually, regression with categorical predictors is not a special problem, and neither is it in SEM. Coding schemes carry the appropriate information on the predictors. But dependent variables are, or can be, another story. A linear regression on a binary dependent variable can lead to predicted scores under the model that are out of range. Theoretically, this is inappropriate (see e.g., MacCallum et al., 2002). Having said this, it has been found in regression that ignoring the categorical nature of a binary dependent variable, when the proportion of "1"s is not too extreme, will lead to just about the same results (e.g., coefficients) as a more appropriate procedure. So, as noted earlier, in some practical circumstances you may be able to avoid using the polyserial/polychoric methodology.

Let us describe the methodology further before giving some other caveats. Imagine a situation in which you have two continuous variables, say, X=height and Y=weight. The Pearson correlation ρ_{XY} is the correct correlation to describe the extent to which taller people are heavier in the population. Now suppose that one of these variables is "cut" into categories. For example, instead of knowing the precise height of respondents, we may only have information that they are "short", "medium height", or "tall." We could code the new variable, say Z, as short=1, medium=2, and tall=3. The correlation of this categorical variable with weight, ρ_{ZY}, will now not equal the true ρ_{XY}. The *polyserial* correlation is an estimate of the correlation of two continuous variables, when one of them has been categorized; here, it estimates ρ_{XY} based on the categorical variable Z rather than the true variable X. If we now also cut weight into, say, five categories ("very light," "light," "normal weight," "heavy," "very heavy") to create a new variable A, the correlation between Z and A, ρ_{ZA}, is no longer equal to, nor necessarily a good approximation to, ρ_{XY}. The correlation between two categorical variables that estimates the correlation of the continuous counterparts to the categorical variables is called the *polychoric* correlation. If we could estimate polyserial and polychoric correlations as needed, and combine them with ordinary correlations (for the continuous-continuous combinations of variables) to generate a correlation matrix R_p, we would simply use this matrix and proceed as we explained.

To get anywhere we have to make assumptions. The primary assumption is a weakness of this methodology. We have to assume that each categorized variable is really based on an underlying continuous *normally distributed* variable that we have not been able to measure. In the case of height and weight, in many populations (e.g., adult males in the US), this is probably a reasonable assumption. Having data on *Z* and *A*, we could certainly imagine the existence of *X* and *Y*, both continuous and normal, and it would make sense to try to estimate how much *X* and *Y* correlate with each other and with other variables – even though we never see *X* and *Y*. The basic idea is the following. Imagine a two-way scatter plot of *X* and *Y*, which is the cigar-shaped set of points representing the relation between the variables. The correlation (or a regression) can be represented by a line in this scatter plot. Now make some cuts as described above on both axes. What is left is a small contingency table, here, a 3×5 table of frequencies. This is, in fact, our ordinal data. The process of estimating polychoric (both variables cut) or polyserial (only one variable cut) correlations is to reverse this process: start with the frequency table, and estimate the correlation. Technical procedures to accomplish this have been known a long time, e.g., Olsson (1979a) and Olsson, Drasgow, and Dorans (1982). In order to do this, we have to know the thresholds, i.e., the places where the continuous variables were cut. In the example above, if height was cut at 5'2" and 5'10", then "short" people are 5'2" and below, "tall" people are above 5'10", and "medium height" people are between 5'2" and 5'10". The proportion of cases in each category is determined by where the cuts are made. If we know the distribution is normal, and we know the proportion of short, medium and tall people, we can reverse the procedure and estimate the thresholds from the univariate marginal distributions. Then the correlation can be estimated in a similar way from the bivariate data.

The assumption of underlying normality may be questionable in some contexts. For example, suppose a variable in the analysis is "psychoticism." Is there a normal distribution of psychoticism in the population? Probably not. It is no doubt a highly skewed variable, with most people having no psychoticism, and a few people having a lot. There are tests for trivariate normality of binary indicators (Muthén & Hofacker, 1988), but general methods for evaluating the assumption of underlying normality for a set of arbitrarily categorized variables are not well developed. No tests are available in EQS. So if a model does not fit, it could be because the normality assumption is unreasonable.

A secondary assumption that can be hard to meet is that sample size should be large enough to enable reliable estimation of these correlations. In the simple example above, there were 15 cells of a 3×5 table of frequencies that require enough cases in each cell so that the corresponding population proportions can be estimated meaningfully. Cells with zero or near zero cases create estimation difficulties. This may occur if the cuts are badly made: if being over 6'5" describes "tall", people who are both tall and "very light", "light", or "normal" weighted may not exist in a sample. In models with a large number of variables, each with many response categories, there can be so many cells of the complete cross-tabulated contingency table that a number of cells will be empty unless the sample size is very large. Stated differently, a common situation in practice is that sample size is small relative to the number of response patterns (specific category scores across all categorical variables). Then the quality of estimates and tests can be easily disturbed. The problem is especially severe when the distributions of different variables are skewed in opposite directions.

An additional consequence is that the weight matrix needed to get χ^2 tests and standard errors is unstable unless sample size is large (e.g., Dolan, 1994). It is hard to give minimum numbers, but samples can be substantially smaller when using the ROBUST option as compared to the AGLS option. In general, we can say that the sample size problem here is even more difficult than it is with AGLS for continuous variables. As noted by Moustaki (2001, p. 477) regarding an AGLS polychoric method, "… that estimation method … can only be used for large samples, otherwise the results will be unstable."

Coenders, Satorra, and Saris (1997) found the polychoric approach to be not very robust to nonnormality. "Further research is needed on the effects of underlying non-normality on parameter estimates and fit measures" (Jöreskog & Moustaki, 2001, p. 385). This is true, but for the user the real issue is comparative, not absolute. Unfortunately, nothing seems to be known about the robustness of this methodology to violation of normality *relative to* the robustness associated with treating ordinal variables (with a certain number of categories, with certain distributions) as continuous at reasonable sample sizes. If at some sample size the polychoric/polyserial methodology is hard to downgrade due to a normality violation, while ordinary correlations are easily destroyed by legitimately recoding scores, then we certainly should use the "polys." If, on the other hand, ordinary correlations yield more robust

results than the ordinal methodology, in practice we should do the technically wrong thing and treat the ordinal variables as continuous. As you can see, we have ambivalent feelings about this methodology.

Example

The data file poon.ess, which is distributed with EQS, can be modeled by a two factor confirmatory factor analysis model with variables 1-4 being indicators of factor 1, and variables 5-8 being indicators of factor 2. The factors are correlated. This leads to an ordinary model file that we do not reproduce. However, variables V7 and V8 are categorical. As is shown in boldface, the key part of the specification is:

```
/SPECIFICATIONS
   DATA='c:\eqs6\examples\poon.ess';
   VARIABLES=8; CASES=200; GROUPS=1;
   METHODS=ML,ROBUST;
   CATEGORY=V7,V8;
   MATRIX=RAW;
   ANALYSIS=CORRELATION;
```

By and large, the output from an EQS run with categorical variables is the same as with any other run. We show some specialized additional output:

```
YOUR MODEL HAS SPECIFIED CATEGORICAL VARIABLES
    TOTAL NUMBER OF VARIABLES ARE         8
    NUMBER OF CONTINUOUS VARIABLES ARE    6
    NUMBER OF DISCRETE   VARIABLES ARE    2
    INFORMATION ON DISCRETE VARIABLES
       V7  WITH    3 CATEGORIES
       V8  WITH    3 CATEGORIES
```

The program informs you that it has counted the categories of each variable. The categorical information is used in the computations. Information on the polyserial correlations is presented first, for each of the variables in turn. The estimated thresholds are given first. The covariance estimates follow next, but since all variables are standardized, the covariances actually are polyserial correlations. Standard error estimates also are provided.

```
RESULTS OF POLYSERIAL PARTITION USING V 7 --  3 CATEGORIES
                                 THRESHOLDS
                          ESTIMATES       STD. ERR
                            -0.5112         0.0741
                             0.4318         0.0767
                                 ESTIMATES
          VARIABLE    COVARIANCE    STD. ERR
             V 1          0.4154      0.0520
             V 2          0.4442      0.0507
             V 3          0.4957      0.0493
             V 4          0.4281      0.0516
             V 5          0.6181      0.0449
             V 6          0.6358      0.0427

RESULTS OF POLYSERIAL PARTITION USING V 8 --  3 CATEGORIES
                                 THRESHOLDS
                          ESTIMATES       STD. ERR
                            -0.4566         0.0747
                             0.4986         0.0779
                                 ESTIMATES
          VARIABLE    COVARIANCE    STD. ERR
             V 1          0.3812      0.0532
             V 2          0.2654      0.0552
             V 3          0.3558      0.0539
             V 4          0.4390      0.0504
             V 5          0.6220      0.0430
             V 6          0.6728      0.0396
```

Information on polychoric correlations is presented next. Again, thresholds are computed (as there are several estimates, these are averaged) and then the polychoric correlation estimates are given.

```
RESULTS OF POLYCHORIC PARTITION
                    AVERAGE THRESHOLDS
                V 7  -0.5044   0.4327
                V 8  -0.4580   0.4854

POLYCHORIC CORRELATION MATRIX BETWEEN DISCRETE VARIABLES
                  V  7       V  8
     V  7       1.000
     V  8       0.583      1.000
```

Following this output, sample statistics for the original data file are given, with a warning `*** NOTE *** CATEGORICAL VARIABLES LISTED ABOVE ARE INDICATORS OF LATENT CONTINUOUS VARIABLES. THEIR UNI-VARIATE AND JOINT STATISTICS MAY NOT BE MEANINGFUL`. For example, the means and SDs of categorical variables are not meaningful in this context. Then, the constructed correlation matrix to be analyzed is given, followed by the usual normal theory correlation statistics given in the standard EQS format. While normal theory estimates are useable, the statistics in the normal theory output cannot be trusted. As discussed above, you should be using the Yuan-Bentler residual-based statistics, and/or the Satorra-Bentler test statistic to evaluate your model. Also, you will see `*** WARNING *** WITH CATEGORICAL DATA, NORMAL THEORY RESULTS WITHOUT CORRECTION SHOULD NOT BE TRUSTED`, reminding you that normal theory standard errors should be ignored, and the robust standard errors should be used to evaluate parameter significance. If you had used ME=AGLS, you would have found the key output under the title `GENERALIZED LEAST SQUARES SOLUTION (LEE, POON, AND BENTLER THEORY)`. The format for the output is exactly that of the continuous variable AGLS method in EQS.

Technical Details

We assume that there exists an underlying but partially unknown observed variable z^*, where z^* contains the continuous variables from the data z and the continuous counterparts to all the ordinal variables z_i. Each observed z_i contains category scores that are ordinal indicators obtained by cutting its corresponding z_i^*. Embedding the variables z^* into the larger set of independent and dependent variables v associated with a model, we now follow the usual Bentler-Weeks (1.2)-(1.3) approach. That is, all the variables of the model are arranged into the vector $v=(\eta',\xi')'$, and we select the z^* using the known [0,1] selection matrix G according to $z^* = Gv$. Then, the covariance matrix of the z^* is Σ_p, the "poly" version of Σ, see (1.4). However, in the polychoric/polyserial situation, the variances of the variables are generally not identified, so the covariances are standardized to the correlation matrix P_p, which is the population counterpart to the sample correlation matrix R_p. R_p, the matrix to be modeled, contains the ordinary sample correlations among continuous variables, sample polyserial correlations between these and categorical variables, and the sample polychoric correlations among the categorical variables.

There are actually several different approaches to modeling ordinal categorical data. Moustaki (2001) and Jöreskog and Moustaki (2001) state that two classes of methods are item response theory (IRT) approaches, and various variants of an underlying variable approach. Among the IRT approaches are models based on logistic regression (e.g., Bartholomew & Tzamourani, 1999; Moustaki, 2000; Moustaki & Knott, 2000) and probit models (Muraki & Carlson, 1995). For dichotomous data, a factor analysis procedure that has IRT and underlying variable origins (McDonald, 1967) is that of Fraser and McDonald (1988), extended statistically by Maydeu-Olivares (2001). An extremely general approach in the underlying variable tradition is the model of Küsters (1987), allowing censored and other variable types (see Arminger & Küsters, 1988; Browne & Arminger, 1995; Schepers & Arminger, 1992). The development in EQS follows the tradition of Muthén (1984), Jöreskog (1994), and Lee, Poon and Bentler (1990b, 1992), in which a sequential approach to estimation is used. Univariate statistics such as thresholds are estimated first, followed by estimation of bivariate statistics such as correlations, and completed by estimation of the SEM model using a GLS approach (1.5) with an appropriate weight matrix. This methodology is more practical than that based on full maximum likelihood (Lee, Poon, & Bentler, 1990a).

What differentiates EQS from other programs for ordinal variables is that we use the theory of Lee, Poon, and Bentler (1995), extended with several ROBUST methodologies. This uses a distinct approach to obtaining the initial estimates. A partition maximum likelihood approach (Lee, Poon, & Bentler, 1994; Poon & Lee, 1987) is used to obtain the estimates of the thresholds and the polyserial and polychoric correlations in the underlying correlation matrix. The joint asymptotic distribution of this estimator, akin to (5.58), is given by Lee, Poon, and Bentler (1995) in their Theorem 1, which is too technical to present in this *Manual*. Their theorem gives details on the matrix

$\tilde{V}$ that is the asymptotic covariance matrix whose inverse is the appropriate weight matrix in the GLS function (5.57) adapted to polychoric/polyserial correlations. Minimizing this yields optimal estimates and standard errors when sample size is large enough. As implemented in EQS, this AGLS methodology is that given in the 1995 publication. Several extensions, e.g., the missing data poly methodology of Lee and Leung (1992), sensitivity or influence analysis (Poon, Wang, & Lee, 1999), and nonlinear ordinal models (Lee & Zhu, 2000) are not yet implemented in EQS. EQS does implement a multiple group correlation structure methodology, see Chapter 7. However, further multiple group specializations (Lee, Poon, & Bentler, 1989; Poon, Lee, et. al., 1990, 1993) are not yet implemented.

The correlation estimates and weight matrices in EQS are similar to those of Muthén (1984; see also Muthén & Satorra, 1995b) and Jöreskog (1994), but they are not identical. There are thus several alternative approaches yielding asymptotically correct results, but, remarkably, it seems that there have been no independent evaluations of the several approaches. It has been reported that sample size requirements for adequate behavior of the test statistics of the Muthén and Jöreskog approaches are substantial (Dolan, 1994; Lee, Poon, & Bentler, 1992, 1995). No doubt the sample sizes required by the AGLS approach in EQS are also substantially larger than for the ROBUST alternatives.

EQS 6 presents for the first time several robust statistics available through the use of Lee et. al's Theorem 1. The ROBUST methodology permits obtaining correct statistics for our extension of Jennrich's (1970) GLS correlation estimators as described in (5.62)-(5.65). These are quite stable even in small samples, though they are not asymptotically optimal when used with categorical variables. However, using the correct asymptotic covariance matrix, we developed the Satorra-Bentler scaled χ^2 statistic and the associated covariance matrix of the estimates in this context. These arise naturally by appropriate redefinition of equations (5.48)-(5.52). Muthén (1993) has similarly adopted the Satorra-Bentler test for his approach. In addition, using the appropriate definitions in (5.53)-(5.56) using the various Lee-Poon-Bentler estimators, we also obtained the residual-based test statistics T_{RES}, $T_{\mathrm{YB(RES)}}$, and $T_{\mathrm{F(RES)}}$. Available in any program for the first time with ordinal methodology, we fully expect the various robust statistics to outperform the AGLS statistics at all but the largest sample sizes.

In correlation structure models, the diagonals of the reproduced model correlation matrix should equal 1.0. You can evaluate whether the constraints $diag(\hat{P}_p) = I$ have been imposed by checking the diagonal of the residual covariance matrix. The Lee-Poon-Bentler approach is to consider the additive unique variances in a latent variable model not as free parameters, but rather as functions of the remaining parameters in the model. This method is convenient computationally, and we use it in EQS but most likely will change it. It has the difficulty that one cannot impose equality constraints on the unique variances, and these variances could be estimated as negative. The alternative, treating all parameters as free and imposing nonlinear constraints on the parameters so that the resulting matrix is in fact a correlation matrix, may be adopted.

Finally, all ordinal approaches to SEM are based on the assumption that the ordinal categories are used the same way by all respondents. If different respondents interpret category labels (such as "short" or "medium height") differently, the ordinal categories may not be interpersonally comparable. Stated differently, different individuals may have different thresholds; this is not allowed in the current model. Even more extremely, the ordinal data may be ipsative, e.g., represent within-individual ranks (Chan & Bentler, 1998). Methods that take interpersonal incomparability into account are not currently available in any well-known SEM program.

Case-Weighted Means and Covariances

Although "weights" of various sorts are mentioned throughout this *Manual*, up to now we have been talking about weights that need to be used in order to properly take into account the sampling distributions of various sample statistics, such as the sampling distributions of means and covariances, or the sampling distributions of test statistics. In this section we talk about quite a different type of weight – a weight associated with each *case* or observation in the data file. Such weights are needed if the cases are not equally representative of the population from which they were drawn. We use such weights to obtain "case-weighted" means and covariances for modeling. As typically applied, these weights are also called design weights, sample weights, or sampling weights, but case weighting can

be applied regardless of the origin of the weights. For example, the weights can be used to control heterogeneous variances across individuals. Meijer and Mooijaart (1996) show how heteroscedastic errors invalidate normal theory χ^2 tests in a factor analysis context.

In virtually all mean and covariance structure modeling, each case (individual, subject, observation) in the data file contributes equally to the sample statistics being modeled. A sample mean is a simple average, and a covariance is a simple average deviation cross-product. Implicitly, such computations use equal weights for all cases in the sample. This standard procedure can be justified under simple random sampling from the population, where each case has an equal probability of appearing in the sample. However, there are situations where some cases should be weighted more heavily than others in order to obtain proper unstructured estimates of population means and covariances. A typical example is when different groups of subjects have different probabilities of appearing in a sample. For example, a proportionately small ethnic group (e.g., American Indians) might be oversampled, i.e., proportionately more such cases appear in the sample than in the population. This may be done in order to obtain stable within-group information or to reliably evaluate possible between-group differences. Weighting each case in such a sample equally would produce population estimates that are closer to those of the small group than the entire population from which all cases are drawn. However, an unbiased estimate of population means and covariances can be obtained by judicious use of weights. In another situation, various groups (e.g., by age, gender, socioeconomic status, etc.) may have appropriate representation in the sampling plan under which data was collected, but chance may have yielded unequal representation in the sample. To illustrate, the sample may contain relatively more females than males, as compared to their representation in the population. In the latter situation, a priori weights may require adjustment through some type of poststratification analysis.

There is almost no literature on the use of case weighted means and covariances in structural modeling. Winship and Radbill (1994) discuss issues in the use of sampling weights in the context of regression analysis, while Kaplan and Ferguson (1999) provide the best introduction to sampling weights for SEM. Stapleton (2002) discusses the use of case weights in multilevel models. However you obtain the necessary weights, EQS will produce case-weighted means and covariances that can be used in structural modeling.

Assumptions and Use

In a simple random sample, each individual in the population has an equal probability of inclusion. This is the typical justification for computing means as unweighted or equally weighted averages, and covariances as unweighted (equally weighted) averages of deviation cross-products. However, large scale studies carried out on probability samples often generate data in which individuals have unequal inclusion probabilities, i.e., the individuals in the sample did not have an equal chance being in the sample. In such a situation, we can use sample or sampling weights to enable the first and second moments of the data to approximate their corresponding parameters in the population. A typical *raw* weight is $w_t' = 1/p_t$, where p_t is the probability of individual t appearing in the sample. Raw weights often appear in public use data tapes. Such raw weights sum up to the population total (which is assumed, in this literature, to represent a finite population rather than an infinitely large population). However, the use of raw weights often will lead to incorrect standard error estimates. Hence, typically, you should use *normalized* weights w_{tN} in which the sum of the weights is sample size. Since the sum of normalized weights is sample size, the average normalized weight is 1.0. If the sum of the raw weights in your file happens to equal sample size N, then the normalization does nothing, and raw and normalized weights are identical. Recently it has been suggested (Stapleton, 2002) that *effective* weights w_{tE} might be even better than normalized weights at yielding correct standard errors. Effective weights sum to a quantity N_E that can be considered as the "effective" sample size. The actual sample size N is greater than or equal to N_E, and the ratio N/N_E, which equals 1.0 when the case weights are equal, is referred to as the design effect. Note that if every individual case's weight is identical, each case would be weighted 1.0 with the consequence that the weighting would have no effect; then, the sample statistics will be the usual estimators.

Weighting is easily accomplished if the weight variable is in your data file. You simply specify the weight variable as WT = V5 (say) when you want to use raw weights w_t' , WTN = V5 if you want the scores given in V5 to be normalized to w_{tN} by EQS prior to computations, and WTE = V5 if you want the scores in V5 to be rescaled to effective weights w_{tE} prior to computations. Either normalized or effective weights typically should be used. In view of the meaning of the weight variable, the scores (on V5, say) must be nonnegative numbers. Any case whose

weight is missing or nonpositive is ignored, and hence it has no effect on the calculations. If MISSING=ML is given, missing values are imputed for these cases, but not for the weight variable itself.

The raw weight scores w_t' for each individual or case t are presumed to be in your data file, and no preprocessing is done on them. The normalized weights w_{tN} are computed as

$$w_{tN} = N w_t' / \sum_{t=1}^{N} w_t' \tag{5.66a}$$

while the effective weights w_{tE} are computed as

$$w_{tE} = w_t' \left(\sum_{t=1}^{N} w_t' / \sum_{t=1}^{N} w_t'^2 \right). \tag{5.66b}$$

If the sum of the raw weights in your file happens to equal sample size N, then the normalization does nothing, and raw and normalized weights are identical. Now let $w_t = w_t'$, $w_t = w_{tN}$, or $w_t = w_{tE}$, that is, it will be either the raw, normalized, or effective weight. EQS follows BMDP and computes the weighted mean of variable z_i as

$$\bar{z}_i = \sum_{t=1}^{N} w_t z_{ti} / \sum_{t=1}^{N} w_t \ . \tag{5.67}$$

Similarly, the weighted covariance between variables z_i and z_j is computed as

$$\mathrm{cov}(z_i, z_j) = \sum_{t=1}^{N} w_t (z_{ti} - \bar{z}_i)(z_{tj} - \bar{z}_j) / N \ , \tag{5.68}$$

where the sample means are the weighted means given by (5.67). In computing (5.67) and (5.68), one of the three weights (raw, normalized, or effective) is used, depending on the user's specification. The covariance formula (5.68), also the BMDP formula, is identical to Kaplan and Ferguson's equation (13) when normalized weights are used, but with raw weights (5.68) differs from their (13), which has $\Sigma_t w_t$ in the denominator. As can be seen from these formulas, if each $w_t = 1.0$, the weighted mean is simply the usual sample mean, the covariance is simply the biased sample covariance. Otherwise, cases with larger w_t have more influence on means and covariances. In the context of heteroscedasticity, if the w_t represent inverse measures of case variance, cases with smallest variances or greatest precision contribute the most to the weighted means and covariances. Whatever the origin of the weights, the sample statistics defined by (5.67) and (5.68) become the input means and covariances to be modeled. Some of the issues involved in modeling with weighted means and covariances are discussed in the previously cited references.

You can use case weighting with covariance structures (with or without structured means), or with correlation structures. As far as we know, case-weighted correlation structures are new to the field. Also, case weighting has historically been used with normal theory statistics. In EQS for the first time you also can use case weighting with elliptical, heterogeneous kurtosis, as well as with robust versions of various of these methods, and distribution free covariance structure methods. When ANALYSIS=ZSCORE with a weight variable, the means and standard deviations are based on the above formulae, and weights are applied to this data set to compute kurtosis etc. as needed.

Technical Details

Although there are several ways to develop a theory for case-weighted structural modeling, it will be done here using the matrix normal distribution (e.g., Arnold, 1981). Following Weng and Bentler (1987, 1997), we consider an N by p data matrix Z to be described by an expected value μ_z, a population covariance matrix Σ for the variables, and a covariance matrix Σ_{row} among the rows of the matrix Z, i.e., among cases or observations. In the usual normal distribution, observations are uncorrelated and one takes $\Sigma_{row} = I$. The log likelihood is given as

$$L = -2^{-1}\{ Np \ln(2\pi) + p \ln | \Sigma_{row} | + N \ln | \Sigma | + tr \Sigma_{row}^{-1} (Z - \mu_z) \Sigma^{-1} (Z - \mu_z)' \} \ . \tag{5.69}$$

In order to apply this to case-weighted observations, we take Σ_{row}^{-1} to be the known diagonal matrix D_w of raw or normalized case weights w_t, and assume that the mean matrix has a structure $\mu_z = 1\mu'$, where 1 is the unit vector. Then the unknown parameters of a saturated model are μ and Σ. Taking derivatives, setting these to zero, and solving, gives the estimated mean vector as

$$\hat{\mu} = Z'D_w 1/1'D_w 1 \tag{5.70}$$

and the estimated covariance matrix as

$$\hat{\Sigma} = (Z - 1\hat{\mu}')'D_w(Z - 1\hat{\mu}')/N . \tag{5.71}$$

The elements of (5.70) are the means given in (5.67). The elements of (5.71) are the covariances given in (5.68). These are then used as sample means and covariances in the usual normal theory likelihood function (see e.g., 5.13 for covariances) to evaluate further structures on (5.70) and (5.71). Using an argument similar to Weng and Bentler (1987), it can be shown that this produces the correct matrix normal likelihood ratio test statistic and information matrix based standard errors.

So that information about nonnormality can be used in case-weighted modeling, to deal with nonnormality in this context we redefine the unweighted skew and kurtosis coefficients given by (4.1)-(4.8). We defined weighted counterparts to these coefficients in a manner consistent with SAS for variance divisor N. The case weighted counterpart to (4.1) is given by

$$g_{1(i)} = N^{1/2}[\Sigma_1^N w_t^{3/2}(z_{it} - \bar{z}_i)^3]/[\Sigma_1^N w_t(z_{it} - \bar{z}_i)^2]^{3/2} , \tag{5.72}$$

where the weighted means (5.67) are used. Similarly, the kurtosis coefficient corresponding to (4.2) is given by

$$g_{2(i)} = N[\Sigma_1^N w_t^2(z_{it} - \bar{z}_i)^4]/[\Sigma_1^N w_t(z_{it} - \bar{z}_i)^2]^2 - 3 . \tag{5.73}$$

For the vector of variables *z*, the weighted Mardia-like coefficient of kurtosis is given by

$$g_{2,p} = N^{-1}\Sigma_1^N[w_t(z_t - \bar{z})'\hat{\Sigma}^{-1}(z_t - \bar{z})]^2 - p(p+2) , \tag{5.74}$$

using the weighted means and covariances. The weighted normalized Mardia coefficient is again given by (4.4). Weighted counterparts to missing data kurtosis (4.5)-(4.6), and elliptical kurtosis coefficients (4.7)-(4.8), are defined similarly and do not need to be presented in detail. These definitions allow elliptical and heterogeneous kurtosis theory to be used in the usual way. Similarly, the fourth-order moments needed to define AGLS methods and various robust methods in EQS are obtained using elements as in (5.7) with case weights w_t^2.

Essentially nothing has been done in evaluating the matrix normal rationale for case weighting. That is, simulation studies do not exist on the quality of structural modeling statistics that result from this approach. In the simulation study of Kaplan and Ferguson (1999), data was generated using a mixture of normals with different covariance matrices. The robustness of the basic matrix normal approach, and the performance of nonnormal case weighting methods, also require study.

Case-Robust Methods

In the previous section, we assumed that the weights used in case-weighting are known, and are to be applied to produce more accurate estimates of population means and covariances, and in turn, the SEM model parameters. In this section, in contrast, we assume that case weights are unknown, but we may wonder whether some cases or individuals should be downweighted when computing our sample statistics. As we noted in Chapter 1, sample means and covariances are easily influenced by a few outlying observations, which can lead to bad measures of fit even when the majority of the data may be well explained by a structural model. This can be illustrated with the file manul7.ess, which contains raw scores of 50 cases on 6 variables. The data were artificially generated to be normal, except that scores for one case, subject number 50, were modified so that it is an outlier. EQS was run with a 2-factor model, with the 1st 3 variables as indicators for F1, and the last 3 variables as indicators of F2. The factors were set to correlate. This model in fact represents how the data was generated. Using ML, results include

```
                    MULTIVARIATE KURTOSIS
                    ---------------------
MARDIA'S COEFFICIENT (G2,P) = 18.9017    NORMALIZED ESTIMATE = 6.8206 ! THIS IS HIGH
```

```
  CASE NUMBERS WITH LARGEST CONTRIBUTION TO NORMALIZED MULTIVARIATE KURTOSIS:
  ----------------------------------------------------------------------------
  CASE NUMBER        23           27           39           42           50
  ESTIMATE        48.3404      19.6590      31.0902      41.3800     546.8184
          !CASE #50 HAS A VERY HIGH CONTRIBUTION TO KURTOSIS -- PROBABLY AN OUTLIER

 CHI-SQUARE =       16.692 BASED ON     8 DEGREES OF FREEDOM !MODEL SHOULD BE REJECTED
 PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS     0.03348
```

Clearly the case contributions to normalized multivariate kurtosis are very good at identifying the outlying case. In order to determine whether case 50 created a problem for the analysis, the job is resubmitted with case 50 deleted. This could be done using the DEL=50; command in the /SPEC section, or using the file manul7a.ess that has eliminated this case. Using the same setup otherwise, we now find

```
                          MULTIVARIATE KURTOSIS
                          ---------------------
  MARDIA'S COEFFICIENT (G2,P) = -2.3786    NORMALIZED ESTIMATE = -0.8497 ! SMALL NOW

  CASE NUMBERS WITH LARGEST CONTRIBUTION TO NORMALIZED MULTIVARIATE KURTOSIS:
  ----------------------------------------------------------------------------
  CASE NUMBER        12           23           27           39           42
  ESTIMATE        37.2676      46.7041      24.9805      62.8271      46.2463
! CASE CONTRIBUTIONS ARE NOW SIMILAR IN SIZE.  NO INDICATION OF ANOTHER OUTLIER.

 CHI-SQUARE =        2.701 BASED ON     8 DEGREES OF FREEDOM   ! BIG IMPROVEMENT
 PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS     0.95169
```

Clearly, a single case can destroy an otherwise excellent model. A theoretical understanding of how this happens was provided by Poon and Poon (2002) for means and covariances, and by Yuan and Bentler (2001a) for structural models. As was pointed out earlier, the influence function associated with sample covariances is quadratic, so that a few influential cases or outliers can lead to inappropriate solutions for virtually all standard statistical methods that rely on sample covariances. This example might suggest that the procedure shown here – outlier identification followed by a standard normal theory run – might serve well in general. This may be true in practice, but in fact the statistical theory to justify such a pragmatic approach does not exist. In fact, there are situations when sets of outliers may be so located in multivariate space that it is hard for any method of outlier detection to unambiguously identify such cases. Case-robust methods make this unnecessary, since a case will be downweighted only when this will help generate more stable mean and covariance matrix estimates.

Assumptions and Use

Case-robust methods are distinct from the robust statistics methodology given in the section **Corrections to Basic Statistics** (above). In the ME=xx,ROBUST approach, each case is weighted equally and hence the ordinary sample means and covariances are accepted as statistics to be modeled; in turn, various types of adjustments, such as the Satorra-Bentler scaled statistic, are used to correct for nonnormality. You use case-robust methodology, in contrast, when you are concerned that some observations in your data file, such as #50 above, might inappropriately impact the sample means and covariances, and hence you do not want to model the ordinary means and covariances. A unique feature of EQS, case-robust methods find those weights that can be applied to individual cases so as to yield robust estimates of means and covariances. Case-robust weighting involves finding a proper weight, usually ranging from 0-1, that can be given to each individual in the data file in calculating a robust mean vector and covariance matrix. The advantages of case-robust weighting are that when the data are normal, virtually nothing is changed because each case gets a weight of about 1.0, while when there are outliers or influential cases, the case-robust means and covariances are no longer highly influenced by these few cases. Of special importance for SEM, these robustness properties of the means and covariances are carried over to the parameter estimator, as well as the standard errors and test statistics. It has also been shown that case-robust weighting can aid in cross-validation of results (Yuan, Marshall, & Weston, 2002).

While Jöreskog (1977) and Browne (1982) had suggested using a case-robust covariance matrix in standard modeling software, Huba and Harlow (1987) provided the first substantial study showing that this methodology works in practice. Yamaguchi and Watanabe (1993) showed that a covariance matrix based on the multivariate *t*-distribution yielded more efficient estimates than those from the ordinary covariance matrix. However, these authors provided no theory to verify that the resulting test statistics, standard errors, and so on were appropriate.

Ke-Hai Yuan and his colleagues (Yuan & Bentler, 1998b,c, 2000c; Yuan, Bentler, & Chan, 2000; Yuan, Chan, & Bentler, 2000) provided several theoretical approaches to finding optimal case weights and for assuring that the resulting statistics behave properly in a SEM context. The range of methods developed by Yuan exceed those currently available in EQS.

At this time, only the weighting methodology developed by Yuan and Bentler (1998c) is implemented in EQS. This computes case-robust means and covariances using Campbell's (1980) recommended weights, and then uses a standard estimation method -- for now you should use ML -- to estimate parameters. Subsequently, the Yuan-Bentler statistics are computed for these estimates. You should use the test called `YUAN-BENTLER SCALED CHI-SQUARE (CASE ROBUST WEIGHTING)` and the parameter tests given as (`YUAN-BENTLER ROBUST STATISTICS IN PARENTHESES`).

You specify this case-robust method in the SPECIFICATION section by specifying CROBUST=#,#, where # is a number. In the Windows program, this is specified in the Miscellaneous Option box, and you will get `CROBUST = 2.00,1.25` if you choose the default weighting. These values are recommended unless you have studied the literature and have a rationale for other values. If you use this method for the manul7.ess data as indicated above, you will get the statement

```
ROBUST ESTIMATES OF MEANS AND COVARIANCE MATRIX
  BASED ON CASE ROBUST METHOD
```

followed by the robust sample statistics. Subsequently, the program informs you

```
CASE NUMBERS WITH SMALLEST CASE ROBUST WEIGHTS
   ----------------------------------------
   CASE NUMBER        50
   WEIGHT             .0000
   CASE ROBUST WEIGHTS FOR ALL OTHER CASES ARE 1.0.
```

As you can see, the methodology downweights the outlying case #50 so much that it has zero contribution to the means and covariances, and, in this special example, all other cases are weighted 1.0. That is, the Yuan-Bentler procedure acts very much like the method of removing the outlier. This is extremely unusual, since more typically, a wide range of case weights will be obtained as various cases may be more-or-less problematic.

In addition to the statistics given by Yuan and Bentler (1998c), you will get a number of new statistics for case-robust methodology in EQS. In particular, the three residual-based test statistics T_{RES}, $T_{\mathrm{YB(RES)}}$, and $T_{\mathrm{F(RES)}}$, see (5.54) – (5.56), from Yuan and Bentler (1998a) and Bentler and Yuan (1999b) were generalized here to apply to case-robust SEM. A methodology for dealing with missing data (e.g., Cheng & Victoria-Feser, 2002) is not implemented in the initial release of EQS 6.

Technical Details

The case-robust means and covariances used in EQS are very similar to the matrix normal case-weighted means and covariances as given earlier in (5.67) and (5.68). In particular, for the means, as before, $\hat{\mu}$ has elements

$$\hat{\mu}_i = \sum_{t=1}^{N} w_t z_{ti} \Big/ \sum_{t=1}^{N} w_t \tag{5.75}$$

where the weights are not given a priori but are determined iteratively as noted below. In contrast to the fixed weight situation, the weights w_t are squared when used in the covariance matrix $\hat{\Sigma}$ with elements

$$\hat{\sigma}_{ij} = \sum_{t=1}^{N} w_t^2 (z_{ti} - \hat{\mu}_i)(z_{tj} - \hat{\mu}_j) \Big/ \left(\sum_{t=1}^{N} w_t^2 - 1 \right), \tag{5.76}$$

where the divisor is based on the sum of the squared weights, not N. See Yuan-Bentler, eqs. (21)-(22). Clearly if the w_t all equal 1.0, (5.76) is just the unbiased sample covariance matrix with denominator (N – 1). The weights w_t depend on the distance that a case is from the robust mean of the distribution in the metric of the inverse robust covariance matrix, that is,

$$d_t = \{(z_t - \hat{\mu})\hat{\Sigma}^{-1}(z_t - \hat{\mu})\}^{\frac{1}{2}}. \tag{5.77}$$

These distances are translated into weights using Campbell's (1980) suggestion that

$$w_t = \begin{cases} 1 & \text{if } d_t \le d_0 \\ \dfrac{d_0}{d_t}\exp\{-.5(d_t - d_0)^2 / b_2^2\} & \text{if } d_t > d_0 \end{cases} \tag{5.78}$$

where $d_0 = \sqrt{p} + b_1 / \sqrt{2}$ and b_1 and b_2 are tuning constants. Notice in particular that if d_t is small, the case receives a weight of 1, and otherwise the weight is smaller than 1. On the basis of empirical experience, Campbell gave three recommendations for choices of tuning constants: (a) $b_1 = \infty$, corresponding to the usual sample covariance; (b) $b_1 = 2$ and $b_2 = \infty$, corresponding to a Huber-type M-estimator (Huber, 1981); and (c) $b_1 = 2$ and $b_2 = 1.25$, corresponding to a Hampel-type redescending M-estimator (Hampel, 1974). We propose that you use case (c) as your default. If you want to experiment, you could use $b_1 = 2$ and vary b_2 from 1.25 to a large number, aiming to control the number of downweighted cases to about 5%. Internally in EQS, for a given start value of (5.75) and (5.76), distances (5.77) provide weights (5.78) that can be used to update (5.75)-(5.76). This procedure is iterated until convergence.

Since we recommend that you use ML on these robust means and covariance matrices, the usual statistics for that method are computed and printed, that is, ML χ^2 test and standard errors and parameter tests are given. However, these should be ignored, since the actual distribution of (5.75) and (5.76) is not accurately described by normal theory ML. Yuan and Bentler (1998c) obtained the asymptotic distributions of (5.75) and (5.76) for use in computing correct statistics. The asymptotic covariances are given in their eq. (20), with the relevant G_1 and G_2 given on p. 377 of their article. The standard errors computed in EQS then come from Yuan-Bentler's eq. (8), with appropriate consistent estimates replacing the population matrices. Equivalently, it can be shown that section 5.1 of Yuan-Bentler (1998b) also provides the asymptotic covariances for this situation. The detailed formulae are too involved to present here, but are computed automatically in EQS. Thus for standard errors and parameter tests you should use the statistics labeled `YUAN-BENTLER ROBUST STATISTICS`.

The Yuan-Bentler scaled χ^2 test (their eq. 25), given as `YUAN-BENTLER SCALED CHI-SQUARE (CASE ROBUST WEIGHTING)`, is similar in form to the Satorra-Bentler test given in (5.51) for an appropriately defined scaling correction. Like the Satorra-Bentler, this test is precisely χ^2 distributed only for elliptical samples with models that are invariant under a constant scaling factor, but you should expect it to perform well under a wide range of conditions. New here for the first time, we also developed for EQS the residual based tests T_{RES}, $T_{\mathrm{YB(RES)}}$, and $T_{\mathrm{F(RES)}}$ given for ordinary covariances in (5.53)-(5.56) to be applicable to the case-robust situation. As these tests are new in this context, nothing is known about their performance in practice. The test T_{RES} can be shown to be an example of the test T_n given by Yuan-Bentler (1998b), eqs. (25)-(26), and proven to be asymptotically χ^2 distributed in their Theorem 5.1. Hence it should perform well in large samples. $T_{\mathrm{YB(RES)}}$ and $T_{\mathrm{F(RES)}}$ are not so well rationalized analytically in this context, and hence empirical work will have to determine if they will perform as well as or better than the scaled test in practice. Since they retain their asymptotic equivalence to T_{RES}, the real question is whether their corrections for small sample performance still work in practice.

In the near future, EQS will be extended to also allow LS and GLS estimation for case-robust weighting. Here we explain how this leads to correct statistics, since these methods were not developed in their original publication (1998c). The Yuan-Bentler scaled χ^2 test (their eq. 25) is simply modified in the usual way, and their eq. (8) is modified by substituting I or S for Σ, and substituting a sandwich estimator of their Γ to obtain the correct standard errors.

Although it is not currently available in EQS, a minimum χ^2 or AGLS estimator for case-robust methodology with case-robust weighting will be implemented in future versions of EQS. This will be based on a version of (5.1) as given by Yuan and Bentler (1998b), eq. (18), using as the weight matrix the inverse of the asymptotic covariance matrix as described above. The standard errors will come from eq. (21) and the χ^2 test is given by Theorem 5.2 (see Yuan-Bentler, 1998b). And, since AGLS estimators often are associated with too large test statistics and too small standard errors, corrections of the sort given in (5.2) and (5.11) will be available.

Factor Score Estimation

Factor scores are the unknown scores of the subjects on the latent factors, the Fs, Es, and Ds in any EQS model. The case scores on the V variables are in the data file. The real interest usually focuses primarily on the Fs – the circles or ovals in a path diagram -- which are presumed to generate the correlations among the observed variables. Unfortunately, true factor scores, i.e., the Fs for individual cases, are always unknown. The best that can be done is to estimate or predict them. There will always be an error in doing this, due to a basic indeterminacy in the model (see e.g., Krijnen, Dijkstra, & Gill, 1998). Thus although we may call them "factor scores", you should realize that we only obtain "factor score estimates." (Also, be aware that this is a somewhat odd notation, because we are predicting or estimating random variables, and not population parameters – an unusual task.)

There are many methods of factor score estimation. These have been developed primarily for exploratory factor analysis. In EQS, we provide only two basic types of factor scores, but these are available for almost any SEM model with latent variables. You use the SAVE section of a model file to specify factor score estimates. There are two possible choices, the GLS or REGRESSION estimators:

FSCORE=GLS or
FSCORE=REG.

The GLS factor score estimator (Bentler & Yuan, 1997) is a modified Bartlett (1937) estimator. Unlike the Bartlett estimator, this optimal unbiased equivariant estimator does not break down when a unique variance is zero. Suppose there are p variables and q factors. Let $\hat{P}$ represent the p x p model-reproduced correlation matrix from factor analysis or any structural model, and let $\hat{\Lambda}$ represent the estimated p x q factor loading matrix. This matrix is either explicit from rotated or unrotated solutions, or it computed from implicit formulas in the structural model. With z_t being the p x 1 vector of *z*-scores for the t^{th} individual, the estimated GLS factor scores for individual *t* are given by

$$(\hat{\Lambda}'\hat{P}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{P}^{-1}z_t. \qquad (5.79)$$

This estimator goes back a long way. Heermann (1963) and Hakstian, Zidek, and McDonald (1977) studied it as a "univocal" estimator of orthogonal factors. A univocal estimator is correlated only with its own factors. Schönemann and Steiger (1979) gave (5.79) as a definitional statement for components, not factors. Bentler and Yuan developed various optimality properties of this estimator under general as well as specialized conditions. For example, the estimator makes a certain residual covariance matrix as small as possible. Also, the GLS estimator becomes the Bartlett estimator when no unique variance is zero. If there are no zero unique variances, (5.79) is equal to $(\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Psi}^{-1}z_t$, which is the classical form of the Bartlett estimator when $\hat{\Psi}$ is the covariance matrix of the unique scores (of the E variables, in EQS).

The GLS estimator is conditionally unbiased, that is $E(\hat{\xi} \mid \xi) = \xi$ where ξ is the vector of true factor scores (Fs in EQS). This may or may not be an important property in some situations. Hence the more standard Thurstone (1935) and Thomson (1936) regression factor score estimator is also available. It is given by

$$\hat{\Phi}\hat{\Lambda}'\hat{P}^{-1}z_t, \qquad (5.80)$$

where $\hat{\Phi}$ is the estimated covariance or correlation matrix of the factors.

As noted in the *EQS 6 User's Guide*, if these factor scores are computed in the process of an exploratory factor analysis done with the Windows program, they can be appended to the data file. Otherwise, they can be saved in an external file. This feature clarifies that estimated factor scores cannot be the true factor scores, since Fs can never be appended to a data file. They are a figment of our scientific imagination.

R^2 Coefficients for Equations

In single-equation regression in which a dependent variable is predicted from a set of variables, the predictability of the dependent variable is typically summarized with the squared multiple correlation R^2. That is, if we have, as in (1.1), $y = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + e$, then we can define the population squared multiple correlation as

$$\rho^2 = corr^2 (y, \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p) , \tag{5.81}$$

namely the simple correlation, squared, of y with the optimal linear combination of its predictors. This gives the proportion of variance in the dependent variable explained by its predictors. Since e is uncorrelated with the predictors x_i, the explained variance and the unexplained variance -- the variance of the residual e – are additive, and $1 - \rho^2$ is the proportion of unexplained variance. The corresponding sample coefficient requires first estimating the parameters β and then computing

$$R^2 = corr^2 (y, \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_p x_p) . \tag{5.82}$$

In practice, of course, simplifying methods are used to actually compute this coefficient. In structural equation models of the multivariate regression and multiple-equation types, especially, recursive models, each equation in turn can yield a sample coefficient of the form (5.82). Such coefficients are routinely computed in EQS and printed along with the standardized solution.

Teel, Bearden and Sharma (1986) pointed out that measurement of explained variance is not at all straightforward when a model contains feedback loops among variables, i.e., it is a nonrecursive model, or if it has disturbances that are correlated across equations. In this situation, standard definitions used in various programs can break down. For example, Hayduk (1996, pp. xx-xxi) noted "I used to think that no acceptable model could contain a concept having a negative R^2, but I was wrong!" Actually, values outside the 0-1 range make no sense. Such values can occur using standard approaches because the residual variable may not be uncorrelated with the predictor variables in the equation, and hence the usual formula of total variance being partitioned into explained and residual variance is questionable. Some approaches to solving this problem were offered by Teel et al. and Hayduck (1996). EQS has adopted the Bentler and Raykov (2000) method, which builds on a basic Teel et al. definition. The Bentler-Raykov approach is to consider a single structural equation from the set of Bentler-Weeks equations (1.2), or their FASEM equivalent, as $\eta_i = \beta_1 \eta_1 + ... + \beta_p \eta_p + \gamma_1 \xi_1 + ... + \gamma_q \xi_q + \zeta$, where the residual variable, one of the independent ξ variables, is denoted ζ for clarity. In EQS, ζ is either a D or E variable. The Bentler-Raykov population squared multiple correlation for the i^{th} variable is defined as

$$\rho_i^2 = corr^2 (\eta_i, \beta_1 \eta_1 + ... + \beta_p \eta_p + \gamma_1 \xi_1 + ... + \gamma_q \xi_q) . \tag{5.83}$$

In practice, the β and γ parameters are estimated first, and the sample values

$$R_i^2 = corr^2 (\eta_i, \hat{\beta}_1 \eta_1 + ... + \hat{\beta}_p \eta_p + \hat{\gamma}_1 \xi_1 + ... + \hat{\gamma}_q \xi_q) \tag{5.84}$$

are computed and printed in the standardized solution. These values must be in the 0-1 range. When the model is recursive, this simplifies to the usual definition.

Difference Test for Satorra-Bentler Statistics

χ^2 difference (D) tests involving the Satorra-Bentler scaled statistics are not computed automatically by EQS and must be hand computed. Normally, the D test is simply given as the difference between two nested test statistics, and the result is evaluated against the difference in degrees of freedom of the two tests. So, for example, consider nested models $\mathcal{M}_0$ and $\mathcal{M}_1$, where $\mathcal{M}_0$ is the restricted model and $\mathcal{M}_1$ is the more general model, with ML estimation. $D = T_{ML(0)} - T_{ML(1)}$ is the algebraic difference between the values of the ML test statistics for restricted and more general ML model runs. Normally, this can be evaluated as a χ^2 test with $d = d_0 - d_1$ degrees of freedom. However,

the simple difference between the corresponding Satorra-Bentler statistics $\bar{D} \neq \bar{T}_{ML(0)} - \bar{T}_{ML(1)}$ is not χ^2 distributed. Satorra (2000) shows how to compute such a test correctly in a general form, but his method cannot be adapted from the output of normal EQS runs. We summarize here the procedure developed by Satorra and Bentler (2001, p. 511) to correctly compute difference tests based on the information available from two separate runs and their ordinary (say, ML) and robust (SB) test statistics.

1. Available from model $\mathcal{M}_0$: T_0, d_0, and the SB scaled statistic $\bar{T}_0$. Calculate the scaling coefficient (see 5.50) k_0 as $k_0 = T_0 / \bar{T}_0$, or take it from the EQS output.
2. Available from model $\mathcal{M}_1$: T_1, d_1, and the SB scaled statistic $\bar{T}_1$. Calculate the scaling coefficient k_1 as $k_1 = T_1 / \bar{T}_1$, or take it from the EQS output.
3. Compute from the model comparison:
   (a) The ordinary difference test on the original unscaled tests as $D = T_0 - T_1$.
   (b) With $d = d_0 - d_1$, compute the SB scaling coefficient for a difference test as

$$k = (d_0 k_0 - d_1 k_1) / d \ . \tag{5.85}$$

4. You are now ready to compute the SB scaled difference test as

$$\bar{D} = D / k \ . \tag{5.86}$$

Note that (5.86) has the usual form of a Satorra-Bentler corrected test, see e.g., (5.51). The trick here is to obtain the correct scaling divisor (5.85) as a weighted combination of the separate divisors for the individual test statistics.

A word of caution: while the population counterpart to k in (5.85) is necessarily nonnegative, it is possible that in some extreme situations (5.85) may be negative. Then (5.86) cannot be used. A probable cause of this could be that $\mathcal{M}_0$ is highly misspecified. If you nonetheless need a difference test, you can compute one from the residual-based test statistics T_{RES} and $T_{YB(RES)}$. Such a test can be computed in the usual way without any corrections.

This completes the basic documentation of statistical methods for EQS 6. Some technical aspects of specialized methods are given in subsequent chapters.

6. LAGRANGE MULTIPLIER AND WALD TESTS

In addition to the general goodness of fit test that tests the adequacy of a given model, tests on the statistical necessity of sets of parameters that might be added to a model, or deleted from the model, are also frequently needed in structural equation and covariance structure modeling. Model comparisons of such *nested* models provide the statistical backbone for deciding whether a simpler model may be acceptable in explaining a phenomenon, or whether a more complex model may be required (Steiger, Shapiro, & Browne, 1985). Generally, simpler models are desired for reasons of parsimony (for a discussion of issues, see Breivik & Olsson, 2001; Cheung & Rensvold, 2001; James, Mulaik, & Brett, 1982; Marsh & Balla, 1994; Mulaik, James, Van Alstine, Bennett, Lind, & Stillwell, 1989; Raykov & Marcoulides, 1999; Williams & Holahan, 1994) as well as for precision of estimates and stability of the solution (Bentler & Mooijaart, 1989; Kano, Bentler, & Mooijaart, 1993). Methods for comparing *nonnested* models are only now being developed (e.g., Lee & Song, 2002; Song & Lee, 2001, 2002a).

The chi-square difference test (*D* test), based upon separate estimation of two nested models, and calculating the difference between the associated goodness-of-fit chi-square statistics and their degrees of freedom (df), historically has been used to compare nested models. Generally, this is the best test to use, but it can be cumbersome to apply.[1] Luckily, there are two equivalent test procedures, known as Lagrange Multiplier (*LM*) and Wald (*W*) tests, that are less difficult to apply and can be used as well (Bentler, 1986a; Chou & Bentler, 1990, 2002). (Note: the "*W*" in "*W* test" should not be confused with the weight matrix in (1.5) and elsewhere.) Actually, because each test has its special virtues, in practice these tests are not used in the same circumstances. The *LM* test is used to evaluate the effect of adding free parameters to a restricted model (i.e., reducing restrictions on the model). The *W* test is used to evaluate the effect of dropping free parameters from a more complete model (i.e., adding restrictions to the model). This chapter provides a theoretical overview of these tests and shows how they are implemented in EQS 6, for which we have generalized these tests to a range of modeling circumstances beyond the published literature.

There are several reasons for considering the use of *LM* and *W* tests in practice. First, statistical theory has verified that the D, *LM*, and *W* statistics are asymptotically equivalent chi-square tests (Bentler & Dijkstra, 1985; Engle, 1984; Lee, 1985b; Lee & Bentler, 1980; Satorra, 1989). Thus, in large samples all three tests have the same size and power, so, from a technical point of view, they can be used interchangeably when testing the same hypothesis (Chou & Bentler, 1990). This equivalence also means that an *LM* or *W* test statistic can be interpreted as if a *D* test had been carried out. In other words, the *LM* statistic can be interpreted as an approximate decrease in model goodness-of-fit chi-square resulting from freeing previously fixed parameters and from eliminating equality restrictions. Similarly, the *W* statistic can be interpreted as an approximate increase in model goodness-of-fit chi-square resulting from fixing previously free parameters. However, there is no need to actually estimate the alternative models to obtain statistics that can be used to form the *D* test: In principle, the actual difference between chi-squares obtained from estimating two models (i.e., the *D* statistic) is no more accurate or meaningful as a test on the model-differentiating parameters than are the *LM* or *W* tests. In practice, some caution must be used on the *W* test (e.g., Chou & Bentler, 1990; Gonzalez & Griffin, 2001) as is explained more fully below.

A second practical reason for using *LM* and *W* tests is that these statistics can be obtained conveniently in a single computer run, which is not true of the *D* test where two runs need to be made, and the results compared. *LM* and *W* tests obtained from a single run, of course, do not provide information on exactly the same parameters: the *LM* test focuses on the effect of freeing parameters that are currently fixed in the given model, while the *W* test focuses on the effect of fixing parameters that are currently free in the given model. A second run with the more restricted model that fixes the specified previously-free parameters would provide the goodness-of-fit information for the *D* test which would be equivalent to the *W* test computed in the first run. A third run with the less restricted model that frees the specified previously-fixed parameters would provide the goodness of fit information for that *D* test which

[1] Normally, the *D* test is simply given as the difference between two nested test statistics, e.g., $D_{ML}=T_{ML(res)} - T_{ML(gen)}$, the simple difference between the values of restricted and more general model runs. With the Satorra-Bentler statistic, the matter is more complicated because the simple difference $\bar{T}_{ML(res)} - \bar{T}_{ML(gen)}$ is not χ^2 distributed. Chapter 5 discusses how to correctly compute the Satorra-Bentler difference test.

would be equivalent to the *LM* test computed in the first run. Thus, three computer runs would be needed to provide the difference tests that are equivalent to the *LM* and *W* tests obtained from a single run.

A third practical reason for considering the use of *LM* and *W* tests is that these tests can be relatively easily implemented in an exploratory mode, so that they may provide some guidance on how a model may be modified to yield improved fit of a model to data (Bentler, 1986a). The reality is that SEM is most commonly used when only a tentative initial model can be specified (Jöreskog, 1993). While *D* tests can be similarly used, because of the large number of computer runs needed, they do not lend themselves easily to model modification. Of course, when used in an exploratory fashion, the statistical theory associated with *LM*, W, and *D* tests is compromised, and the computed probability levels may not be accurate. These probabilities may also not be accurate if the more general model is not correct (Yuan & Bentler, 2002d). When using these tests in an exploratory mode, cross-validation of results is highly recommended (see Cudeck & Browne, 1983); this requires evaluating the modified model in a new or split sample, not just computing a single-sample cross-validation index (Browne & Cudeck, 1993). Similar cautionary statements have been made regarding related univariate methods for model modification. A good discussion of problems associated with post-hoc model modification is given by Chou and Bentler (2002), Cliff (1983), Green, Thompson, and Poirier (1999), Kaplan (1988, 2000), Luijben, Boomsma, and Molenaar (1988), MacCallum (1986, 1995), MacCallum, Roznowski, and Necowitz (1992), and Silvia and MacCallum (1988). In order to get a successful model search, MacCallum and his colleagues proposed carefully formulating the initial model, using a large sample, adopting restricted or theoretically guided searches, continuing searches even after a model with a nonsignificant χ^2 has been obtained, and focusing only on a pre-specified set of fixed parameters that are considered substantively meaningful. Methods for controlling Type I error in search procedures have only recently been developed, but also would seem to be critical to the process in small samples (Green & Babyak, 1997; Green, Thompson, & Babyak, 1998; Green, Thompson, & Poirier, 1999, 2001; Hancock, 1999).

As will be seen below, the *LM* and *W* tests are implemented in EQS to give you great control in the specification of hypotheses to test, in adding exploratory information to confirmatory tests, and in yielding guided exploratory results about the importance of parameters and constraints. As usual, there are defaults that will cover many standard cases, but the tests will have their greatest value when specialized to your particular application. The statistical theory for the *LM* test covers almost all estimation methods in EQS (LS is excluded), but if several methods are obtained in a single run, *the LM test is provided only on the last method of estimation requested*. Thus if both ML and ERLS estimators are computed, the *LM* test is provided only for the ERLS method. In multisample analysis, the *LM* test evaluates all constraints, both within-group and cross-group. The *W* test is provided for every method in EQS, and is now available in multisample analysis.

Although EQS 6 provides *LM* and *W* tests as aids for improving models, it must be recognized that any process of model improvement based on an existing model is liable to be fraught with danger. In particular, if the current model is not remotely a subset of the "true" model in the population, then you should have little hope that dropping parameters (*W* test) or adding parameters (*LM* test) will yield the "true" model. Of course, if you take the viewpoint that there is no single "true" model (e.g., Cudeck & Henly, 1991; MacCallum, Tucker, & Briggs, 2001), and that all models are just approximations to reality, then if sample size were large enough, *LM* and *W* tests have to yield better approximations to reality than given by the starting model. Of course, you should be humble, because in practice sample size is rarely large enough and the model you select may be the wrong model (see e.g., Raykov & Marcoulides, 1999). A second problem to recognize is that the set of variables being analyzed is given, and conceivably a model could be improved by adding or dropping *variables* from the analysis (e.g., Breivik & Olsson, 2001). Although exciting new methods for doing this meaningfully have been developed (Kano, 2002; Kano & Harada, 2000), this is not accomplished by the *parameter* tests currently implemented in EQS. A third problem to recognize is that model modifications always are made to your current model specification, even though in principle they are equally relevant to *equivalent* models (see Hershberger, 1994; Lee & Hershberger, 1990; Luijben, 1991; MacCallum, Wegener, Uchino, & Fabrigar, 1993; Raykov & Marcoulides, 2001; Raykov & Penev, 2001; Stelzl, 1986; Williams, Bozdogan, & Aiman-Smith, 1996; Wood, 2000). As an example, X→Y→Z and X←Y←Z are equivalent models. Equivalent models have identical fit, but modifications of such models typically would no longer keep them equivalent and it is possible that starting with one of a set of equivalent models could yield the "true" model while starting with the other might not. A final problem to recognize is that EQS implements a specific methodology for adding and dropping parameters, based on a type of stepwise regression. This may not necessarily be the best possible methodology. A dozen years ago, there were hints that the TETRAD approach

could work better in some situations (Bentler & Chou, 1990; Scheines, Spirtes, Glymour, & Meek,1994; Spirtes, Scheines, & Glymour, 1990), and it is certainly possible that improved search methods can be found (e.g., Marcoulides & Drezner, 2001, 2003; Marcoulides, Drezner, & Schumacker, 1998; Shipley, 2000a, 2002; Spirtes, Glymour, & Scheines, 2001). We are experimenting with some of these new methods. Nonetheless, you will find the current EQS methodology to be quite serviceable.

There is also a philosophical issue. If you are using SEM to test a theory or series of theories, once you have specified the theory and its alternatives and evaluated them against data, one approach is to say that there is nothing else you can or should do. Your theory or theories will be found to be adequate to a certain extent. Another point of view is that data can be extremely expensive to obtain, and if your model is inadequate, you have an obligation to try to find out how it might be improved. The methods of this chapter are given in the latter spirit. In our view, if your model is attempting to explain the growth of a certain cancer, for example, but your starting model is inadequate, we would hope that you not give up and would try to find a better model.

A review of the ideas that form a basis of the *LM* and *W* tests is provided in the next sections. These sections explain the basis for computations used in EQS. If you are interested primarily in applications, you can skip the more technical parts of these sections. A good introductory overview of *LM*, W, and *D* tests can be found in Buse (1982) and a thorough statistical analysis can be found in Satorra (1989). Not all extensions of these tests are automatically computed in EQS (Yuan & Bentler, 1997a). Some relations of these tests to Akaike's AIC are given in Chou and Bentler (1996); see also Kaplan (1991). The last part of this chapter contains some of the technical details for the various model modification statistics computed in EQS 6.

LM Test: Theory

If a structural model does not fit sample data adequately, theory may suggest that certain fixed parameters, or restrictions, be released. The *LM* test evaluates the statistical necessity of the restrictions, based on calculations that can be obtained on the restricted model alone. A test equivalent to the *LM* test in a maximum likelihood context, called the score test, was first introduced by Rao (1948). Aitchison and Silvey (1958) rationalized Rao's test by the use of Lagrange Multipliers. The LM principle is quite general, and it was adopted for normal theory GLS estimation in covariance structure models by Lee and Bentler (1980). Using results in Bentler and Dijkstra (1985), we have made extensions of their approach so that the theory is applicable to the wide variety of estimation methods and statistics available in EQS 6 (see Chapter 5). Here are a few examples. The *LM* test can be based on information matrices or on Hessian matrices, depending whether the expected or observed information matrix is specified for obtaining standard errors. Another innovation is that robust versions of *LM* tests on equality constraints are also available to handle distributional misspecification (see e.g., Satorra, 1989). And, although there is no literature that we know of regarding the *LM* test with case-robust methods, this methodology is available in EQS 6. Similarly, although typically developed in a single-group context, in EQS 6 you also get *LM* tests with multisample analysis. The following discussion emphasizes the common basis for all *LM* tests in EQS.

It is helpful to think of all the restrictions involved in a model in a sequence, and to give a name to each restriction. The designation $c_1, c_2,..., c_r$ will be used to describe the r restrictions in a model. Actually the number r is somewhat arbitrary. It includes all explicit equalities specified in the /CONSTRAINT section, plus a lot of additional restrictions that are implicit in the model and not listed in the input, such as "missing paths" or "fixed zero" parameters. The latter parameters are handled in the theory by first considering the parameters as "free", but then also subject to a constraint, e.g., θ_i is free, but also $\theta_i = 0$. (So of course it must be estimated at $\hat{\theta}_i = 0$.) Thus the i^{th} constraint c_i states $\hat{\theta}_i = 0$.

Corresponding to each constraint c_i is a number called a Lagrange Multiplier λ_i. In a sample, λ_i is estimated as $\hat{\lambda}_i$. With r constraints, there is an $r \times 1$ vector $\hat{\lambda}$ of Lagrange Multipliers corresponding to the $r \times 1$ constraint vector $c(\theta)' = (c_1,..., c_r)$, and one usually tests $c(\theta)=0$. The *LM* test translates a test on the constraints c_i into a test on the significance of the Lagrange Multiplier $\hat{\lambda}_i$. The test is available in two versions, the multivariate version, which tests all r constraints simultaneously, and a univariate version, often called the modification index (Sörbom, 1989),

which tests a single constraint. In particular, the multivariate *LM* statistic is distributed, in large samples, as a chi-square variate with r df

$$LM \sim \chi^2_{(r),}$$

enabling the statistic to be used to evaluate the restrictions. For a specified single i^{th} restriction, the univariate *LM* statistic is an asymptotic 1 df chi-square variate

$$LM_i \sim \chi^2_{(1)}.$$

The univariate test provides an evaluation of the necessity of a given restriction. This test can be applied repeatedly to test a variety of single restrictions. However, results based on such univariate tests cannot be used to determine what the simultaneous effect of several restrictions may be. The reason is, of course, that the univariate *LM* statistics are correlated. Thus two restrictions may be, for all practical purposes, synonymous, and releasing one of the two restrictions may improve the fit of the model as much as releasing both restrictions, as could be seen by the multivariate test: The two-restriction *LM* statistic may be barely larger than one of the *LM* statistics.

Unique to EQS is a method of breaking down the multivariate test into a series of incremental univariate tests, based on the work of Bentler and Chou (1986). That is, standard theory is used to partition the total chi-square statistic based on r df, into r separate components, each of 1 df. Although there are many ways to accomplish such a partition, the APRIORI method used in EQS is based on forward stepwise inclusion, a standard procedure in regression analysis (see, e.g., Jennrich, 1995; Jennrich & Sampson, 1968). The starting point for this process is based on the restriction having the largest univariate *LM* statistic. Then, a fixed parameter is freed in turn, or a constraint released, by picking the associated largest univariate increment to cumulative chi-square. This process is repeated until all r constraints have been included, yielding the final multivariate *LM* statistic. The resulting partition of the multivariate *LM* statistic might show, for example, that the total chi-square is largely a function of only one or two restrictions. Consequently, releasing all of the hypothesized restrictions would have no greater impact on model fit than releasing only the few key restrictions. Thus the partition of chi-square can provide potentially useful, somewhat exploratory, information to an *a priori* confirmatory multivariate test. The total chi-square can also be decomposed by adding parameters (or, releasing restrictions) in a precise, *a priori* hierarchical order specified by the investigator. This "HAPRIORI" procedure, of course, also partitions the multivariate χ^2 statistic.

It is implicit in the *LM* test, as described above, that you know exactly which r restrictions should be subjected to statistical evaluation. In the early stages of model building, however, the model under investigation may be so highly restricted that it is extremely poor at describing the data being modeled; that is, the goodness-of-fit χ^2 may be very large compared to df. In such a case you may not have precise enough knowledge to specify particular restrictions to evaluate, since typically there will exist a very large number of parameters that are erroneously fixed or omitted from the model, or are inappropriately constrained. For such applications, EQS provides you with an option to specify sets of parameters in various matrices of the Bentler-Weeks model to be evaluated by *LM* test. While all of the relevant fixed parameters are evaluated by the univariate *LM* test, in this more exploratory context the multivariate *LM* test will contain only those parameters that provide significant univariate increments to the cumulative chi-square statistic. The multivariate test prints parameters in order of decreasing contribution to χ^2. *If a meaningless parameter appears somewhere in this list, tests on parameters below that point should not be trusted.*

An interesting question is whether the multivariate test, called the *LM*-incremental method by Green, Thompson, and Poirier (1999), or repeated application of the univariate test should be used in exploratory model modification. Kaplan and Wenger (1993) discuss implications of correlations among parameter estimates, and note that when parameter estimates are mutually asymptotically independent, they (and associated *LM* and *W* tests) can safely be evaluated separately. Nonetheless, following Sörbom (1989) and others, they recommend using the univariate test even when asymptotic independence does not hold, i.e., respecifying the model, doing another univariate test, and so on in a sequential manner. This is also the recommended approach of Green, Thompson, and Poirier (1999), who call this the *LM*-respecified approach. There is certainly nothing wrong with the *LM*-respecified approach, though it can be tedious to implement. An obvious compromise is to add a few parameters at a time based on the multivariate test.

In an exploratory context, a set of univariate *LM* tests cannot inform you about the potential adequacy of your model, while the multivariate test can do this. Suppose that your initial model has a very large χ^2, say 1000, and to get an acceptable model this χ^2 would have to drop about 800 points. The multivariate test can immediately tell you whether this could conceivably happen, given your starting model. If your specification for the *LM* test has allowed all conceivable reasonable parameters to be evaluated by the *LM* test, and the multivariate test with its significantly contributing parameters shows an *LM* χ^2 value of, say, 300 points, you immediately know that you can only achieve a χ^2 fit of approximately 700 = 1000 – 300. Hence the multivariate test immediately informs you that you are working within an unacceptable class of models, and that you must try a different initial model structure.

Specific procedures for implementing the *LM* test are described in a subsequent section.

Parameter Change

The *LM* test is very useful for evaluating whether, from a statistical point of view, a model could be improved substantially by freeing a previously fixed parameter. Associated with each such *LM* test is an estimate of the value that the parameter might take if it were to be freely estimated rather than constrained. This estimate is called the Parameter Change, or expected parameter change. The estimate can be used to help determine whether a parameter, when freed, would be of such a magnitude that it is substantively important. Clearly, a parameter associated with a large *LM* test and also a large Parameter Change should be evaluated for model improvement. It also has been argued that even if an *LM* test does not suggest statistically that a fixed parameter should be freed, a large Parameter Change statistic implies that the model is badly misspecified and the parameter should be freed for that reason. This is the point of view taken by Saris, Satorra, and Sörbom (1987), who developed the concept. See also Kaplan (1989, 2000) and Luijben, Boomsma, and Molenaar (1988). However, the size of the Parameter Change can be affected by how variables and factors are scaled or identified, so the absolute size of this statistic is sometimes hard to interpret. Hence, a completely standardized parameter change statistic is also computed in EQS. This is based on the work Chou and Bentler (1993), who extended the results of Kaplan (1989).

A technical development for the expected parameter change statistic is given by Satorra (1989), and a summary can be found below. In addition to change expected on fixed parameters, it is possible to obtain expected changes on free parameters (Bentler & Chou, 1993), but this methodology is not available automatically in EQS. You can obtain study change in free parameters with judicious use of the *W* test as noted below.

W Test: Theory

It was pointed out above that *LM*, W, and *D* tests are equivalent in large samples. Thus the *W* test could be used to test the same restrictions as discussed under the *LM* test above. If this were done, one would start with a different model as compared to the *LM* test, namely, the more complete model that contained the r restricted parameters as free parameters, in addition to the free parameters of the previously restricted model. The *W* test then can be used to evaluate whether the previously considered r restrictions are statistically significant. This can indeed be done in EQS, and, again, both univariate and multivariate tests are provided. In practice, however, the *W* test has better uses than simply evaluating the identical restrictions as the *LM* test. One would start with the given model, whatever that might be, and use the *W* test to evaluate whether some of the free parameters in the current model could be restricted. In contrast, the *LM* test would be used to evaluate whether some of the fixed parameters in the current model could be freed. When *LM* and *W* tests are both applied to the same model, they address different restrictions, and thus deal with different substantive questions.

The *W* test was developed by Wald (1943) many years ago. This test may be considered to be a multivariate generalization of the square of the normal *z*-test, which tests the null hypothesis that a single free parameter is zero in the population, and was introduced to structural modeling by Bentler and Dijkstra (1985) and Lee (1985b). The *W* test has a natural application in evaluating whether a set of free parameters can be simultaneously constrained, usually, set to zero. This might be done, for example, to test an *a priori* hypothesis, to obtain a more simplified model for purposes of interpretation, or to improve model fit by gaining degrees of freedom at only minimal loss in

overall goodness of fit. The theory of the *W* test is based on the following ideas. A more complete description is given at the end of this chapter.

From a theoretical point of view (which would permit *LM* and *W* tests to evaluate the same restrictions), the *W* test differs from the *LM* test in that only the more complete, less restricted model is estimated to yield estimates $\hat{\theta}$ of the parameter vector θ. The hypothesis $c(\theta) = 0$, involving r constraints, is evaluated on the basis of the distribution of the estimated constraint $\hat{c} = c(\hat{\theta})$. As an example, suppose the constraint is to evaluate the hypothesis that $\theta_1 - \theta_2 = 0$, i.e., that the two parameters are equal. When using the *W* test, $\hat{\theta}_1$ and $\hat{\theta}_2$ are freely estimated and hence they may well not be equal, so the *W* test procedure is to evaluate the distribution of $\hat{c} = c(\hat{\theta}) = \hat{\theta}_1 - \hat{\theta}_2$. The statistical question is whether their inequality is just chance (i.e., supporting the null hypothesis) or representative of a true difference (i.e., rejecting the null hypothesis). Thus in general, whether or not $c(\theta) = 0$, typically $\hat{c} \neq 0$ in the *W* test because the restrictions are not imposed during the estimation. This is in contrast to the *LM* test, where always $c(\hat{\theta}) = 0$ since the restrictions are imposed.

The multivariate *W* statistic is distributed asymptotically as a chi-square variate

$$W \sim \chi^2_{(r)}.$$

The test, which evaluates whether r nondependent restrictions can be imposed, has r df. For the specific single i^{th} free parameter, the univariate *W* test of the hypothesis $\theta_i = 0$ is

$$W_i \sim \chi^2_{(1)}.$$

As noted above, the univariate *W* test is the square of the (asymptotic) normal *z*-test often used to evaluate the significance of a single parameter. Unique to EQS, robust versions of *W* tests are obtained automatically when robust sandwich estimates of standard errors are obtained (see Chapter 5).

While the *W* test, as developed above, can be used to evaluate any restrictions, in structural modeling practice it would usually not be used to evaluate the same restrictions as the *LM* test. Thus, conceptually, it may be clearer to think of the *LM* test evaluating the restrictions $c_1(\theta)$ and the *W* test evaluating the restrictions $c_2(\theta)$. The restrictions c_1 are imposed during estimation (e.g., some parameters are fixed), and the restrictions c_2 are not imposed (e.g., some parameters are free). The *LM* test is concerned with the restrictions c_1 that may be harmful to model fit, while the *W* test is concerned with the restrictions c_2 that may be able to be added without also disturbing model fit. *W* test restrictions will typically be simple, namely, fixing to zero a previously free parameter.

In EQS, the multivariate *W* test is carried out in a novel stepwise fashion, again, by analogy to well-known procedures in the regression model. The stepping is backward, rather than forward as in the *LM* test. In backward stepping, the least significant parameter (potential restriction) is chosen first. This free parameter can be eliminated from the model without degradation of fit if the *W* test is not significant. Next, the parameter that provides the smallest univariate increment to the χ^2 gets added to create the 2 df *W* test. This process is repeated in sequence, where at each step the parameter adding minimally to the current cumulative χ^2 gets included in the current *W* test. Of course, parameters at the end of the sequence will have larger χ^2 increments than those at the beginning of the sequence. The process terminates when all r free parameters have been included, yielding the simultaneous multivariate *W* test on whether r restrictions can be imposed on the parameters, usually, whether r free parameters can be simultaneously set to zero. The value of the stepwise implementation of the simultaneous test is that it may indicate that only a few parameters carry all of the weight in the multivariate test. For example, a significant 10 df *W* test may indicate that 10 free parameters cannot simultaneously be set to zero. However, the stepwise *W* test may reveal that 8 of the 10 parameters could, in fact, be set to zero, since the 8 df *W* test based on those parameters may not be significant. However, the final 2 parameters may be unable to be set to zero, since their increment to χ^2 may be very large. If the stepwise procedure is performed on a set of free parameters specified by the researcher, the procedure is called "APRIORI" in EQS. This procedure adds exploratory information to an otherwise completely *a priori* multivariate test. If the backward stepping is to be performed in the exact hierarchical sequence specified by

the researcher, the procedure is called “HAPRIORI". The theory for these tests was developed by Bentler and Chou (1986; Bentler, 1986a). See Satorra (1989) for more technical detail.

A purely exploratory implementation of the *W* test has also been provided in EQS. In this approach, all free parameters automatically are candidates for inclusion in the *W* test to evaluate whether they could be simultaneously set to zero. Parameters are added to the *W* test (i.e., dropped from the model) as long as none of the univariate increments associated with inclusion of a given parameter in the test becomes significant, provided that the most current multivariate *W* test is not significant. If the next parameter in the sequence were to lead to a significant univariate increment in chi-square, or to a significant multivariate statistic, the process stops and the new parameter is not included in the multivariate test, i.e., the free parameter is not fixed to zero. At this point, since the multivariate *W* test is not significant, a set of parameters will have been located that can be simultaneously dropped from the model with only a trivial loss of fit.

Now we issue a few words of caution about the *W* test. First, as noted above, research by MacCallum and others (Green, Thompson, & Poirier, 1999) has established that in an exploratory context it is wise to overfit i.e., to add perhaps more parameters than minimally needed, before considering parameters to drop from a model. We recommend dropping parameters only at a near-final step in model modification. Second, obviously the results of any *W* test procedure have to be examined with caution, since the test is blind as to the meaning of parameters and you should not eliminate meaningful parameters. For example, we would recommend never deleting variance parameters, even though they may be non-significant. Usually such parameters are important conceptually to the model structure. Third, if the multivariate test yields an uninterpretable result at a given step, you should not trust results below that step since they are conditional on the earlier results. Fourth, the *W* test on some parameters may not be invariant to certain possibly arbitrary rescalings of variables, e.g., to alternative identification constraints. In a related vein, the power function of the *W* test has been found not to be monotonic with the distance between null and alternative hypotheses (Chou & Bentler, 1990). As a result, the *W* test may yield misleading conclusions, especially in small samples (Chou & Bentler, 1990; Gonzalez & Griffin, 2001). If you have marginally significant results, you may wish to follow up a *W* test with a *D* test of the same restrictions. The *D* tests holds up in situations where the *W* test could fail. This can be accomplished easily with the COMPARE command (see below).

/WTEST

To continue this thread, we discuss the *W* test first, then the *LM* test. This test is implemented if the keyword /WTEST (or /WTE) is included in the input section. If only the section heading /WTEST is provided, with no further specification, the program will compute the sequential, backward-stepping, multivariate *W* test based on all free parameters currently in the model. Free parameters that are associated with a nonsignificant univariate chi-square increment will be added to the *W* test, but a parameter will not be added if the current multivariate test has become significant.

You are urged initially simply to provide the specification:

```
/WTEST
```

in the input model file, which will provide for a default testing procedure. When this basic method is understood, additional features may become valuable as indicated next.

The *W* test can also be implemented under specified additional conditions given by the researcher. These conditions control the operation of the *W* test. They are as follows:

KEYWORD	ABBREVIATION	DEFAULT	MEANING
PVAL	PV = #;	.05	Criterion probability value of statistic
NOFIX	NO = (P,P);	None	List of parameters not to be dropped
APRIORI	AP = (P,P);	None	Parameters to be dropped first
HAPRIORI	HA = (P,P);	None	Parameters to be dropped in given sequence

KEYWORD	ABBREVIATION	DEFAULT	MEANING
PRIORITY	PR = ZERO;	None	Sequences tests within APRIORI
COMPARE	CO = YES;	No	Runs reduced model after dropping parameters in APRIORI and HAPRIORI

The information in the /WTEST section can be placed in any order, but each separate type of information must be delineated by the semicolon (;). Some additional information on each of the operands is given below.

PVAL

PVAL refers to the probability value to be used in evaluating the statistics computed in the *W* test. This value must be less than 1.0, and, by default, it is .05. Thus a parameter will be added to the *W* test (as a free parameter that can be fixed to zero) if the current univariate chi-square increment in the backward step has an associated probability that is larger than PVAL, and the current multivariate test has a probability larger than PVAL. As a consequence, more parameters will be dropped (included in the *W* test) if PVAL is made smaller (say, to .01), and fewer parameters will be dropped if PVAL is made larger (say, to .10).

If an APRIORI or HAPRIORI *W* test is made, PVAL will be ignored in these computations, and all parameters specified by you will be tested. Additional parameters, beyond those specified by you, may also be included in the *W* test if they meet the PVAL criterion as noted above.

NOFIX

The stepwise multivariate *W* test is carried out on all free parameters unless you specify that certain free parameters not be considered in the test. This statement permits you to specify up to 999 free parameters that should not be used in the *W* test. The specification is done by using the usual double-label convention to describe each parameter, enclosing each parameter in parentheses as is done in the /CONSTRAINT section. Thus, if a researcher wants to have certain free parameters in the model whether or not they are needed by a statistical significance criterion, this can be assured by an NOFIX specification. Examples are the following:

```
/WTEST
NOFIX = (E1,E1), (V3,F2), (F3,F2);

/WTEST
PVAL = .10; NOF = (V1,V4),(V1,V5);
```

The first example specifies three free parameters that should be excluded from the *W* test, while the second forces two free parameters to be excluded from the test while also changing the criterion probability value used to make the test. *Note that only free parameters are allowed to be in the list of NOF parameters*. The list of parameters can extend across several lines, but care must be taken that a given parameter is on only one line, and that a comma is used as the last character of each line except the final line.

APRIORI

In the APRIORI procedure, the *W* test is carried out on the specific parameters you designate. An example is the following:

```
/WTEST
APRIORI = (V3,F2), (V5,F3), (V8,F1), (F2,F1);
```

which designates four free parameters to be used in the *W* test. Note that the usual double-label convention is used to specify the parameters, and that *all parameters must be free.* A maximum of 999 parameters can be included in such a list.

When the APRIORI procedure is used, parameters are added into the *W* test using the backward selection procedure: the least significant parameter from the list is included first, then the next least significant parameter (as given by the

univariate increment to chi-square) is added, etc. The complete multivariate *a priori* test is completed after all the designated parameters have been included in the test. PVAL is ignored in this procedure.

EQS does not necessarily stop dropping parameters when the *a priori* test is completed. In addition to the parameters specified in the APRIORI list, EQS will search for additional free parameters to drop. If these additional parameters meet the PVAL criterion, they will be included in the test. However, they are added subsequent to completion of the multivariate APRIORI test, so that each parameter added in this way must provide a nonsignificant increment to the APRIORI test. Of course, if the multivariate *a priori* test is significant to begin with, no further parameters will be added to the test.

HAPRIORI

The hierarchical *a priori W* test is specified in virtually the same way as the APRIORI test, except that the letter H is added as a prefix:

```
/WTEST
HAPR = (F2,F1), (F3,F1), (F4,F1);
```

The given free parameters (up to a maximum of 999 parameters) are all included in one *W* test. In this case, however, the parameters are added stepwise in the particular sequence you give. In the example, (F2,F1) is added first, (F3,F1) is added next, and (F4,F1) is added third. The given sequence is adhered to even if the first parameter is significantly different from zero, while the other two are not. (In contrast, the APRIORI procedure would have searched for the least significant parameter on the list to include first.) The HAPRIORI procedure is a completely confirmatory procedure, which, additionally, provides a partitioning of the total chi-square. The HAPRIORI and APRIORI procedures cannot be used at the same time.

Wald Test On User-Specified Constants

The /WTEST feature of the program can also evaluate whether or not parameters are equal to pre-specified nonzero constants. The specific parameters involved in such hypotheses must be given: hence, this feature is available only under the APRIORI and HAPRIORI implementations of the *W* test. The format is as follows:

```
/WTEST
APRIORI = (V5,F1): 0.98, (E2,E3): .33, (F2,F2): 1.0;
```

where the parenthesized expressions represent free parameter names in the usual double-label convention, and the colon ":" indicates that the number which follows represents the hypothesized value of the parameter. Commas separate the various parameters, and a semicolon completes the statement. Note that if a constant is not specified after the double-label parameter name, the *W* test will be based on a hypothesized constant of zero, as before. Thus if ": 0.98" is not specified for (V5,F1), then (V5,F1) would be tested against 0.0. This implementation works with both HAPRIORI and APRIORI variants of the Wald test. Results from the *W* test are printed in the usual way, under the heading `WALD TEST ON ** USER SPECIFIED PARAMETER CONSTANTS`. In addition, two correlational statistics are reported.

When parameter constants are specified, EQS computes two correlation coefficients to summarize the degree of similarity, across all free parameters, between the hypothesized constants and the final parameter estimates in the current run. These are printed out under the heading `CORRELATION BETWEEN A PRIORI CONSTANTS AND FINAL ESTIMATES IS **` and `RANK CORRELATION BETWEEN A PRIORI CONSTANTS AND FINAL ESTIMATES IS **`. The first correlation reported is a standard product-moment correlation; the second is the Spearman rank-order correlation. The rank-order correlation may be preferred for interpretation because it is based on weaker assumptions. Obviously, both of these correlations will be 1.0 if each estimate is exactly equal to its corresponding constant, and will be somewhat less otherwise.

The correlations between constants and current estimates can be used as indicators of similarity between two sets of estimates when the constants in the W test represent previously estimated free parameters. That is, the list of fixed constants given in /WTEST can be based on the final estimates from the previous run's results. These final estimates

can be obtained from the new WPAR=YES; feature in /PRINT, used with RETEST, as described in Chapter 3. Suppose, for example, that the current model being estimated is identical to the previous model, except that it contains additional parameters. Then the free parameters from the previous model are also free parameters in the current model, and the correlation coefficients summarize the degree of similarity of the final estimates from the two runs. If the correlations are high, the estimates of the common parameters are very similar even if the overall models differ substantially otherwise. Such a result implies that the added parameters in the new model did not distort the pattern of results from the previous model. On the other hand, if the correlations are intermediate to low, substantial changes in estimates have occurred.

It is also possible to use the backward-stepping feature of the *W* test with constants to locate those free parameters that remain stable under change in model, and those that change substantially. The earliest steps in the *W* test will be associated with parameter estimates that change little or not at all, while the last steps that reflect significant increments in the *W* test will be associated with parameters whose estimates change substantially due to model change. In contrast to the correlational index of parameter estimate similarity, the *W* test uses the estimated sampling variabilities and covariances of the estimates to help gauge the extent of change, and it helps to locate where nontrivial changes in estimates have occurred.

These types of analysis also can be used in comparing results from different samples. Furthermore, specified constants in the *W* test can be useful to evaluating the power of the model against a specified alternative. Bollen's text (1989b, pp. 338-349 – soon to be updated) provides a good discussion of power based on the work of Satorra and Saris (1985) and Matsueda and Bielby (1986). See also Satorra (2003) and Yuan and Hayashi (in press).

PRIORITY

While the HAPRIORI feature of the *W* test permits tests on the parameters to be sequenced, there are times when one may not be able to provide a logical sequencing of all parameters. Of course, APRIORI accomplishes a test based only on the empirical sequencing of least significant increments to the χ^2 test. Something intermediate, akin to the SEQUENTIAL option in the *LM* test (see below), is sometimes useful. PRIORITY provides such an option; it must be used with the APRIORI or HAPRIORI specification, since it indicates which of two types of parameters in a specified list are to be given priority in the test. One option is that tests based on the null hypothesis that a parameter is zero are to be performed before tests based on the null hypothesis that a parameter is equal to some fixed nonzero value. The reverse priority is also possible. Implementation is as follows.

```
/WTEST
PRIORITY = ZERO; APRIORI = (F2,F1):0, (V6,F2):1.0, (E3,E2):0;
```

indicates to the program to first select from the APRIORI list all parameters that are to be evaluated against the fixed value of zero, perform the *W* test on these parameters as a group in the specified APRIORI way, and thereafter select from the list the remaining parameters to be tested as a group in the specified way. Thus, (F2,F1) and (E3,E2) would be tested first in the example, and (V6,F2) would only be tested subsequently. Within each group of such parameters, the test will be conducted in the usual APRIORI or HAPRIORI way, in accord with the given specification. In the example, either (F2,F1) or (E3,E2) would enter first, depending on their univariate χ^2 values. If HAPRIORI had been specified, of course (F2,F1) would be entered first.

PRIORITY = NONZERO; is the other alternative. In this case, the *W* test would give priority to the fixed nonzero constants in the APRIORI or HAPRIORI list of parameters. The parameters associated with nonzero constants would thus be tested first as a group. The parameters associated with zero constants would be tested subsequently.

The output from this procedure is printed under the usual heading `WALD TEST (FOR DROPPING PARAMETERS)` in the output file. The additional line `WALD TEST PRIORITIZED ON ZERO (or NONZERO) CONSTANT PARAMETERS` is printed subsequently. Then, there are two parts to the *W* test printout. The first part would correspond to the hypotheses given priority; and the second part, to the hypotheses that remain.

The PRIORITY procedure is intended to provide a more useful way to partition the overall χ^2 into components. The final cumulative multivariate χ^2 statistic is, of course, the same no matter what sequence of parameters is chosen for

the *W* test. However, PRIORITY permits evaluating the contribution to the total χ^2 of the tests associated with zero parameters, as well as those associated with nonzero parameters.

COMPARE

When COMPARE=YES; is added to the WTEST section, EQS will run the reduced model, after deleting all the parameters specified in the APRIORI or HAPRIORI list. The goodness-of-fit section of the output for the reduced model includes the chi-square difference test (*D* test), with degrees of freedom, and tail probability. This comparison may be used in multisample analyses as well as in a single sample.

Wald Test in Multisample Analyses

In multisample analyses, cross-sample /WTEST should be specified in the last sample only. Parameters that appear in NOFIX, APRIORI, and HAPRIORI specifications must include the group number, as in the CONSTRAINT section, e.g.

```
/WTEST
 APR = (1,V5,F2), (2,E5,E5);
```

/LMTEST

The *LM* test uses several commands in common with the WTEST, but it can be much more complicated to implement because it has many more options. You specify the LMTEST in its own input section, using the keyword /LMTEST (or /LMT). Initially, you are urged simply to insert the line:

```
/LMTEST
```

in the input file. This produces a default *LM* test, which provides information on univariate *LM* tests for specific parameters fixed at zero, and for all parameters fixed at a nonzero value. It also provides for a forward stepwise procedure that, at any stage, selects as the next parameter to be added to the multivariate test that single fixed parameter that provides the largest increase in the multivariate chi-square statistic. A second *LM* test may also be provided automatically. If the current model contains a /CONSTRAINT section specifying one or more equality constraints, a separate *LM* test is performed to evaluate the statistical necessity of each constraint. This test is provided in both a univariate version, and a forward-stepping multivariate version. After working with these basic tests for a while, you may wish to modify the various program defaults in accord with the specifications discussed next.

Several characteristics of the test can be controlled by using similar operands as in the *W* test, but additional control is provided to permit detailed specification of the test when used in an exploratory mode. The following operands are applicable to the *LM* test.

KEYWORD	ABBREVIATION	DEFAULT	MEANING
PVAL	PV = #;	.05	Criterion probability value of statistic
NOFREE	NO = (P,P);	None	List of parameters not to be included
APRIORI	AP = (P,P);	None	Parameters to be included first
HAPRIORI	HA = (P,P);	None	Parameters to be included in given sequence
SET	SE = PFF,GVF; = NO;	#	Submatrices to be used Use no submatrices
PROCESS	PR = SEP; = SIM; = SEQ;	SIM	Type of *LM* test to be performed: SEPARATE, SIMULTANEOUS, or SEQUENTIAL
BLOCK	BL =(V1,…), …,(V9,…);	None	Groups variables into sets
LAG	LA = 0, 1, …;	None	Specifies time or block lag

Default for SET is PVV, PFV, PFF, PDD, GVV, GVF, GFV, GFF, BVF, BFF.

PVAL

PVAL is the probability value to be used as a criterion for inclusion of fixed parameters, or constraints, in the *LM* test. It must be a number such that 0 < PVAL < 1.0. The default is .05.

When the /LMTEST is implemented in a forward stepwise manner, the most significant parameter based on the univariate increment to chi-square will be added first. That parameter will have a low observed probability associated with the *LM* statistic, since the lower the probability value of the test, the more significant the parameter, and hence the more critical statistically it will be to a model. Parameters continue to be added to the test if the associated univariate increment in chi-square at that step is significant when compared to PVAL. Thus, when the probability level for the univariate increment becomes greater than PVAL, the multivariate *LM* test will stop.

Changing PVAL will generally alter the number of parameters added to the multivariate test. Lowering PVAL, to .01, say, will have the effect that fewer parameters will be added to the test. Increasing PVAL, to .10, say, will have the effect that a greater number of parameters will be added to the test.

NOFREE

Since the *LM* test can be implemented in a relatively blind manner, parameters could possibly be added to the test that, logically or on the basis of theory, should not be added. Stated differently, certain fixed parameters in a model should perhaps always remain fixed, no matter how the fit of the model might be improved if these parameters were free to be estimated. For example, it would be foolish to free a fixed parameter that corresponds to a causal action that operates backwards in time. NOFREE permits you to specify parameters that should not be included in an *LM* test, using the usual double-label convention. An example is the following:

```
/LMTEST
NOF = (V1,V2), (V2,V3), (V3,V4);
```

In this example, the three fixed parameters given in the list are excluded from the *LM* test. Even if one of them were to have a significant univariate *LM* statistic, or provide a significant univariate increment to the multivariate *LM* statistic, it would not be added. The NOFREE statement overrides the SET option, discussed below. *The NOFREE list may not contain any free parameters*. The list may extend across several lines of input, but each parameter should be stated on one line; and the last character on a line should be a comma if the input continues to the next line.

APRIORI

You may wish to obtain an *LM* test for certain fixed parameters to see whether freeing them simultaneously would significantly improve the fit of the model. This would imply that the given fixed parameters would be associated with a significant multivariate *LM* test. The test is specified by providing the list of parameters:

```
/LMTEST
APR = (F1,F3), (F5,F4), (V2,V1);
```

In this example, the multivariate test is specified to be performed on the given three fixed parameters. *Each of the parameters in the list must be fixed (not free) in the basic model.* The maximum number of parameters in such a list is 999.

When an APRIORI *LM* test is specified, parameters are added to the test in a forward stepwise manner. The parameter having the largest univariate *LM* statistic is taken first. Succeeding parameters are added to the multivariate test one at a time, based on the size of the univariate increment to the then-current multivariate chi-square. PVAL is ignored in this process, which continues until all the parameters on the specified list have been included.

At the completion of the *a priori* testing procedure, the program checks to see whether additional parameters beyond those on the list could be usefully added to the *LM* test. Unless SET = NO, additional fixed parameters will be selected to add to the multivariate test if they are associated with a significant univariate increment in chi-square, using PVAL to evaluate significance. This procedure is implemented in a forward stepwise manner; it terminates when no fixed parameter can be found that is associated with a significant univariate increment in chi-square. If the phrase SET = NO is added to the specification:

```
/LMTEST
APR = (D4,E5), (F2,F3); SET = NO;
```

then the *LM* test will be based only on the given parameters and no others. Otherwise, additional parameters on the list of matrices given by SET will be searched.

HAPRIORI

If HAPRIORI is specified, rather than APRIORI, the parameters on the list will be included in the multivariate *LM* test in the particular sequence described. Thus,

```
/LMTEST
HAPR = (F2,F3), (F1,F2), (F3,F4);
```

will perform the forward stepwise inclusion of parameters in the order: F2,F3 first; F1,F2 second; F3,F4 third. As a consequence, the intermediate test statistics will generally not be the same as with APRIORI, but the final multivariate test will be the same as that obtained from APRIORI. In all other respects, the HAPRIORI procedure is the same as the APRIORI procedure.

The following aspects of the LM test (SET and PROCESS) represent advanced features of the procedure. These features should be ignored until you are familiar with the other features.

SET

Univariate *LM* test statistics can be determined for a very large potential set of parameters, subject to storage limitations of the program. The phrase SET determines which ones will be computed. In turn, this also specifies the potential parameters to be included in the multivariate forward stepwise procedure, in addition to those specified in an APR or HAPR statement. The specification of these parameters with SET is done with the matrices of the Bentler-Weeks model as given in eqs. (1.2)-(1.4).

It will be remembered that all variables used in EQS are either *dependent* or *independent* variables. The designation of each variable as dependent or independent is needed to run the EQS program, since dependent variables have

equations, and only independent variables may have variances and covariances. It is possible to specify that the *LM* test should be computed for certain types of structural regression coefficients, or for certain types of covariances. First we shall address the covariances.

The possible independent variables in any model are V, F, E, or D variables (e.g., V2, F3, E11, or D1), hence their combinations gives the possible variances and covariances. These are given as elements of the matrix PHI.

Covariance Matrix PHI of Independent Variables

Independent Variables	V	F	E	D
V	**VV**			
F	**FV**	**FF**		
E	EV	EF	EE	
D	DV	DF	DE	**DD**

DEFAULT: (VV, FV, FF, DD) OR PHI

In any given application, there may not be any independent variables of a certain sort. For example, in factor analysis, all V variables will be dependent variables, so the first row and column, above, will be empty.

Specific parts of PHI are addressed by the two letters shown. Thus, the EE submatrix of PHI will contain potential correlated errors (actually, variances and covariances of error variables). The FF submatrix of PHI will contain variances and covariances of factors. The DE submatrix will contain covariances between D and E variables. (These are usually present only in nonstandard models.) As noted above, the *LM* test will be applied as a default (unless otherwise specified) to the VV, FV, FF, and DD parts of PHI only, i.e., to the matrices given in boldface above. *Note that LM tests on correlated errors would not be computed unless you explicitly specify that you want these.* Such specifications are based on adding the prefix P (for PHI) to the submatrix names, as in:

```
/LMTEST
SET = PEE, PVV;
```

which would select only the EE and VV parts of PHI for the computation of *LM* tests. *Note that any specification of "SET" overrides all default matrices.* Thus covariances among factors, i.e., PFF, would not be tested in this example.

The program *output* will provide a submatrix number for each parameter considered. These numbers are not used in the input. This information can help clarify the meaning of a given parameter. The submatrices are numbered from 1 to 10 from upper left to lower right in the figure above. That is, VV = 1, FV = 2, FF = 3, ..., DE is 9, and DD = 10. Thus the default submatrices are 1, 2, 3, and 10.

Structural regression coefficients are specified similarly, using the GAMMA and BETA matrices, see eq. (1.2). Regressions of dependent on independent variables are given in GAMMA, and dependent on other dependent variables, in BETA.

Regression Matrix GAMMA

Dependent Variables	Independent Variables			
	V	F	E	D
V	**VV**	**VF**	VE	VD
F	**FV**	**FF**	FE	FD

DEFAULT: (VV, VF, FV, FF) OR GAMMA

The letter G for GAMMA is used to specify particular submatrices above, such as GVV for the VV part of GAMMA, or GFV for the regression of dependent Fs on independent Vs. Note that only V and F variables can be dependent variables. If no submatrices of GAMMA are specified, the default parts noted in bold above will be used in the *LM* test. The submatrices of GAMMA are numbered, in the *output* only, from 11 to 18, in the sequence VV = 11, VF =12, ,..., FE = 17, and FD = 18. Thus the default submatrices are 11, 12, 15, and 16.

The BETA matrix contains the coefficients for the structural regression of dependent on other dependent variables. The matrix is as follows.

Regression Matrix BETA

Dependent Variables	V	F
V	VV	**VF**
F	FV	**FF**

DEFAULT: (VF, FF) OR BETA

This matrix contains only four submatrices, and the two bolded ones are used as a default in the *LM* test. The default matrices cover F variables as predictors. The submatrices can be referred to with the letter B, as in BFF, which would compute *LM* tests for regressions of dependent factors on dependent factors. In the *output*, VV = 19, VF = 20, FV = 21, and FF = 22. An example of the use of SET with the *LM* test is the following.

```
/LMTEST
SET = P, GVF, BETA;
```

It will be remembered that use of the word SET overrides all defaults, so only the specified matrices will be searched for fixed parameters to free. The letter "P" indicates the use of the default submatrices of PHI, namely VV, FV, FF, and DD. GVF indicates use of the VF submatrix of GAMMA, which would usually represent factor loadings (paths from Fs to Vs). BETA indicates use of the default submatrices of BETA, namely VF and FF. Note that when only the default submatrices of a given matrix are to be specified, it is not necessary to use the three-letter code for the specific submatrices. Using either the full name (e.g., PHI) or the abbreviation (P) for the matrix will do the job.

There is a final feature of SET that works in cooperation with the PROCESS specification discussed next. This feature, based on parentheses, permits grouping of parameters from various submatrices into equivalent units or sets. Thus SET = PEE, BVF; creates two parameter sets. So does SET = (PDD,PFF),GFF; but here, PDD and PFF are in the same set, while GFF is in another set. The default (no SET specification) yields the 10 sets of parameters described previously. As another example,

```
/LMTEST
PROC = SEP;
SET = (PEE,P),(GFV,GFF),BFF;
```

groups PEE and the default matrices of PHI into one unit or set for analysis; groups the matrices GFV and GFF into another set for analysis; and keeps BFF as a third unit for analysis. Note that commas separate the various units or sets. The phrase PROC = SEP indicates that several *LM* tests are to be performed, as noted in the next section.

PROCESS

This statement controls the type of *LM* test to be performed and permits, in fact, several *LM* tests to be computed in a single run. When the statement is not included in the specification, the default method SIM or SIMULTANEOUS is used. The simultaneous *LM* test is a single test that uses all of the parameters from all matrices at once: all parameters specified in "SET" compete for simultaneous inclusion in the forward stepwise test. Thus, the most significant parameter may come from PFF, the next most significant from BVF, the next from GFF, etc., in accord with the forward stepping procedure described previously. If SET is not specified, the default submatrices will be used.

The statement PROC = SEP; above, indicates that SEPARATE *LM* tests are to be performed for each set of parameters, as specified in SET. Thus each set of parameters has its own forward stepwise multivariate test. The above example with SET = (PEE,P),(GFV,GFF),BFF; in effect specifies that three multivariate *LM* tests are to be performed in a single run: one test based on PEE and the default matrices in P; another based on GFV and GFF; and a third based on BFF. Note that parameters from one set cannot compete for inclusion in the *LM* test for another set. Separate tests may be useful when attempting to clarify several different parts of a model, e.g., factor loadings or correlated errors, without letting these parts compete for importance in a single test. (As was noted above, EQS always automatically provides a separate *LM* test for equality constraints: this test does not need to be specified, and cannot be modified by you.)

The statement PROC = SEQUENTIAL; (or SEQ) provides a way of adding a hierarchical, theory-based procedure to a single exploratory *LM* test. With this statement, sets of parameters are considered for inclusion in the multivariate *LM* test in accordance with the grouping of matrices specified in the SET statement. In general, SET would thus be specified to reflect an *a priori* view of the importance of various types of parameters. Thus, the specification

```
/LMTEST
PROC = SEQ; SET = PFF, (GVF, BVF), PEE;
```

indicates that any significant covariances among factors (PFF) are to be entered first in the *LM* test; that factor loadings for both independent (GVF) and dependent (BVF) factors are to be considered next, equally; that no factor loading is to be entered until all significant factor covariances are first entered; and that no correlated errors (PEE) are to be considered until all significant PFF, GVF, and BVF parameters have first been included. Thus, even if the largest univariate *LM* statistic were to be found in PEE, the corresponding error covariance or variance parameter would not be added to the stepwise *LM* test until all significant parameters in PFF, GVF and BVF were first included. It is, of course, entirely possible that when the PEE parameter is considered later in the sequence, it might no longer provide a significant increment to chi-square.

BLOCK

The BLOCK command permits you to group variables into blocks. It also partitions the matrices specified in SET into smaller submatrices for analysis, then specifies the direction of possible paths. In addition, it specifies possible covariance linkages among variables to be included or eliminated.

Only V and F types of variables can be listed; the program will search for E and D types of variables and group them appropriately based on their correspondence to V and F variables. BLOCK will group together into a single block all of the V and F variables that are listed in a statement, whether the variables are listed individually or included in a sequence via a TO or dash convention. When listed separately, each variable must be separated from other variables by a comma. Each block must be identified by a pair of parentheses ().

If there are to be several blocks of variables, each set must be surrounded by parentheses and a comma must separate the blocks. So, for example: BLOCK = (V1-V3,F1), (V4-V6,F2), (V7-V9, F3); creates three blocks of variables corresponding, for example, to three measurement times. V1, V2, V3, and F1 are in the first block. V4, V5, V6, and F2 are in the second block. V7, V8, V9, and F3 are in the third block. *The listing sequence of the blocks indicates the directional sequence in which paths across blocks are permitted to be evaluated*; i.e., only forward paths or covariances will be analyzed. (If you want to shift the direction of the paths, you must reverse the sequence listing of the blocks.) Still greater control is made possible by the LAG command.

LAG

The LAG specification defines the "time" lag desired for paths between variables in the *LM* test. Possible values are LAG = 0; up to LAG = *b-1*, where *b* is the number of blocks created by the BLOCK statement. LAG = 0 means that only variables within the same block will be selected; with 3 blocks, there would be 3 possible sets of within-block paths or covariances to evaluate. With LAG = 1; only paths or covariances across adjacent blocks would be evaluated, for example, from T_1 to T_2 and from T_2 to T_3. If you wanted to study the cross-block effects from T_1 to

T_3, you would write LAG = 2;. In typical practice, one might consider only LAG = 0; in one analysis, LAG = 1; in another analysis, and so on. However, several lags can also be specified simultaneously, for example, LAG = 1,2,4;. When LAG is not specified, a default is implemented, which is ALL, i.e. 0,1, up to *b-1*.

Examples of the directional blocking feature are as follows:

```
/LMTEST
   BLOCK = (V1,V7,V9,F1), (V2,V3), (F2,V4,V5,V6), (F3,V10 TO V15);
   SET = BFF, BVV;
   LAG = 1;
```

In this example, there are 4 blocks, and only paths between adjacent blocks will be evaluated. Paths are to be of the type involving regression of dependent Vs on other dependent Vs, and dependent Fs on other dependent Fs.

```
/LMTEST
   BLOCK = (V1-V5), (V6-V10), (V11-V15);
   SET = PEE;
   LAG = 0;
```

In this example, correlated errors are evaluated, but only covariances within blocks are to be searched. If LAG = 1; had been used instead, only cross-time covariances with lag one would be evaluated.

Output Code

The results of an EQS run will also contain, in the printout, a numbered CODE for each parameter, representing information utilized during computation of the *LM* test. This information may be useful to clarifying what the *LM* test did, if standard *LM* output seems inadequate. The CODE contains two numbers. The first number refers to whether the parameter is included in an *a priori* test (in which case it is 1), or which set of parameters it belongs to (2, 3, etc.). In a simultaneous test, all parameters will be labeled as belonging to set 2. The second number refers to the submatrix number in which the parameter resides (1 to 22, as described above). However, if the parameter has been fixed at a nonzero value, its code is 0, and if the parameter is part of an *a priori* test, the code represents the parameter's sequence number in the *a priori* list.

A Word of Warning

While the *LM* test can be trusted when its assumptions -- including the a priori specification of the constraints to be tested -- are met, in reality it is often used as a model building tool. It is a remarkably powerful tool, but it also can mislead. We already have noted the problems associated with small samples, but it should be clear that the size of the model space being searched also has an impact. In a carefully restricted search, only a few meaningful parameters are evaluated. But in a blind search, say with only the /LMTEST specification, you may be testing hundreds of possible parameters. Clearly Type I errors are liable to be a serious problem, and can perhaps be mitigated, especially in small samples, using methods to control error rates as noted by Green and Babyak (1997), Green, Thompson, and Babyak (1998), Green, Thompson, and Poirier (1999, 2001), and Hancock (1999). But even in huge samples there is a problem: the *LM* test, being a forward stepping procedure, depends on the starting model as noted theoretically (Steiger, Shapiro, & Browne, 1985) and empirically (MacCallum, 1986). Thus you should be prepared not necessarily to find the "true" model. As noted earlier, your chances are improved if you first overfit, and only then reverse course and remove the least significant parameters (e.g., Bentler & Chou, 1987; Chou & Bentler, 1990, 2002; Green, Thompson, & Poirier, 1999; MacCallum, Roznowski, & Necowitz, 1992; Silvia & MacCallum, 1988).

If you want to be conservative in your testing strategy, you should be serious about controlling Type I error in exploratory model modification. To help you do this, EQS provides *LM* test probabilities based on Hancock's (1999) finite intersection multiple comparison rationale. This is a criterion to use when you (1) have huge misspecifications and (2) want to correct only the most serious ones. In Hancock's procedure, a 1-df *LM* statistic is evaluated not with 1 df, but with the df of the current model. Clearly, in practice, one will have to have a large χ^2/df

ratio for this approach to be useful, but since badly misspecified models may have χ^2/df ratios of 5 or more, Hancock's criterion will help to safely find the larger misspecifications. It will not help you find small misspecifications. If χ^2/df = 2, for example, there could be a parameter to add that would lead to an *LM* test value of size df or larger. But if χ^2/df = 1.3, say, it would be hard to find an *LM* test value of about df or larger, since if one had the correct model, the expected value of the test χ^2 is df; thus here one might expect an *LM* test of only .3df, which would then not meet Hancock's criterion. EQS also implements Hancock's sequential method in the context of the multivariate *LM* test. The sequential test, based on univariate increments to the multivariate χ^2, decrements the starting df by one with each new tested parameter.

Hancock (personal communication) has also suggested a less conservative modification strategy that would aim to avoid Type I errors in the structural relations but would be more liberal in the measurement model. When refining the measurement model in Step 1 of a two-step process, one would use the ordinary df test to find relevant cross-loadings or error covariances since they often have little impact interpretively on the structural relations parameters. However, when entertaining structural modifications in Step 2 of this two-step process, one would use Hancock's (1999) conservative approach since structural parameters are more closely tied to substantive theory.

Illustrative Examples

Constrained Latent Variable Model

The latent variable model shown in Figure 2.11 of this *Manual* will be used to illustrate some features of the *LM* testing procedure. It will be remembered that this is a very restricted model. The input in Figure 2.12 was modified in two ways: 1) The equality constraint (F1,F3) = (F2,F3); was added to the constraint section, and 2) the statement /LMTEST was added. As a consequence, only the default *LM* procedures are implemented. This model was submitted to EQS. The run was uneventful, yielding no condition codes and appropriately meeting the constraints at the solution. The resulting model had a chi-square of 48.592 with 10 df, $p < .001$. Apparently, adding this one additional constraint changed the model from adequate to inadequate. It would be interesting to see if /LMTEST could locate the problems with the resulting model.

As was noted above, when constraints exist in a model, EQS produces univariate tests and a multivariate test for all of the constraints. The original model contained three constraints, so the current model contains four. EQS first prints a title, and then echoes back information on the constraints and numbers them:

```
CONSTRAINTS TO BE RELEASED ARE:

        CONSTR:    1   (E1,E1)-(E3,E3)=0;
        CONSTR:    2   (E2,E2)-(E4,E4)=0;
        CONSTR:    3   (E3,E1)-(E4,E2)=0;
        CONSTR:    4   (F1,F3)-(F2,F3)=0;
```

Notice that constraint #4 is the new constraint added to the model. Next, EQS prints out the statistics for each of the separate constraints:

```
           UNIVARIATE TEST STATISTICS:

   NO     CONSTRAINT     CHI-SQUARE    PROBABILITY

    1     CONSTR:   1         0.151         0.698
    2     CONSTR:   2         0.548         0.459
    3     CONSTR:   3         1.524         0.217
    4     CONSTR:   4        32.669         0.000
```

Based on these results, we can conclude that only constraint #4 is associated with a significant *LM* test ($p < .001$). Apparently, constraints #1 to #3 are reasonable constraints, but #4 is not consistent with the data: the null hypothesis

that #4 is zero in the population must be rejected. However, to be sure of our conclusions, we should look at the multivariate test, which is printed next:

	CUMULATIVE MULTIVARIATE STATISTICS				UNIVARIATE INCREMENT	
STEP	PARAMETER	CHI-SQUARE	D.F.	PROBABILITY	CHI-SQUARE	PROBABILITY
1	CONSTR: 4	32.669	1	0.000	32.669	0.000
2	CONSTR: 3	33.750	2	0.000	1.081	0.299
3	CONSTR: 2	34.114	3	0.000	0.365	0.546
4	CONSTR: 1	34.115	4	0.000	0.001	0.977

The multivariate test is a forward stepwise procedure. Since there are four constraints, there are four steps. The constraint with the largest univariate chi-square is entered first. It yields the same statistic as before. Next, the constraint adding the most to the current value (32.669) of chi-square is added to the test. This is #3, which yielded a chi-square increment of 1.081 for a total 2 df multivariate test of 33.750. Two more forward steps are taken, with constraints #2 and then #1 added in sequence. In this particular case, the sequence for adding constraints to the *LM* test was the same as the relative size of the univariate statistics, but this will not generally be the case.

It will be noted that the final multivariate test, based on 4 df, is statistically significant. Thus, keeping all four constraints simultaneously implies a degraded model fit. However, from the multivariate test (and also in this instance, the univariate tests) we see that constraint #4 accounts for virtually all of the 34.115 value of the statistic. Clearly, only constraint #4 is unreasonable. In fact, none of the univariate increments to the multivariate chi-square, beyond the first, was significant. For example, the increment in step 2, for constraint #3, had p = .299, which is not significant at the .05 level, showing that the constraint is acceptable, i.e., cannot be rejected.

Next, the default procedure provides a series of tests on the fixed parameters of the model, starting with univariate *LM* statistics. The program prints these in order of size, starting with the largest:

LAGRANGE MULTIPLIER TEST (FOR ADDING PARAMETERS)
ORDERED UNIVARIATE TEST STATISTICS:

NO	CODE	PARAMETER	CHI-SQUARE	PROB.	HANCOCK 10 DF PROB.	PARAMETER CHANGE	STANDARDIZED CHANGE
1	2 10	D2,D1	32.669	0.000	0.000	4.612	0.879
2	2 22	F1,F2	32.669	0.000	0.000	0.341	0.041
3	2 20	V5,F2	23.757	0.000	0.008	0.778	0.083
4	2 12	V2,F3	14.360	0.000	0.157	-0.125	-0.017
5	2 20	V6,F2	12.518	0.000	0.252	0.394	0.062
6	2 20	V5,F1	12.023	0.001	0.284	-0.211	-0.025
7	2 12	V3,F3	8.030	0.005	0.626	0.112	0.013
8	2 20	V2,F2	5.224	0.022	0.876	0.063	0.007
9	2 20	V1,F2	2.236	0.135	0.994	0.047	0.005
10	2 12	V1,F3	1.857	0.173	0.997	-0.050	-0.006
11	2 0	V2,F1	1.520	0.218	0.999	0.045	0.006
12	2 0	V3,F2	1.399	0.237	0.999	-0.045	-0.004
13	2 12	V4,F3	0.598	0.439	1.000	0.027	0.003
14	2 0	V1,F1	0.437	0.509	1.000	0.029	0.003
15	2 20	V6,F1	0.365	0.546	1.000	-0.024	-0.004
16	2 20	V4,F1	0.228	0.633	1.000	-0.015	-0.002
17	2 20	V3,F1	0.228	0.633	1.000	0.018	0.002
18	2 0	V4,F2	0.213	0.644	1.000	-0.015	-0.002
19	2 0	V5,F3	0.000	1.000	1.000	0.000	0.000

There were 19 fixed parameters evaluated by the univariate *LM* statistic. These parameters come from the fixed elements of the 10 default submatrices of PHI, GAMMA, and BETA as noted above, and as printed out below. In each of the 19 rows of printout, we have the following information: First, the sequence number of the test, starting with the parameter having the largest value. Second, two items labeled CODE. The first of these indicates that the parameters all belong to set number 2, that is, they will be considered all together in the multivariate test. The next number indicates the matrix from which the fixed parameter is taken. For example, the parameter D2,D1 comes from matrix 10, namely, PDD. The only exception to this numbering is when matrix 0 is indicated: *fixed nonzero*

parameters are not indicated by matrix number, but by the code 0. Thus, for example, the 11th parameter, V2,F1, is a fixed nonzero parameter in the model: its actual location in the submatrices is not shown. Third, the parameter is given its usual double-label name, as was just noted. Fourth, the column labeled CHI-SQUARE gives the univariate 1 df *LM* statistic for testing the null hypothesis that the restriction on that particular parameter is valid in the population. Fifth, the corresponding probability is printed out. Small values indicate high significance, i.e., that the restriction is unreasonable and perhaps the parameter should be freed. Sixth, the probability associated with Hancock's (1999) conservative test, here based on 10 df, is printed out. Note that the p-values are generally larger than the 1-df based probability. The next column, PARAMETER CHANGE, provides a projection of the change in parameter value that would occur when releasing the constraint. For example, D2,D1 is estimated to take on the value of 4.612. The parameter V4,F2, which was fixed at .833, is expected to be smaller by .015 if estimated freely. These values can be used as start values in subsequent runs, if they are needed. The final column gives the standardized value of the parameter change, which is usually more interpretively useful. As just noted, the D2,D1 covariance 4.612 is hard to interpret, but in the standardized correlation metric, the correlation of .879 is quite high.

In the example, the first 8 of 19 univariate *LM* tests are significant ($p < .05$) by the standard 1 df test. Thus, one might be tempted to release all eight constraints. Hancock's (1999) conservative test shows that only the first 3 of these are significant. An even more stringent conclusion results from the multivariate test.

```
MULTIVARIATE LAGRANGE MULTIPLIER TEST BY SIMULTANEOUS PROCESS IN STAGE  1
   PARAMETER SETS (SUBMATRICES) ACTIVE AT THIS STAGE ARE:
      PVV PFV PFF PDD GVV GVF GFV GFF BVF BFF

       CUMULATIVE MULTIVARIATE STATISTICS      UNIVARIATE INCREMENT
       ----------------------------------      ----------------------------
                                                                  HANCOCK'S
                                                                  SEQUENTIAL
STEP   PARAMETER   CHI-SQUARE  D.F.   PROB.   CHI-SQUARE   PROB.  D.F.   PROB.
----  -----------  ----------  ----   -----   ----------   -----  ----   -----
   1       D2,D1       32.669     1   0.000       32.669   0.000    10   0.000
```

The simultaneous process evaluates the contributions to the multivariate *LM* test of individual restrictions, using a forward stepwise procedure. The parameter having the largest univariate *LM* statistic is entered first, and then, parameters are added sequentially providing that they make a significant univariate increment to the multivariate test. In this case, the univariate *LM* statistics are highly misleading, since only one parameter is needed (i.e., one restriction needs to be released), whether based on standard or Hancock degrees of freedom. EQS has adapted the Hancock sequential testing procedure in the multivariate context to the univariate increments. The df for each univariate increment changes as parameters are added to the multivariate test. The first step begins with the model df, and in each succeeding step the df is decremented by one. As there is no 2nd step in this example, this is not shown. The remaining 7 (standard) or 2 (Hancock) restrictions that are significant by univariate test are evidently so highly correlated with the D2,D1 restriction that they become nonsignificant when D2,D1 is entered first. This is why there is no second step to the multivariate test.

Note: The multivariate probabilities are always based on the standard degrees of freedom, without any correction for exploratory model building. When conducting APRIORI and HAPRIORI *LM* tests, which are not exploratory, or tests of equality constraints, Hancock's sequential procedure is not printed.

It is very important that you note that the first and second parameters in the univariate LM test list essentially have the same χ^2 value of 32.669. Thus it is a toss-up as to which parameter should be entered first in the multivariate test as the two possible choices may be precisely equivalent statistically. EQS uses an arbitrary decision rule to decide which parameter to include first – but your interpretation of the meaningfulness of the parameters always should override any default procedure. In this case, the path F1←F2 is the alternative to the covariance D2,D1. If this parameter were more interpretively useful, you should ignore the multivariate test output as the "wrong" parameter was used.

To conclude this example, one might now be tempted to reestimate the model with the two changes noted by the *LM* tests: 1) Release constraint #4, the artificial equality constraint, and 2) release the constraint that the covariance of D2,D1 is zero, i.e., let D2,D1 be a free covariance. There is a problem, however. We did not perform a simulta-

neous test on these two restrictions, and it is possible that these constraints are also redundant. Unfortunately, in the current EQS program, *there is no way of simultaneously testing equality constraints and fixed parameters with a single LM test*. We could now use either theory or practice to evaluate the situation. Theoretically, it can be shown that if we eliminate constraint #4, the inner part of the model describing relations among the three factors is saturated. Then additionally permitting D2,D1 to be a free covariance would produce an underidentified model. In practice, if we make the two changes, EQS will indicate that the parameters in the new model are linearly dependent, and it will point to parameter D2,D1 as creating the problem. Thus, in this case, the program would show us how to avoid making a mistake.

Factor Analysis Model

A factor analysis model on file manul7a.ess can be used to illustrate both the *W* and *LM* tests. This is the manul7.ess file studied previously without the outlier case. The basic model for this data set is a 2-factor model, with the 1st 3 variables as free indicators for F1, the last 3 variables as free indicators of F2, the variances of F1 and F2 fixed to 1.0 for identification, and F1,F2 freely correlated. This model fits well, with the T_{ML} = 2.7 for 8 df, with a p-value of .95. This model was modified as follows: 1) Two factor loadings (V5,F1 and V6,F1) were added to the model, 2) /WTEST was added as a section, with an APR = (V5,F1),(V6,F1); specification, and 3) /LMTEST was added as a section, with SET = PEE; added to specify this test further. Since the original model fits the data quite well, we would expect that the *W* test would show that the two added parameters are not needed, and furthermore, we would expect that no correlated errors would help to improve the fit.

The new model obtained T_{ML} = 2.55, which, based on 6 df, indicated excellent fit (p = .86). The *W* test section yielded the following result. Note that the title indicates that the test is *a priori*.

```
WALD TEST (FOR DROPPING PARAMETERS)
MULTIVARIATE WALD TEST BY   APRIORI    PROCESS
```

	CUMULATIVE MULTIVARIATE STATISTICS				UNIVARIATE INCREMENT	
STEP	PARAMETER	CHI-SQUARE	D.F.	PROBABILITY	CHI-SQUARE	PROBABILITY
1	V6,F1	0.028	1	0.867	0.028	0.867
2	V5,F1	0.169	2	0.919	0.141	0.707

The two parameters that were added to the model turn out to be unnecessary. Each univariate increment to chi-square is nonsignificant (p > .05), and the simultaneous 2 df test value of .169 is also nonsignificant (p = .919). The *W* test next proceeds to a second stage, in which an attempt is made to find *additional* free parameters that may not be needed in the model. The PROCESS is SIMULTANEOUS by default, that is, all free parameters will be considered.

```
MULTIVARIATE WALD TEST BY SIMULTANEOUS PROCESS
```

	CUMULATIVE MULTIVARIATE STATISTICS				UNIVARIATE INCREMENT	
STEP	PARAMETER	CHI-SQUARE	D.F.	PROBABILITY	CHI-SQUARE	PROBABILITY
1	E3,E3	1.146	3	0.766	0.977	0.323
2	E5,E5	2.870	4	0.580	1.724	0.189
3	F2,F1	6.387	5	0.270	3.516	0.061

Three parameters are *added* to the previously considered two free parameters, so the multivariate test has 3, 4, and 5 degrees of freedom successively as each parameter is added. Each parameter adds a small univariate increment to the current cumulative chi-square statistic, low enough not to be significant (each p > .05). The final 5 df test value of 6.387 is not significant so that, from a statistical point of view, one could drop the two error variances (of E3 and E5), and the factor correlation F2,F1, that is, fix them to zero. However, *in a latent variable model it will not make sense to fix error variances to zero*! Since the factor correlation is nearly significant, it may be desirable to check this parameter again after dropping the unnecessary factor loadings, above. Next, the program provides *LM* test information on the correlated errors.

```
LAGRANGE MULTIPLIER TEST (FOR ADDING PARAMETERS)
     ORDERED UNIVARIATE TEST STATISTICS:
                                                    HANCOCK              STANDAR-
                                    CHI-              6 DF   PARAMETER   DIZED
 NO    CODE      PARAMETER        SQUARE     PROB.    PROB.    CHANGE    CHANGE
 --   ------    ---------        ------     -----  --------  ---------  --------
  1    2   6        E3,E1         1.329     0.249     0.970    -0.546    -1.254
  2    2   6        E5,E2         0.660     0.417     0.995    -0.080    -0.183
  3    2   6        E2,E1         0.566     0.452     0.997     0.131     0.188
  4    2   6        E6,E1         0.565     0.452     0.997     0.074     0.125
  5    2   6        E5,E3         0.505     0.477     0.998     0.098     0.360
  6    2   6        E6,E2         0.463     0.496     0.998    -0.067    -0.109
  7    2   6        E4,E2         0.095     0.758     1.000     0.032     0.048
  8    2   6        E4,E1         0.044     0.834     1.000    -0.022    -0.034
  9    2   6        E6,E3         0.044     0.835     1.000    -0.026    -0.068
 10    2   6        E5,E1         0.027     0.870     1.000    -0.017    -0.040
 11    2   6        E3,E2         0.013     0.908     1.000     0.033     0.073
 12    2   6        E4,E3         0.000     0.987     1.000    -0.002    -0.005
 13    2   6        E6,E4         0.000     1.000     1.000     0.000     0.000
 14    2   6        E5,E4         0.000     1.000     1.000     0.000     0.000
 15    2   0        F1,F1         0.000     1.000     1.000     0.000     0.000
 16    2   0        F2,F2         0.000     1.000     1.000     0.000     0.000
 17    2   6        E6,E5         0.000     1.000     1.000     0.000     0.000

***** NONE OF THE UNIVARIATE LAGRANGE MULTIPLIERS IS SIGNIFICANT,
***** THE MULTIVARIATE TEST PROCEDURE WILL NOT BE EXECUTED.
```

By traditional 1 df and Hancock (1999) 6 df criteria, all the univariate tests have probabilities greater than .05, so no fixed parameter should be freed. Notice the odd feature that the error covariance E3,E1 is predicted to take on a value of -.546, corresponding to a standardized correlation of -1.254! Clearly, such an estimate is out of range and, even if the *LM* test were significant, you should be suspicious about whether adding this correlated error could destroy your model. As shown in the column titled CODE, most of the parameters come from the 6th submatrix PEE. However, some parameters have the code 0. This indicates that EQS also computed statistics for the fixed nonzero parameters of the model, in this case, the fixed variances of F1 and F2. *Since these two parameters were fixed for identification purposes, the corresponding univariate LM tests should be zero*, and they are. Since none of the univariate *LM* statistics is significant, the program prints the message:

```
***** NONE OF THE UNIVARIATE LAGRANGE MULTIPLIERS IS SIGNIFICANT,
***** THE MULTIVARIATE TEST PROCEDURE WILL NOT BE EXECUTED.
```

and stops. If some parameters had been significant, the forward stepwise procedure would have been implemented.

Nonstandard Model

The nonstandard model given in Chapter 2, Figure 2.15, will provide the final example. It has $T_{ML} = 15.4$ based on 15 df, an excellent model. The program input was modified in several ways: 1) Three free parameters (see APR next) were eliminated from the model, 2) an APR = (V1,V8), (F1,V1), (F3,V8); statement was added under an /LMTEST, and 3) a /WTEST statement was added. If all of the parameters of the original model are needed, the *a priori LM* test should verify that the three parameters should be added to the model. However, some of the other existing free parameters may also not be needed. The results yielded a model with an ML chi-square of 47.3 with 18 df ($p < .001$). The *W* test yielded:

```
MULTIVARIATE WALD TEST BY SIMULTANEOUS PROCESS

        CUMULATIVE MULTIVARIATE STATISTICS              UNIVARIATE INCREMENT

STEP   PARAMETER   CHI-SQUARE     D.F.   PROBABILITY     CHI-SQUARE   PROBABILITY

   1      F2,V8        2.486        1        0.115          2.486        0.115
   2      F2,F2        5.251        2        0.072          2.766        0.096
```

indicating that two of the original free parameters could potentially be dropped from the model. However, the F2,F2 parameter represents the variance of factor 2, and it would almost surely not be desirable to eliminate this parameter since that would have the effect of wiping out the factor by creating a factor with no variance.

The univariate *LM* test results were as follows.

```
LAGRANGE MULTIPLIER TEST (FOR ADDING PARAMETERS)
     ORDERED UNIVARIATE TEST STATISTICS:
                                                                  PARAMETER  STANDARDIZED
 NO    CODE     PARAMETER    CHI-SQUARE   PROBABILITY   CHANGE       CHANGE
 --   ------   -----------   ----------   -----------  ---------  ------------
  1    1   3       F3,V8        13.534        0.000      -2.260      -0.380
  2    1   2       F1,V1        12.778        0.000       0.615       0.108
  3    2  16       F1,F3        10.498        0.001       1.164       0.263
  4    2  20       V1,F1         7.011        0.008       0.201       0.035
  5    1   1       V1,V8         6.995        0.008      -0.109      -0.014
  6    2  11       V3,V8         4.368        0.037       0.120       0.012
  7    2  12       V3,F3         4.214        0.040       0.438       0.096
  8    2  12       V5,F3         3.626        0.057      -0.417      -0.124
  9    2  15       F1,V8         3.150        0.076       0.155       0.016
 10    2  12       V1,F2         2.905        0.088       0.543       0.275
```

plus another 18 lines of output corresponding to less significant constraints. The *a priori* multivariate *LM* test showed:

```
MULTIVARIATE LAGRANGE MULTIPLIER TEST BY   APRIORI    PROCESS IN STAGE  0

       CUMULATIVE MULTIVARIATE STATISTICS              UNIVARIATE INCREMENT

STEP  PARAMETER  CHI-SQUARE   D.F.  PROBABILITY     CHI-SQUARE  PROBABILITY

  1     F3,V8       13.534      1      0.000           13.534       0.000
  2     F1,V1       26.312      2      0.000           12.778       0.000
  3     V1,V8       28.819      3      0.000            2.507       0.113
```

indicating that only two of the three parameters that had been dropped from the model are actually needed. The V1,V8 parameter does not provide a significant increment in chi-square, indicating that freeing this parameter will not improve the fit of the model. In this case, this could have been determined from the normal *z*-test for that parameter in the original model. Note that the current value of the test statistic, T_{ML} = 47.3 with 18 df is predicted by the multivariate *LM* test to go to about 47.3 – 28.8 = 18.5 if the 3 fixed parameters were freed; in fact, we know that such a model with 15 df has T_{ML} = 15.4. So the drop in χ^2 is approximately, but not exactly, predicted.

The program next searches the default list of fixed parameters to determine whether freeing any of these would help improve the fit. As can be seen, no other such parameter yields a significant improvement to the current 3 df multivariate chi-square. This effect occurs in spite of the fact that seven parameters were significant ($p < .05$) by the univariate test, as shown above.

```
MULTIVARIATE LAGRANGE MULTIPLIER TEST BY SIMULTANEOUS PROCESS IN STAGE  1

PARAMETER SETS (SUBMATRICES) ACTIVE AT THIS STAGE ARE:
   PVV PFV PFF PDD GVV GVF GFV GFF BVF BFF

       CUMULATIVE MULTIVARIATE STATISTICS              UNIVARIATE INCREMENT

STEP  PARAMETER  CHI-SQUARE   D.F.  PROBABILITY     CHI-SQUARE  PROBABILITY

   ************
NONE OF THE PARAMETERS IN THIS SECTION IS SIGNIFICANT, THIS STAGE IS SKIPPED.
```

The size of the parameter change statistics was not emphasized in the above discussion. As noted earlier, these statistics show the value that the parameter is estimated to take, when subsequently freed. This information can be used to set start values in a modified model. It may also be helpful in model modification, with larger values for

fixed values of χ^2 indicating greater model misspecification. If one were to worry about larger standardized parameter changes with a significant univariate *LM* test, then the path F3→F1 with an expected value of .263 might also be investigated.

Multisample LM Test

In addition to the test on constraints described previously, EQS also allows *LM* tests on parameters when several samples are analyzed simultaneously (see Chapter 7). /LMTEST may be used in the last sample only. Parameters that appear in NOFREE, APRIORI, and HAPRIORI specifications must include the group number, as in the CONSTRAINT section, e.g.

```
/LMTEST
 APR = (1,V5,F2), (2,E5,E5);
```

When /LMTEST appears in the final group's input stream, within-group and cross-group equality constraints are tested. This test is printed out after the Goodness of Fit Summary, which appears near the end of the output stream. The constraint test is essentially the same in the multiple-group context as in its single-group version. First the title `LAGRANGE MULTIPLIER TEST (FOR RELEASING CONSTRAINTS)` is printed, and the specific equality constraints to be tested are listed. The parameters involved in these constraints have a group identification number; see /CONSTRAINTS, Cross-group Constraints, Chapter 3. Thereafter, the `UNIVARIATE TEST STATISTICS` are printed, followed by the `CUMULATIVE MULTIVARIATE STATISTICS` and the associated `UNIVARIATE INCREMENT`. Interpretation of the χ^2 statistics is handled in the same way in multisample analysis as in any single group. Thus, no example is provided here.

This concludes our overview of the features of the *LM* and *W* tests. Next we provide some technical details that will be of interest primarily to statisticians.

Technical Details

Lagrange Multiplier (*LM*) Test

We describe the basic *LM* statistics for typical covariance structure models. Appropriate redefinitions must be made in other contexts, e.g., correlation structures, case-robust methods, etc. When minimizing Q in (1.5) subject to the constraints $c(\theta) = 0$, a Lagrangian Multiplier λ_i corresponding to each of the r constraints $c_i(\theta) = 0$ is created. The vector λ of these constraint multipliers is defined by the equation

$$g + \dot{c}'\lambda = 0\,, \tag{6.1}$$

where g is $\partial Q/\partial\theta = -\dot{\sigma}'W(s-\sigma(\theta))$, $\dot{c}(\theta) = \partial c/\partial\theta'$, and λ is the vector of Lagrange Multipliers. A similar expression arises when minimizing other functions subject to constraints. Given an estimator $\hat{\theta}$ that minimizes Q and meets the constraint, the sample vector $\hat{\lambda}$ must meet (6.1), evaluated at $\hat{\theta}$. An explicit expression for $\hat{\lambda}$ is

$$\hat{\lambda} = D'\dot{\sigma}'W(s-\sigma(\hat{\theta}))\,, \tag{6.2}$$

where $D' = (\dot{c}\Delta\dot{c}')^{-1}\dot{c}\Delta$, $\Delta = (\dot{\sigma}'W\dot{\sigma})^{-1}$ as given in (5.8) or any equivalent matrix (e.g., a Hessian approximation) that is used for computing standard errors, and the matrices and vectors are evaluated at $\hat{\theta}$. Simple computation, with M defined in (5.9) and the fact that $D'\Delta^{-1}M = 0$, shows that the asymptotic covariance matrix of $\sqrt{n}\hat{\lambda}$ is given by $(D'\dot{\sigma}'WVW\dot{\sigma}D)$. Using the usual definition of a χ^2 variate in terms of quadratic forms, it follows that

$$LM = n\hat{\lambda}'(D'\dot{\sigma}'WVW\dot{\sigma}D)^{-1}\hat{\lambda} \xrightarrow{D} \chi^2_{(r)}. \qquad (6.3)$$

In practice, the matrix in parentheses is evaluated at $\hat{\theta}$. This test statistic, which follows directly from Bentler and Dijkstra (1985), is the Lagrange Multiplier or *LM* test that evaluates the statistical necessity of the r constraints. We may call it the robust *LM* test. Satorra (1989) calls it the generalized score test. It is not the classical test statistic discussed, say, by Rao (1948) and Aitchison and Silvey (1958), and introduced into structural modeling by Lee and Bentler (1980), because it can be applied under misspecification of the distribution of the variables, i.e., even when *W* is not the optimal weight matrix for the distribution involved. *At the present time, the robust LM test (6.3) is implemented in EQS only for tests on equality constraints*; in the future, we expect it to be computed automatically in EQS whenever robust asymptotic covariance matrices are specified, i.e., whenever ME= *, ROBUST. Currently, (6.3) is applied under LS estimation with $W = I$ when the distribution of the variables is normal and $V = V_N$ as given in (5.22). When *W* is a consistent estimator of V^{-1}, in particular, when the correct distribution of variables is specified, the covariance matrix of $\hat{\lambda}$ simplifies, and (6.3) simplifies to

$$LM = n\hat{\lambda}'(\dot{c}\Delta\dot{c}')\hat{\lambda}\ , \qquad (6.4)$$

where, in practice, $(\dot{c}\Delta\dot{c}')$ is evaluated at $\hat{\theta}$. This statistic, developed by Lee and Bentler (1980) for covariance structure analysis under normality assumptions, is available in EQS for all distributional assumptions, including elliptical, heterogeneous kurtosis, and arbitrary distributions, as well as for evaluating cross-group equality constraints in multisample models. More generally, it is available for the various methods discussed in Chapter 5. Appropriate substitution of the relevant estimator $\hat{\theta}$ and the associated Lagrange Multiplier $\hat{\lambda}$ into the equations above produces the correct statistic. Variants of (6.3)-(6.4) are obtained when the inverse expected information matrix Δ is replaced with its corresponding observed information matrix counterpart.

As described earlier in this chapter, Bentler and Chou (1986; Bentler, 1986a) developed a procedure to partition the multivariate *LM* test into a sequence of 1-df hierarchical and forward stepwise tests, to help separate the statistically important constraints from the statistically unnecessary constraints, as well as a procedure for conducting restricted tests for subsets of constraints, given that other constraints are accepted. Satorra (1989) gives explicit expressions for the noncentrality parameters involved when the null hypothesis $c(\theta) = 0$ is minimally false. Satorra's discussion of *LM* tests in relation to other test procedures for structural modeling is highly recommended reading, though technical in nature.

Simple univariate tests that are not part of a sequence of tests, are also of interest. The *LM* test for a single constraint, say, the i^{th} constraint (often the simple constraint that a given parameter equals zero), is a special case of the multivariate test. The i^{th} test corresponding to (6.3) is

$$LM_i = n\hat{\lambda}_i^2 / (D'\dot{\sigma}'WVW\dot{\sigma}D)_{ii} \xrightarrow{D} \chi^2_{(1)}, \qquad (6.5)$$

where the matrix in parentheses is evaluated at $\hat{\theta}$. This is the robust *LM* test. It is currently available only for equality constraints. The corresponding test for (6.4) that is not robust to distributional violation is

$$LM_i = n\hat{\lambda}_i^2 / [(\dot{c}\Delta\dot{c}')^{-1}]_{ii} \xrightarrow{D} \chi^2_{(1)}, \qquad (6.6)$$

evaluated at $\hat{\theta}$. This is the univariate *LM* test known as the modification index (Sörbom, 1989). In practice, large numbers of univariate *LM* tests may be scanned to suggest constraints to release, especially, to locate fixed parameters that may best be freed. In such a case the $\chi^2_{(1)}$ distribution may not describe the behavior of LM_i tests.

It is sometimes helpful to estimate the value that a parameter, fixed at a known value such as zero, might take if estimated freely. Conceptually, this problem is approached by considering the fixed parameter as in fact a free parameter, subject to the simple constraint that its value is a known constant, and asking what would happen if the constraint were released. Bentler and Dijkstra's (1985) development of linearized estimation provides an answer (see Bentler, 1995, Ch. 10, for details). Let $\hat{\theta}$ be the estimator obtained under the restricted model with $c(\hat{\theta}) = 0$, where the constraints guarantee that some elements of $\hat{\theta}$ are fixed at a known value. Let $\hat{\theta} = \theta_1$ be an initial

consistent estimator, and the linearized improved estimator as $\bar{\theta}$. Applying linearized GLS, the improvement in the estimator is given by

$$\bar{\pi} = \bar{\theta} - \theta_1 = \Delta_1 \dot{\sigma}_1' W (s - \sigma_1), \tag{6.7}$$

where the subscript indicates evaluating at the constrained estimate (see Bentler & Dijkstra, 1985, eq. 2.1.10). While all parameters may change when the constraints are released (see Bentler & Chou, 1993), the main interest will focus on those fixed parameters for which the corresponding element of θ_1 is known a priori. The relevant element of (6.7) then gives the parameter value to be expected when freeing the parameter. The effects of the multivariate parameter change $\bar{\pi}$ can also be evaluated on all of the constraint equations, to suggest which constraints might be worth releasing. Multiplying (6.7) by the derivatives of the constraint function $\dot{c}_1$ yields $\dot{c}_1(\bar{\theta} - \theta_1) = \dot{c}_1 \bar{\pi}$. But, according to Bentler and Dijkstra (1985, p. 25), asymptotically $c(\bar{\theta}) = c(\theta_1) + \dot{c}_1(\bar{\theta} - \theta_1)$. In this application, $c(\theta_1) = 0$, so that

$$c(\bar{\theta}) = \dot{c}_1 \bar{\pi}. \tag{6.8}$$

This gives the expected change for all constraint functions, and, as a consequence, the relative size of this multivariate constraint change can give a clue as to which constraints might be fruitful to release. This information can supplement the LM test on the corresponding constraint. The multivariate constraint change $c(\bar{\theta})$ was developed independently by Satorra (1989), though not through the steps given above. It will be noted that when the constraints being evaluated involve fixing a set of parameters to a constant, $\dot{c}_1$ can be permuted to the form (*I, 0*), so that for these fixed parameters (6.7) yields the same result as (6.8). The multivariate parameter change and constraint change statistics are not printed out in EQS. Only the constraint change statistic is computed, applied in univariate fashion to constraints that involve fixed parameters. In this instance, the constraint change and parameter change are identical, that is,

$$\bar{\pi}_i = c(\bar{\theta})_i = \hat{\lambda}_i / [(\dot{c} \Delta \dot{c}')^{-1}]_{ii}, \tag{6.9}$$

where, as usual all elements are evaluated at the restricted solution. A more general approach to evaluating parameter change, not yet incorporated into EQS, is given by Yuan, Marshall, and Bentler (in press).

It was pointed out in previous versions of the *Manual* that $LM_i = n\hat{\lambda}_i c(\bar{\theta})_i$ and, more generally, that $LM = n\hat{\lambda}' c(\bar{\theta})$. See also Bentler (1990b). These expressions make clear that constraints associated with small changes (in absolute value), or small Lagrange Multipliers, do not contribute to a large *LM* test statistic. Stated differently, in the multivariate test constraints associated with products $\hat{\lambda}_i c_i(\bar{\theta})$ that are large will have a big impact on the *LM* test. The relative size of these products provides an indication of the importance of the constraints. Saris, Satorra, and Sörbom (1987) emphasized studying the size of elements in $c(\bar{\theta})$ (in a univariate version) to detect model misspecifications. See also Kaplan (1988) and Luijben, Boomsma, and Molenaar (1988). The current analysis makes clear that the size of a Lagrange Multiplier $\hat{\lambda}_i$ is just as important as the size of the constraint change in determining the statistical importance of misspecifications. But when a significantly large LM_i test is associated with a huge $\hat{\lambda}_i$ but a very small $c(\bar{\theta})_i$, this may indicate primarily that there is a large power to detect a very tiny misspecification, and it may be practically unimportant to release the constraint or to free the fixed parameter. This could well be what is going on in the situation described by Browne, MacCallum, Kim, Anderson, and Glaser (2002), where a restricted model does not fit the data but the residuals indicate only tiny misspecifications. Those tiny misspecifications (e.g., missing correlated errors) may be associated with very large $\hat{\lambda}_i$ and large LM_i tests.

Wald (*W*) Test

Let $\hat{\theta}$ be an estimator for a model, and assume that the constraints c(θ) = 0 have not been imposed during the estimation. In practice, this means that c($\hat{\theta}$) ≠ 0, and it may be desirable to test the hypothesis that c(θ) = 0. In the simplest case, where the constraint functions simply select given parameters, this hypothesis evaluates whether the selected parameters are zero in the population. From Bentler and Dijkstra (1985, p. 24 last line), correcting a misprint, and the asymptotic equivalence of linearized and fully iterated estimators, we have the asymptotic equivalence

$$\sqrt{n}c(\hat{\theta}) = \dot{c}_+ \sqrt{n}(\hat{\theta} - \theta_+) + o_p(1) , \tag{6.10}$$

where "+" denotes the true values. Thus, under the null hypothesis, the large sample distribution of (6.10) is given by

$$\sqrt{n}c(\hat{\theta}) \xrightarrow{D} \mathcal{N}(0, [\dot{c}M\dot{\sigma}'WVW\dot{\sigma}M\dot{c}']_+) , \tag{6.11}$$

where *M* is given in (5.9). It follows that a quadratic form based on (6.11) can be used to test the constraints. Specifically, letting c($\hat{\theta}$) = $\hat{c}$ the Wald statistic is given by

$$W_{Wald} = n\hat{c}'(\dot{c}M\dot{\sigma}'WVW\dot{\sigma}M\dot{c}')^{-1}\hat{c} \xrightarrow{D} \chi^2_{(r)} , \tag{6.12}$$

where the matrix in parentheses is evaluated at $\hat{\theta}$. (Note: here we write W_{Wald} so as not to confuse the *W* test with the weight matrix *W*.) In principle, the test given in (6.12) does not depend on the distribution of variables when *V* is estimated in a distribution-free way. The test is called the generalized Wald test by Satorra (1989), who provides the noncentrality parameter when the null hypothesis is minimally false. For consistency of notation in EQS, we should call this the robust Wald test. In EQS 6, it is available whenever robust statistics are computed (i.e., with METHOD=*, ROBUST. This also is the computation used for least squares under a specified distribution.

When the weight matrix *W* is chosen optimally, $W = V^{-1}$ in the typical case asymptotically, and the covariance matrix in (6.11) simplifies. As a result

$$W_{Wald} = n\hat{c}'(\dot{c}\Delta\dot{c}')^{-1}\hat{c} \xrightarrow{D} \chi^2_{(r)} , \tag{6.13}$$

where, in practice, the matrix in parentheses is evaluated at the unrestricted estimate $\hat{\theta}$. This test is also used when $\hat{\theta}$ is based on linearized estimation, or 1-iteration updated estimation (see Bentler & Dijkstra, eq. 2.2.3). See also Lee (1985a), and Satorra (1989). When requested, EQS also makes use of the observed information matrix instead of the expected information matrix when defining Δ for use in (6.12) and (6.13).

As was noted above, Bentler and Chou (1986; Bentler, 1986a) developed a procedure to partition the multivariate *W* test into a sequence of 1-df hierarchical and backward stepwise tests, to help determine which constraints might be added to a model without significant loss of overall fit. In EQS 6, this partitioning is available for both (6.12) and (6.13).

The *W* test also exists in univariate versions. The *W* test for a single constraint, say, the i^{th} constraint, under robust statistics is given by

$$W_i = n\hat{c}_i^2 / (\dot{c}M\dot{\sigma}'WVW\dot{\sigma}M\dot{c}')_{ii} \xrightarrow{D} \chi^2_{(1)} , \tag{6.14}$$

and otherwise by

$$W_i = n\hat{c}_i^2 / (\dot{c}\Delta\dot{c}')_{ii} \xrightarrow{D} \chi^2_{(1)} . \tag{6.15}$$

As usual, the matrices in brackets are evaluated at the unconstrained estimate $\hat{\theta}$, and the χ^2 distribution is appropriate if the tests are a priori. The W_i tests simplify if the constraint is a simple test on whether a free parameter is zero in the population. In that case, (6.15) becomes $n\hat{\theta}_i^2 / \Delta_{ii}$ i.e., it is the square of an asymptotic *z*-test for evaluating the significance of a single parameter.

7. MULTISAMPLE COVARIANCE AND CORRELATION STRUCTURES

In the typical application of structural modeling it is presumed that all the individuals whose data are being analyzed represent a random sample of observations from a single population. This assumption implies that data from the various individuals provide comparable information about a hypothesized process represented by the model. This assumption will not always be reasonable. For example, data are frequently gathered from individuals who can be identified as belonging to certain groups, such as males and females, age cohorts, ethnic communities, geographical regions such as cities, states, or nations, and so on. In such cases it may be appropriate to inquire whether there are multiple populations rather than a single population, and multiple structural models rather than a single model. Hypotheses on multiple populations can be evaluated when data on the same variables exist in several samples. This chapter primarily addresses covariance structure models. At the end of this chapter, you will see that the principles are largely the same, or similar in obvious ways, for correlation structure models. Models that also involve means or intercepts are discussed in Chapter 9.

Multiple sample (or multiple group or multiple population) analysis can best be understood by first considering two extremes. At the one end, suppose that the populations are completely different as far as the measured variables are concerned. Then one would expect the covariance matrices in the various samples or groups to be different. For example, the correlation between a pair of variables in one sample would be different in magnitude and perhaps in sign as compared to another sample. Or, the variances of a variable may differ substantially across samples. Structural models that generate the corresponding covariance matrices would then also be completely different. In such a case there is not much reason for doing a multisample analysis: one might as well analyze each sample separately using the appropriate model for each sample. At the opposite extreme, suppose that the populations are actually indistinguishable as far as the measured variables are concerned. In that case, the same population covariance matrix would describe all populations, and different sample covariance matrices obtained from the various samples would simply be estimates of the same single-population covariance matrix. Hence, structural models evaluated on data from the different samples should describe the same population, so the models should be identical except perhaps for misspecification due to chance variations. In that case one would like to verify that there exists a single model that accurately describes each of the populations, and to obtain a single set of parameter estimates for the model. Finding this one model would be difficult if one analyzed the data from each sample separately, since optimal parameter estimates from each sample would surely not be precisely identical across samples even though the corresponding population parameters are the same. A multisample analysis analyzes data from all samples simultaneously and, in this case, would verify that a model, identical in all groups, reproduces the sample data of each group to within sampling accuracy. The goodness-of-fit χ^2 test can be used to describe the adequacy of the model.

Symbolically, assume that there are m populations. Each one has a population covariance matrix, say $\Sigma_1, \Sigma_2, \ldots, \Sigma_m$. If all of the populations are identical, $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_m$. If the populations differ, a covariance matrix from at least one group, say, the g^{th} group, Σ_g, will be different from the others. Other covariance matrices may be the same. In the extreme case, none of the covariance matrices may be the same. It is the task of multisample covariance structure analysis to evaluate such similarities and differences, using the sample covariance matrices $S_1, S_2, \ldots, S_m$ based on samples of size $N_1, N_2, \ldots, N_m$. As in the case of a single population, only the lower-triangular (non-redundant) elements of Σ_g, assembled into the vector σ_g, and of S_g, assembled into the vector s_g, are relevant to the analysis.

In EQS, of course, interest centers on the covariance matrices that are generated by linear structural models. The parameters of any such model are the coefficient matrices β and γ of the matrix equation $\eta = \beta\eta + \gamma\xi$ and the covariance matrix Φ of the independent variables; see eqs. (1.2) and (1.3) of Chapter 1. To simplify notation, all of the parameters from β, γ, and Φ are placed into the vector θ. The parameters θ of a model in a single population

give rise to the vector σ of non-redundant elements of Σ via the general form $\sigma = \sigma(\theta)$. When there are m populations, there will be m linear structural models. The g^{th} model is given by the equation $\eta_g = \beta_g\eta_g + \gamma_g\xi_g$ and the covariance matrix Φ_g. Since there are m such models, there are m vectors of parameters $\theta_1, \theta_2, \ldots, \theta_m$, which may or may not be the same; hence the corresponding covariances $\sigma_1 = \sigma(\theta_1)$, $\sigma_2 = \sigma(\theta_2)$, …, $\sigma_m = \sigma(\theta_m)$, may or may not be the same. The statistical problem is one of estimating the various parameters in each sample and evaluating whether or not the model with these parameters adequately accounts for the sample covariances in each of the samples. Formally, the simultaneous hypothesis to be tested is that

$$\sigma_1 = \sigma(\theta_1), \sigma_2 = \sigma(\theta_2), ..., \sigma_m = \sigma(\theta_m). \tag{7.1}$$

Another way to state this hypothesis is to put all population moments into a single vector, end to end, and to do the same with all free parameters of a model. Let $\sigma = (\sigma_1', \ldots, \sigma_m')'$ and similarly $\theta = (\theta_1', \theta_2', \ldots, \theta_m')'$. Then (7.1) is equivalent to the usual null hypothesis form $\sigma = \sigma(\theta)$. Of course, substantively, one is concerned with the extent to which the parameters $\theta_1, \theta_2, \ldots, \theta_m$ are the same; thus different applications of (7.1) will contain different constraints on the parameters θ_g of the g^{th} group. One variant of a more complete statement of the model includes sets of specifications of the form

$$\theta_{1(i)} = \theta_{2(j)} = ... = \theta_{m(k)}, \tag{7.2}$$

giving the constraints of certain parameters across groups, specifically, the equality of the i^{th} parameter in the first group with the j^{th} parameter in the second group, and so on. More abstractly and more generally, we can write

$$c(\theta) = 0, \tag{7.3}$$

where the constraint functions can be across-sample constraints or within-sample constraints. The hypothesis (7.1) under (7.3) is evaluated by the χ^2 goodness-of-fit statistic.

In practice, multiple sample analysis is done by fitting an ordinary EQS model in each sample, simultaneously for all groups, taking into account that some parameters are the same in each of the samples (using equality constraints across groups) while others are allowed to be different. Thus there are m models rather than one model, but there is a single goodness-of-fit χ^2 test to evaluate the joint hypothesis given in (7.1) under the particular cross-group (cross-sample) constraints (7.2) or (7.3). The hypothesis (7.1)-(7.3) has the same form whether all of the parameters of a given group, θ_g, are the same in all groups, or whether certain parameters in θ_1, θ_2, etc. are different for the various groups; the χ^2 test evaluates the particular hypothesis that is implemented in the analysis. If a model having identical parameters in all groups can fit acceptably, the resulting model covariance matrices are identical (since all of the free parameters in a given group are constrained to take on the same values in the other groups), and the samples can be treated as arising from the same population. If the models of the various groups have parameters that are different, the resulting model covariance matrices will be different and the various samples must be treated as arising from different populations. One can then conclude that there is an interaction between population membership and structural model. Of course, in the context of multisample correlation structures, certain hypotheses are not available.

Hypotheses In Multisample Analysis

The general models (7.1)-(7.3) are used in practice to evaluate several popular hypotheses about multiple populations. These hypotheses concern the invariance of key parameters across populations (see e.g., Alwin & Jackson, 1981; Cheung & Rensvold, 1999; Jöreskog, 1971a; Little, 1997, 2000; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000; Werts, Rock, Linn, & Jöreskog, 1976, 1977; Werts, Rock, & Grandy, 1979). Technical aspects are discussed by Meredith (1993), Meredith and Horn (2001), Meredith and Millsap (1992), Millsap (1997, 1998), and Muthén, 1989). While any free parameter, or any set of free parameters, can be evaluated for invariance or equality across populations, certain types of parameters tend to be evaluated together (see, e.g., Cheung & Rensvold, 2002, Table 1). As you study the possibilities, you should recognize that full invariance of all parameters, and even invariance of only key parameters, is only rarely achievable (Byrne, Shavelson, & Muthén, 1989).

Equality of Sets of Parameters of a Linear Structural Model

(1). Equal factor loadings (F → V paths). If the observed variables measure the same factors in each of the groups, the regression of the variables on these factors (the factor loadings) ought to be the same. Equality of loadings can occur for some factors, but perhaps not others. Additionally, cross-group equality in loadings can exist even when the factors have different variances or covariances in the various groups, and when unique or error variances are not the same.

(2). Equal factor variances and covariances (F ↔ F variances and covariances). If these are equal across groups while factor loadings are also equal, the factors are more specifically similar in the various groups. It is unlikely that the factor variances and covariances will be equal across groups when the factor loadings are unequal; and, if such a result were found, it would be difficult to interpret since it would be hard to argue that the factors are the same. Equality of factor variances may hold for the independent factors, while the covariances may not be equal. Furthermore, variances of the residuals (Ds) of dependent factors may remain unequal.

(3). Equal factor regression coefficients (F → F paths). If path coefficients among latent factors are the same across groups, the latent causal process being modeled is similar across groups. Such equality can occur even if variances or covariances of latent residual factors are unequal.

(4). Equal factor residual variances and covariances (D ↔ D variances and covariances). In a model with latent dependent factors, the equality of these variances and covariances across groups is a less important hypothesis to evaluate. Of course, in a standard factor analytic simultaneous equation model (FASEM), if all factor variances and covariances, factor regressions, and factor residual variances and covariances are equal, then all reproduced model factor variances and covariances will be equal as well.

(5). Equal unique or error variances and covariances (E ↔ E variances and covariances). Except in special models, equality of error variances or covariances is probably the least important hypothesis to test. Typically, it is also the last hypothesis to test in a sequence of tests on nested hypotheses.

Following upon a sequence of tests as outlined in (1)-(5), acceptance of the hypotheses in a standard model would imply that all of the parameters of the model are equal across groups. Such a test is possible directly in all models, even in unusual or nonstandard models.

Equality of All Parameters of a Model

This hypothesis is very restrictive and implies not only that all covariance matrices are equal, but also that the linear structural model that generates these matrices is identical in all respects across populations. It is unlikely to be strictly true except in very special circumstances. For example, models may be essentially equivalent, but a few nuisance parameters may differ trivially across samples.

Equality of Covariance Matrices

The hypothesis that covariance matrices Σ_g are identical may be true even when the underlying structural model is unknown. For example, the covariance matrices may be equal but no factor analysis model may be found to model these matrices. Exact equality of all Σ_g is hard to verify in large samples.

Equality of Correlation Matrices

This hypothesis allows covariance matrices Σ_g to differ between groups, as long as the underlying correlation matrices P_g of the measured variables are the same. Then the covariance matrices are unequal only because variables need not have the same variances in all the groups.

Tests of equality of correlation matrices can be set up directly using the correlation structure option in EQS 6, with ANALYSIS=CORRELATIONS. More will be said on this below. In addition, covariance structure analysis can be used with a special setup. This is useful if you are also interested to see what variance differences may exist across groups. This can be done as follows, using a two-group example. Let y_1 be the vector of observed (V) variables in the first group, and let the population covariance be $E(y_1 - \mu_1)(y_1 - \mu_1)' = \Sigma_1$. The population correlation matrix is P_1, and hence $\Sigma_1 = D_1 P_1 D_1$, where D_1 is the diagonal matrix of standard deviations of the variables. Thus if we take $y_1 = D_1 x_1$, it is apparent that the covariance matrix of x_1 is P_1. This transformed model can be analyzed by taking the y_1 variables as V variables, and the x_1 variables as F variables. Thus for the first V, there is an equation of the form V1 = *F1; without an error term, and there will be as many such equations as variables. The free parameter * in each equation is, in fact, just the standard deviation of the V variable involved, i.e., an element of D_1. Each of these standard deviations will be a free parameter since its value is unknown in the sample. The covariance matrix of the F variables is given by P_1, but since this is a correlation matrix, the variance of each F variable needs to be fixed at one. The covariances of all F variables are free parameters, and, corresponding to the off-diagonal elements of P_1, they are correlations between the measured variables. The same type of analysis and setup is done in the second group, where $\Sigma_2 = D_2 P_2 D_2$. To evaluate the hypothesis that $P_1 = P_2$, i.e., that the correlation matrices are equal, cross-group constraints will have to be made on the covariances of the F variables (which are in fact correlations because of the variance constraint). In the most general version of this setup, the diagonal matrices D_1 and D_2, i.e., the regression coefficients V ← F, are not constrained to be equal across groups. If this constraint were also made, the covariance matrices Σ_1 and Σ_2 would be forced to be equal as well. Thus a stronger hypothesis would be tested. In between these endpoints is the situation where some elements of D_1 equal some elements of D_2, but others are unequal.

Statistical Considerations

The original development of multiple population covariance structure analysis was based on the assumption that the observed variables are multivariate normally distributed in all groups (Jöreskog, 1971a; Lee & Tsui, 1982). This is a very restrictive assumption. Populations may differ not only in terms of the structural models that generate their covariance matrices, but also because the distribution of the variables in the models, and hence the distribution of the measured variables, differ across populations. As a consequence, EQS has implemented and even extended a very general theory that allows different populations to have different distributions of variables (Bentler, Lee, & Weng, 1987). One population may be normal, another elliptical, another heterogeneous kurtose, and still another arbitrarily distributed. All may be taken to be normal, or all may be taken to be arbitrary in distribution. While we do not recommend unrationalized use of unusual combinations of fitting functions, this can be done in EQS. One normal population may involve estimation by GLS, while another normal population may involve estimation by ML, and a third population may be estimated by AGLS. The overall χ^2 test takes into account the appropriate function and distributional theory for each group. Good recent discussions of certain aspects of multisample statistics are given in Neale (2000), Satorra (2001, 2002, 2003), and Wothke (2000). Yuan and Bentler (2001b) provide a wide-ranging overview that covers some of the newly available methods in EQS 6.

As noted earlier, in multisample analysis there are m sample covariance matrices $S_1, S_2, \ldots, S_m$ and m corresponding model matrices $\Sigma_1, \Sigma_2, \ldots, \Sigma_m$. The nonredundant elements of these matrices are collected into the sample and model vectors s_g and σ_g , one for each of g groups. The multisample χ^2 test is most easily understood in the context of generalized least squares estimation for covariance structures. In a single group, say the g^{th}, the function paralleling eq. (1.5)

$$Q_g = (s_g - \sigma_g(\theta_g))' W_g (s_g - \sigma_g(\theta_g)) \qquad (7.4)$$

is to be minimized with respect to θ in the model $\sigma_g = \sigma(\theta_g)$. At the minimum, the estimator $\hat{\theta}$ yields the function $n_g \hat{Q}_g = n_g (s_g - \hat{\sigma}_g)' W_g (s_g - \hat{\sigma}_g)$ which in large samples, under the model and assumed distribution, is distributed as a χ^2 variate with p* - q + r degrees of freedom, where p* is the number of sample variances and covariances, q is the

number of free parameters to be estimated, r is the number of equality restrictions, and $N_g = (n_g + 1)$ is the sample size. When there are m samples to be analyzed, this basic setup is extended, along with appropriate constraints. Now there are m vectors of parameters $\theta_1, \theta_2, \ldots, \theta_m$ of the separate models (7.1) $\sigma_1 = \sigma(\theta_1)$, $\sigma_2 = \sigma(\theta_2)$, …, $\sigma_m = \sigma(\theta_m)$. As a result, there are m functions such as (7.4) that must be minimized simultaneously. If the model and the assumed distributions for the variables are correct, the test statistic

$$T = n_1\hat{Q}_1 + n_2\hat{Q}_2 + \ldots + n_m\hat{Q}_m \qquad (7.5)$$

is distributed as an asymptotic χ^2 variate with p* - q + r degrees of freedom as before, but now p* is the number of sample variances and covariances in all groups (usually, m times the number in a single group), q is the total number of free parameters in all groups, and r is the total number of within-group and between-group equality restrictions. While (7.4)-(7.5) are stated in terms of GLS estimation, if Q_g in (7.4) is replaced by the single sample ML function F given in (5.13), the multisample ML test statistic is also given as in (7.5) with Q_g replaced by F_g. In all cases, standard error estimates are also computed as an extension of the one-sample approach. If there are distributional violations, as in Chapter 5, corrections to tests and standard errors should be made. More information is given in the last section of this chapter.

For a given lack of fit $\hat{Q}_g$ in a single group, it is apparent from (7.5) that larger samples influence the overall test T more than smaller samples. It is also apparent that if $n_g\hat{Q}_g$ is especially large in any one sample, it is quite likely that the goodness-of-fit statistic T (7.5) would be large overall, and the null hypothesis (7.1)-(7.3) most likely would be rejected. Thus even minor model misspecifications, especially in larger samples, translate into large values of T, i.e., into large χ^2 statistics. In one sense, this is desirable as it implies that the test has power to detect model misspecification. Given how hard it can be to fit a model in a single group, since multiple group analysis consists of several such models as well as additional cross-group constraints that make model-fitting harder yet, this difficulty should come as no surprise. Stating this to emphasize the practical, you should not expect a multisample analysis to yield an acceptable model fit unless you are quite sure ahead of time that your model fits well in one or more of your separate samples.

The calculation of standard errors for free parameters is similar to single-group analysis. In particular, the basic formula for the covariance matrix of the estimator $\hat{\theta}$ has a divisor n_g in it. Since different groups may have different sample sizes, the standard error estimates will generally be smaller in larger samples. However, a parameter estimate that is constrained to be equal across groups will have a standard error estimate that is the same in all groups.

The function Q_g in (7.4) is written in a general form. W_g is a weight matrix appropriate to the particular group and estimation method. The weight matrix that gives AGLS, the arbitrary distribution estimators, makes the fewest assumptions about the distribution of variables. In practice, of course, it becomes difficult to implement the distribution-free theory when the number of variables gets large, and when there are more than a few groups. *You will always be better off selecting a more restricted distributional theory if it is appropriate.* As you know, elliptical and HK methods, also available in the multisample context, do not require substantially more computation than normal theory methods. Robust corrections, including Satorra-Bentler and residual-based statistics and sandwich estimators of standard errors, are available to handle distributional misspecification if necessary, and will be a better choice than AGLS. But the computational demands remain heavy.

Since multisample covariance structure analysis involves parameter estimation methods within each group that are available in the one-group context, the principles associated with estimation in a single group carry over directly to the multiple group case. Thus sample covariance matrices (computable from raw scores) typically are the data to be analyzed. The considerations discussed in the introductory chapters regarding scaling of the variables to achieve approximately similar variances are equally appropriate in this context. Of course, the same scaling must be used in all the samples if cross-sample constraints are to be meaningful. However, for most models, it is permissible in practice for the variables in one sample to be rescaled so that the resulting covariance matrix is close to (or, cheating somewhat, identical to) a correlation matrix, providing that the variables in all other samples are rescaled by the scaling constants from the reference sample. Such rescaling must be done prior to analysis with EQS.

The practical goodness-of-fit indices (see Chapter 14) typically available in a single sample have been generalized to multisample analysis. In most cases, the generalization to multiple samples is direct. For the record, EQS 6 does not – at least in its initial release -- use Steiger's (1998, 2000) multiple sample version of the RMSEA index since it no longer adheres to the basic definition of the original RMSEA fit index as discrepancy per degree of freedom. Fit indices such as the Bentler-Bonett (1980) NFI and NNFI indices and Bentler's (1990a) comparative fit index CFI rely on a null or baseline model. The null model used in multisample analysis is the same one in all groups, namely, that of uncorrelated variables. No cross-group restrictions are imposed in this baseline model. Not much is known about the performance of fit indices in multisample contexts. Cheung and Rensvold (2002) studied 20 fit indices with regard to their ability to reflect changes in fit when measurement invariance is added to a model, and found that CFI was one of the best indices for this purpose.

Lagrange Multiplier and Wald Tests

The Lagrange Multiplier test, or *LM* test, is a valuable test for evaluating whether restrictions on a model are appropriate or not. Chapter 6 provided an overview of the *LM* test, but in that context the emphasis was on applications to single-sample analysis. All of the functions of the one sample test are available in the multiple sample context. In addition, the multi-sample *LM* test evaluates cross-group constraints on the equality of parameters. That is, cross-group constraints (7.2)-(7.3) are specifically evaluated whenever a multiple sample model is run. A univariate *LM* test is provided for each of the cross-group equality constraints in the model. Each test provides evidence on the null hypothesis that the constraint is true in the populations involved. If the probability value of the *LM* statistic for such a constraint is low, this indicates that the constraint is unreasonable. In this case, one might consider another analysis in which the constraint is released. If the *LM* statistic for a cross-group constraint yields a probability value that is high, the null hypothesis of the equality of some parameters across populations cannot be rejected. In that case it makes sense to maintain the constraint in further analyses.

The forward-stepping multivariate *LM* test is produced automatically and printed out. The multivariate test is important to scan since the univariate tests will in general not be independent, and actions regarding one constraint can influence another. However, you should always be careful about nonsense parameters that might creep into such a test as noted in Chapter 6.

The Wald test, or *W* test, is helpful for evaluating whether new restrictions could be added to a model without significantly degrading model fit. The features of the test described in Chapter 6 for a single sample are maintained and extended in obvious ways in the multiple sample context.

Practical Considerations

Associated with the fact that multiple group models often have large χ^2 values is the corollary that they are hard to estimate. Adequate single-group models, and good parameter start values are critical to obtaining good convergence behavior during iterations. The greater the number of samples, the more difficult it will be to find an acceptable fit to the data from all groups simultaneously.

It is unrealistic to expect a multisample analysis to work well when one of the constituent models for a particular sample is very inadequate. Assuming that the sample size is not too small, lack of fit in one sample would almost surely lead to lack of fit of the multiple population model since, as seen in eq. (7.5), the lack of fit in any sample contributes to the overall χ^2 test. In virtually all applications of multisample analysis, the model that would be run in just one sample is a less restricted version of the model that is run in several samples, since multisample analysis implies cross-group constraints and a greater number of constraints imply a worse fit. As a consequence, *if the model without the cross-group constraints fits very badly in a one-group analysis, it is quite likely that the model will not fit when part of a larger multisample analysis.* Of course, if the lack of fit is minimal in one sample, the added cross-group constraints could help to yield a multiple population model that is statistically acceptable due to

the degrees of freedom that are gained, and because one sample's excellent fit can offset, to some extent, a relatively bad fit in another sample.

EQS Job Setup

The basic idea of multisample analysis is simple. When analyzing a model for a single group, there is an input file that specifies the analysis to be performed. In a multisample analysis, there will be as many such input files as there are groups. The input file for any one group is, with a few critical exceptions to be explained shortly, identical to the file that would be created in a one-group analysis. However, all of these files are stacked end to end into a single file for job submission. A maximum of 100 groups can be analyzed simultaneously, subject to computer limitations on the amount of memory needed. For clarity, the process of creating the job setup will be described in two parts.

The Unmodified Multiple Group File

One master file that contains all of the information typically present in a single-group job, stacked end to end for all the groups, should be created first. This file, without the special information needed to specify a multiple group run, looks as follows:

```
/TITLE
…          } Input file for group 1
/END
/TITLE
…          } Input file for group 2
/END
…
…
/TITLE
…          } Input file for group m
/END
```

Within each group's input file, information is provided with the same content and format as in a one-group analysis: the covariance matrix can be within the file or elsewhere, within-group equality and inequality constraints can be specified, the method of estimation to be used in that sample is given, and so on. This information may, of course, be different in the various groups. For example, in the standard application, the same method of estimation will be used in all groups, but in specialized applications, different methods may be used in some groups. In that case the ME = xx; statement in /SPEC of the relevant group must be altered appropriately.

An important practical point is the following. Although the Windows program will create this stacked input file in a convenient way, if you use EQS 6 in other contexts this can require a lot of typing. The solution is to use a good program (ASCII) editor that permits blocks of material to be copied. Thus you can create the input file for the first group; then duplicate or copy the input stream just created as many times as there are groups; and finally, modify the records of each subsequent group in critical places in accord with the true specification for that group (i.e., where parameters differ, the input /MATRIX, constraints, etc.). Since multisample models typically contain submodels that are very similar, this approach only requires typing the changes between groups. Also, start values for parameters that will be constrained across groups will automatically be equal, as they must be.

When this stacked input file is created by block copying with an editor, it is important to remember that each group's input file must reflect the characteristics of that particular group. Thus some modifications to the block-copied file are inevitable, especially with regard to the number of CASES in the sample (specified in the /SPEC section of each

group), as well as the location of any external data file. If the data (raw data, covariance, or correlation matrix) for each group resides in a separate external file, this file is specified in the /SPECIFICATION section of each sample's input file. Since different data sets are being analyzed for the various groups, each group's DATA = 'INPUT.FIL' statement must reflect the appropriate location of that group's data. In typical mainframe environments, all of the data from all groups must be stacked sequentially, end to end, in a single external file. (This can be done also in a PC environment, in which case the filename must be given in the first group's /SPEC paragraph, and other groups must have no filename given in their /SPEC section. But this is an awkward procedure that is best avoided.)

Modifications for Multisample Analysis

1. GROUPS in the first group. The number of samples to be analyzed must be specified in the /SPECIFICATION section of the first group. Specifically, the following statement must be added in /SPEC of the input file of the first sample:

```
GROUPS = n;
```

where n is an integer between 2 and 100. (Without this statement, the program assumes n = 1, namely, that a single group analysis is being undertaken. Thus GROUPS = 1; can be used in a one-sample analysis without any effect.) The GROUPS (or GR) statement need not be included in the input files of subsequent groups.

2. /CONSTRAINTS in the last group. Cross-group equality constraints must be specified in the last group. *In many cases, these constraints may be given via the easy-to-use SET command.* The section **Cross-Group Constraints** in Chapter 3 describes how to specify such constraints, by using SET and/or writing individual constraints. The use of SET is illustrated below.

Within-group constraints are not affected by this procedure. Such constraints are stated within each group in the standard way. Within-group constraints for the last group can also be stated in the usual way within the same /CON section that specifies cross-group constraints.

3. /LMTEST in the last group. The multisample *LM* test automatically tests the appropriateness of the cross-group equality constraints. An evaluation of these constraints is highly recommended in all applications, and thus the test should be specified in the last group's input file.

EQS Output

Multisample output is basically identical to the standard output. However, just as there are several sets of input files stacked end-to-end, there are several sets of output files stacked end-to-end. The output file for each group contains the standard output from a single-group analysis, with one major exception. There is no iterative summary and no χ^2 test for each group. The iterative summary and the goodness-of-fit χ^2 test are presented at the end of the output stack, after the end of the output from the last group, in a section titled `STATISTICS FOR MULTIPLE POPULATION ANALYSIS`. This section will also indicate whether all cross-group equality constraints were correctly imposed. (Information on within-group constraints is printed within each group's output file, as usual.)

Example: Equality Of Factors

Werts, Rock, Linn, and Jöreskog (1976) presented covariance matrices on four variables (a 40-item verbal aptitude score; a separately timed 50-item verbal aptitude score; a 35-item math aptitude score; and a separately timed 25-item math aptitude score) from two random samples of individuals who took the Scholastic Aptitude Test. Since the first two variables were designed to reflect verbal aptitude, and the other two, mathematical aptitude, a two-factor hypothesis with correlated factors for these variables is rather obvious. Further, since the data were obtained from

two random samples, presumably from the same population, it is reasonable to expect that the factor loadings and factor variances and covariances (or, depending on the method of identifying the scale of factors, the factor correlations) are equal in the two samples. It is also possible that the unique or error variables have equal variances across the two samples; this hypothesis is not tested initially, but will be evaluated later. A highly edited version of the EQS output file is shown below.

```
  1  /TITLE
  2  2 GROUP EXAMPLE FROM WERTS ET AL 1976 - GROUP 1
  3  1 FACTOR MODEL WITH UNEQUAL FACTOR CORRELATIONS
  4  /SPECIFICATIONS
  5  CASES = 865; VARIABLES = 4; GROUPS = 2;  ! TWO GROUPS ARE SPECIFIED
  6  /MODEL
  7   (V1,V2) ON F1 = 5*;             ! TWO FACTOR MODEL, WITH FREE FACTOR LOADINGS
  8   (V3,V4) ON F2 = 5*;
  9   VAR F1,F2;                      ! FACTOR VARIANCES FIXED FOR IDENTIFICATION
 10   COV FF = .5*; VAR E1-E4 = 50*;
 11  /MATRIX
 12  63.382
 13  70.984 110.237
 14  41.710  52.747 60.584
 15  30.218  37.489 36.392 32.295
 16  /END

    16 CUMULATED RECORDS OF INPUT MODEL FILE WERE READ (GROUP   1)

 17  /TITLE
 18  2 GROUP EXAMPLE FROM WERTS ET AL 1976  - GROUP 2
 19  /SPECIFICATIONS
 20  CASES = 900; VARIABLES = 4;  ! DIFFERENT SAMPLE SIZE
 21  /MODEL
 22   (V1,V2) ON F1 = 5*;            ! SAME MODEL AS IN GROUP 1.  REMEMBER THAT
 23   (V3,V4) ON F2 = 5*;            ! START VALUES MUST BE THE SAME IF CROSS-
 24   VAR F1,F2;                     ! GROUP EQUALITIES ARE IMPOSED.
 25   COV FF = .5*; VAR E1-E4 = 50*;
 26  /MATRIX
 27  67.898
 28  72.301 107.330
 29  40.549  55.347 63.203
 30  28.976  38.896 39.261 35.403
 31  /CONSTRAINTS                   ! BETWEEN GROUP CONSTRAINTS GIVEN HERE
 32  (1,F2,F1)=(2,F2,F1);           ! FACTOR CORRELATION  EQUAL ACROSS GROUPS.
 33  SET = GVF;                     ! EACH FACTOR LOADING EQUAL ACROSS GROUPS.
 34  /LMTEST                        ! SEE LM TEST, BELOW, FOR LIST OF CONSTRAINTS.
 35  /END

    35 CUMULATED RECORDS OF INPUT MODEL FILE WERE READ (GROUP   2)

MULTIPLE POPULATION ANALYSIS, INFORMATION IN GROUP  1

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

PARAMETER ESTIMATES APPEAR IN ORDER,
NO SPECIAL PROBLEMS WERE ENCOUNTERED DURING OPTIMIZATION.

                AVERAGE ABSOLUTE STANDARDIZED RESIDUALS     =        0.0193
     AVERAGE OFF-DIAGONAL ABSOLUTE STANDARDIZED RESIDUALS     =        0.0159

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

  V1   =V1  =   7.443*F1    + 1.000 E1  ! BE SURE TO COMPARE WITH VALUES IN
                 .156                   ! GROUP 2 IF EQUALITIES ARE IMPOSED.
               47.854@
  V2   =V2  =   9.682*F1    + 1.000 E2
                 .199
               48.596@
  V3   =V3  =   7.287*F2    + 1.000 E3
                 .152
               47.976@
```

```
   V4   =V4  =   5.195*F2    + 1.000 E4
                  .114
                45.481@

 VARIANCES OF INDEPENDENT VARIABLES
 ----------------------------------
 STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

                 E                                D
                ---                              ---
E1  -   V1               7.599*I                                 I
                         1.025 I                                 I
                         7.413@I                                 I
                               I                                 I
E2  -   V2              19.740*I                                 I
                         1.887 I                                 I
                        10.463@I                                 I
                               I                                 I
E3  -   V3               9.581*I                                 I
                         1.099 I                                 I
                         8.717@I                                 I
                               I                                 I
E4  -   V4               6.191*I                                 I
                          .590 I                                 I
                        10.489@I                                 I
                               I                                 I

 COVARIANCES AMONG INDEPENDENT VARIABLES
 ---------------------------------------
 STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

                 V                                F
                ---                              ---
                               I F2  -   F2                .767*I
                               I F1  -   F1                .012 I
                               I                         62.148@I
                               I                                 I

 MULTIPLE POPULATION ANALYSIS, INFORMATION IN GROUP  2

 MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

 PARAMETER ESTIMATES APPEAR IN ORDER,
 NO SPECIAL PROBLEMS WERE ENCOUNTERED DURING OPTIMIZATION.

                   AVERAGE ABSOLUTE STANDARDIZED RESIDUALS     =        0.0172
        AVERAGE OFF-DIAGONAL ABSOLUTE STANDARDIZED RESIDUALS     =        0.0138

 MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
 STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

   V1   =V1  =   7.443*F1    + 1.000 E1
                  .156
                47.854@
   V2   =V2  =   9.682*F1    + 1.000 E2
                  .199
                48.596@
   V3   =V3  =   7.287*F2    + 1.000 E3
                  .152
                47.976@
   V4   =V4  =   5.195*F2    + 1.000 E4
                  .114
                45.481@
```

```
VARIANCES OF INDEPENDENT VARIABLES
 --------------------------------
 STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

                E                                D
               ---                              ---
E1  -   V1           13.507*I                              I
              1.156 I                                      I
             11.683@I                                      I
                    I                                      I
E2  -   V2           11.479*I                              I
              1.705 I                                      I
              6.731@I                                      I
                    I                                      I
E3  -   V3            8.190*I                              I
              1.067 I                                      I
              7.674@I                                      I
                    I                                      I
E4  -   V4            7.521*I                              I
               .622 I                                      I
             12.083@I                                      I
                    I                                      I

 COVARIANCES AMONG INDEPENDENT VARIABLES
 --------------------------------------
 STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

                V                                F
               ---                              ---
                           I F2  -   F2                .767*I
                           I F1  -   F1                .012 I
                           I                         62.148@I
                           I                               I

 STATISTICS FOR MULTIPLE POPULATION ANALYSIS     ! OVERALL STATISTICS APPEAR AFTER LAST GROUP.

 ALL EQUALITY CONSTRAINTS WERE CORRECTLY IMPOSED ! IMPORTANT MESSAGE

 GOODNESS OF FIT SUMMARY FOR METHOD = ML

 INDEPENDENCE MODEL CHI-SQUARE =       5472.640 ON    12 DEGREES OF FREEDOM

 CHI-SQUARE =       10.870 BASED ON     7 DEGREES OF FREEDOM
 PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS     0.14440 ! GOOD MODEL FIT

 BENTLER-BONETT NORMED    FIT INDEX=       0.998  ! ALL FIT INDICES INDICATE
 BENTLER-BONETT NONNORMED FIT INDEX=       0.999  ! GOOD MODEL FIT.
 COMPARATIVE FIT INDEX (CFI)       =       0.999

                      ITERATIVE SUMMARY         ! ITERATIONS CONVERGED QUICKLY

                  PARAMETER
 ITERATION        ABS CHANGE         ALPHA               FUNCTION
     1            18.141560         0.50000              1.13827
     2             9.708140         1.00000              0.00888
     3             0.264657         1.00000              0.00617
     4             0.011735         1.00000              0.00617
     5             0.001032         1.00000              0.00617
     6             0.000080         1.00000              0.00617

 LAGRANGE MULTIPLIER TEST (FOR RELEASING CONSTRAINTS) ! THIS IS AN IMPORTANT TEST

 CONSTRAINTS TO BE RELEASED ARE:

        CONSTRAINTS FROM GROUP  2

        CONSTR:    1   (1,F2,F1)-(2,F2,F1)=0;
        CONSTR:    2   (1,V1,F1)-(2,V1,F1)=0; ! CONSTRAINTS 2-5 GENERATED FROM
        CONSTR:    3   (1,V2,F1)-(2,V2,F1)=0; ! SET COMMAND IN CONSTRAINT
        CONSTR:    4   (1,V3,F2)-(2,V3,F2)=0; ! SECTION.
        CONSTR:    5   (1,V4,F2)-(2,V4,F2)=0;
```

```
        UNIVARIATE TEST STATISTICS:

 NO    CONSTRAINT    CHI-SQUARE   PROBABILITY
 --    ----------    ----------   -----------
  1    CONSTR:   1        3.240       0.072   ! UNIVARIATE TESTS INDICATE THAT
  2    CONSTR:   2        2.519       0.113   ! CONSTRAINTS ARE REASONABLE,
  3    CONSTR:   3        3.158       0.076   ! SINCE EACH P EXCEEDS .05.
  4    CONSTR:   4        1.168       0.280
  5    CONSTR:   5        0.205       0.651

       CUMULATIVE MULTIVARIATE STATISTICS              UNIVARIATE INCREMENT
       ----------------------------------              --------------------

STEP  PARAMETER   CHI-SQUARE  D.F.  PROBABILITY       CHI-SQUARE  PROBABILITY
---- ----------  ----------  ----  -----------       ----------  -----------
  1  CONSTR:   1      3.240     1      0.072             3.240       0.072
  2  CONSTR:   2      6.728     2      0.035             3.488       0.062
  3  CONSTR:   3      7.366     3      0.061             0.637       0.425
  4  CONSTR:   4      8.189     4      0.085             0.824       0.364
  5  CONSTR:   5      8.627     5      0.125             0.437       0.508
! ALL INCREMENTS ARE NOT SIGNIFICANT. THE MULTIVARIATE TEST COMES TO THE SAME CONCLUSION
```

The *LM* tests generally provide very valuable information on the adequacy of constraints in a multisample model. In this case, the tests verify that the cross-group equality constraints on the factor correlation and factor loadings are all reasonable. If the model had not been statistically acceptable, i.e., the χ^2 goodness-of-fit test implied rejection of the model, lack of overall fit could possibly be due to an inappropriate cross-group constraint. The *LM* test is set up to detect such problems, as shown next.

Equal Error Variances Evaluated by /LMTEST

The model of equality of factor loadings and correlations, evaluated above, proved to be acceptable. Thus one might consider the more restricted hypothesis that adds to the above model the further constraints that each variable's error variance is equal between groups. This was done by changing the SET command, above, to SET = PEE, GVF; the job was run with the following result. Only the last segment of output is shown, consisting of the model fit and *LM* test.

```
 STATISTICS FOR MULTIPLE POPULATION ANALYSIS

 ALL EQUALITY CONSTRAINTS WERE CORRECTLY IMPOSED

 GOODNESS OF FIT SUMMARY FOR METHOD = ML                    ! MODEL MUST BE REJECTED AS TOO
                                                            ! RESTRICTED, SINCE P < .05.
 CHI-SQUARE =       34.895 BASED ON    11 DEGREES OF FREEDOM
 PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS     0.00026

 LAGRANGE MULTIPLIER TEST (FOR RELEASING CONSTRAINTS)

 CONSTRAINTS TO BE RELEASED ARE:

        CONSTRAINTS FROM GROUP  2

        CONSTR:    1   (1,F2,F1)-(2,F2,F1)=0;
        CONSTR:    2   (1,E1,E1)-(2,E1,E1)=0;
        CONSTR:    3   (1,E2,E2)-(2,E2,E2)=0;
        CONSTR:    4   (1,E3,E3)-(2,E3,E3)=0;
        CONSTR:    5   (1,E4,E4)-(2,E4,E4)=0;
        CONSTR:    6   (1,V1,F1)-(2,V1,F1)=0;
        CONSTR:    7   (1,V2,F1)-(2,V2,F1)=0;
        CONSTR:    8   (1,V3,F2)-(2,V3,F2)=0;
        CONSTR:    9   (1,V4,F2)-(2,V4,F2)=0;
```

```
UNIVARIATE TEST STATISTICS:

  NO    CONSTRAINT    CHI-SQUARE   PROBABILITY
  --    ----------    ----------   -----------
   1    CONSTR:   1        2.520        0.112
   2    CONSTR:   2        7.398        0.007   ! #2 IS PROBLEMATIC.
   3    CONSTR:   3        2.854        0.091   ! #3 CAUSES TROUBLE IN
   4    CONSTR:   4        0.041        0.840   ! MULTIVARIATE TEST, BELOW.
   5    CONSTR:   5        1.722        0.189
   6    CONSTR:   6        1.971        0.160
   7    CONSTR:   7        1.553        0.213
   8    CONSTR:   8        0.000        0.997
   9    CONSTR:   9        1.630        0.202

        CUMULATIVE MULTIVARIATE STATISTICS              UNIVARIATE INCREMENT
        ----------------------------------              --------------------

 STEP  PARAMETER   CHI-SQUARE  D.F.  PROBABILITY       CHI-SQUARE  PROBABILITY
 ---- -----------  ----------  ----  -----------       ----------  -----------
   1  CONSTR:   2       7.398     1       0.007            7.398       0.007
   2  CONSTR:   3      20.735     2       0.000           13.337       0.000
   3  CONSTR:   7      23.733     3       0.000            2.997       0.083
   4  CONSTR:   1      26.049     4       0.000            2.316       0.128
   5  CONSTR:   9      27.675     5       0.000            1.626       0.202
   6  CONSTR:   6      28.773     6       0.000            1.097       0.295
   7  CONSTR:   5      29.588     7       0.000            0.815       0.367
   8  CONSTR:   4      30.406     8       0.000            0.818       0.366
   9  CONSTR:   8      31.531     9       0.000            1.125       0.289
```

The new constraints #2-#5 handle equality of error variances. The results indicate that constraints #2 and #3, concerned with equal variances of (E1,E1) and (E2,E2) across groups, were statistically unlikely to be true in the population. The equality across groups of the variance (E3,E3) and the variance (E4,E4) could not be rejected. Thus the model was resubmitted by removing the cross-group constraints of equal error variances for E1 and E2, with the following result. Again, only the final part of the printout is shown.

```
 STATISTICS FOR MULTIPLE POPULATION ANALYSIS

 ALL EQUALITY CONSTRAINTS WERE CORRECTLY IMPOSED

 GOODNESS OF FIT SUMMARY FOR METHOD = ML ! NOW THE MODEL FITS FINE.

 CHI-SQUARE =       13.863 BASED ON     9 DEGREES OF FREEDOM
 PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS     0.12729

 LAGRANGE MULTIPLIER TEST (FOR RELEASING CONSTRAINTS)

 CONSTRAINTS TO BE RELEASED ARE:

        CONSTRAINTS FROM GROUP  2

        CONSTR:    1   (1,F2,F1)-(2,F2,F1)=0;
        CONSTR:    2   (1,E3,E3)-(2,E3,E3)=0;
        CONSTR:    3   (1,E4,E4)-(2,E4,E4)=0;
        CONSTR:    4   (1,V1,F1)-(2,V1,F1)=0;
        CONSTR:    5   (1,V2,F1)-(2,V2,F1)=0;
        CONSTR:    6   (1,V3,F2)-(2,V3,F2)=0;
        CONSTR:    7   (1,V4,F2)-(2,V4,F2)=0;
```

```
UNIVARIATE TEST STATISTICS:

  NO    CONSTRAINT     CHI-SQUARE    PROBABILITY
  --    ----------     ----------    -----------

   1    CONSTR:   1         2.980        0.084 ! NO CONSTRAINT IS SIGNIFICANT.
   2    CONSTR:   2         0.067        0.796
   3    CONSTR:   3         1.906        0.167
   4    CONSTR:   4         2.406        0.121
   5    CONSTR:   5         3.045        0.081
   6    CONSTR:   6         0.001        0.981
   7    CONSTR:   7         1.594        0.207

        CUMULATIVE MULTIVARIATE STATISTICS              UNIVARIATE INCREMENT
        ----------------------------------              --------------------

 STEP  PARAMETER    CHI-SQUARE  D.F.  PROBABILITY       CHI-SQUARE  PROBABILITY
 ----  ----------   ----------  ----  -----------       ----------  -----------
   1   CONSTR:   5      3.045     1      0.081             3.045       0.081
   2   CONSTR:   1      5.403     2      0.067             2.358       0.125
   3   CONSTR:   3      7.003     3      0.072             1.600       0.206
   4   CONSTR:   2      8.562     4      0.073             1.559       0.212
   5   CONSTR:   4     10.160     5      0.071             1.598       0.206
   6   CONSTR:   6     10.986     6      0.089             0.826       0.363
   7   CONSTR:   7     11.423     7      0.121             0.437       0.509
```

This final model is now acceptable, and there are no further indications of inappropriateness of cross-group constraints.

Since multiple group models are only minor extensions, conceptually, of single-group models, additional examples of multisample analysis are not presented.

Further Statistical Aspects

EQS 6 starts with the general approach to multisample covariance structure analysis given by Bentler, Lee, and Weng (1987), and extends it to a variety of corrected test statistics that are multisample generalizations of those given in Chapter 5 for the one sample case. The basic approach generalizes the distribution-free work of Browne (1982, 1984) and Chamberlain (1982) to multiple groups, while also permitting the use of specific distributions, e.g., normal, elliptical, heterogeneous kurtosis, in each of the samples. When a specific distributional method, such as normal theory maximum likelihood, is used under distributional violations, the resulting statistics can be corrected to take this misspecification into account.

The function to be optimized in multisample analysis can be written in the form (1.5) or (7.4), redefined to refer not to a single sample but to a set of simultaneous samples. Satorra (1993, 2000, 2001, 2002, 2003) consistently prefers this approach. Its advantage is that all the statistical theory developed for the single-group situation carries over directly to the multisample situation with a simple modification of the weight matrix involved. Collecting all parameters θ_g into the vector θ, the models σ_q into the combined model $\sigma = \sigma(\theta)$, and the sample vectors s_q into the combined vector s, the statistical problem is one of fitting $\sigma(\theta)$ to s while estimating θ subject to within-sample and between-sample constraints (7.2) or, more simply, $c(\theta) = 0$. Then W in (1.5) is a block-diagonal matrix containing diagonal blocks $(n_g/n)W_g$, where $n = N - m$ and $N=\Sigma_1^m N_g$. With this notation, the fit function (1.5) remains $Q = (s-\sigma(\theta))'W(s-\sigma(\theta))$, and its minimization yields $T = n\hat{Q}$, the same as (7.5). In the case of normal theory weight matrices, Lee and Jennrich's (1979) results on reweighted least squares (RLS) extend immediately to the multiple group case, and, at the RLS minimum $T = n\hat{Q}$, one obtains the maximum likelihood estimator $\hat{\theta}$. The ML test statistic is obtained by taking the one-group ML function (5.13) and extending it to multiple samples. Of course, the generality afforded by minimizing Q permits ML estimators in some groups, and other estimators in other groups.

As in the one sample situation, the correctness of the statistics depends on either correctly specifying the distribution of variables in each sample, or making adjustments for misspecification. As noted earlier, the weight matrices W_g used in the fit function (7.4) or in Q can be taken in EQS 6 as any of those described for a single sample in Chapter 5. These include the distribution-free weights (5.6)-(5.7), the normal theory weights (5.15) and their specialization (5.17), the elliptical weights (5.28) and their specialization (5.30), and the heterogeneous kurtosis weights (5.42) and their specialization (5.43). At this time, the general skew/kurtosis methodology with weights based on (5.47) is not implemented. With a correct choice of weights, if sample size is large enough, under the null model hypothesis the test statistic T will have the assumed χ^2 distribution, and standard error estimators follow from the general theory given in Chapter 5 with the redefined multiple sample data s and model $\sigma(\theta)$. Note in particular that the Satorra-Bentler test statistic $\bar{T} = T/k$ given in (5.51) and the various residual based test statistics T_{RES} (5.54), T_{YBRES} (5.55), and F_{RES} (5.56) become available as well. Satorra (2003) discusses $\bar{T}$, T_{RES}, and some additional tests that are not yet incorporated into EQS.

Other statistical methods and results described for the single group carry over immediately to multisample analysis. For example, Lagrange Multiplier and Wald theory as described in Chapter 6 with equations (6.1)-(6.15) follows immediately, although EQS currently only evaluates cross-group equality constraints with the *LM* test and prints no *W* test in this situation. It is possible for a sample to contain missing data, in which case the statistics of Chapter 12 also become available for that sample in the context of multisample analysis.

Multisample Correlation Structures

Multisample analysis is available for models involving ordinary Pearson correlations as well as for models requiring polyserial/polychoric correlations. Unstructured estimates of these correlations in each sample are obtained in the first step, and the structural model for these correlations is evaluated in the second step. The sampling distributions of the sample correlation coefficients described in Chapter 5 are simply extended to multiple populations using devices similar to those described earlier. Thus multiple sample correlation structure analysis is accomplished by redefining the single sample correlational methods (5.57)-(5.65) so that they apply to data in multiple samples in much the same way as noted for covariance matrices above. For example, for an appropriately defined combined sample size N, the distribution of sample correlations given in (5.57) is redefined such that r is the vector of correlations from all samples, taken end to end, ρ is the corresponding vector of population correlations, and the statistics are based on the distribution of $\sqrt{n}(r-\rho)$ under the general multisample model $\rho = \rho(\theta)$. This yields appropriately weighted correlation structure functions that are optimized to obtain parameter estimates, model tests, and standard errors.

A less abstract way to think about this is that, with m samples, there are potentially one to m different populations. If the population correlation matrices, say P_1, P_2, …, P_m are all different, there is no point to doing a multisample analysis. If all of the populations have identical correlation matrices, $P_1 = P_2 = \ldots = P_m$, and thus a common correlation matrix P, and a single correlation structure $P(\theta)$ for the common matrix, would be possible. In reality, something in-between is more likely, where some of the parameters are identical across populations but others differ. A multisample analysis is done to evaluate null hypotheses associated with models that may imply identical, to slightly varying, to severely varying, structural models. If the populations are largely similar, most of the correlation matrices may be the same, but the matrix from at least one group, say, the g^{th} group, P_g, will be different. This may be because some of the parameters in that group, say θ_g, differ from those of the other groups. The ways that correlation structures are identical or different across groups is evaluated with multisample correlation structure analysis.

The hypotheses described in association with (7.1)-(7.3) can thus be modified to apply to correlations rather than covariances. However, since sample correlation matrices R_1, R_2, …, R_m are the data that is modeled in multisample correlation structures, information on differences in variances across samples or populations cannot be evaluated. *Correlation structure analysis only makes sense if hypotheses related to variances are not important.* As a result, not all hypotheses described for covariance structure analysis are meaningful in the context of correlation structures. For example, differences in variances of factors across populations typically result in differences in variances of

variables – but as the variances of variables are not available in correlation matrices, factor variance differences are probably not meaningful. Also, if you compare factor loading (pattern) matrices across groups that have equal factor variances, you will be comparing standardized factor loadings. You should do this only if you have a good rationale for standardized invariance, as the larger literature on factor invariance (e.g., Meredith, 1993; Millsap, 1997, 1998) would not lead one to expect standardized loadings to be invariant, even if the covariance structure non-standardized loadings are invariant.

In EQS, methods used to compute "poly" correlations in one sample are repeated in each of the several samples being studied. This means that, in each group, thresholds are freely estimated and the variances of the variables are fixed to 1.0. This is not the only possible approach. An alternative approach is to force the estimated intercepts of the variables to be equal across groups before estimating unstructured polychoric and polyserial correlations. An advantage of this approach is that it allows for variance differences across samples, yielding polyserial and polychoric covariances for further structural analysis (Lee, Poon, & Bentler, 1989). This methodology is not yet implemented in EQS.

To use correlation structures in practice, you implement it the same way as multiple sample covariance structures but make some minor modifications. The EQS job setup is as described earlier, but, as with single sample correlation structures, the phrase ANALYSIS=CORRELATION must be added in the SPECIFICATION section. If there are categorical variables, these must be specified in the usual way. To handle distributional misspecification, or misspecification of the fit function, a multiple group run with ANALYSIS = CORRELATION can be used with METHOD=AGLS as well as with METHOD=ML, ROBUST. As usual, the latter provides robust corrections such as the Satorra-Bentler and residual-based test statistics and sandwich estimators of standard errors.

8. MEAN AND COVARIANCE STRUCTURES

In the typical application, the parameters of a structural equation model are the regression coefficients as well as the variances and covariances of independent variables, given in EQS as elements of the matrices on the right side of $\Sigma = G(I-B)^{-1}\Gamma\Phi\Gamma'(I-B)^{-1}'G'$ (see eq. 1.4). In such applications, the means of the measured variables are not decomposed into more basic parameters. As a consequence, the covariance structure of the variables Σ carries the critical parametric information, and the model can be estimated and tested via the sample covariance matrix *S*. This approach does not work when the population mean vector μ is also expressed in terms of parameters, given in EQS as the elements of the matrices on the right side of $\mu = G(I-B)^{-1}\Gamma\mu_{\xi}$ (see eq. 1.4). Then the sample mean vector $\bar{z}$ also carries statistical information and both *S* and $\bar{z}$ are best analyzed simultaneously. In addition, the model must be specified in a somewhat different form.

Consider the simple bivariate regression model $y = \beta x + \varepsilon$, which has one dependent variable and two independent variables (in the sense of Bentler/Weeks). In covariance structure analysis, the means of y and x are irrelevant, and, since x and ε are assumed to be uncorrelated, the coefficient β in the equation and the variances σ_x^2 of x and σ_ε^2 of ε are sufficient to describe the model. Implicitly it is assumed that all variables have zero means. Now suppose that the means are not zero. Then we might consider writing

$$y = \alpha + \beta x + \varepsilon , \tag{8.1}$$

where α is an intercept parameter. The intercept helps to define the mean of y, but it is not, in general, equal to the mean. Taking expectations of both sides, and assuming that the mean of ε is zero, we have

$$\mu_y = \alpha + \beta\mu_x , \tag{8.2}$$

where μ_y is the mean of y and μ_x is the mean of x. Thus μ_y can be expressed in terms of the model parameters α, β, and μ_x. This decomposition of the mean of the dependent variable provides an illustration of the nomenclature "structured means" as a way to designate a model in which the means of dependent variables are expressed or structured in terms of structural coefficients and the means of independent variables. As usual, moments of dependent variables (here, the mean μ_y) are not parameters, but moments of independent variables (here, μ_x) are. Thus incorporating a mean structure into the model (8.1) introduces the new parameters α and μ_x into the model. This illustration shows that models with means extend a basic idea used in covariance structure analysis.

The parameters of any linear structural model with structured means are the regression coefficients, the variances and covariances of the independent variables, the intercepts of the dependent variables, and the means of the independent variables.

Intercepts of dependent variables are parameters because, as will be seen next, they are basically regression coefficients associated with a constant "variable" that is in fact an independent variable in the sense of Bentler-Weeks. This extension of the Bentler-Weeks model was given by Bentler (1983a,b) and is used in EQS. The ideas can also be extended to provide a structure for higher-order moments such as skewness and kurtosis (Bentler, 1983a), but such generality is not needed here.

The use of mean and covariance structure analysis requires rethinking the system of linear equations that describes a model. In (8.1), there is just one equation. As usual, the equation represents a structural equation for the dependent variable. While it contains an intercept term, this intercept is nothing more than a regression coefficient on a constant "variable." That is, (8.1) can also be written as

$$y = \alpha 1 + \beta x + \varepsilon , \tag{8.3}$$

where 1 is an independent variable that takes on the same value 1.0 for all observations and hence has no variance. As a consequence, intercepts are associated with an independent variable. Thus we have the following key idea, which will be useful in all types of models with constants.

The coefficient for regression on a constant is an intercept.

This simple fact permits introducing structured means into a model by adding a constant "variable" into a path diagram. In EQS, the constant is called V999. In principle V999 could literally be a column of 1.0s in the input stream, i.e., it is a measured variable, but in practice this input is not needed since the program creates the constant by default. The designation V999 was chosen so that the constant will always be in the same position, the last, relative to the measured variables in the input file.

A standard structural equation, even with an intercept as in (8.1), is insufficient to fully describe a structured means model. It is apparent that (8.1) does not contain information on the mean of the independent variable x. One solution would be to amend standard structural modeling procedures by incorporating into the system a new type of equation (such as (8.2)) that does not describe relations between *variables*, but rather describes relations between *means.* This approach is feasible. However, another approach permits using standard EQS techniques, including the use of equations to describe relations among variables only.

> In a structured means model, each independent variable can be decomposed into two new variables: the mean, and a deviation-from-mean variable. As a consequence, the variable becomes a dependent variable in a new equation, in which the intercept is the mean of the variable.

As a result, the model will be set up in such a way that *all measured and latent variables having nonzero means will be dependent variables.* In the example, we may write

$$x = \mu_x + x_d , \tag{8.4}$$

where x_d is the deviation-from-mean variable whose mean is zero. As before, (8.4) can be written equivalently as

$$x = \mu_x 1 + x_d \tag{8.5}$$

where 1 is the constant, and μ_x is the coefficient for the regression of x on the constant variable 1. As a result, *all means in a model are treated as intercepts.* Since x is a dependent variable in (8.4) or (8.5), it has no variance as a parameter. Both 1 and x_d are independent variables, with 1 having no variance, while the variance σ_d^2 of x_d is actually the variance of x (i.e., $\sigma_x^2 = \sigma_d^2$). Of course, since 1 is a constant, it will have zero covariance with other independent variables, and x_d will have x's covariances with other variables.

The constant 1 (V999 in EQS) is an independent "variable" that has no variance, and no covariances with other variables in a model.

A diagrammatic representation of the ideas just discussed is given in Figure 8.1.

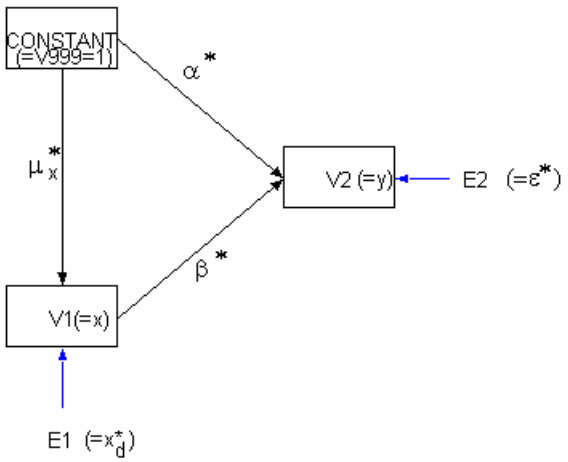


Figure 8.1

Since the complete model consists of equations (8.1) and (8.4), or (8.3) and (8.5), as well as the usual variances and covariances of the independent variables (here only σ_d^2 and σ_ε^2), the parameters associated with the original model (8.1)-(8.2) are captured correctly with this setup. Figure 8.1 shows the model of equations (8.1)-(8.5) in both the original notation and the parallel EQS notation. In the EQS notation, V999 replaces the constant 1.0, V2 replaces y, E2 replaces ε, V1 replaces x, and E1 replaces x_d. As usual, regression coefficients μ_x, α, and β have "*" attached to show that they are free parameters, and also independent variables have variances as parameters. What is new in the diagrammatic representation of a structured means model is that (1) the constant V999 does *not* have a variance as a parameter, more exactly, it has a variance that is fixed at zero, (2) coefficients for regressions on the constant, V999, are interpreted as intercepts, (3) the intercept μ_x is also the mean of x, as there is no indirect effect of V999 on x (see **Interpretation and Output** below), and (4) the *residual* of a variable that is a dependent variable solely by virtue of being regressed on the constant V999 carries the variance and covariance information of such a variable. Here, x_d (or E1) carries the variance of x (or V1). If x in (8.1) had covaried with other independent variables in a larger model, those covariances would now be specified as covariances with this residual. This is natural, because *the residual is the deviation variable, and covariances or variances are always defined on deviation variables.*

The model of Figure 8.1 has two intercept and mean parameters, or paths from V999. The data has two sample means, the means of variables x (or V1) and y (or V2). This implies that the mean structure is just identified, and that this part of the model has zero degrees of freedom and hence is not testable. A similar issue will arise in models with latent variables. To be able to test latent variable mean structures, you will have to have fewer intercept and mean parameters than you have means of observed variables.

A testable mean structure with positive degrees of freedom for the mean part of the model requires fewer intercept and mean parameters in the model than there are means of observed variables.

EQS JOB SETUP

Once the diagrammatic representation of a structured means model has been created, the model is set up in EQS in the usual way. However, because the constant V999 is not in an input data file, the program must be told to create it. In addition, the means of the measured variables must be available for analysis, and the program must be informed that the means, as well as the variances and covariances, are to be analyzed.

In the /SPECIFICATION section, the keyword ANALYSIS = MOMENT (or MOM, or MOMENTS) tells EQS to analyze the 1st moments of the data, namely the means, as well as the 2nd moments of the data about the means, namely the variances and covariances. In principle, higher-order moments could also be analyzed (Bentler, 1983a), but EQS has not implemented such a procedure. The data to be analyzed are the sample means and covariances. These can be input to EQS as: 1) raw score data, using MA = RAW in /SPEC, in which case the program will compute the summary statistics, 2) means and covariances using /MEANS to give the means and MA = COV to give the covariance matrix, or 3) means using /MEANS, and MA = COR to specify the correlation matrix as input, with /STANDARD DEVIATIONS to give the standard deviations, so that EQS will compute the covariances from the correlations and standard deviations. As usual, raw score data must be in an external file, while covariances and correlations may be in the model file or an external file. In graphical environments like Windows, it is simplest to use the raw score data file that is easily made active before setting up the model. Otherwise, the file can be specified using the DATA_FILE statement of /SPEC (except possibly in some mainframe environments, where other methods must be used).

Whenever ANALYSIS=MOMENT is specified, EQS requires that the constant variable V999 be included in the /EQUATIONS or /MODEL section of the model file. V999 will be an independent variable having zero variance.

INTERPRETATION AND OUTPUT

Since structured means models are basically variants of standard linear structural models, all of the general information found in other chapters of this manual related to EQS input and output is relevant to these models. Special attention should, however, be addressed to the means in these models.

The intercepts of the dependent variables are parameters of the model, and hence the estimated values of these parameters are available in the standard EQS output. The model-implied means of the dependent variables, however, are not parameters, and hence these estimated means must be computed separately. One way is by computing total effects using EFFECT = YES in /PRINT.

In the example of equations (8.1)-(8.2), the estimated model means must obey $\hat{\mu}_y = \hat{\alpha} + \hat{\beta}\hat{\mu}_x$, i.e., the model-implied mean of the dependent variable y is obtained as a sum of the intercept $\hat{\alpha}$ and the product $\hat{\beta}\hat{\mu}_x$. It is hoped that the model-implied mean $\hat{\mu}_y$ is close to $\bar{y}$, the sample mean. In Figure 8.1, $\hat{\alpha}$ is given by the path V999 $\rightarrow$ V2, i.e., it is a direct effect whose value can be found in the output equations. The product $\hat{\beta}\hat{\mu}_x$ is not a single path,; rather, it is the product of the coefficients in the path traced from V999 $\rightarrow$ V1 $\rightarrow$ V2. Thus it is an *indirect effect.* EQS computes such nonstandardized indirect effects, and prints out their values, as well as their estimated standard errors. As usual, the ratio of estimate/(standard error) provides a *z*-test to evaluate whether an indirect effect is significant. This is printed out as well. EQS also computes standardized indirect effects that can be more useful when the scales of the variables are arbitrary.

A *total effect* is the sum of direct and indirect effects. The model-implied mean $\hat{\mu}_y$ corresponding to Eq. (8.2) is thus $\hat{\alpha} + \hat{\beta}\hat{\mu}_x$ as can be seen from Fig. 8.1. This is printed out under nonstandardized total effects. The following important principle summarizes these ideas.

An intercept of a variable is its model-implied mean whenever there are no indirect effects of the constant (V999) on the variable. When there are such indirect effects, the total effect is the model-implied mean.

The mean of an independent variable is a parameter of the model. In EQS, however, all independent variables have been transformed into dependent variables by the addition of new equations; see eqs. (8.4)-(8.5). The means of such "independent, but set-up as dependent" variables remain as parameters of the model. Specifically, they are intercepts that are found in the equations as coefficients of V999 variables,; see, e.g., the V999 $\rightarrow$ V1 path in Figure

8.1. Since these variables are not affected indirectly by V999, the intercepts are the means. Thus they appear in the standard EQS output.

The model-implied means of V and F variables, whether dependent or independent variables, can also be obtained by the COVARIANCE = YES statement of the /PRINT paragraph. The covariance/mean model matrix

	V	F
V	$\hat{\Sigma}_{VV}$	$\hat{\Sigma}_{VF}$
F	$\hat{\Sigma}_{FV}$	$\hat{\Sigma}_{FF}$

is obtained by this statement. This matrix contains variances and covariances, printed in matrix form. Additionally, *if the constant V999 is in the model (as it will be for structured means models), the corresponding row and column contains means and not covariances.* Specifically, in this output $\hat{\Sigma}_{VV}$ is the reproduced covariance/mean matrix of all V variables, $\hat{\Sigma}_{FF}$ is the reproduced covariance matrix of all F variables, and $\hat{\Sigma}_{FV}$ is the reproduced cross-covariance/mean matrix of F and V variables. Within each segment of V and F variables, the variables are ordered sequentially by number, so reproduced model values for the intercept V999 are given in the last row of $\hat{\Sigma}_{VV}$. These values represent $\hat{\mu}'$, the vector of means for all the V variables. Since $\hat{\Sigma}_{VF}$ and $\hat{\Sigma}_{FV}$ contain the same information, only $\hat{\Sigma}_{FV}$ is printed. The final column of $\hat{\Sigma}_{FV}$ contains $\hat{\mu}_F$, the vector of means for all the F variables.

A simpler way can be used to interpret the model-implied matrix. The general rule for interpreting elements of the matrix is straightforward: Any single element in the reproduced model matrix indexes a pair of variables, say (V999, V7), (V3, F2), etc. If one of the pair of variables is V999, the corresponding element is the mean of the other variable. Thus the entry corresponding to (V999, V7) is the mean of V7, and the entry corresponding to (F3, V999) is the mean of F3. If neither variable is V999, the entry is a variance or covariance in accord with the usual convention: (F6, F6) is the variance of F6, (F1,V17) is the covariance of F1 and V17. Of course, if V999 is not included in the model, the reproduced model matrix is just a variance-covariance matrix. The standardized model matrix, i.e. the reproduced correlation matrix, is also available via /PRINT, but this matrix is not of special relevance here.

A SIMPLE REGRESSION EXAMPLE

Consider the following data on intellectual performance at four age points.

V1	V2
6.066	21.074
6.946	25.819
8.800	35.267
10.788	46.569

Variable V1 represents the four ages, and V2 represents four scores from an intelligence test. We are interested in the regression of intelligence on age, and want to include the means as part of the analysis. The conceptual model is that given in Eqs. (8.1)-(8.2), as represented in Figure 8.1 and set up in EQS using Eqs. (8.3) and (8.5). It is apparent from the figure that there are 5 parameters to be estimated. As data to be modeled, there are the two means of V1 and V2, the two variances of V1 and V2 and their covariance. Thus there are 5 data points to be modeled. The number of degrees of freedom is 5 - 5 = 0, i.e., the model is saturated and one should expect that the model will perfectly fit the data. It can also be calculated that the sample correlation between V1 and V2 across the 4 observations is .9995, so that the variables are almost perfectly linearly related, and thus one might expect some

computational problems to arise due to near-multicollinearity of these variables. The source of these data is given at the end of this chapter.

The output from an EQS run, heavily edited, is as follows.

```
     1  /TITLE
     2   REGRESSION WITH MEAN
     3  /SPECIFICATIONS
     4   CASES = 4; VARIABLES = 2; MATRIX = RAW;
     5   DATA = 'AGE.DAT';         ! THIS IS WHERE THE DATA RESIDES
     6   ANALYSIS = MOMENT;        ! SPECIFIES MEAN AND COVARIANCE STRUCTURE
     7  /EQUATIONS
     8   V1 =   8*V999 + E1;       ! EXCEPT FOR CONSTANT, V1 = E1
     9   V2 = -10*V999 + 5*V1 + E2;
    10  /VARIANCES
    11   E1 = 5*; E2 = 1*;
    12  /PRINT
    13   EFFECT = YES;             ! ONE WAY TO COMPUTE MEAN OF V2
    14   COVARIANCE = YES;         ! ANOTHER WAY TO COMPUTE MEAN OF V2
    15  /END

  SAMPLE STATISTICS BASED ON COMPLETE CASES

                          UNIVARIATE STATISTICS
                          ---------------------
   VARIABLE            V1         V2         V999

   MEAN               8.1500    32.1823     1.0000 ! THE MEAN OF THE CONSTANT IS 1
   SKEWNESS (G1)      0.3278     0.3717     0.0000
   KURTOSIS (G2)     -1.3966    -1.3469     0.0000
   STANDARD DEV.      2.0956    11.2602     0.0000

  MATRIX CONTAINS SPECIAL VARIABLE V999, THE UNIT CONSTANT
  COVARIANCE MATRIX IS IN UPPER TRIANGLE; MEANS ARE IN BOTTOM ROW OF MATRIX
  COVARIANCE/MEAN MATRIX TO BE ANALYZED:
     2 VARIABLES (SELECTED FROM   2 VARIABLES), BASED ON     4 CASES.

                         V1         V2         V999
                        V  1       V  2       V999
       V1    V  1       4.391                      ! THE USUAL COVARIANCE MATRIX IN 1ST TWO ROWS
       V2    V  2      23.589    126.793
       V999  V999       8.150     32.182      1.000 ! SAMPLE MEANS (SEE SAMPLE STATISTICS ABOVE)
       V1    V  1       4.391                          ! THE USUAL COVARIANCE MATRIX IN 1ST TWO
ROWS
       V2    V  2      23.589    126.793
       V999  V999        8.150      32.182       1.000    ! SAMPLE MEANS (SEE SAMPLE STATISTICS,
ABOVE)

  MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

  PARAMETER ESTIMATES APPEAR IN ORDER,
  NO SPECIAL PROBLEMS WERE ENCOUNTERED DURING OPTIMIZATION.
  ! LUCKY, IN VIEW OF HIGH CORRELATION BETWEEN V1 AND V2.

  RESIDUAL COVARIANCE/MEAN MATRIX  (S-SIGMA) :

                         V1         V2         V999
                        V  1       V  2       V999
       V1    V  1       0.000                      ! COVARIANCES AND MEANS ARE REPRODUCED
       V2    V  2       0.000      0.000           ! PERFECTLY WITH NO ERROR
       V999  V999       0.000      0.000      0.000
```

```
 MODEL COVARIANCE MATRIX FOR MEASURED AND LATENT VARIABLES

                        V1         V2         V999  ! THE MODEL COVARIANCES AND MODEL MEANS
                       V  1       V  2       V999   ! PEFECTLY REPRODUCE THE SAMPLE
       V1    V  1       4.391                       ! COVARIANCE/MEAN MATRIX.
       V2    V  2      23.589    126.793
       V999  V999       8.150     32.182      1.000  ! V999,V2 IS PREDICTED MEAN OF V2.

  CHI-SQUARE =        0.000 BASED ON     0 DEGREES OF FREEDOM   ! SATURATED MODEL

  MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
  STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

    V1   =V1  =    8.150*V999  +  1.000 E1  ! THE COEFFICIENTS OF V999 ARE INTER-
                   1.210                    ! CEPTS.  BOTH INTERCEPTS ARE
                   6.736@                   ! STATISTICALLY SIGNIFICANT.

    V2   =V2  =    5.372*V1    - 11.597*V999  +  1.000 E2
                    .076           .643
                  70.323@       -18.042@

  VARIANCES OF INDEPENDENT VARIABLES        ! NO VARIANCE FOR THE CONSTANT, V999.
  ---------------------------------
  STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

                  E                                D
                 ---                              ---
 E1  -   V1               4.391*I                                 I
                          3.586 I                                 I
                          1.225 I                                 I
                                I                                 I
 E2  -   V2                .077*I                                 I
                           .063 I                                 I
                          1.225 I                                 I
                                I                                 I

  DECOMPOSITION OF EFFECTS WITH NONSTANDARDIZED VALUES

  PARAMETER TOTAL EFFECTS ! EQUATION-LIKE FORMAT IS FOR SUMMARY PURPOSES
  ----------------------
    V1   =V1  =    8.150*V999  +  1.000 E1

    V2   =V2  =    5.372*V1    + 32.182*V999  +  5.372 E1    +  1.000 E2
                    .076          6.501           .076
                  70.323          4.950         70.323
! V2,V999 COEFFICIENT IS THE REPRODUCED MEAN OF V2, CALCULATED AS THE SUM OF
! THE DIRECT EFFECT ABOVE, AND THE INDIRECT EFFECT BELOW (-11.597 AND 43.780).

  PARAMETER INDIRECT EFFECTS ! INDIRECT EFFECT OF V999 ON V2 IS SIGNIFICANT
  -------------------------
    V2   =V2  =   43.780*V999  +  5.372 E1
                   6.529           .076
                   6.706         70.323

  STANDARDIZED SOLUTION:                                          R-SQUARED

    V1   =V1  =     .000*V999  +  1.000 E1                             .000
    V2   =V2  =    1.000*V1    +   .000*V999  +   .025 E2              .999

! INTERCEPTS ARE ZERO IN A STANDARDIZED SOLUTION
! V2 IS ESSENTIALLY PERFECTLY PREDICTED FROM V1 BECAUSE R-SQUARED IS CLOSE TO 1.0
```

This concludes the simple regression example.

SOME HISTORY AND CAUTIONS

Structured means models were introduced into the literature primarily by Dag Sörbom (1974) about 30 years ago. Although in principle mean structures could be considered for any SEM model (e.g., Oort, 1999; Raykov, 2001a), in practice today they are applied mainly to modeling changes across time (e.g., Dolan, Molenaar, & Boomsma, 1991; Duncan, Duncan, Strycker, Li, & Alpert, 1999; Meredith & Tisak, 1990). The growth curve aspect is discussed further in Chapter 10. In addition, as shown by Sörbom (1978, 1982), structural modeling with means can be very informative in multiple samples (e.g., Little, 1997; Whiteside-Mansell, Bradley, Little, Corwyn, & Spiker, 2001). The multiple group mean and covariance structure aspect is discussed further in Chapter 9. A further area where mean structures are important is in multilevel modeling (e.g., Bentler & Liang, 2003). This aspect is covered in Chapter 11. Models with missing data also may have mean structure aspects, as discussed in Chapter 12. A final area of application that we note here, but do not explore further, involves the analysis of experimental data. Although this has a long history (e.g., Bagozzi, 1977; Bagozzi & Yi, 1989; Kühnel, 1988), only recently has a careful technical development been accomplished (Kano, 2001).

In the analysis of variance and similar classical statistical methods, the primary focus of the analysis is on the means; the variances and covariances are only incidental to the methodology. Structural equation modeling can be used to model means alone, that is, you could consider a mean structure $\mu = \mu(\theta)$ without any further structure on covariances. For an example, see **Mean Structure with Free Covariances**, below. Of course, such mean structures cannot have latent variables, as these are generated only with covariance structures. In reality, structured means models will usually be considered simultaneously with a covariance structure model of some sort, i.e., $\sigma = \sigma(\theta)$, or specifically in EQS, $\Sigma = G(I-B)^{-1}\Gamma\Phi\Gamma'(I-B)^{-1}'G'$. The accompanying mean structure will be of the standard form given by $\mu = G(I-B)^{-1}\Gamma\mu_{\xi}$. It is important to note that in this formulation, the regression coefficients in *B* and Γ are involved in both the mean and covariance structure, that is, $G(I-B)^{-1}\Gamma$ is common to both structures. This means that estimates of *B* and Γ have to reproduce the sample means $\bar{z}$ as well as the covariance matrix *S*. Hence estimates that might be optimal for the covariance structure alone need to be adjusted so that they also are good for the mean structure, leading to a consequent degradation in fit of the covariance structure. So, if you have done an initial covariance structure model and it has a marginal or unacceptable fit, adding the mean structure will have the effect that the fit of the covariance structure will be even worse. Thus, in practice, it may be worthwhile to verify that your covariance structure model is not too misspecified before adding a mean structure.

The overall model χ^2 statistic in mean and covariance structures tests both structures simultaneously. While a large χ^2 implies rejection of the model, this could be due to misspecification of the structured means part of the model, the covariance structure, or both. Analysis of the residuals and *LM* tests will help you locate the nature of the misspecification.

While it is usually good practice to have a decent covariance structure before considering the addition of a mean structure, there are special circumstances where valid inferences about a mean structure can be made even though the covariance structure is misspecified. Yuan and Bentler (2002d) have shown that in linear growth curve models, nonzero intercept and slope parameters can be estimated consistently even when the covariance structure is misspecified. They also showed that mean structures can be estimated consistently with misspecified covariance structures under null conditions in multiple group models.

When structured means models are used with latent variables, some care must be taken to ensure that the model is identified. In general, the part of the model excluding V999 that is an ordinary latent variable model should be identified using the usual criteria for covariance structure analysis. In addition, care must be taken that the number of intercepts being estimated is less than the number of sample means. It is possible, of course, to set up a model in which the number of intercepts and sample means is equal, so that V999 imposes no special restrictions on the model. In that case, the model means will be estimated at the sample means, and the covariances can be modeled separately. Then there is not much point to doing a structured means analysis.

An important characteristic of mean structure models is that the addition of a mean structure may allow – but will not necessarily guarantee – some parameters of the covariance structure to be estimated with greater precision (Bentler & Chou, 1989), i.e., with smaller standard errors. The parameters involved are those regression coefficients B and Γ, as noted above, which affect both the means and covariances. Estimators of these parameters common to both structures can have smaller standard errors than when these parameters are estimated in covariance structure models alone. As was first noted by Bentler and Chou (1989), this gain in efficiency can be appreciable with nonnormal data. Results on the conditions under which smaller standard errors are possible are somewhat technical (Kano, Bentler, & Mooijaart, 1993; Yung & Bentler, 1996b, 1999).

EXAMPLE: GROWTH IN WISC SCORES

An interesting application of a model with structure on the means as well as the covariances was given by McArdle and Epstein (1987), using data from R. T. Osborne and collaborators on the Wechsler Intelligence Scale for Children administered to 204 children at four times from about age 6 to 11. McArdle and Epstein re-scored the WISC variables into a maximum percent format, so that each variable mean is interpretable as the percent of items answered correctly on the average. Summary means, correlations, and standard deviations were presented by McArdle and Epstein in their Table 1. These were transformed into a "cross-products moment matrix" for analysis. They developed several models for these data, none of which was statistically acceptable. A different model is developed below for the same means and covariances. See Figure 8.2.

A one-factor model for these four variables may not be adequate. A two-factor model was hypothesized, with scores from the test at the earliest two ages as indicators of the first factor (F1), and scores at the latest two ages serving as indicators of the second factor (F2). Of course, F2 can be expected to be strongly influenced by F1, but a one-factor model implies that F1 and F2 are identical, which may not be true at these ages. The hypothesized measurement model is a standard covariance structure model, as shown in the upper part of the following figure. Note that the covariance structure has 1 df, since there are $4(5)/2 = 10$ variances and covariances, and 9 parameters.

In addition to the factor structure of the measured variables, the means of the factors and variables are of interest. The bottom part of the figure shows that the constant V999 affects V1 and the factors F1 and F2. There are 3 parameters in the mean structure. With 4 sample means for 3 variables, the mean structure has 1 df. Hence the overall model will have 2 degrees of freedom.

As usual, V999 is an independent variable, but it has zero variance. The paths from V999 are intercepts, since a regression on a constant is an intercept. Without V999, F1 would be an independent variable; hence the intercept path from V999 is also the mean of the factor F1. The other variables influenced directly by V999, namely V1 and F2, are dependent variables even without V999, so the intercepts for these variables are not sufficient to define these variables' means. The means will be obtained as the sum of the direct effects and indirect effects, namely nonstandardized total effects. They are also obtainable in the reproduced model matrix.

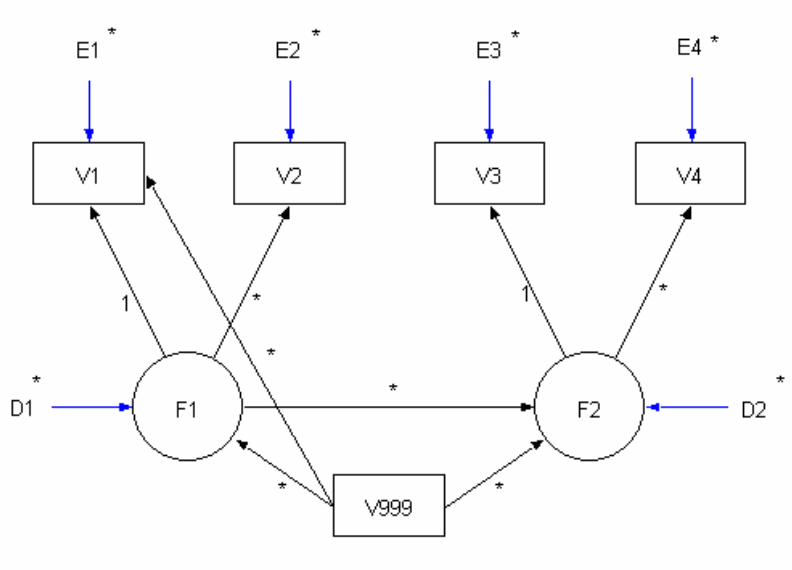


Figure 8.2

The EQS setup follows the diagram exactly. No new principles are involved. Every variable that has a one-way arrow aiming at it is a dependent variable in an equation. The remaining variables are independent variables and may have variances and covariances. All independent variables will have variances, except V999, as noted above. There are no covariances (two-way paths) in the figure. The EQS setup is shown in the first part of the edited output below. The "cross-products moment matrix" required by McArdle and Epstein (1987, Table 1C) in their use of the LISREL program is irrelevant to the setup of this model in EQS, or to the interpretation of any results. (Their matrix is not the simple covariance/mean matrix analyzed in EQS when ANAL = MOM is specified.)

```
 1  /TITLE
 2   GROWTH IN WISC SCORES
 3   OSBORNE DATA (MCARDLE & EPSTEIN, 1987, P.113)
 4  /SPECIFICATIONS
 5   CASES = 204; VAR = 4; ANALYSIS=MOMENT;  MATRIX = CORRELATIONS;
 6  /EQUATIONS
 7   V1 =   F1 +        E1 - 3*V999;
 8   V2 = 1*F1 +        E2;
 9   V3 =         F2 +  E3;
10   V4 =       1*F2 +  E4;
11   F1 =                   21*V999 + D1;
12   F2 = 1*F1 +             8*V999 + D2;
13  /VARIANCES
14   D1 = 30*; D2 = 3*;
15   E1 TO E3 = 8*; E4 = 16*;
16  /PRINT
17   EFFECT = YES;
18  /MEANS
19      18.034 25.819 35.255 46.593
20  /STANDARD DEVIATIONS
21       6.374  7.319  7.796 10.386
22  /MATRIX
23   1.000
24    .809 1.000
25    .806  .850 1.000
26    .765  .831  .867 1.000
27   /END
```

```
MATRIX CONTAINS SPECIAL VARIABLE V999, THE UNIT CONSTANT
  COVARIANCE MATRIX IS IN UPPER TRIANGLE; MEANS ARE IN BOTTOM ROW OF MATRIX
  COVARIANCE/MEAN MATRIX TO BE ANALYZED:
     4 VARIABLES (SELECTED FROM   4 VARIABLES), BASED ON   204 CASES.
! COVARIANCE MATRIX IS CREATED FROM INPUT CORRELATIONS AND STANDARD DEVIATIONS.

                      V1         V2         V3         V4         V999
                     V  1       V  2       V  3       V  4       V999
      V1    V  1     40.628
      V2    V  2     37.741     53.568
      V3    V  3     40.052     48.500     60.778
      V4    V  4     50.643     63.169     70.200    107.869
      V999  V999     18.034     25.819     35.255     46.593     1.000 ! THE USUAL MEANS

  MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

  PARAMETER ESTIMATES APPEAR IN ORDER,
  NO SPECIAL PROBLEMS WERE ENCOUNTERED DURING OPTIMIZATION.

  RESIDUAL COVARIANCE/MEAN MATRIX  (S-SIGMA) : ! RESIDUALS LOOK QUITE SMALL COMPARED TO S, ABOVE.

                      V1         V2         V3         V4         V999
                     V  1       V  2       V  3       V  4       V999
      V1    V  1     0.000
      V2    V  2     0.000      0.000
      V3    V  3     0.682      0.266      0.858
      V4    V  4    -1.343     -0.523     -0.279     -2.233
      V999  V999     0.000      0.000     -0.012      0.024      0.000

  STANDARDIZED RESIDUAL MATRIX:               ! RESIDUALS ARE QUITE SMALL IN THIS METRIC, ALSO

                      V1         V2         V3         V4         V999
                     V  1       V  2       V  3       V  4       V999
      V1    V  1     0.000
      V2    V  2     0.000      0.000
      V3    V  3     0.014      0.005      0.014
      V4    V  4    -0.020     -0.007     -0.003     -0.021
      V999  V999     0.000      0.000     -0.002      0.002      0.000

  GOODNESS OF FIT SUMMARY FOR METHOD = ML

  INDEPENDENCE MODEL CHI-SQUARE =        820.575 ON     6 DEGREES OF FREEDOM

  CHI-SQUARE =        1.465 BASED ON     2 DEGREES OF FREEDOM
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS     0.48075
  THE NORMAL THEORY RLS CHI-SQUARE FOR THIS ML SOLUTION IS           1.460.

  BENTLER-BONETT NORMED    FIT INDEX=       0.998   ! GOOD FIT BY ALL CRITERIA
  BENTLER-BONETT NONNORMED FIT INDEX=       1.0020.997   ! FIT BASED ON COVARIANCE MATRIX ONLY
  COMPARATIVE FIT INDEX (CFI)       =       1.0000.999

                        ITERATIVE SUMMARY      ! ITERATIONS CONVERGED, DESPITE BAD START VALUES

                  PARAMETER
  ITERATION       ABS CHANGE         ALPHA               FUNCTION
      1           31.079317         0.50000              6.90004
      2           13.844810         0.50000              5.47317
      3            7.562433         0.50000              2.48269
      4            6.870684         0.50000              2.17760
      5            3.926286         1.00000              1.23002
      6            2.825303         1.00000              1.08885
      7            0.970395         1.00000              0.41689
      8            2.706371         0.50000              0.21070
      9            2.046119         1.00000              0.14210
     10            0.173178         1.00000              0.01576
     11            0.371534         1.00000              0.00739
     12            0.025409         1.00000              0.00722
     13            0.002004         1.00000              0.00722
     14            0.000118         1.00000              0.00722
```

```
MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
  STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

    V1   =V1  =    1.000 F1    -  3.040*V999  +  1.000 E1
                                   1.123
                                  -2.706@

    V2   =V2  =    1.225*F1    +  1.000 E2
                     .063
                   19.325@

    V3   =V3  =    1.000 F2    +  1.000 E3

    V4   =V4  =    1.320*F2    +  1.000 E4
                     .010
                  126.650@

  CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
  STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

    F1   =F1  =   21.074*V999  +  1.000 D1
                   1.168
                  18.037@

    F2   =F2  =    1.278*F1    +  8.334*V999  +  1.000 D2
                     .065          1.214
                   19.532@         6.862@
  ! THERE IS A SIGNIFICANT GAIN IN THE MEAN OF F2, SINCE THE INTERCEPT IS SIGNIFICANT.

  VARIANCES OF INDEPENDENT VARIABLES
  ----------------------------------
  STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

                  E                                D
                 ---                              ---
 E1  -   V1               9.823*I D1  -   F1              30.805*I
                          1.223 I                          3.982 I
                          8.030@I                          7.736@I
                               I                                I
 E2  -   V2               7.329*I D2  -   F2               3.058*I
                          1.326 I                          1.458 I
                          5.526@I                          2.098@I
                               I                                I
 E3  -   V3               6.546*I                               I
                          1.172 I                               I
                          5.584@I                               I
                               I                                I
 E4  -   V4              17.037*I                               I
                          2.399 I                               I
                          7.103@I                               I
                               I                                I

  DECOMPOSITION OF EFFECTS WITH NONSTANDARDIZED VALUES

  PARAMETER TOTAL EFFECTS ! HERE IS WHERE TO FIND THE PREDICTED MEANS
  ----------------------
    V1   =V1  =    1.000 F1    + 18.034*V999  +  1.000 E1    +  1.000 D1
                                     .447
                                   40.311

    V2   =V2  =    1.225*F1    + 25.819 V999  +  1.000 E2    +  1.225 D1
                     .063          .514                          .063
                   19.325         50.262                       19.325

    V3   =V3  =    1.278 F1    +  1.000 F2    + 35.267 V999  +  1.000 E3    +  1.278 D1 +1.000 D2
                     .065                          .543                          .065
                   19.532                        64.963                        19.532
```

```
V4   =V4  =    1.688 F1    +  1.320*F2    + 46.569 V999  +  1.000 E4    +  1.688 D1 +1.320 D2
                  .087           .010           .735                          .087      .010
                19.486         126.650         63.337                       19.486   126.650

   V3   =V3  =     1.278 F1     +  1.000 F2     + 35.267 V999  +  1.000 E3     +  1.278 D1    +
1.000 D2
                   .065                           .543                          .065
                 19.532                         64.963                        19.532

   V4   =V4  =     1.688 F1     +  1.320*F2     + 46.569 V999  +  1.000 E4     +  1.688 D1    +
1.320 D2
                   .087              .010              .735                                  .087
.010
                 19.486           126.650            63.337                               19.486
126.650

   F1   =F1  =   21.074*V999  +  1.000 D1  ! THE MEAN OF F1, WHICH IS IN THE METRIC OF V1

   F2   =F2  =    1.278*F1    + 35.267*V999  +  1.278 D1    +  1.000 D2
                   .065           .543           .065
                 19.532         64.963         19.532

 PARAMETER INDIRECT EFFECTS
 -------------------------

   V1   =V1  =   21.074*V999  +  1.000 D1
                  1.168
                 18.037

   V2   =V2  =   25.819 V999  +  1.225 D1
                   .514           .063
                 50.262         19.325

   V3   =V3  =    1.278 F1    + 35.267 V999  +  1.278 D1    +  1.000 D2
                   .065           .543           .065
                 19.532         64.963         19.532

   V4   =V4  =    1.688 F1    + 46.569 V999  +  1.688 D1    +  1.320 D2
                   .087           .735           .087           .010
                 19.486         63.337         19.486        126.650

   F2   =F2  =   26.933*V999  +  1.278 D1
                  1.302           .065
                 20.687         19.532

 STANDARDIZED SOLUTION:                                                R-SQUARED

   V1   =V1  =    .871 F1    +  .000*V999  +  .492 E1                      .758
   V2   =V2  =    .929*F1    +  .370 E2                                   .863
   V3   =V3  =    .944 F2    +  .331 E3                                   .891
   V4   =V4  =    .919*F2    +  .393 E4                                   .845
   F1   =F1  =    .000*V999  + 1.000 D1                                   .000
   F2   =F2  =    .971*F1    +  .000*V999  +  .239 D2                     .943
 ! F1 = D1 WITHOUT V999, AND F2 IS HIGHLY PREDICTED BY F1.
```

The growth in WISC intelligence scores across these four ages is successfully described by the two-factor model in Figure 8.2. In contrast to the models considered by McArdle and Epstein, this model fits the data well by statistical criteria and fit indices. The results show that the variables can be well described by the simple two-factor model shown, with a substantial stability to the factors across time (a coefficient of .971 in the standardized solution, associated with an R-squared of .943). Furthermore, the means of the factors shift significantly over time. The factor mean for F2 is significantly greater than the mean of F1, i.e., there was substantial intellectual growth in the time period indicated. The means of all V variables can be expressed in terms of the means of the factors, as indirect effects. However, V1 acts differently from the other variables in this regard, needing an additional direct component not explainable by the factors in the model. Evidently the V1 mean of the WISC scores contains a specific component not mediated by the intellectual factor F1.

Although the output indicated that there were no problems during the optimization, in fact less optimal start values for this problem created some difficulty for the iterative procedure, as can be seen from the large absolute changes in parameters in the iterative summary. Such a result ought perhaps to be expected since there are only two indicators for each of the factors, and the factors are very highly correlated. However, structured means models often have greater difficulty in reaching a stable optimum as compared with covariance structure models. This is no doubt one reason for their relative unpopularity. You may want to experiment with the difficulties that non-optimal start values can have with the model of Figure 8.2 applied to these data. One thing you can do is to modify the tolerance, i.e., the criterion that specifies when different parameters become identical. Does the program converge with /TECHNICAL and TOL=0.00001?

To conclude this example, it may be interesting to consider the variable means that are implied by the factors in the model. For all variables, they are given by the indirect effects of V999 operating through the factors on the variable in question. These indirect effects appear in the printout above, but they can also be obtained by path tracing rules, for example, for V1, this is the product 21.074×1, tracing the paths V999 $\rightarrow$ F1 $\rightarrow$ V1. For the other variables V2-V4, the results are 25.819, 35.267, and 46.569. These were the values taken as the four scores on V2 in the simple regression example discussed earlier in this chapter. In fact, that example was constructed to predict the factor-based IQ scores, obtained from the model associated with Figure 8.2, from the means of the ages of the children. Those age means, taken from McArdle and Epstein (1987, p. 113), were given as V1 in the first example of this chapter. Remembering the result of the simple regression, it is apparent that in this data set there is an almost perfect linear prediction of factor-based mean intellectual performance from mean age.

MEAN STRUCTURE WITH FREE COVARIANCES

As noted previously, it is possible to do a structured means model without a covariance structure model. To do this, we treat the covariance structure as saturated, that is, allow all the covariances to be free parameters. To illustrate this, consider again the growth in WISC scores. One might hypothesize that the mean differences from occasion to occasion are equal, but it seems from the means in the data file that this is only approximately true for the last 3 occasions. Perhaps we can evaluate whether the mean difference $\mu_4 - \mu_3$ is the same as that of $\mu_3 - \mu_2$. Setting this equality and rearranging, we obtain the linear constraint $\mu_4 - 2\mu_3 + \mu_2 = 0$. To set this up in EQS, we let the means of the 4 variables be free parameters and impose one constraint on them. The key part of the model file is the following.

```
/EQUATIONS
V1 = 18*V999 + E1;
V2 = 25*V999 + E2;
V3 = 35*V999 + E3;
V4 = 45*V999 + E4;
/VARIANCES
E1 TO E4 = 80*;
/COVARIANCES
E1 TO E4 = *;
/CONSTRAINTS
(V4,V999)-2(V3,V999)+(V2,V999)=0;
```

Now each variable has been decomposed into a mean plus a deviation from mean, as shown in (8.4). The deviations from means are E variables, and they have free variances and covariances. Hence the covariance matrix of the E and V variables will be identical, and these covariances should reproduce the sample covariances. The 4 intercepts here are means. They are free parameters. Start values were chosen to obey the equality constraint. In addition to the usual EQS output, such as optimal parameter estimates and standard errors, the model test concluded

```
CHI-SQUARE =       12.689 BASED ON       1 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS      .00037
```

The results show that the proposed mean structure has to be rejected statistically, even though the largest mean residual $\bar{z}_4 - \hat{\mu}_4$ is 1.04. Mean structures can be very hard to fit, as there is usually high power to detect small misspecifications. Here, the 1df model test is equivalent to a test of the equality restriction.

TEST FOR A POTENTIAL MEAN STRUCTURE

As noted previously, the vast majority of structural equation models contain no mean structure. In fact, one way to think about the typical SEM is that the variables' population means are estimated at the sample means and are then eliminated from modeling consideration. Now suppose that you are doing only a covariance structure analysis, $\sigma = \sigma(\theta)$, but at the same time you wonder whether the structure $\mu = \mu(\theta)$ could be added to the model. The direct way to approach this is to modify your model by adding its new mean structure, estimate the resulting model, and see what you get. EQS provides another approach, based on Bentler and Yuan (2000), in which you run only the covariance structure but add a command to evaluate a possible structured means model.

Whenever a covariance structure model (ANALYSIS=COV) with latent factors (F variables) is run in EQS, and you provide raw data input or make the variables' sample means available, and you further specify FMEANS=BGLS in the PRINT section, then, in addition to the usual computations and printouts, EQS estimates factor means and computes the Bentler-Yuan (2000) modified test for a potential structured means model. As an example, EQS computes and prints

```
BENTLER-YUAN MODIFIED TEST STATISTIC FOR A POTENTIAL STRUCTURED MEANS MODEL IS:
  CHI-SQUARE =          .624 BASED ON       4 DEGREES OF FREEDOM
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS      .96032
 IF PROBABILITY IS LARGE, A STRUCTURED MEANS MODEL MAY BE APPROPRIATE.
```

In this instance, a structured means model would be appropriate, and you could follow this up by setting up the complete model with a mean structure. This test statistic and its df and probability are printed following the basic model goodness of fit. Subsequently, and immediately before the iterative summary, you also find the estimated factor means that are computed using the Bentler-Yuan best generalized least squares method:

```
FACTOR MEANS FOR A POTENTIAL STRUCTURED MEANS MODEL
(BENTLER-YUAN BEST GLS ESTIMATES)
F1         F2
.1266      .1959
```

These estimated means are based on a two-stage estimation procedure and hence will not be precisely the same as estimates you would obtain from a simultaneous mean and covariance structure analysis. But the test and estimates should give you a good idea of whether it would be worthwhile to do a simultaneous analysis.

For the technically minded, internally EQS computes the best GLS (BGLS) factor means as

$$\hat{\mu}_\xi = ([G(I-\hat{B})^{-1}\hat{\Gamma}]'\hat{\Upsilon}^{-1}[G(I-\hat{B})^{-1}\hat{\Gamma}])^{-1}[G(I-\hat{B})^{-1}\hat{\Gamma}]'\hat{\Upsilon}^{-1}\bar{z} \tag{8.6}$$

based on the covariance structure estimated parameters $\hat{B}$ and $\hat{\Gamma}$, an optimal weight matrix $\hat{\Upsilon}$ given in Bentler-Yuan's eq. (13), and the sample means $\bar{z}$. The factor mean estimator (8.6) can be thought of as an inverse of the Bentler-Weeks mean structure (1.4), or as a variant of the factor score estimator (5.79). A basic test statistic is computed (but not printed in EQS) as

$$\hat{T}_{\mu|\Sigma} = n(\bar{z} - G(I-\hat{B})^{-1}\hat{\Gamma}\hat{\mu}_\xi)'\hat{\Upsilon}^{-1}(\bar{z} - G(I-\hat{B})^{-1}\hat{\Gamma}\hat{\mu}_\xi)\,, \tag{8.7}$$

which is used to produce the final modified χ^2 statistic (Bentler-Yuan, eq. 18) that is printed out by EQS

$$\hat{T}^*_{\mu|\Sigma} = \hat{T}_{\mu|\Sigma}\,/\{1+\hat{T}_{\mu|\Sigma}\,/\,n\}\,. \tag{8.8}$$

This test statistic is of the same form as the Yuan-Bentler test statistic T_{YB} given in (5.2). Compared to (8.7), it is designed to yield improved small sample performance.

STATISTICAL BASIS

Assumptions and Use

The general model $\sigma = \sigma(\theta)$ and corresponding data vector s that are used in the general fit function (1.5) must be reconceptualized when both means and covariances are being analyzed. Since the data that are being modeled include means and covariances, both data and model vectors must take this into account. (The word "covariances" is used here to represent both variances and covariances, since a variance is a special case of a covariance.) In particular, the vector of parameters σ cannot be considered as containing only the population covariances; rather it must contain both the means and covariances. Suppose we use the designation $\sigma_1 = \mu$ for the 1st-order moments, the means, and σ_2 for the 2nd-order moments about the means, the covariances. Then $\sigma(\theta)$ really contains two parts, $\sigma_1(\theta)$ and $\sigma_2(\theta)$. That is, both population means and covariances depend on some parameter vector θ, whose elements in EQS are the elements of the matrices B, Γ, Φ and μ_ξ as shown in eq. (1.4). The corresponding sample vectors are $s_1 = \bar{z}$, the vector of sample means, and s_2, the vector of non-redundant sample covariances from the covariance matrix *S*.

Maximum likelihood is the typical estimation method used for structured means models. In EQS we minimize

$$F_{ML} = F + (s_1 - \sigma_1)'\hat{\Sigma}^{-1}(s_1 - \sigma_1)\,, \tag{8.9}$$

where *F* is the ML covariance structure fit function given in (5.13) and $\hat{\Sigma}^{-1}$ is iteratively updated. This is the same function as in ordinary covariance structure ML estimation except for the added weighted sum of squares resulting from the discrepancy between $\bar{z}$ and $\hat{\mu}$. At the minimum of (8.9), the mean and covariance structure model test statistic is obtained as

$$T_{ML} = n\hat{F}_{ML} \xrightarrow{D} \chi^2_{(p^*-q+r)}\,. \tag{8.10}$$

As a natural extension of the covariance structure test, (8.10) is typically used in mean structure models (e.g., Widaman & Thompson, 2003, eq. 6), although slight variants are possible (see below). The degrees of freedom for the test are given by the number of sample means plus the number of non-redundant sample variances and covariances, minus the number of free parameters estimated, plus the number of equality restrictions imposed during the estimation.

Several additional estimation methods available in EQS (see Chapter 5) also may be used to analyze mean and covariance structure models. Typically this requires assuring that the weight matrix used in the generalized least squares function, eq. (1.5), $Q(\theta) = (s - \sigma(\theta))'W(s - \sigma(\theta))$ is appropriately redefined. The weight matrix has to represent the inverse of the sampling covariances of s_1 and s_2, which in turn has to take into account whether sample means and sample covariances are asymptotically independent or not. As a result, overall goodness of fit will depend on a weighted function of the extent of fit of both the mean and covariance structures. It reflects the extent to which the sample mean vector $\bar{z}$ is reproduced by the estimated model mean vector $\hat{\mu}$ as well as how closely the sample covariance matrix *S* is reproduced by the estimated model covariance matrix $\hat{\Sigma}$. As a result, a model may fit badly if the means are modeled poorly, if the covariances are modeled poorly, or both.

A check on the normality of the variables to be analyzed can be made by the usual EQS diagnostics, e.g., univariate skewnesses and kurtoses, and Mardia's normalized multivariate kurtosis coefficient. There are some conditions under which the test statistic (8.10), and some parameter estimates and standard errors, are correct under

nonnormality (Satorra, 1992), but as these conditions are hard to verify and EQS provides no such diagnostics, you should probably not assume such asymptotic robustness. As usual, if the variables are not normal, you could use a distribution-free methodology (e.g. Yuan & Bentler, 1997d), but unless you have a very large sample size, you will be better off accepting the ML estimates and correcting test statistics and standard errors for nonnormality using METHOD=ML,ROBUST. For example, you can use the Satorra-Bentler corrected test statistic $\bar{T}$ or the Yuan-Bentler residual tests $T_{YB(RES)}$ or $T_{F(RES)}$ for more reliable inference about the model. These statistics were discussed in Chapter 5. In EQS, they have been extended in natural ways to mean and covariance structures. If you are suspicious about outliers and influential cases, i.e., you do not trust $\bar{z}$ and *S* as good statistical summaries of the main data cloud (see e.g., Cheng & Victoria-Feser, 2002; Poon & Poon, 2002; Yuan & Bentler, 1998bc, 2000c; Yuan, Chan & Bentler, 2000), you might want to consider the use of case-robust methods. These are described in Chapter 5, and have been implemented in EQS to be relevant to mean and covariance structures. All of the methods for nonnormal data require the use of raw data.

Technical Details

Consider the covariance/mean sample and model matrices

$$\begin{bmatrix} S & \bar{z} \\ \bar{z}' & 1 \end{bmatrix} \text{ and } \begin{bmatrix} \Sigma & \mu \\ \mu' & 1 \end{bmatrix} \tag{8.11}$$

in which *S* and $\bar{z}$ are the sample covariance matrix and mean vector, respectively, and $\Sigma = \Sigma(\theta)$ and $\mu = \mu(\theta)$, that is, both covariances and means are functions of a basic parameter vector θ. The data to be modeled are s_2, the vector of lower triangular elements of *S*, and $s_1 = \bar{z}$. The subscripts indicate the order of moments involved, s_2 referring to 2nd moments about the mean, and s_1 referring to 1st-order moments. The constant 1 in (8.11) is irrelevant to both data and model. The matrices given in (8.11) are used in EQS for models involving intercepts and means, as well as covariances. Let $s = (s_1', s_2')'$ be the data vector and $\sigma = (\sigma_1', \sigma_2')'$ be the corresponding model vector. Then, we take $\sigma = \sigma(\theta)$ as usual, and the general statistical theory summarized in Chapter 5 for estimating parameters, testing goodness of fit, and so on also applies as usual, providing the weight matrix *W* is appropriately redefined. The number of data points to be modeled, the sample elements of *s*, are now redefined as p* = p(p + 1)/2 + p, but q and r can still describe the number of free parameters and restrictions $c(\theta) = 0$, respectively. As developed by Bentler (1989), the theory specializes, depending on the sampling covariance of s_1 and s_2. See also Browne and Arminger (1995).

Dependence of Sample Covariances and Means

The asymptotic distribution of $\sqrt{n}(s_2 - \sigma_2)$ has already been described abstractly in eq. (5.5), where the covariance matrix *V* in the distribution-free case of covariance structures was given by (5.6)-(5.7). Specialized formulae were given for the normal theory V_N (5.22), elliptical V_E (5.27), and heterogeneous kurtosis V_{HK} (5.39) special cases. Redefining *V* in these equations as V_{22} to designate the second-order moments, the distribution of sample covariances in (5.5) becomes, in the new notation,

$$\sqrt{n}(s_2 - \sigma_2) \xrightarrow{D} \mathcal{N}(\delta_2, V_{22}). \tag{8.12}$$

For simplicity we suppress the obvious notation $\sigma_2 = \sigma_2(\theta)$. Although typically one assumes that $\delta_2 = 0$, here we allow some misspecification of the mean of the asymptotic distribution (see Bentler & Dijkstra, 1985 for rationale and details). Also, we assume that the probability limit of s_2 is σ_2, i.e., the sample covariance matrix is a consistent estimator of its population counterpart. The distribution of the sample mean can be specified without any distributional assumptions. The expected value and probability limit of s_1 is μ, and the asymptotic distribution of the sample mean is

$$\sqrt{N}(s_1 - \sigma_1) \xrightarrow{D} \mathcal{N}(\delta_1, \Sigma). \tag{8.13}$$

Again, we suppress the obvious notation $\sigma_1 = \sigma_1(\theta)$. When there is no misspecification $\delta_1 = 0$. Comparing (8.12) and (8.13), the sample size multiplier is not the same for covariances and means since n = N-1. They would be consistent if n were replaced by N in (8.12), i.e., if s_2 were the vector of biased rather than unbiased sample covariances. However, unbiased covariances are typically used in structural modeling. This creates a small technical problem, discussed next. In practice, this technical problem is ignored in EQS as it disappears as sample size gets large.

We may rescale the distribution of the mean in (8.13) to be consistent with the n multiplier. That is, we multiply the variable in (8.13) by $\sqrt{a_n} = \sqrt{n/N}$, yielding

$$\sqrt{n}(s_1 - \sigma_1) \xrightarrow{D} \mathcal{N}(\sqrt{a_n}\delta_1, V_{11}). \tag{8.14}$$

where

$$V_{11} = a_n \Sigma. \tag{8.15}$$

In a sense this is irrelevant as we are discussing the asymptotic distribution (8.13) as $N \to \infty$. In that case, it follows that $a_n \to 1.0$, i.e., that $V_{11} = \Sigma$. So, (8.13) or (8.14) can be used interchangeably. With $\delta = (a_n \delta_1', \delta_2')'$, the means in (8.14) and the covariances in (8.12) are in fact jointly asymptotically normally distributed (Kendall & Stuart, 1969) as

$$\sqrt{n}(s - \sigma) \xrightarrow{D} \mathcal{N}(\delta, V), \tag{8.16}$$

where *V* is the supermatrix

$$\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \tag{8.17}$$

The submatrices V_{11} and V_{22} of (8.17) have been given in (8.14)-(8.15) and (8.12). The elements of V_{12} describe the asymptotic covariances of sample means and sample covariances. The elements of V_{12} are

$$[V_{12}]_{ijk} = \sqrt{a_n} E(z_i - \mu_i)(z_j - \mu_j)(z_k - \mu_k) = \sqrt{a_n}\sigma_{ijk}. \tag{8.18}$$

This describes the sampling covariance of a sample covariance and a sample mean. Again, as $n \to \infty$, $a_n \to 1.0$ so that the constant a_n is essentially irrelevant. A consistent estimator of σ_{ijk} is the sample mean of the product of individuals' third-order deviations about the mean

$$s_{ijk} = N^{-1} \sum_1^N (z_{it} - \bar{z}_i)(z_{jt} - \bar{z}_j)(z_{kt} - \bar{z}_k). \tag{8.19}$$

A slightly adjusted divisor could be used in (8.19) to obtain an unbiased estimator, but as shown by Chang, Yung, and Bentler (1995) in the case of covariance structures, such an adjustment is too small to matter in practice. Further, in EQS we assume that sample size is large enough, and the constant a_n is simply set to 1.0.

The asymptotic distributions given here specialize to yield Browne's (1982, 1984) and Chamberlain's (1982) distribution-free covariance structure analysis when $\sigma_1 = \sigma(\theta_1)$ with $\mu = \theta_1$, and $\sigma_2 = \sigma(\theta_2)$. When both means and covariances are structured in terms of some of the same parameters in the vector θ, as shown in (8.16), both sample means and covariances are modeled. In view of the formal similarity of (8.16) and its covariance structure counterpart (5.5), it will be clear that all the statistical approaches, methods, and results described in Chapter 5 and 6 for covariance structures are directly applicable to mean and covariance structure analysis with obvious minor modifications. The test statistics, standard errors, Lagrange Multiplier and Wald tests, and so on, are computed in mean and covariance structures using the modifications described above. The function $Q = (s - \sigma(\theta))' W (s - \sigma(\theta))$

given in (1.5), for example, remains an appropriate general function to minimize subject to an appropriate choice of *W* based on *V* in (8.17) under the relevant distributional assumption.

Independence of Sample Covariances and Means

If σ_{ijk} = 0 in (8.18), then *V* is a block diagonal matrix since

$$V_{12} = 0. \tag{8.20}$$

As a consequence, when minimizing *Q*, one may take *W* in (1.5) to be block diagonal with blocks W_{11} and W_{22}. It follows that the generalized least squares function (1.5) specializes to

$$Q = (s_1 - \sigma_1)' W_{11} (s_1 - \sigma_1) + (s_2 - \sigma_2)' W_{22} (s_2 - \sigma_2)\,. \tag{8.21}$$

Thus if W_{11} consistently estimates V_{11}^{-1} in (8.15), and W_{22} is consistent for V_{22}^{-1} based on (5.6), a distribution-free mean and covariance structure analysis remains. Of course V_{22} may have a still more specialized structure, as is described for normal, elliptical, and heterogeneous kurtose distributions in Chapter 5.

Now suppose that the variables *z* are multinormally distributed. Then V_{22} is just V_N given in (5.22), and (8.21) can be specialized further to

$$Q = (s_1 - \sigma_1)' W_{11} (s_1 - \sigma_1) + Q_N\,, \tag{8.22}$$

where the normal theory covariance structure GLS function Q_N is given by (5.16) in Chapter 5. Under more a general elliptical assumption, in (8.22) one would use Q_E in (5.29) in place of Q_N, or under heterogeneous kurtosis theory, Q_{HK} given in (5.41). A complete LS analysis results when $W_{11} = I$, and $W_{22} = I$ in (8.21), or, as seen in (8.22), when $W_2 = I$ in (5.15)-(5.16). Optimal choices for W_{11} are more meaningful, especially V_{11}^{-1} as given above. Combined with optimal choices for W_2 in (5.15)-(5.16) as $W_2 = S^{-1}$ for GLS, or $W_2 = \hat{\Sigma}^{-1}$ for RLS or ML, we obtain two different mean and covariance structure GLS forms. If $W_2 = S^{-1}$ we have a GLS generalization of Browne's (1974) covariance structure function, namely

$$Q_{N(GLS)} = (s_1 - \sigma_1)' S^{-1} (s_1 - \sigma_1) + 2^{-1} tr[(S - \Sigma) S^{-1}]^2\,, \tag{8.23}$$

while if $W_2 = \hat{\Sigma}^{-1}$ is updated iteratively,

$$Q_{N(RLS)} = (s_1 - \sigma_1)' \hat{\Sigma}^{-1} (s_1 - \sigma_1) + 2^{-1} tr[(S - \Sigma) \hat{\Sigma}^{-1}]^2\,. \tag{8.24}$$

As before, since asymptotically, $a_n \rightarrow 1.0$, EQS ignores a_n when taking (8.22) into (8.23)-(8.24). Standard errors, χ^2 tests, tests on restrictions $c(\theta) = 0$ and related statistical results follow immediately from the general theory. For example, at the minimum of (8.24), $n\hat{Q}_{N(RLS)} \xrightarrow{D} \chi^2_{(p^*-q+r)}$. Minor adjustments to Lee and Bentler's (1980) work can verify that GLS and ML estimators, standard errors, and test statistics are asymptotically equal. Computationally, the work of Lee and Jennrich (1979) can be extended directly to show that $\hat{\theta}$ based on minimizing (8.24) gives the ML estimator. In addition to minimizing (8.9) or (8.24), ML and RLS estimation also can be accomplished (and is in EQS) by using an augmented moment matrix in which $\bar{z}\bar{z}'$ is added to S (S_b) and $\mu\mu'$ is added to Σ in (8.11), and treating the supermatrix as a covariance structure model (McDonald, 1980). Additionally, (8.22) would permit mixed LS/GLS/RLS estimation as well, where different estimators for W_2 are used in the two parts of the overall function. Such combinations are not provided in EQS.

In (8.9) we provided the ML function optimized in EQS. For completeness we note that the associated test statistic (8.10) is trivially different from the standard likelihood ratio test as developed under maximum likelihood theory. The standard ML function maximizes, aside from a constant, the log-likelihood of the N observations

$$\mathrm{L} = -\mathrm{N}\left\{\ln|\Sigma| + tr(S_b \Sigma^{-1}) + (\bar{z} - \mu)' \Sigma^{-1} (\bar{z} - \mu)\right\} / 2 \tag{8.25}$$

under choice of $\hat{\mu} = \mu(\hat{\theta})$ and $\hat{\Sigma} = \Sigma(\hat{\theta})$, where S_b is the biased (divisor N) sample covariance matrix. The corresponding log-likelihood of the saturated model that obtains when the estimates $\bar{z}$ of μ and S_b of Σ are placed in (8.25) is given by

$$\hat{L}_s = -N\{\ln|S_b| + p\}/2 \; . \tag{8.26}$$

The statistic $2\left(\hat{L}_s - \hat{L}\right)$, evaluated at $\hat{\theta}$, is given by the likelihood ratio test

$$LR = N\left\{\ln|\hat{\Sigma}| + tr(S_b\hat{\Sigma}^{-1}) - \ln|S_b| - p + (\bar{z} - \hat{\mu})'\hat{\Sigma}^{-1}(\bar{z} - \hat{\mu})\right\}. \tag{8.27}$$

If you compare (8.27) to (8.10), you will see that they differ in the use of sample size multiplier (n or N) and the sample covariance matrix (unbiased or biased). EQS follows typical practice (e.g., Widaman & Thompson, 2003) and uses the unbiased sample covariance matrix S in place of S_b and uses the multiplier n instead of N in (8.27). These corrections vanish as N gets large. For additional detail, see Browne and Arminger (1995).

With incomplete data, for which $\bar{z}$ and S_b or S do not exist, these statistics cannot be used. The estimation method, test statistic, and degrees of freedom are redefined appropriately.

9. MULTISAMPLE MEAN AND COVARIANCE STRUCTURES

Multisample structured means models combine the techniques of multisample analysis and structured means models. These techniques were described separately in Chapters 7 and 8. When combined, these types of models raise some new ideas that are worth discussing and illustrating. For example, a structured means model that is not identified in one group may become identified when analyzed in two groups simultaneously and subject to cross-group constraints. More importantly, structured means models in several groups that involve latent factors require a new concept of factor identification. Before we turn to these topics, however, let us offer some words of caution.

If you have two groups, such as males and females, there will be two sets of means and two sets of covariance matrices. If the covariance matrices as well as means for males and females are very different, you may not be able to specify a mean and covariance structure model that is largely similar in both groups. Completely different models may be needed. In such a situation, applying the methods of this chapter will be an exercise in frustration.

Most researchers hope, of course, that relatively invariant processes exist across groups, and hence that these invariant processes can be captured by parameters that are invariant across groups. This is a relevant question in several areas of application, such as cross-cultural research (Cheung & Rensvold, 1999, 2000; Little, 1997, 2000), differential item functioning (Chan, 2000), and intervention research (Bentler, 1991). The main point of doing a simultaneous mean/covariance structure analysis is to evaluate or establish the degree of invariance of key parameters across groups. If the means and covariances for the two groups are quite similar, it should be possible to estimate a structured means model that is identical or very similar in the two groups. You can accomplish this with the approach covered in this chapter. Even if the means and covariances are different, they may be similar enough to allow a largely identical model to describe the several samples. For example, perhaps only a few factor loadings, or a few factor means, may differ across groups while the other parameters are invariant. Because differences in structural parameters can propagate throughout a model, simple differences can show up as complex patterns of differences in reproduced means and covariances. For example, if a single factor mean differs across groups, this influence will be felt in all the measured variables that are indicators of this factor. Thus it is easy to be optimistic about the practical relevance of multisample mean and covariance structures.

However, you should be realistic. You already know that multisample covariance structures are hard to fit, and also that single-group structured means models are hard to fit. Hence you should expect that merging these two features in a single combined model also will be hard to do in practice. Is there a simpler alternative?

The MIMIC Alternative

If you expect the main differences between groups to show up in the means of factors and variables, there is a simpler alternative approach. This consists of a single sample model containing dummy-coded indicators of group membership as predictors of other key latent and observed variables in the model. This approach is an appropriate substitute for a multisample mean/covariance structure model if you expect the covariances to be substantially the same for your groups and the differences to primarily represent mean differences. For example, in considering height and weight, we certainly would expect males to be taller and heavier on average. But it may be possible that the variances of height and weight, and their correlation, is the same among males as among females. In such a case, you might consider setting up a single group model that allows for mean differences by gender.

In this approach, you set up your fundamental model without any concern for group membership. Then you add dummy variable predictors to account for group differences. In the simplest case, suppose initially you have a one-factor model with the specifications:

```
/EQUATIONS
V1 =  F1 + E1;
V2 = *F1 + E2;
V3 = *F1 + E3;
/VARIANCES
F1 = *; E1 TO E3 = *;
```

Now with V4 as your gender variable, scored, for example, "1" to indicate female and "2" to indicate male, you account for gender differences in the data by letting V4 predict the key variable in the model as follows:

```
/EQUATIONS
V1 =  F1 + E1;
V2 = *F1 + E2;
V3 = *F1 + E3;
F1 = *V4 + D1;
/VARIANCES
V4 = *; D1 = *; E1 TO E3 = *;
```

Of course, V4 is an independent variable and so its variance is a free parameter, and now F1 is a dependent variable whose residual variate D1 has a free variance. This model is estimated in the combined male/female data, so that a 2-group model has been simplified into a 1-group form. If the estimated coefficient for V4→F1 is positive, this means that males have the higher scores on F1; and if the factor loadings are all positive, through path tracing it also follows that males have the higher scores on V1-V3. If the V4→F1 path does not sufficiently explain the gender differences in, say, V1, you could also consider adding a direct path from V4 to V1 providing the model is identified.

A model of this form is sometimes called a MIMIC model (Jöreskog & Goldberger, 1975), meaning, a model with multiple indicators (V1-V3) and multiple causes (here, only V4) of a single latent variable (here, F1). See Muthén (1989), Kaplan (2000) and Hancock (2001) for further discussion. Clearly the idea of using control variables including dummy group code variables in a one-group model alternative to multisample mean and covariance structures extends naturally to many predictors (V5, V6, … in addition to V4), none or many (F2, F3, … in addition to F1) factors, and many groups (e.g., 5 different age groups). Care must be taken that the predictors are dummy coded appropriately to represent group membership, as the coding will determine the interpretation of group effects. A good discussion of dummy coding can be found in Cohen, Cohen, West and Aiken (2003). This approach can also be used with experimental data (Kano, 2001), or as an alternative to the analysis of covariance (Sörbom, 1978). It is more powerful than manova (Hancock, 2001; Hancock, Lawrence, & Nevitt, 2000; Kano, 2001).

MIMIC-type models are not appropriate if the correlations or covariances among variables are very different in the several groups. Then there is no alternative to running a multisample model.

UNDERIDENFIRIED ONE-GROUP MODEL

We are almost ready to study multisample mean and covariance structures. Before doing so, let us consider a model that is not identified as a one-group model, because one of the interesting features of multisample models is that they can become identified in a multisample model that contains appropriate cross-group constraints. The model we will consider has 6 variables and two factors, as shown below.

The left part of Figure 9.1 shows two factors, with the factor regression F1 → F2. The measured variables are in the middle of the figure. Variables V1-V4 are indicators of F1, while V5-V6 are indicators of F2. Each variable has an error residual, with E1 and E2 being correlated. This is a standard factor analysis model, except

that the constant (V999), shown in the right part of the figure, transforms the model from a covariance structure model into a mean/covariance structure model. In this model, intercepts are hypothesized for all variables V1-V6, as well as for both factors, F1 and F2. As usual, parameters that have an asterisk are considered free parameters. At this point, the "=" and "0" next to "*" should be ignored.

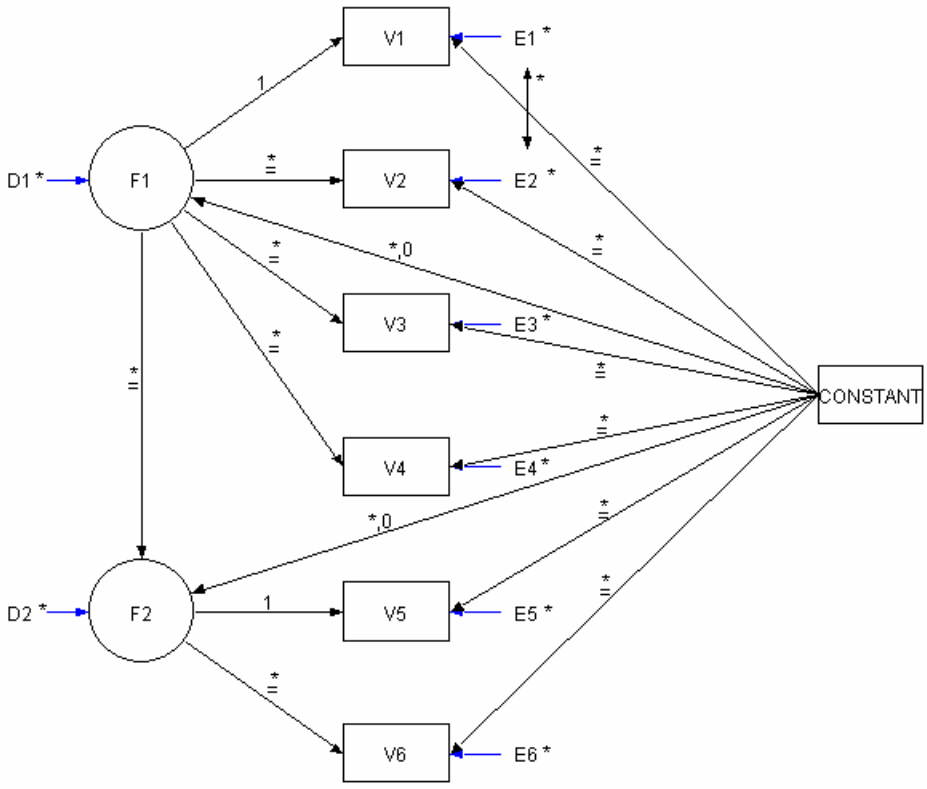


Figure 9.1

If you run this model, you will get output such as the following, which is highly edited:

```
 1  /TITLE
 2     HEAD START ONE-GROUP MODEL
 3  /SPECIFICATIONS
 4      CASES=148; VARIABLES=6; ANALYSIS=MOMENT; MATRIX=CORRELATION;
 5      METHOD=ML;
 6  /EQUATIONS                              ! AN INTERCEPT FOR EACH DEPENDENT VARIABLE
 7      V1 =  3.9*V999 +      F1 + E1;
 8      V2 =  3.3*V999 + 0.85*F1 + E2;
 9      V3 =  2.6*V999 + 1.21*F1 + E3;
10      V4 =  6.4*V999 + 2.16*F1 + E4;
11      V5 = 20.3*V999 +      F2 + E5;
12      V6 = 10.1*V999 + 0.85*F2 + E6;
13      F2 =     *V999 + 2.10*F1 + D2;
14      F1 = -0.4*V999 + D1;
15  /VARIANCES
16       E1 TO E6 = 1.5*;
17       D1 TO D2 = 0.3*;
18  /COVARIANCES
19       E2,E1=*;
20  /PRINT
21    EFFECT=YES;
22  /MATRIX
```

```
   23     1.000
   24     0.441  1.000
   25     0.220  0.203  1.000
   26     0.304  0.182  0.377  1.000
   27     0.274  0.265  0.208  0.084  1.000
   28     0.270  0.122  0.251  0.198  0.664  1.000
   29  /STANDARD DEVIATIONS
   30     1.332  1.281  1.075  2.648  3.764  2.677
   31  /MEANS
   32     3.520  3.081  2.088  5.358 19.672  9.562
   33  /END

      33 RECORDS OF INPUT MODEL FILE WERE READ

  MATRIX CONTAINS SPECIAL VARIABLE V999, THE UNIT CONSTANT
  COVARIANCE MATRIX IS IN UPPER TRIANGLE; MEANS ARE IN BOTTOM ROW OF MATRIX
  COVARIANCE/MEAN MATRIX TO BE ANALYZED:
     6 VARIABLES (SELECTED FROM   6 VARIABLES), BASED ON   148 CASES.

                         V1           V2           V3           V4           V5           V6
V999
                        V  1         V  2         V  3         V  4         V  5         V  6
V999
        V1    V  1       1.774
        V2    V  2       0.752      1.641
        V3    V  3       0.315      0.280      1.156
        V4    V  4       1.072      0.617      1.073      7.012
        V5    V  5       1.374      1.278      0.842      0.837     14.168
        V6    V  6       0.963      0.418      0.722      1.404      6.691      7.166
        V999   V999        3.520       3.081       2.088       5.358      19.672       9.562
1.000

  PARAMETER        CONDITION CODE
    V1,V999        LINEARLY DEPENDENT ON OTHER PARAMETERS
    V5,V999        LINEARLY DEPENDENT ON OTHER PARAMETERS

  CHI-SQUARE =       15.655 BASED ON     5 DEGREES OF FREEDOM
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS     0.00790
```

Although the program finds a minimum of the function, and computes a χ^2 statistic, there are two condition codes indicating linear dependencies among parameters. Such messages indicate theoretical or empirical under-identification (see e.g., Rindskopf, 1984a). In fact, both linear dependencies involve the intercept V999, suggesting that a careful theoretical analysis of the part of the model dealing with V999 should be made. In Chapter 7 it was noted that the number of intercepts must be less than or equal to the number of measured variables. In this case, there are 6 Vs, but 8 intercepts. There are two too many mean parameters in this model, as logic and Condition Codes inform us. EQS does not uniquely indicate which parameters are underidentified since many sets of two would serve equally well. Logically, the intercepts of the V variables will account for the means of the V variables, so there is no need to also have intercepts for the F variables.

The job is modified slightly to delete as free parameters (or set to fixed zero) the intercepts for F1 and F2, which are the coefficients of V999 in lines 13 and 14 of the above output. Removing V999 from lines 13 and 14, and resubmitting the job, yields the following result.

```
  PARAMETER ESTIMATES APPEAR IN ORDER,
  NO SPECIAL PROBLEMS WERE ENCOUNTERED DURING OPTIMIZATION.

  CHI-SQUARE =       15.655 BASED ON     7 DEGREES OF FREEDOM
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS     0.02846
```

Now there is no sign of linear dependency among parameters. The same function value and χ^2 statistic are obtained, but the probability changes because the degrees of freedom have increased by two. An analysis of the parameter estimates, not shown here, reveals that the intercepts have been estimated precisely at the observed sample means. This result does not surprise us. There is exactly one intercept for each measured variable, and

since there are no indirect paths from V999 to any of V1-V6s, the intercepts are the means. Thus the means are reproduced exactly.

Next we evaluate what happens to the model of Figure 9.1 when two groups are considered simultaneously. The importance of doing this was first developed by Sörbom (1974).

MULTISAMPLE PATH DIAGRAM AND FACTOR MEANS

The model of Figure 9.1 is actually a representation of a two-group model studied by Sörbom (1982). It deals with two samples of children. One sample had been given a Head Start educational program; the other sample was a matched Control group. V1-V4 represent mother's education, father's education, father's occupation, and family income. These variables were taken as indicators of F1, a latent socioeconomic status or SES factor. They were background variables, essentially to be controlled. The more important variables are V5 and V6, scores on the Metropolitan Readiness Test and the Illinois Test of Psycholinguistic Abilities, respectively. These were taken by Bentler and Woodward (1978) as indicators of a construct of cognitive ability, shown in the figure by F2. It is presumed that SES (F1) will affect ability (F2), but the real interest is whether the mean of F2 is higher in the Head Start group as compared to the Control group, taking into account the SES background of the children. As shown in the figure, all Vs and Fs are presumed to have intercepts that are parameters to be estimated. Next, Figure 9.1 is studied in greater detail, both for its specific meaning in this particular model, and for some principles for setting up multisample structured means models.

The figure, which is not a conventional path diagram, attempts to show the two models of the Head Start and Control samples simultaneously. A key to distinguishing the two groups is as follows:

*	Denotes a parameter that is free in both groups.
* =	Denotes a parameter that is free in both groups, but constrained to be equal in the two groups.
*,0	Denotes a parameter that is free in one group (Head Start), and zero in the other (Control) group.

Figure 9.2

Thus the path diagram makes clear that:

- The variances of E and D variables are free parameters in each group, and not constrained to be equal across groups.
- The covariance of E1 and E2 is a free parameter in each group, and not constrained to be equal across groups.
- The free factor loadings of all variables (F $\rightarrow$ V paths) are free parameters in each group, but each loading is constrained to be equal in the two groups.
- The fixed 1.0 factor loadings are fixed for identification in both groups.
- The factor regression (F1 $\rightarrow$ F2) is a free parameter in each group, but the value is constrained to be equal across the two groups.
- The intercepts of the measured (V) variables are free parameters in each group, but each intercept is constrained to be equal in the two groups.
- The intercepts of the factors (Fs) are free parameters in the Head Start group, but set to a fixed zero value in the Control group.

Some of these points require further discussion.

Regarding the left part of the figure, if the same factor model holds for both samples, then it is helpful to have the factor loadings and factor regressions be equal across groups. Equality of residual variances and covariances is not so important, and is not imposed. A good discussion of restrictions in this context is given by Dolan and Molenaar (1994).

Regarding the right part of the figure, *the paths from V999 to the Vs, each of which is held to be equal across groups, represent a kind of baseline level for the variables. Differences in means of the variables across groups, if they exist, must arise from other sources.* In the diagram, these sources can be traced back to the intercepts of F1 and F2. For example, the path V999 → F1 → V4 makes clear that the intercept of F1 will affect V4, and the path V999 → F1 → F2 → V6 shows that it will also affect V6. (The final means of the V variables, are, of course, total effects, or model moments, that can be obtained with /PRINT). The paths V999 → F1 and V999 → F2, the intercepts for F1 and F2, represent coefficients that are free to be estimated in one group, but are held to zero in the other group. Consequently, if the freely estimated coefficients are nonzero and large, the differences in factor intercepts will be reflected in mean differences of the observed variables across groups. In this model, the Head Start factor intercepts are estimated, but the Control intercepts are set to zero.

The rationale for parameters that are free to be estimated in one group, but fixed in another group, arises as follows:

In multisample latent variable models having measured variables with intercepts that are constrained to be equal across groups, the intercepts of the latent factors have an arbitrary origin. A standard way to fix this origin is to set the factor intercepts of one group to zero.

When there are two groups, as here, all F intercepts are set to zero in one group, and freely estimated in the other group. As a consequence, these intercepts can be interpreted only in a relative sense. If the Head Start intercept for F1 is significantly higher than 0, which is the value of the intercept in the Control sample, then the Head Start sample has more of whatever F1 is measuring. If the intercept is significantly negative, then the Head Start sample has less of whatever F1 is measuring. (What F1 represents, and the meaning of "more" or "less", is not a new issue in these types of models; the simplest way to determine this is by studying the meaning and direction of scoring of the measured variables that are indicators of the factor. In this example, F1 measures SES, with higher numbers indicating higher status.) It is arbitrary as to which group has an intercept set to zero, but the estimated intercept of another group is always interpreted relative to the value of 0.0 used in some reference group. (Actually, the value 0.0 is also chosen arbitrarily.) In theory, different factors could have their zero intercepts set in different groups, but this is not a practical suggestion. Let the sample that has all factor intercepts set at zero be called the reference sample. There is a consequence of fixed zero factor intercepts:

In the reference sample, the coefficient for regression of a measured variable on a constant (V999) is the expected (model-reproduced) mean of the measured variable. Thus the mean is a direct effect of the constant (V999).

How this works in the reference sample, but differs in other samples, can be seen by path tracing in Figure 9.1. The expected mean of a measured variable in all non-reference groups is represented by a total effect, since there will be various indirect ways for V999 to affect a final V, as shown above (also see the figure). Since the expected means in the reference group should be close to the observed sample means if the model is correct, the observed means in the reference sample provide a good set of start values for the cross-group constrained intercepts of the V variables.

EXAMPLE: EFFECTS OF HEAD START

The quasi-experimental Head Start study investigated whether children who received a special Head Start educational program subsequently perform better on cognitive tasks than Control children who did not receive

this extra training. The study did not succeed in randomly assigning children to treatment conditions, and, as might be expected from the goals of a social intervention, the Head Start program was given to children who were relatively more disadvantaged (lower SES) than the Controls. The means of the variables, shown for the Head Start children above and shown for the Controls below, are not encouraging: the Controls actually have the higher means on V5 and V6, the cognitive ability measures. (Compare line 32 of program output above with line 66 of output below.) However, as just noted and as is clear from the sample means, it is also true that the Head Start children are from lower SES families. Possibly the higher SES of the controls can explain their higher cognitive performance ability. So it is worthwhile to see whether the Head Start children score higher than the controls on F2, the cognitive ability factor, if SES is controlled as a factor (F1).

The job setup for the multisample structured means model is given below, in an edited version of the EQS output. The first 9 lines show that the model for the Head Start group is identical to that shown previously, except that a two-group model is being analyzed; the rest of this group is omitted. Subsequently, the model is given for the Control group.

```
     1  /TITLE
     2     MULTIPLE GROUPS AND STRUCTURED MEANS -- GROUP 1
     3     HEAD START DATA -- LISREL 7 MANUAL, P. 254
     4     HEAD START GROUP
     5  /SPECIFICATIONS
     6      CASES=148; VARIABLES=6; ANALYSIS=MOMENT; MATRIX=CORRELATION;
     7       METHOD=ML; GROUPS=2;                                          ! GROUPS=2; ADDED
HERE
     8  /EQUATIONS
     9      V1 =  3.9*V999 +      F1 + E1;

       (SAME AS PREVIOUS FILE GIVEN ABOVE, BUT WITH LINES 20 AND 21 REMOVED)

    34  /TITLE
    35     MULTIPLE GROUPS AND STRUCTURED MEANS -- GROUP 2
    36     CONTROL GROUP
    37  /SPECIFICATIONS
    38      CASES=155; VARIABLES=6; ANALYSIS=MOMENT; MATRIX=CORRELATION;
    39      METHOD=ML;
    40  /EQUATIONS
    41      V1 =  3.9*V999 +      F1 + E1;
    42      V2 =  3.3*V999 + 0.85*F1 + E2;
    43      V3 =  2.6*V999 + 1.21*F1 + E3;
    44      V4 =  6.4*V999 + 2.16*F1 + E4;
    45      V5 = 20.3*V999 +      F2 + E5;
    46      V6 = 10.1*V999 + 0.85*F2 + E6;
    47      F2 =   2.10*F1 + D2;            ! NO V999 HERE
    48      F1 =     0V999 + D1;            ! FIXED 0 IS EQUIVALENT TO NO V999
    49  /VARIANCES
    50       E1 TO E6 = 1.5*;
    51       D1 TO D2 = 0.3*;
    52  /COVARIANCES
    53       E2,E1=*;
    54  /PRINT
    55    EFFECT=YES;
    56  /MATRIX
    57     1.000
    58     0.484  1.000
    59     0.224  0.342  1.000
    60     0.268  0.215  0.387  1.000
    61     0.230  0.215  0.196  0.115  1.000
    62     0.265  0.297  0.234  0.162  0.635  1.000
    63  /STANDARD DEVIATIONS
    64     1.360  1.195  1.193  3.239  3.900  2.719
    65  /MEANS
    66     3.839  3.290  2.600  6.435 20.415 10.070
    67     ! COMPARE TO HEAD START GROUP -- ALL VARIABLES HAVE HIGHER MEANS,
    68     ! HIGHER SES, HIGHER COGNITIVE ABILITY
    69  /CONSTRAINTS              ! SET GENERATES EQUALITY CONSTRAINTS ON ALL FREE PARAMETERS
IN
    70      SET = GVV, BVF, BFF;   ! MEASUREMENT EQUATIONS.  SEE LIST OF CONSTRAINTS IN LM
TEST.
```

```
   71  /LMTEST                  ! TO CHECK ADEQUACY OF CONSTRAINTS
   72  /END

  COVARIANCE/MEAN MATRIX TO BE ANALYZED:                              ! GROUP 2 MATRIX TO BE
ANALYZED
     6 VARIABLES (SELECTED FROM   6 VARIABLES), BASED ON   155 CASES.

                         V1          V2          V3          V4          V5          V6
V999
                        V  1        V  2        V  3        V  4        V  5        V  6
V999
       V1    V  1       1.850
       V2    V  2       0.787      1.428
       V3    V  3       0.363      0.488      1.423
       V4    V  4       1.181      0.832      1.495     10.491
       V5    V  5       1.220      1.002      0.912      1.453     15.210
       V6    V  6       0.980      0.965      0.759      1.427      6.734      7.393
       V999   V999        3.839       3.290       2.600       6.435      20.415      10.070
1.000

  MULTIPLE POPULATION ANALYSIS, INFORMATION IN GROUP  1

  PARAMETER ESTIMATES APPEAR IN ORDER,
  NO SPECIAL PROBLEMS WERE ENCOUNTERED DURING OPTIMIZATION.

  MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
  STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

    V1   =V1  =   1.000 F1    + 3.869*V999  + 1.000 E1
                                  .094
                                41.087@

    V2   =V2  =    .852*F1    + 3.339*V999  + 1.000 E2  ! ALL INTERCEPTS ARE SIGNIFICANT
                   .144          .083
                  5.923@       40.315@

    V3   =V3  =   1.208*F1    + 2.573*V999  + 1.000 E3  ! FACTOR LOADINGS ALSO
                   .222          .090
                  5.429@       28.643@

    V4   =V4  =   2.760*F1    + 6.421*V999  + 1.000 E4
                   .517          .229
                  5.333@       28.094@

    V5   =V5  =   1.000 F2    +20.357*V999  + 1.000 E5
                                  .287
                                70.885@

    V6   =V6  =    .850*F2    +10.085*V999  + 1.000 E6
                   .141          .217
                  6.016@       46.442@

  CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
  STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

    F1   =F1  =   -.382*V999  + 1.000 D1  ! HEAD START GROUP SIGNIFICANTLY
                   .104                   ! LOWER IN SES
                 -3.685@

    F2   =F2  =   2.137*F1    +  .184*V999  + 1.000 D2
                   .552          .378
                  3.874@         .487

   ! HEAD START GROUP HIGHER THAN CONTROLS ON ABILITY, GIVEN THAT SES IS
   ! CONTROLLED, BUT EFFECT IS NOT SIGNIFICANT
```

The regression of F2 on F1, ability on SES, is significant (and equal in both groups, see below). Higher SES children do better than lower SES children. Also, Head Start children were lower in SES to begin with, when -.382 is compared to the Control's value of fixed 0. Note that the experimental Head Start program produced a

positive impact (.184) on ability, though the effect is not significant by z-test. This positive impact stands in contrast to the raw variable means (where the Controls had the higher means on the ability indicators V5 and V6 -- see input data of both groups), and to the total effects of V999 on V5 and V6, shown below, which also verify that the Controls have the higher expected variable means in the model.

To conserve space, the estimates of variances of the E and D variables are not shown below. These variances were not constrained to be equal across groups, though a more restricted model that imposes such a constraint could also have been considered. Similarly, the estimated covariance of E1 and E2 is not shown for either group. Next we look at the effect decomposition, which will provide the model-reproduced variable means.

```
  DECOMPOSITION OF EFFECTS WITH NONSTANDARDIZED VALUES

  PARAMETER TOTAL EFFECTS   ! COEFFICIENTS OF V999 IN V-EQUATIONS ARE REPRODUCED MEANS OF
VARIABLES.
  ----------------------  ! IN F-EQUATIONS, THEY ARE MODEL FACTOR MEANS.

    V1   =V1  =   1.000 F1    + 3.487*V999  + 1.000 E1    + 1.000 D1
                                  .092
                                 37.880

    V2   =V2  =    .852*F1    + 3.014*V999  + 1.000 E2    +  .852 D1
                   .144          .085                      .144
                  5.923        35.446                     5.923

    V3   =V3  =   1.208*F1    + 2.112*V999  + 1.000 E3    + 1.208 D1
                   .222          .084                      .222
                  5.429        25.207                     5.429

    V4   =V4  =   2.760*F1    + 5.367*V999  + 1.000 E4    + 2.760 D1
                   .517          .202                      .517
                  5.333        26.537                     5.333

    V5   =V5  =   2.137 F1    + 1.000 F2    +19.724*V999  + 1.000 E5    + 2.137 D1    + 1.000
D2
                   .552                        .286                      .552
                  3.874                      68.950                     3.874

    V6   =V6  =   1.817 F1    +  .850*F2    + 9.548*V999  + 1.000 E6    + 1.817 D1    +  .850
D2
                   .415          .141          .217                      .415          .141
                  4.375         6.016        43.946                     4.375         6.016

    F1   =F1  =   -.382*V999  + 1.000 D1

    F2   =F2  =   2.137*F1    -  .632*V999  + 2.137 D1    + 1.000 D2
                   .552          .365          .552
                  3.874        -1.733         3.874

  ! THE ESTIMATED MEAN FOR THE ABILITY FACTOR F2 UNDER THE MODEL IS LOWER (BY .632) FOR HEAD
  ! START CHILDREN COMPARED TO CONTROLS, BUT THIS IS BASICALLY DUE TO THE DIFFERENTIAL SES OF
THE
  ! CHILDREN, I.E. THE INDIRECT EFFECT OF V999 ON F2, WHICH IS -.816, SIGNIFICANT WITH Z = -
3.089
  ! (NOT SHOWN HERE).

  MULTIPLE POPULATION ANALYSIS, INFORMATION IN GROUP  2
  PARAMETER ESTIMATES APPEAR IN ORDER,
  NO SPECIAL PROBLEMS WERE ENCOUNTERED DURING OPTIMIZATION.

  ALL EQUALITY CONSTRAINTS WERE CORRECTLY IMPOSED

  ! MEASUREMENT EQUATIONS ARE NOT SHOWN, SINCE THEY ARE IDENTICAL TO THOSE IN GROUP 1
```

```
CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
  STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

    F1   =F1  =   1.000 D1

    F2   =F2  =   2.137*F1    + 1.000 D2  !COEFFICIENT IS THE SAME IN BOTH GROUPS
                   .552
                  3.874@

  DECOMPOSITION OF EFFECTS WITH NONSTANDARDIZED VALUES

  PARAMETER TOTAL EFFECTS ! MODEL PREDICTS HIGHER MEANS ON ALL Vs FOR CONTROLS
  ----------------------
    V1   =V1  =   1.000 F1    + 3.869*V999  + 1.000 E1    + 1.000 D1

    V2   =V2  =    .852*F1    + 3.339*V999  + 1.000 E2    +  .852 D1
                   .144                                     .144
                  5.923                                    5.923

    V3   =V3  =   1.208*F1    + 2.573*V999  + 1.000 E3    + 1.208 D1
                   .222                                     .222
                  5.429                                    5.429

    V4   =V4  =   2.760*F1    + 6.421*V999  + 1.000 E4    + 2.760 D1
                   .517                                     .517
                  5.333                                    5.333

    V5   =V5  =   2.137 F1    + 1.000 F2    +20.357*V999  + 1.000 E5    + 2.137 D1    + 1.000
D2
                   .552                                                    .552
                  3.874                                                   3.874

    V6   =V6  =   1.817 F1    +  .850*F2    +10.085*V999  + 1.000 E6    + 1.817 D1    +  .850
D2
                   .415          .141                                      .415          .141
                  4.375         6.016                                     4.375         6.016

    F1   =F1  =   1.000 D1

    F2   =F2  =   2.137*F1    + 2.137 D1    + 1.000 D2
                   .552          .552
                  3.874         3.874

  ! BECAUSE GROUP 2 IS THE CONTROL GROUP, THE TOTAL EFFECTS OF V999 ON Vs ARE THE DIRECT
EFFECTS.
  ! SEE MEASUREMENT EQUATIONS ABOVE.  THEY ALSO SHOULD REPRODUCE GROUP 2'S SAMPLE MEANS WELL.

  STATISTICS FOR MULTIPLE POPULATION ANALYSIS

  ALL EQUALITY CONSTRAINTS WERE CORRECTLY IMPOSED   ! ALWAYS NICE TO SEE

  GOODNESS OF FIT SUMMARY FOR METHOD = ML

  CHI-SQUARE =       27.451 BASED ON    23 DEGREES OF FREEDOM      ! GOOD FIT
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS     0.23726

  BENTLER-BONETT NORMED    FIT INDEX=       0.932   ! FIT INDICES ARE OK ALSO
  BENTLER-BONETT NONNORMED FIT INDEX=       0.968
  COMPARATIVE FIT INDEX (CFI)       =       0.980
```

```
                    ITERATIVE SUMMARY          ! FUNCTION CONVERGED OK

                  PARAMETER
 ITERATION        ABS CHANGE        ALPHA              FUNCTION
     1             1.101052        1.00000             0.22199
     2             0.542323        0.50000             0.13418
     3             0.270606        1.00000             0.09364
     4             0.073611        1.00000             0.09156
     5             0.028961        1.00000             0.09123
     6             0.011527        1.00000             0.09120
     7             0.003917        1.00000             0.09120
     8             0.001521        1.00000             0.09120
     9             0.000542        1.00000             0.09120

 LAGRANGE MULTIPLIER TEST (FOR RELEASING CONSTRAINTS)

 CONSTRAINTS TO BE RELEASED ARE: ! CONSTRAINTS GENERATED BY SET COMMAND IN CONSTRAINT
SECTION

         CONSTRAINTS FROM GROUP  2

         CONSTR:    1   (1,V1,V999)-(2,V1,V999)=0;
         CONSTR:    2   (1,V2,V999)-(2,V2,V999)=0;
         CONSTR:    3   (1,V3,V999)-(2,V3,V999)=0;
         CONSTR:    4   (1,V4,V999)-(2,V4,V999)=0;
         CONSTR:    5   (1,V5,V999)-(2,V5,V999)=0;
         CONSTR:    6   (1,V6,V999)-(2,V6,V999)=0;
         CONSTR:    7   (1,V2,F1)-(2,V2,F1)=0;
         CONSTR:    8   (1,V3,F1)-(2,V3,F1)=0;
         CONSTR:    9   (1,V4,F1)-(2,V4,F1)=0;
         CONSTR:   10   (1,V6,F2)-(2,V6,F2)=0;
         CONSTR:   11   (1,F2,F1)-(2,F2,F1)=0;

          UNIVARIATE TEST STATISTICS:

  NO    CONSTRAINT    CHI-SQUARE   PROBABILITY
  --    ----------    ----------   -----------
   1    CONSTR:   1        0.035       0.851   ! NO CONSTRAINT IS SIGNIFICANT,
   2    CONSTR:   2        0.781       0.377   ! I.E. THEY ARE ALL ACCEPTABLE.
   3    CONSTR:   3        0.594       0.441
   4    CONSTR:   4        0.012       0.912
   5    CONSTR:   5        0.196       0.658
   6    CONSTR:   6        0.196       0.658
   7    CONSTR:   7        1.694       0.193
   8    CONSTR:   8        0.019       0.891
   9    CONSTR:   9        0.008       0.927
  10    CONSTR:  10        0.001       0.971
  11    CONSTR:  11        0.108       0.742
```

The multivariate test verified the univariate LM test results. Also, the LM test for adding parameters did not find any parameter with a significant p-value. This concludes the 2-group structured mean Head Start/Controls example.

An issue raised at the beginning of the chapter was that a model that may not be identified in one group could become identified when several groups are considered simultaneously and cross-group equality constraints are imposed. The Head Start (only) analysis with V999 failed as a model because there were more intercept parameters to be estimated than there were means to be modeled. Of course this fact is not changed when the 2-group model is considered. However, both groups provided 12 sample means. Parameters involving V999 that were estimated in both groups included the 12 free parameters for the common level of the variables, which, with 6 equalities, implies 12 - 6 = 6 effective free parameters. Add the two intercepts for F1 and F2 in the Head Start sample, for a total of 8 free parameters. Thus there are 12 - 8 = 4 effective degrees of freedom for the part of the model concerned with the means. Of course, having fewer parameters in a model than there are data points is a necessary, but not sufficient, condition for model identification.

PRACTICAL AND STATISTICAL ASPECTS

It was noted in Chapters 7 and 8 that multisample models and structured means models can both be difficult to estimate and test. Structured means models may be adequate for reproducing sample means or covariances, but perhaps not both. Multisample models may be adequate in one sample, but inadequate in another sample. As a consequence, the type of model discussed in this chapter can also be difficult to apply in practice, especially when many variables and parameters are involved.

The most serious difficulty is probably the multisample aspect of an analysis. Since a model must be at least marginal to adequate in all of the samples, a model that fits very poorly in one sample, especially a large sample, can have an inordinate influence on the overall fitting function value in the multisample version of the same model. Of course, a model that fits marginally in one sample may be good enough in a multisample context because its poor fit is offset by an especially good fit in the other samples, and because there may be a significant gain in degrees of freedom (df). The Head Start data illustrates this phenomenon. When the model of Figure 9.1 was considered only in the Head Start sample, the model with 7 df yielded a fit probability of .028, which is below the standard .05 cutoff for model acceptance. Yet the same model in a 2-sample context yielded an acceptable overall probability of .237 based on 23 df.

Convergence difficulties will tend to occur when theoretical or empirical under-identification exists in a model. A clue to such under-identification is given by EQS' error messages, especially, that one or more parameter estimates may be linearly dependent on others. But a more serious problem occurs if the model is extremely inadequate at describing the data, or if extremely bad starting values are used for an otherwise potentially acceptable model, for the iterations may not converge at all. Then it may be best to evaluate the models in their separate groups before attempting a combined multisample analysis. After all, the multisample analysis will, in general, impose additional restrictions on the models as compared to a single-sample analysis. If these restrictions are extremely inappropriate, even an otherwise acceptable model will become statistically inadequate. The /LMTEST that evaluates the cross-sample constraints can be extremely helpful in locating inappropriate constraints. If it makes sense to do so, such constraints could then be released in future runs.

Multisample models can be used in nonstandard ways to tackle a variety of interesting questions. For example, if the number of variables to be analyzed, or the number of factors, is intended to be different in the different groups, it is possible to introduce pseudo-Vs or pseudo-Fs so as to make the number of such variables equal in the job set-up (Jöreskog, 1971a). This type of procedure has been used in the multisample approach to dealing with missing data (Allison, 1987; Muthén, Kaplan, & Hollis, 1987), where cases exhibiting a given pattern of present/missing data are treated as a sample, to be distinguished from other patterns, and hence, samples. As shown in previous versions of this *Manual* (Bentler, 1995), analysis of missing data using multisample analysis can be done easily with EQS, providing you pay attention to the idea that a sample with a missing variable can contain specifications of equations, intercepts, variances, and covariances only for variables that are actually observed in that sample, as well as for the hypothesized factors and residual variables relevant to those variables. Today, however, missing data methods do not need to utilize the clever multisample approach of Allison and Muthén, Kaplan, and Hollis. See Chapter 12. Good discussions of related issues are given by Neale (2000) and Wothke (2000). Of course, if you do have natural multiple samples, and then also have missing data in one or each of these samples, the methods of Chapter 12 can be applied to these multiple samples.

Multisample structured means models were historically estimated in LISREL, and discussed in applications, using a cross-products sample and model matrix sometimes called a moment matrix. This is not to be confused with EQS' use of the term "ANALYSIS = MOMENT" referring to the analysis of means (1st moments) and covariances (2nd moments about the means). Hayduk (1987, Appendix D; see also, 1996) summarizes the relevance of this matrix to LISREL's procedures. This moment matrix is irrelevant to either model specification or the statistical theory in EQS.

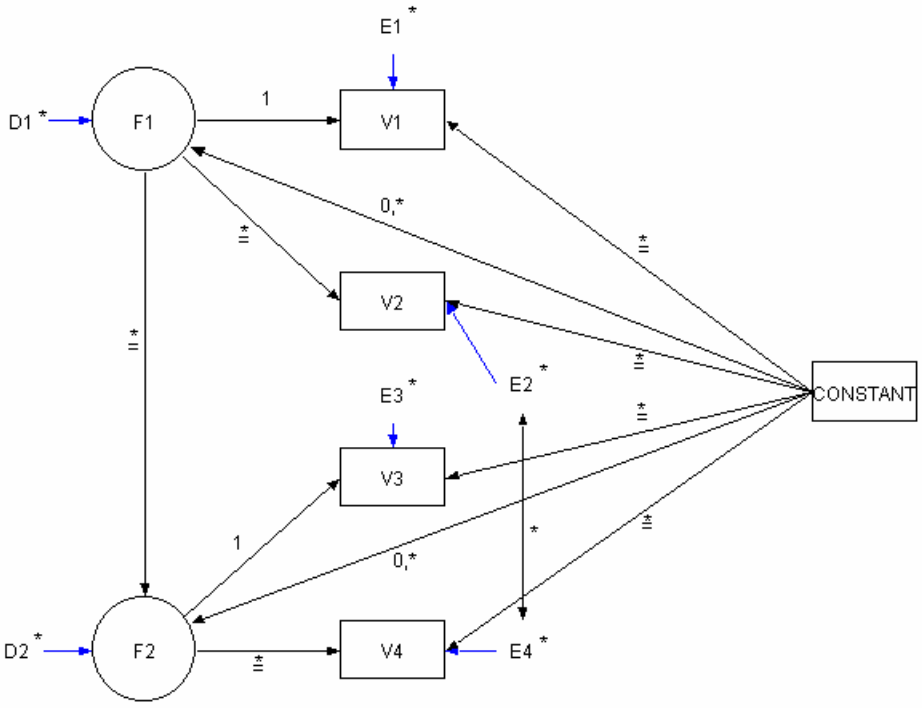
The statistical theory used in EQS assumes that sample covariances and means are being analyzed in all groups, i.e., that the variables have not been standardized. Additionally, scales of the latent variables should be fixed in

each group by fixing an F → V path for the same variable, to assure that the factor is on a common scale across all groups. For reasons outlined by Cudeck (1989), EQS does not provide a scaling in which the solution is standardized to a common metric across groups.

In earlier versions of EQS, the statistical theory of multiple group models was limited to maximum likelihood estimation. This is still the appropriate theory to use if one can assume asymptotic robustness for the model with nonnormal data (Satorra, 2001, 2002). However, this assumption is not easily evaluated, and so EQS 6 provides extensions to the multiple group case of all estimation methods that allow a mean structure in a single group (see Chapter 8). This includes, for example, arbitrary distribution theory and its various special cases. An overview of the relevant extended statistical theory is given by Yuan and Bentler (2001b). However, Bayesian approaches (Song & Lee, 2001, 2002b) are not yet implemented.

EXAMPLE: EXPERIMENTAL DATA

Sörbom (1978) reported on an experiment by Olsson on the effects of training on abilities to perform verbal tasks. In a pre-test, 11-year-old children were assessed for their verbal ability with two kinds of verbal material, synonyms and opposites (presumably antonyms). Thereafter, they were randomly assigned to experimental and control conditions. The experimental group received training on similar materials, while the control group did not. Both groups were then retested, yielding post-test data. The pre-test and post-test data are to be analyzed.



Sörbom studied several models for these data. A path diagram for his final model is given in Figure 9.3, where V1 and V2 represent the pre-test scores on the synonyms and opposites tasks, respectively, and V3 and V4 represent the post-test scores. As seen in the left part of the diagram, it is hypothesized that synonyms and opposites, at each time point, can be conceived as indicators of a latent factor, say, verbal ability. Ability at post-test, F2, is expected to be a function of ability at pre-test, F1. In addition, the residuals of opposites at the two time points, E2 and E4, are expected to be correlated from pre-test to post-test.

Figure 9.3

The right part of the figure gives the constant V999. This is presumed to affect each of the variables V1-V4, reflecting their intercept, a general level for these variables. V999 also directly affects the factors F1 and F2, representing an intercept for these factors. The figure maintains the convention of Figure 9.2, above. Thus, factor loadings, factor regressions, and variable intercepts are constrained to be equal across groups. Factor intercepts are fixed at 0 for identification in the Control group, and are free to be estimated in the Experimental group. All residual variances, and the covariance of E2, with E4 are free to be estimated in each group, without any constraints.

Consider the model in one group only, say, the Controls. In this group, the factor intercepts (V999 → F1, V999 → F2) are set to zero. Then, counting parameters, there are 4 intercept parameters and 10 additional parameters. With 4 input variables, there are 4 means and 10 variances and covariances. It is apparent that the model is just identified for both the means and the covariances. Thus, running this model in the Control group alone would lead to a perfect fit, with zero residuals, zero χ^2, and zero degrees of freedom. Running the same model in the Experimental group would have the same effect, while including the factor intercept paths will lead to an under-identified model. These identification problems disappear when both groups are considered simultaneously. The cross-group restrictions serve to make the model estimable and testable. An abbreviated EQS output for the 2-group run is given below.

```
 1  /TITLE
 2     OLSSON'S DATA (SORBOM, 1978), CONTROL GROUP IS FIRST
 3  /SPECIFICATIONS
 4      VARIABLES = 4; CASES = 105; ANALYSIS = MOMENT; GROUPS = 2;
 5  /LABELS
 6      V1 = SYNONYM1; V2 = OPPOSIT1; V3 = SYNONYM2; V4 = OPPOSIT2;
 7      F1 = ABILITY1; F2 = ABILITY2;
 8  /EQUATIONS
 9      V1 =  20*V999 +     F1 + E1;
10      V2 =  20*V999 + 0.9*F1 + E2;
11      V3 =  20*V999 +     F2 + E3;
12      V4 =  20*V999 + 0.9*F2 + E4;
13      F1 =    0V999          + D1;
14      F2 =    0V999 + 0.9*F1 + D2;
15  /VARIANCES
16      D1 = 30*; D2 = 5*;
17      E1 TO E4 = 10*;
18  /COVARIANCES
19       E2,E4=5*;
20  /MATRIX                        ! LOW VARIANCES WHEN COMPARED TO EXPERIMENTALS
21    37.626
22    24.933 34.680
23    26.639 24.236 32.013
24    23.649 27.760 23.565 33.443
25  /MEANS
26    18.381 20.229 20.400 21.343  ! LOW MEAN OF V1, COMPARED TO EXPERIMENTALS
27  /END
28  /TITLE
29     OLSSON'S DATA (SORBOM, 1978), EXPERIMENTAL GROUP
30  /SPECIFICATIONS
31      VARIABLES = 4; CASES = 108; ANALYSIS = MOMENT;
32  /LABELS
33      V1 = SYNONYM1; V2 = OPPOSIT1; V3 = SYNONYM2; V4 = OPPOSIT2;
34      F1 = ABILITY1; F2 = ABILITY2;
35  /EQUATIONS
36      V1 =  20*V999 +     F1 + E1;
37      V2 =  20*V999 + 0.9*F1 + E2;
38      V3 =  20*V999 +     F2 + E3;
39      V4 =  20*V999 + 0.9*F2 + E4;
40      F1 =   2*V999          + D1;
41      F2 =   2*V999 + 0.9*F1 + D2;
42  /VARIANCES
43      D1 = 30*; D2 = 5*;
44      E1 TO E4 = 10*;
45  /COVARIANCES
```

```
46       E2,E4=5*;
47   /MATRIX
48     50.084
49     42.373 49.872
50     40.760 36.094 51.237
51     37.343 40.396 39.890 53.641
52   /MEANS
53     20.556 21.241 25.667 25.870 ! HIGHER POST-TEST MEANS THAN CONTROLS
54   /CONSTRAINTS                  ! SET COMMAND GENERATES EQUALITY CONSTRAINTS ON
55     SET = GVV, BVF, BFF;        ! ALL FREE PARAMETERS IN MEASUREMENT EQUATIONS.
56   /LMTEST
57   /END
```

Before examining the output, some comments on the data are in order. There would seem to be some question as to whether the two groups were in fact randomly assigned to conditions, or, if so, whether randomization may have failed. In particular, one would expect the pre-test means and variances and covariances for the two groups to be equal, but this does not appear to be so. Such a hypothesis can be tested by another structural model, but this will not be done here because, in the end, a complete model of all the data must be obtained. The results of Sörbom's model are presented below.

```
 MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS ! CONTROLS
 STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

SYNONYM1=V1  =   1.000 F1    +18.619*V999  + 1.000 E1
                               .597
                              31.205@

OPPOSIT1=V2  =    .878*F1    +19.910*V999  + 1.000 E2
                  .051          .544
                17.286@       36.603@

SYNONYM2=V3  =   1.000 F2    +20.383*V999  + 1.000 E3
                               .538
                              37.882@

OPPOSIT2=V4  =    .907*F2    +21.203*V999  + 1.000 E4
                  .053          .534
                17.031@       39.719@
```

The factor loadings look good, and the intercepts for the Vs are nearly equal to the means in the Control group.

```
 CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS ! CONTROLS
 STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

ABILITY1=F1  =   1.000 D1

ABILITY2=F2  =    .895*F1    + 1.000 D2
                  .052
                17.145@
```

The factor is quite stable from pre-test to post-test. The variances and covariances are not presented to save space, though it should be noted that the variances for D1 and D2 are substantially higher in the Experimental as compared to the Control group. In the Experimental group, the following Construct Equations are obtained (Measurement Equations are the same as for Controls).

```
 CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS ! EXPERIMENTALS
 STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

ABILITY1=F1  =   1.875*V999  + 1.000 D1
                  .899
                 2.085@

ABILITY2=F2  =    .895*F1    + 3.628*V999  + 1.000 D2
                  .052          .480
                17.145@       7.558@
```

Since the Control's V999 → F1 path is set to zero, the Experimental's comparable path shows that Experimental subjects were significantly higher in the verbal ability factor F1 at pre-test ($z = 2.085$). Thus there is some reason to doubt that the two groups were initially equal in ability. (The intercept for F1 is the mean, since there are no indirect paths from V999 to F1.) Given that they may have been higher in ability, nonetheless the experimental training on the verbal materials improved that group's subsequent performance, when compared to the controls. This can be seen in the intercept for F2, which is significantly greater than zero, reflecting the observed mean differences between the groups on V3 and V4. Overall, the model is also acceptable, with $\chi^2 = 3.952$, based on 5 df, having an associated probability of .556.

An obvious question is how this model might perform when the intercepts for F1 are forced to be equal for Controls and Experimentals. Such a specification would be consistent with a randomized assignment of subjects to conditions. Rerunning the above model with the path V999 → F1 set to zero in both groups, the following results are obtained. While the model is statistically acceptable by the χ^2 goodness-of-fit test, the LM test shows a definite problem.

```
  CHI-SQUARE =        8.234 BASED ON     6 DEGREES OF FREEDOM
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS     0.22144

  LAGRANGE MULTIPLIER TEST (FOR RELEASING CONSTRAINTS)

  CONSTRAINTS TO BE RELEASED ARE: ! CONSTRAINTS GENERATED BY SET COMMAND IN CONSTRAINT
SECTION

          CONSTRAINTS FROM GROUP  2

          CONSTR:    1   (1,V1,V999)-(2,V1,V999)=0;
          CONSTR:    2   (1,V2,V999)-(2,V2,V999)=0;
          CONSTR:    3   (1,V3,V999)-(2,V3,V999)=0;
          CONSTR:    4   (1,V4,V999)-(2,V4,V999)=0;
          CONSTR:    5   (1,V2,F1)-(2,V2,F1)=0;
          CONSTR:    6   (1,V4,F2)-(2,V4,F2)=0;
          CONSTR:    7   (1,F2,F1)-(2,F2,F1)=0;

           UNIVARIATE TEST STATISTICS:

   NO    CONSTRAINT    CHI-SQUARE   PROBABILITY
   --    ----------    ----------   -----------
    1    CONSTR:   1        5.173       0.023
    2    CONSTR:   2        1.204       0.272
    3    CONSTR:   3        0.003       0.954
    4    CONSTR:   4        0.003       0.954
    5    CONSTR:   5        0.167       0.683
    6    CONSTR:   6        0.076       0.782
    7    CONSTR:   7        1.120       0.290
```

It is apparent that the first constraint (equal intercepts for V1 across the two groups) may be implausible. Consequently a final model will be considered in which this constraint is released. Note particularly that, as a result, the basic level of V1 across groups is not forced to be the same. This final model shows the following results.

```
  MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS ! CONTROLS
  STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

 SYNONYM1=V1  =   1.000 F1    +18.738*V999  + 1.000 E1
                                 .541
                               34.637@

 OPPOSIT1=V2  =    .888*F1    +20.651*V999  + 1.000 E2
                    .051          .442
                 17.353@       46.736@

 SYNONYM2=V3  =   1.000 F2    +20.768*V999  + 1.000 E3
                                 .465
                               44.659@
```

```
OPPOSIT2=V4  =    .891*F2    +21.607*V999  + 1.000 E4
                  .055          .461
                16.307@       46.880@

 CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
 STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

ABILITY1=F1  =   1.000 D1

ABILITY2=F2  =    .906*F1    + 1.000 D2
                  .053
                17.165@

 MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS ! EXPERIMENTAL
 STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

SYNONYM1=V1  =   1.000 F1    +20.002*V999  + 1.000 E1
                                 .515
                               38.811@

OPPOSIT1=V2  =    .888*F1    +20.651*V999  + 1.000 E2
                  .051          .442
                17.353@       46.736@

SYNONYM2=V3  =   1.000 F2    +20.768*V999  + 1.000 E3
                                 .465
                               44.659@

OPPOSIT2=V4  =    .891*F2    +21.607*V999  + 1.000 E4
                  .055          .461
                16.307@       46.880@

 CONSTRUCT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS
 STATISTICS SIGNIFICANT AT THE 95% LEVEL ARE MARKED WITH @.

ABILITY1=F1  =   1.000 D1

ABILITY2=F2  =    .906*F1    + 4.342*V999  + 1.000 D2
                  .053          .538
                17.165@        8.075@

 CHI-SQUARE =        2.962 BASED ON     5 DEGREES OF FREEDOM
 PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS     0.70590

 BENTLER-BONETT NORMED    FIT INDEX=       0.998
 BENTLER-BONETT NONNORMED FIT INDEX=       1.010
 COMPARATIVE FIT INDEX (CFI)       =       1.000
```

The equal intercepts for F1 across groups suggests that the children in the two groups may well have been equal in mean verbal ability at pre-test, but that, for reasons that cannot be ascertained from the analysis, the Controls had a lower mean on V1 as well as lower variance on D1, and hence, on verbal ability F1 (variances not shown above). While the model fits even better than the model proposed by Sörbom, ideally, a further understanding of the experimental procedure would be needed in order to explain the pre-test imbalances between conditions. Nonequivalence of groups should occur relatively rarely by chance alone under random assignment at the given sample sizes (e.g., Hsu, 1989).

10. GROWTH CURVE MODELS

Hopefully you already will have encountered trend analysis or orthogonal polynomials in the analysis of variance, and similarly, the use of powers of predictor variables to model nonlinear relations among predictors and a quantitative dependent variable. Maxwell and Delaney (2004, Ch. 6) and Cohen, Cohen, West, and Aiken (2003, Ch. 6) provide a good introduction to trend analysis and polynomial regression, respectively. See also Cudeck and du Toit (2002). This chapter discusses trend analysis in the context of structural equation modeling. The basic idea is that the observed variables z are a function of time, that is, $z = z(t)$, where t is a discrete index or measure of time, and hence any model for z also will be a function of t. In SEM, growth curve models represent models for trends, trajectories, or growth in a repeatedly measured variable, or a set of such variables, at the level of the individual. Thus we might write for individual i, $z_i = z_i(t)$, not only emphasizing the trend for each individual, but also explicitly recognizing that such trends may differ among individuals. Since one can treat time t as a fixed explanatory variable on which one regresses outcome variables for individuals and groups, it is thus not surprising to know that growth curve models also can be represented as multilevel models (e.g., Chou, Bentler, & Pentz, 1998; MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997; Raudenbush, 2001a,b; Singer & Willett, 2003; Willett & Sayer, 1994), and vice versa (Bauer, 2003). We do not discuss multilevel model variants of growth curve models in this chapter, and concentrate on the SEM approach. In the SEM approach, these models are parameterized in a typical mean and covariance structure way, with the further feature that time is modeled via the factor loadings. The special feature of these models is that the parameters can be interpreted to represent what happens at the level of the individual as well as at the level of the group.

The kind of data that growth curve modeling is intended to model can be illustrated with the following data on the growth of 13 male mice from days 3 to 21 after birth, taken from Appendix A1 of Pan and Fang (2002). The 13 by 7 7-variable data matrix is reproduced in the file mouse.ess.

Weight of Mice (in lbs.) on Day No.							
	3	6	9	12	15	18	21
	V1	V2	V3	V4	V5	V6	V7
1	0.190	0.388	0.621	0.823	1.078	1.123	1.191
2	0.218	0.393	0.568	0.729	0.839	0.852	1.004
3	0.211	0.394	0.549	0.700	0.783	0.870	0.925
4	0.209	0.419	0.645	0.850	1.001	1.026	1.069
5	0.193	0.362	0.520	0.530	0.641	0.640	0.751
6	0.201	0.361	0.502	0.530	0.657	0.762	0.888
7	0.202	0.370	0.498	0.650	0.795	0.858	0.910
8	0.190	0.350	0.510	0.666	0.819	0.879	0.929
9	0.219	0.399	0.578	0.699	0.709	0.822	0.953
10	0.225	0.400	0.545	0.690	0.796	0.825	0.836
11	0.224	0.381	0.577	0.756	0.869	0.929	0.999
12	0.187	0.329	0.441	0.525	0.589	0.621	0.796
13	0.278	0.471	0.606	0.770	0.888	1.001	1.105
Mean	.211	.386	.551	.686	.805	.862	.950
SD	.024	.035	.057	.106	.139	.141	.124

Table 10.1

In this data set, t is the number of days since birth, and varies from 3 to 21. We could transform the time scale, if desired, to any other meaningful scale, e.g., we could let t be the integers 1, 2, …, 7 representing the occasions of measurement. Each row of the matrix represents a given individual i whose trajectory we would like to model. Clearly, each individual mouse increases in weight across the 7 measurement occasions, so there is a monotonic increase for all mice. But the mice are not identical in their growth patterns. At day 3, the typical mouse weighs around $1/5^{th}$ of a pound, and at day 21, it weighs close to one pound, but there are certainly individual differences. Some mice (e.g., #5) never reach a full pound; others (#1) reach this weight quickly. In fact, the individual

differences grow with time, as the SDs increase with time (except at day 21). Growth curve modeling aims to explain the individual trends as well as the average trend.

It may also be helpful to visualize these data by plotting the trajectories of the individual mice across the measurement occasions. A plot is given below.

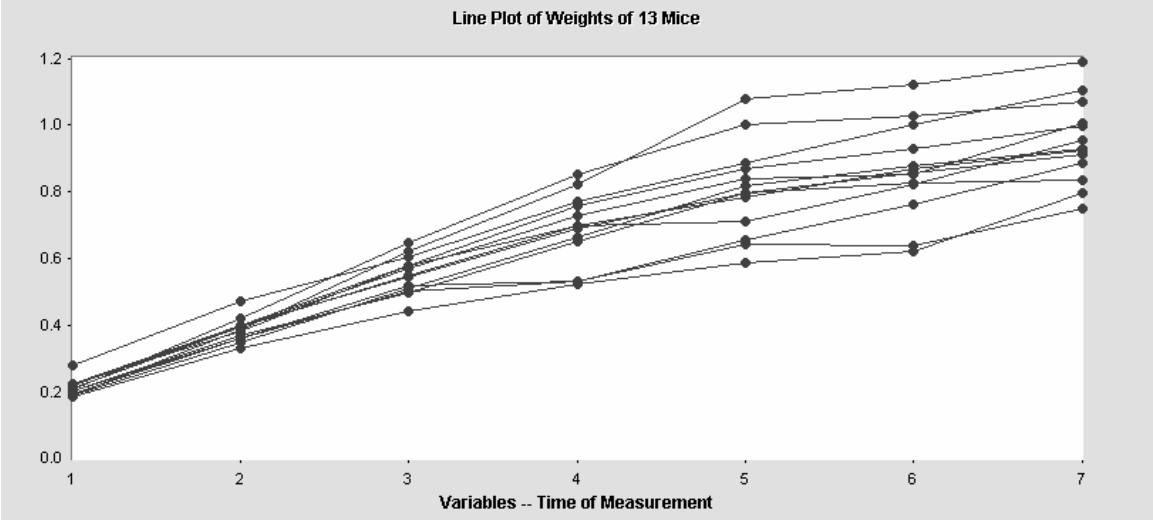


Figure 10.1

In EQS, the figure is in color, and the 13 individual trends are color coded. Even in black and white, the figure clearly reveals an upward trend in weight of all mice, an average trend that is linear except that for some mice there may be a leveling off at the last few measurements. Perhaps this implies an added quadratic trend. Growth curve modeling is intended to parameterize the average trend shown in the figure, and also to characterize the individual differences, especially, individual differences in the starting point, and individual differences in the parameters of the trends. For example, some mice appear to have a larger slope than others, in the sense that they grow somewhat more quickly. A SAS macro for plotting individual curves is given by Carrig, Wirth, and Curran (2004).

Individual and Group Curves

If each trajectory in Figure 10.1 is primarily a straight line, we might consider modeling the weight of an individual mouse *i* at time *j* using an *individual* linear regression equation such as

$$z_{ij} = \alpha_i + \beta_i t_j + \varepsilon_{ij} \ , \qquad (10.1)$$

where α_i is the intercept and β_i is the slope of that individual's line. The data – say, the top trajectory in Figure 10.1) -- may not precisely lie on the line, so there is a residual error term ε_{ij}. The predictor variable t_j is some coding for the seven time points, such as 0, 1, 2, 3, 4, 5, 6. If the coding has $t_j = 0$ for the initial time point j=1, then α_i is the initial weight of mouse *i* according to the linear model (the observed weight may deviate from α_i due to the error). In principle, we can have one equation such as (10.1) for each individual. In the mouse example, there are 13 mice, so we could have 13 straight lines characterized by equations of the form (10.1). As the data of Day 3 in Table 10.1 show, mice have different initial weights, so the individual α_i (i=1,…,13) will differ across mice. But on average, α_i should be near .21, with a bit of variance. Similarly, as Figure 10.1 shows, different mice have different trends, with some having a rapid growth (larger β_i values), and others growing more slowly (smaller β_i values). But all the β_i will be positive as the trend is increasing.

It is clear that the α_i and β_i thus have a distribution. They are sometimes called the *random components of growth*. If we treat them like other random variables, we can inquire about their means and variances. This implies that we can consider the simple model

$$\alpha_i = \mu_\alpha + \delta_{\alpha i} \text{ and } \beta_i = \mu_\beta + \delta_{\beta i}\,. \tag{10.2}$$

All we have done is to decompose each individual's intercept and slope into a group mean and the individual's deviation from mean. The mean carries the group information and the deviations carry the individual difference information. Taking expectations of (10.1) under the further structure (10.2), we get the straight line given by the linear model for the means

$$\mu_j = \mu_\alpha + \mu_\beta t_j\,, \tag{10.3}$$

where μ_j is the model-based estimate of the observed means of the data at time *j* (that is, μ_j should approximate the j^{th} sample mean at the bottom of Table 10.1). This is the group average curve relating the model based variable means to the growth curve mean parameters μ_α and μ_β of the model, which are sometimes called the *fixed components of growth*. The variances of the deviation from means in (10.2) carry the individual differences among mice at the initial time point (the variance of $\delta_{\alpha i}$) and of the slopes (the variance of $\delta_{\beta i}$). Possibly, $\delta_{\alpha i}$ and $\delta_{\beta i}$ may correlate; that is, across individuals, initial level and slope may covary.

As the random variables α_i and β_i are not in our data file, in growth curve modeling we will treat them like latent common factors in factor analysis with factor loadings given by the coded time variable t_j. Under the decomposition (10.2), however, we clearly have a mean structure as well. So we have a factor model with a mean structure. The variances and correlations of $\delta_{\alpha i}$ and $\delta_{\beta i}$, and variances of the residuals ε_{ij} that we can treat like unique variates in the factor model, generate a covariance structure. Thus this growth curve model is a type of model discussed in Chapter 8. As discussed in that chapter, some modifications to the setup are used to run such a model in EQS. This type of model, and its extensions, is discussed further below.

Advantages and Limitations

With their emphasis on both individual- and group-level phenomena, growth curve models accomplish very old goals in the measurement of growth or change. In developmental psychology, a long-standing concern has been to avoid confusing individual differences with intraindividual differences, that is, to separate intra-individual patterns of change from inter-individual patterns of change (e.g., Daly, Bath, & Nesselroade, 1974; Nesselroade, 1991). In personality theory, there has been a concern with short term and long term mean, variance, cyclical, and dynamic change across time (e.g., Biesanz, West, & Kwok, 2003). Similarly, in experimental psychology, individual trends in learning or memory processes often are plotted as a function of time, and because such trends vary substantially across individuals, such trends sometimes were averaged in order to obtain a smooth representation of the process involved. However, it was recognized a long time ago that the average curve might not faithfully summarize the curves of the individuals (e.g., Bakan, 1954; Sidman, 1952). Yet, in spite of this, artifacts stemming from the assumption that average curves represent the curves of individuals have continued until today. For example, for a long time it was thought that curves of practice are better described by power functions than by exponential functions. "However, this entire development depended on the fitting of curves to group data, averaged over subjects…there is a pervasive tendency for averaged data to be fit better by power functions even when performance of the individuals is known to conform to exponential functions" (Estes, 2002, p. 6). By assuring that the same functional form holds for the mean structure as the covariance structure, growth curve models within the SEM tradition can assure that such an artifact does not occur.

If you want trajectories of scores across time to be consistent between the individual and group levels of analysis, it makes sense to concentrate on models that are *linear in the parameters.* Such models include polynomial functions of time, e.g., linear, quadratic and so on. As described by Singer and Willett (2003), these functions have *dynamic consistency* (Keats, 1983), meaning that the functional forms or shapes for individual and average trajectories are equivalent, or, the "curve of averages" is the same as the "average of the curves." If individual curves are quadratic,

so will be the group average curve. Of course, polynomial functions of time also have limitations that may make them unsuitable to certain applications. Specifically, polynomial trajectories are unbounded, that is, they do not reach upper or lower limits, or asymptotes, below infinity. Yet, this may be a logical requirement in some contexts. If you need to consider a wider classes of functional forms, and are willing to give up the property of dynamic consistency, you could consider functions such as logistic, hyperbolic, inverse polynomial, and exponential. These are not discussed in this introductory chapter; Singer and Willett (Ch. 6) provide an excellent discussion. See also du Toit and Cudeck (2001).

Even though it makes conceptual sense to talk about the curves of individuals, and they are easy to visualize as in Figure 10.1, in SEM it is statistically impossible to extract parameters for the curve of each individual, e.g., the α_i and β_i in equation (10.1) or the $\delta_{\alpha i}$ and $\delta_{\beta i}$ in eq. (10.2). Rather, consistency among individual curves is described by the variances of the parameters associated with the average curve. Further, as noted above, variances and other parameters of most structural equation growth curve models turn out to be subsets of mean and covariance structure models, previously introduced in Chapter 8. Since $z = z(t)$ and $\mu = \mu(\theta)$ and $\Sigma = \Sigma(\theta)$, in growth curve models $\mu = \mu(\theta, t)$ and $\Sigma = \Sigma(\theta, t)$. In EQS, the usual Bentler-Weeks form applies, namely, the model-reproduced means are $\mu = G(I-B)^{-1}\Gamma\mu_\xi$ and the model-implied covariances are $\Sigma = G(I-B)^{-1}\Gamma\Phi\Gamma'(I-B)^{-1}'G'$ (see eq. (1.4)). A specialized growth curve variant of the Bentler-Weeks model with $\Gamma = I$ yields eqs. (64) and (65) of Bollen and Curran's (2004) autoregressive latent trajectory (ALT) model. Typically, the simpler confirmatory factor structure $\mu = \Lambda\mu_\xi$ and $\Sigma = \Lambda\Phi\Lambda' + \Psi$ is employed as a base model. In any case, the SEM approach implies that *the data requirements for growth curve modeling are the standard ones in mean and covariance structure analysis* – a reasonable but not too large number of variables (with at least one repeatedly measured variable) and a reasonably large sample of subjects. The mouse data in Table 10.1 are certainly on the very low end of sample size. These data requirements set growth curve modeling apart from time series analysis, in which there are one or few subjects and a huge number of repeated measures (e.g., Cudek & Klebe, 2002; Li & Shedden, 2002; Velicer & Fava, 2003); functional data analysis, with a similar structure but with repeated measures so frequent as to be essentially continuous in nature (e.g., Lee, Zhang, & Song, 2002; Ramsey & Silverman, 1997; 2002); and differential structural equation models, with a huge number of essentially continuous differences measured on a few to a large number of subjects (e.g., Arminger, 1986; Boker, 2001; Dwyer, 1992).

While growth curve models are valuable tools to model trajectories over time, they also have limitations. The basic design for measurement has to be *balanced on time* (Ware, 1985), meaning that all subjects' data is obtained at about the same time and that the repeated measurements are spaced the same way for all subjects. This implies that the coding t_j is the same for everyone. This requirement can be loosened, e.g., by coding for the smallest possible spacing and treating absent measurements as missing data (see Chapter 12). Also, it was pointed out in Chapter 8 that mean and covariance structure models can be very hard to fit to data, since both the mean structure and the covariance structure must be correct. Growth curve models additionally add very strong restrictions on these parameters -- for example, factor loadings in Λ contain fixed instead of free parameters -- implying that these models can be even harder to fit than ordinary mean and covariance structure models. Of course, the typical requirement still holds that the statistical theory invoked should adequately reflect the possible non-normal distribution of the variables being analyzed. Unfortunately, many reports of growth curve modeling use normal theory methods without justification and provide little in the way of traditional measures of model fit. In such cases, it is appropriate to be skeptical about the adequacy of the models.

An important point not encountered in standard mean and covariance structure models is that *growth curve models contain arbitrary features* that may impede generalizing conclusions. Time *t* usually enters the structure of a given model via familiar model parameters such as factor loadings. As a result it seems that nothing new is involved. However, the particular interpretation of the parameters in a given model depends on how *t* is coded. An excellent discussion is given in Biesanz, Deeb-Sossa, Papadakis, Bollen, and Curran (2004). Different coding schemes imply different parameters. "The contention that conclusions drawn from latent growth curve analysis are sensitive to the way time is incorporated in the model is less well documented. This lack of invariance of the growth parameters can be seen as a consequence of the time scale involved. When different time scales are incorporated in the models, different values of the initial status and the growth rate will be obtained, as well as of their (residual) variances and

covariances (Garst, 2000; Mehta & West, 2000; Rogosa & Willett, 1985; Rovine & Molenaar, 1998a; Rudinger & Rietz, 1998; Stoolmiller, 1995). In the literature on latent growth curve models, however, no attention has been paid to the fact that estimated effects of exogenous covariates (e.g. IQ) on the initial status and growth rate may also depend on the time scale involved. Merely a short comment (Stoolmiller, 1995, p. 118) suggests the interference" (Stoel, 2003, p. 28). Stoel documents the latter limitation in detail in his excellent book; see also Garst (2000) and Stoel and van den Wittenboer (2003). Aspects of this issue are discussed further below.

The literature on growth curve models has a long history in a variety of fields, where different approaches and emphases have prevailed. In the statistics field, an important originator was Rao (1958). Generalized multivariate analysis of variance models (see Pan & Fang, 2002) and generalized linear models (Liang & Zeger, 1986; Diggle, Liang & Zeger, 1994) have provided a main focus, while random effects (Laird & Ware, 1982), random coefficient (Longford, 1993b), or mixed effects (Cudek & Klebe, 2002; Wu & Zhang, 2002) models, closely related to multilevel models, also have enjoyed extensive research and application. These types of models can be related to SEM models (e.g., Rovine & Molenaar, 2000, 2001). Rogosa, Brandt, and Zimowski (1982), Rogosa and Willett (1985), and Strenio, Weisberg, and Bryk (1983) were influential in helping to refocus attention on the importance of individual trajectories. Tucker (1958) was a key originator of the latent variable or latent curve approach, which was further extended by McArdle (1986, 1988), McArdle and Epstein (1987), and especially Meredith and Tisak (1990).

Good introductions to the structural modeling approach to growth curve models can be found in Byrne and Crombie (2003), Curran (2000), Curran and Hussong (2002), Duncan, Duncan, Strycker, Li, and Alpert (1999), McArdle and Nesselroade (2003), Mehta and West (2000), and Stoel (2003), all of whom provide many additional references and applications. Only the basics can be covered in this chapter. Some early applications that include tutorial material include Duncan and Duncan (1995), Duncan, Duncan, and Stoolmiller (1994), and Stoolmiller (1994). Interesting recent applications can be found, for example, in Bentler, Newcomb and Zimmerman (2002), Chan, Ramey, Ramey, and Schmitt (2000), Curran (2000), Curran and Bollen (2001), Curran and Muthén (1999), Duncan, Duncan, Biglan and Ary (1998), du Toit and Cudeck (2001), Ferrer & McArdle (2003), Figueredo, Brooks, Leff, and Sechrest (2000), Frosch, Stein, and Shoptaw (2002), Ghisletta and McArdle (2001), Khoo (2001), Khoo and Muthén (2000), Li, Harmer, McAuley, Fisher, Duncan, and Duncan (2001), Llabre, Spitzer, Saab, and Schneiderman (2001), McArdle (2001), Muthén and Curran (1997), Raykov (2000), Sayer and Cumsille (2001) and Scott and Hancock (2001). Chapter 1 gives many further references. The role of factorial invariance is discussed by Meredith and Horn (2001). Power can be a problem when trying to isolate small group differences (Fan, 2003). Although recently popular, extensions to mixture models in which there are different types of growth for different people (e.g., Colder, Campbell, Ruel, Richardson, & Flay, 2002; Li, Duncan, Duncan, & Acock, 2001; Muthén, 2000, 2001a, b, 2002) remain controversial, as mixtures of normal distributions can be artificially discovered even when data conform to a single nonnormal population (Bauer & Curran, 2003). It is not yet known whether newer clustering approaches also have the same weakness (James & Sugar, 2003).

Although this chapter deals with growth curve models, you should not assume that this type of model is necessarily the best one to describe change across time in every application of SEM. Some alternatives you can consider are the hierarchical linear or multilevel models discussed in Chapter 11, the latent state-trait models of Steyer and his colleagues (e.g., Steyer & Partchev, 2001; Steyer, Partchev, & Shanahan, 2000; Steyer, Schmitt, & Eid, 1999), the stable trait and autoregressive trait and state model (Kenny & Zautra, 1995; 2001), varieties of autoregressive and cross-lag panel models (e.g., Bentler, Newcomb, & Zimmerman, 2002; Burkholder & Harlow, 2003; Kessler & Greenberg, 1981), SEM-type dynamic factor or state-space models (e.g., Molenaar, 1985; Molenaar, de Gooijer, & Schmitz, 1992; Oud & Jansen, 2000; Wood & Brown, 1994), and combinations of model types (e.g., Bollen & Curran, 2004; Curran & Bollen, 2001; Green & Hershberger, 2000). Any model that is only covariance structure based, can, if desired, be extended with a mean structure. Of course, a still wider class of models exists outside the SEM tradition (e.g., event occurrence models, Singer & Willett, 2003; nonparametric mixed effect models, Wu & Zhang, 2002).

Structural Models

Although as shown in equation (10.1) above, the motivation for growth curve models comes from tracking the trajectories of individual as well as groups, in the end these models are structural models. In this section we expand on this aspect.

*A basic growth curve **mean structure** is modeled by a confirmatory factor model with*

(1) fixed factor loadings based on the degree of a polynomial; and

(2) factor means as free parameters.

This is enough to model the means. Thus, specifically, for a given individual we have the confirmatory factor model

$$z = \Lambda\xi + \varepsilon \tag{10.4}$$

where the factor loading matrix Λ contains fixed parameter codings for the growth curve factors (explained further below), the latent factors ξ have expectation μ_ξ, and the residual variables ε have expectation zero. What makes (10.4) different from any arbitrary covariance structure model is that, *in the context of growth curve models, the factors ξ are conceptualized as the scores of an individual on the particular specific hypothesized growth factors, such as intercept, slope, or quadratic factors.* Referring back to the illustrative model given in (10.1)-(10.2), in that example there are two factors α_i and β_i, so ξ would contain two elements. In addition to (10.4), there is a mean structure of the observed repeatedly measured variables, given as

$$\mu = \Lambda\mu_\xi \,. \tag{10.5}$$

So, under the model, the means of the observed variables are a linear combination of the coded growth curve loadings and the means of the factors. A particular example was shown in (10.3). Equation (10.5) is often referred to as describing the *fixed* component of growth.

Under standard SEM assumptions, the covariance structure based on (10.4) is well known to be $\Sigma = \Lambda\Phi\Lambda' + \Psi$, but various further structures can be considered. In the context of the general manova approach to growth curves, Pan and Fan (2002, p. 20) give 6 logical covariance structures. To their six, we add two more. So we have:

*A basic growth curve **covariance structure** is modeled by a confirmatory factor model with*

M_1: Σ is an arbitrary positive definite matrix (saturated model)

M_2: $\Sigma = \Lambda\Phi\Lambda' + \Lambda_c\Theta\Lambda_c{}'$

M_3: $\Sigma = \sigma^2\{(1-\rho)I_p + \rho 11'\}$

M_4: $\Sigma = \Lambda\Phi\Lambda' + \sigma^2 I_p$ (10.6)

M_5: $\Sigma = \sigma^2(\rho^{|i-j|})_{1\le i,j\le p}$

M_6: $\Sigma = \sigma^2 G$ for a given matrix G.

M_7: $\Sigma = \Lambda\Phi\Lambda' + \Psi$

M_8: $\Sigma = \Sigma(\theta)$.

M_1 is appropriate when you are only interested in modeling mean growth. M_2 is Rao's simple covariance structure, an unusual structure for SEM in which the residual matrix is structured as the orthogonal complement Λ_c of the factor loading matrix Λ. M_3 is the uniform or compound symmetry model of equal variances and equal covariances. M_4 is called a random coefficient regression structure, but amounts to a factor model with equal residual variances (Hayashi & Bentler, 2000), which may be useful when sample size is small. M_5 is a serial covariance structure. M_6 allows the covariance matrix to be proportional to a known matrix. M_7 is the usual confirmatory factor structure, which is the typical structure used in SEM growth curve modeling. M_8 is clearly the most general model, stating that any structure is possible. In between M_7 and M_8 are a whole series of additional models. For example, if we take $\Phi = \Gamma\Psi\Gamma'$ in M_7, we have a second-order factor model, and if we take $\Lambda = G(I-B)^{-1}\Gamma$ we have the Bentler-

Weeks model. In any case, the random variables in (10.4) and their covariance structure (10.6) describe what is sometimes called the *random* component of growth.

The various covariances structures (10.6) are not often discussed in latent growth curve models. The list should assure you that the usual structure $\Sigma = \Lambda\Phi\Lambda' + \Psi$ is not the only possibility, though it is probably the most logical choice under dynamic consistency. Also note that the mean structure (10.5) may or may not be true somewhat independently on the covariance structure M_7, and that these aspects can be tested by appropriate use of chi-square difference tests. For example, the mean structure can be specifically tested by comparing to a model with an unstructured set of means (e.g., Mandys, Dolan, & Molenaar, 1994).

Obviously, the basic growth curve factor model can be extended in many ways, e.g., by making ξ in (10.4) dependent variables predicted by other variables. This would make what is sometimes called a *conditional* growth curve model. The varieties of model extension are huge, as any SEM structure can be accommodated. For example, if *z* is a vector of first-order factors (e.g., $z = \eta$ in a model for observed scores *x*, as in $x = \Delta\eta + \zeta$), then the decomposition above provides a second-order growth curve model; see Hancock, Kuo and Lawrence (2001) and Sayer and Cumsille (2001) for a detailed discussion.

The model given by (10.4)- (10.6) *is not scale invariant*. This means that you cannot arbitrarily rescale the variables and expect to maintain the hypothesized structure. The observed scores, or repeated measures, must all be quantified in the same metric.

The Initial Status Model

A variety of options for Λ in (10.4) are available in the SEM approach to growth curve modeling. This has to do with how you want to code time. The most popular way is with the *Initial Status Model*. In the mouse data of Table 10.1, there is one conceptual variable measured at 7 time points, meaning that there are 7 data variables organized sequentially in the data file (V1-V7). We noted that this data represented largely linear growth with perhaps a quadratic component. But if we had a very large sample size, and we suspected the need for higher order growth factors, we could model this data with 5 factors (F1-F5) with a 7 by 5 factor loading matrix. Its elements are the numbers shown in Table 10.2.

Variable or Time	Intercept F1	Slope F2	Quadratic F3	Cubic F4	Quartic F5
V1	1	0	0	0	0
V2	1	1	1	1	1
V3	1	2	4	8	16
V4	1	3	9	27	81
V5	1	4	16	64	256
V6	1	5	25	125	625
V7	1	6	36	216	1296

Table 10.2

In this factor loading matrix, there are 5 factors (F1-F5, or Intercept, Slope, Quadratic, Cubic, and Quartic factors). The factor loadings of all variables on the intercept factor are fixed 1.0 values. V1 has a loading of zero on the slope factor, and subsequent variables have loadings 1, 2, etc. incremented in equal units. The loadings on the quadratic factor are just the squares of those of the slope factor. In turn, the loadings on the cubic and quartic factors are the cubes and 4th-power, respectively, of those on the slope factor. So, for example, the specific equation for V3 with all of the factors plus an added residual error term is given by

V3 = 1F1 + 2F2 + 4F3 + 8F4 + 16F5 + E3;

This is ready to be input into the EQS program. Notice there is no "*" since the coefficients are known.

In a given study, you probably would not use all of the above factors. The number and kind of growth factors you use would depend on the type of growth that you expect, as well as on the data available. More variables are needed to identify higher-order trends, four for a quadratic trend, five for a cubic trend, and so on. If the data show that there is no increment across time in mean scores, i.e., a horizontal flat line, then only the intercept is needed. If there are straight line increments or decrements, both intercept and slope factors most likely would be needed. If there was an upward (downward) trend with some leveling off, a quadratic factor might be appropriate.

Cubic and quartic growth are almost never proposed as models for real data in practice, as this would require extremely reliable data on very large sample sizes in order to be able to distinguish – as well as a good theory to interpret – the oscillating trends in growth. Notice furthermore how the elements of the loading matrix for these factors (F4 and F5) would grow to be remarkably large numbers with only a few more measurement occasions. The impact of the later measures thus becomes especially large.

With regard to the slope factor, an equal spacing of loadings is appropriate only if the data given by V1-V7 are obtained equally spaced in time, say, 1 week apart (or 1 month, or 3 months or 5 years). In the mouse data, measurements are 3 days apart. With unequal spacing, the *loadings on the slope factor have to be made proportional to the time differences*. If the measurement had occurred at weeks 1, 2, 4, 5, 6, 9, 10, then the slope loadings can be taken as 0, 1, 3, 4, 5, 8, 9. How to measure time in growth curve models can be a tricky issue, as there may be no unique metric. Also, in the SEM approach it is assumed that all subjects are measured at the same time point, that is, that the data is *time-structured*. Although exceptions can be handled (see Bauer, 2003; Mehta & West, 2000), multilevel modeling can deal better with *time-unstructured* data, in which various individuals are measured at different times. One approach in standard growth curve modeling is to have the data matrix as well as the factor loading matrix parameterize all possible times of measurement, implying that data will be missing for some subjects on some occasions. Missing data can be handled flexibly in EQS. See Chapter 12.

Now let us look at equation (10.5), $\mu = \Lambda\mu_\xi$, in some detail under the factor loading matrix given in Table 10.2. In Table 10.3 we show the elements that enter the product (10.2) based on the 5 factors (where $\xi_i = F_i$). Each row shows how a given model-reproduced variable mean is obtained based on the contributions of the factors. For example, for the mean of V3, we have

$$\mu_3 = 1\,\mu_{F1} + 2\,\mu_{F2} + 4\,\mu_{F3} + 8\,\mu_{F4} + 16\,\mu_{F5}.$$

In all cases except for V1, there are as many terms involved in reproducing a variable mean as there are factors in the model. Thus a model that contained only F1-F3 would have just those three additive parts generating the observed mean of any variable.

The row for V1 shows that $\mu_1 = \mu_{F1}$, no matter whether additional factors are included in the model. If only an intercept factor is included in the model, the model-reproduced means of all variables would be equal, and would be equal to μ_{F1}. In practice, of course, sample means will differ slightly and randomly from strict equality, but if they systematically differ, a polynomial of higher-order degree, and its corresponding factor, would be needed.

Variable Mean	Intercept F1	Slope F2	Quadratic F3	Cubic F4	Quartic F5
μ_1 =	$1\,\mu_{F1}$	0	0	0	0
μ_2 =	$1\,\mu_{F1}$ +	$1\,\mu_{F2}$ +	$1\,\mu_{F3}$ +	$1\,\mu_{F4}$ +	$1\,\mu_{F5}$
μ_3 =	$1\,\mu_{F1}$ +	$2\,\mu_{F2}$ +	$4\,\mu_{F3}$ +	$8\,\mu_{F4}$ +	$16\,\mu_{F5}$
μ_4 =	$1\,\mu_{F1}$ +	$3\,\mu_{F2}$ +	$9\,\mu_{F3}$ +	$27\,\mu_{F4}$ +	$81\,\mu_{F5}$
μ_5 =	$1\,\mu_{F1}$ +	$4\,\mu_{F2}$ +	$16\,\mu_{F3}$ +	$64\,\mu_{F4}$ +	$256\,\mu_{F5}$
μ_6 =	$1\,\mu_{F1}$ +	$5\,\mu_{F2}$ +	$25\,\mu_{F3}$ +	$125\,\mu_{F4}$ +	$625\,\mu_{F5}$
μ_7 =	$1\,\mu_{F1}$ +	$6\,\mu_{F2}$ +	$36\,\mu_{F3}$ +	$216\,\mu_{F4}$ +	$1296\,\mu_{F5}$

Table 10.3

Now consider that only intercept and slope factors are used. Such linear growth is frequently used in practice, especially when there are only a few measurement occasions (at least 3) and it is not be possible to use more complicated curves since the model would not be identified. With both factors, it remains true that $\mu_1 = \mu_{F1}$, i.e., this represents the initial level. However, since $\mu_2 = \mu_{F1} + \mu_{F2}$, the increment in means is $\mu_2 - \mu_1 = \mu_{F2}$. This is the size of average growth from occasion to occasion. That is, μ_{F2} is the increment in means between any two subsequent occasions. This growth is constant, that is, the growth is equal in magnitude between any two equally spaced time points, as it should be with linear growth. Of course, if μ_{F2} is negative, there would be a systematic linear decrement in observed means across time.

With the additional inclusion of a quadratic factor, increments in means from occasion to occasion are no longer equal. The precise values need to be calculated for the pairs of occasions involved as an odd multiplier (which grows 3, 5, 7,…) of μ_{F3} is involved. For example, the difference in means between the 5th and 6th occasions is $\mu_{F2} + 9\mu_{F3}$. Suppose now that μ_{F2} is positive, meaning positive incremental growth across time, but that μ_{F3} is negative. Then, as time goes on, the quadratic factor has the effect of slowing down the growth that would be observed without it. The overall effect of the quadratic factor with a negative mean is that, as time goes on, because of the increasing multiplier (3, 5, 7…), the slowdown in average growth would become larger. Of course, if the mean μ_{F3} is positive, growth would increase faster than expected based on the linear model alone.

Clearly, interpreting the trends in variable means becomes more difficult to understand with cubic or quartic factors. Luckily, such higher order functions are not often used to model growth in the social sciences, where limited numbers of observations, data quality, sample size, and lack of continuity of measures limit exploration of complicated functions. But if your theory and data demand it, you should now be able to understand the growth curve approach to doing so.

The Restricted Covariance Structure

The mean and covariance structure of the initial status model is quite restricted. With only intercept and slope factors, the typical situation, the means grow in a meaningful and orderly way, but the variances and covariances under the model are peculiar. In order to get some more insight into the general mean structure (10.5) and typical covariance structure $\Sigma = \Lambda\Phi\Lambda' + \Psi$ under M_7 in (10.6) in this situation, suppose we were modeling an initial status model for just the first three variables of the mouse data in Table 10.1. Graphically, the model would look something like this.

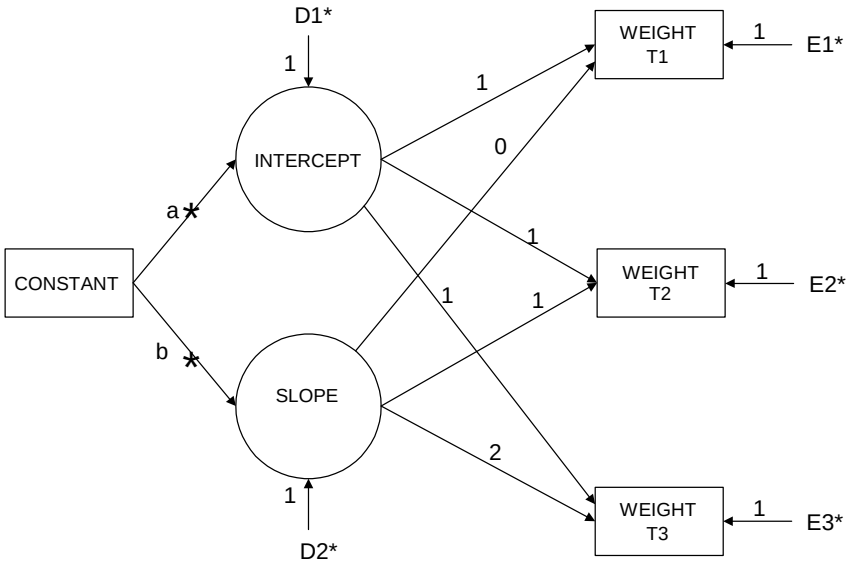


Figure 10. 2

In this figure, the variables V1-V3 are called T1-T3 to emphasize the time sequence of the data. The constant is V999 in EQS, the intercept is F1 with its fixed 1.0 loadings, and the slope is F2 with its incrementing loadings 0,1,2. Since intercept and slope are otherwise (except for V999) independent variables, *a* is the mean intercept and *b* is the mean slope; both are free parameters. Individual differences in intercept are shown in the variance of D1, a free parameter, and individual differences in slope are shown in the variance of D2, also a free parameter. In this particular model, D1 and D2 are not correlated, though this would be an option to consider. Finally, the residuals E1, E2, E3 are unique variates with variances to be estimated. The Es are not correlated. In sum, for some values of the free parameters, this model is meant to reproduce the means and covariances of three observed variables.

Without belaboring the point, if you have followed many of the examples of this *Manual*, it will be obvious that a model of the form of Figure 10.2 can be embedded into a much larger model. Times of measurement can be extended, there could be other set of variables that are predictors of the factors above, or consequences of them, or there could be other variables that have similar or different growth structures at the same time, where the growth factors in one predict, or are predicted by, the others. But this simple prototype can be used to make a few points.

Rather than estimate this model, following Bentler, Newcomb and Zimmerman (2002), we present next the abstract mean and covariance structure of the model, under the hypothesis that the model is correct.

	V1 **Weight T1**	**V2** **Weight T2**	**V3** **Weight T3**
V1	var(D1) + var(E1)		
V2	var(D1)	var(D1) + var(D2) + var(E2)	
V3	var(D1)	var(D1) + 2var(D2)	var(D1) + 4var(D2) + var(E3)
Means	**a**	**a + b**	**a + 2b**

Table 10.4

The mean structure at the bottom of Table 10.4 is easiest to see, and it makes a lot of sense. The mean weight at T1 is *a* of Figure 10.2. The mean weight at T2 has obviously incremented over that at T1 by *b*, and the mean at T3 has incremented again by *b*. Thus clearly *a* is the initial level of the means, and *b* is the increment, which is constant from time point to time point.

The covariance structure is more complicated. Since var(D1) is a part of reproducing var(V1) as well as that of var(V2) and var(V3), and since variances are nonnegative, typically (but not necessarily) the variances of the variables will be increasing over time. Suppose the error variances are equal at all 3 time points. Then the variances definitely increase across time, but the increment is uneven. The increment from T1 to T2 is then given by var(D2), but the increment from T2 to T3 is given by 3var(D2). So the variances increase more rapidly at later time points.

The observed variable V1 at T1 has equal covariances with V2 at T2, and also with V3 at T3. On the other hand, the covariance between 2nd and 3rd measurements increases over these values by 2var(D2).

So a linear growth model of this sort would not fit empirical data if (1) means did not increment more or less equally across time; (2) covariances did not meet the particular pattern shown in Table 10.4. This model may thus be somewhat difficult to fit. Of course, it can be modified by allowing D1 and D2 to covary, or errors to correlate, paths to be freed, or whatever. An aspect of this topic is discussed next.

Freeing Fixed Parameters

By now you will have seen that this class of models is extremely restricted. Structured mean factor models with freely estimated parameters often are hard to fit to real data. It follows that models that also have fixed a priori known factor loadings inevitably will be harder to fit, although of course the special data structure of repeated measurements may improve this poor prognosis.

Consider first that the coefficients in Table 10.2 are raw polynomial coefficients. Since the coefficients on the later factors are all positively correlated (e.g., across V1-V7, there are increasing positive loadings on F2-F5), it can be expected that the factor scores associated with these trends also may be highly correlated. This can lead to estimation problems. In order to avoid this, it is possible to use so-called orthogonal polynomials instead, as we discuss further below.

In principle, you can make your model less restricted by freeing up some fixed parameters. Each freed parameter will use up a degree of freedom, but the possible improvement in fit may be substantial. With regard to the factor loadings, this process can be rationalized by the concept of establishment of a new time scale and allowing the variables to be nonlinearly related to the latent basis functions. When you do this, in general you will need to keep a minimum of two fixed loadings in each column in order to identify the model, and there may be some arbitrariness as to which parameters you free (see e.g., Stoel, 2003; McArdle & Bell, 2000). What freeing fixed loadings does is to maintain the estimation of a linear model, but the interpretation becomes nonlinear because a new reference time scale is established with the resulting new parameter values. This time scale does not relate linearly to the actual time scale. Another way to think about this is that when you free up loadings on the linear slope factor, "we are determining the best-possible-fitting linear spline [see below] under the constraint that the shape and slope are confounded" (Stoolmiller, 1995, pp. 119-120).

It is easiest to see the change by focusing on the means. In the example of Tables 10.2-10.3, we concluded that μ_{F2} was the mean increment in growth between any two occasions when the slope factor loading coding was 0, 1, 2, 3, 4, 5, 6. Suppose now that the loading value 3 is estimated freely, and its estimated value is 2.5. Look at Table 10.3 and imagine 2.5 replacing 3 in the column for F2. This implies that the mean growth between the time corresponding to V2 to V3 remains μ_{F2}, the growth between V3 and V4 is .5 μ_{F2}, the growth between V4 and V5 is 1.5 μ_{F2}, and the growth between subsequent occasions is μ_{F2} again. *If growth were actually linear on some time scale* [note that this is a counterfactual statement], this implies that the actual time scale expanded between V3 and V4 (i.e., there was not enough time to get a whole μ_{F2} increment), and contracted between V4 and V5 (i.e., there

was more time than needed to get only the μ_{F2} increment). In principle, we could compute the time point between V3 and V5 before which, and after which, exactly one unit of μ_{F2} change occurred. For more information, see McArdle and Bell (2000), McArdle and Hamagami (1996), Muthén and Khoo (1997) or Stoel (2003).

If you free a loading on the slope factor, an interesting question is whether you also need to adjust the corresponding variable's loading on quadratic and later factors. After all, in the basic model the quadratic loadings are the squares of the loadings on the slope factor, the cubic's are the cube of those, etc. For coherence of meaning, these relationships should be maintained. In principle, this requires freeing up the loadings on subsequent factors and adding the nonlinear constraint that the quadratic loading should be the square of the freed slope loading, the cubic loading should be the cube of the freed slope loading, and so on. Unfortunately, the current version of EQS does not allow these types of nonlinear constraints.

Researchers who are opposed to post-hoc model modification will require an a priori justification for freeing one or more previously fixed factor loading parameters. This would require the researcher to understand in an a priori way where the actual time scale may be misleading relative to a hypothesized time scale on which change is linear. It would seem to be very hard or impossible with real data, before looking at the data or doing the analysis, to be able to justify which loadings should be freed in a standard growth curve model. Hence, it is probably fair to always be suspicious about such rationales and to entertain the hypothesis that freeing the fixed loadings occurred subsequent to the lack of fit of an a priori model. After all, the Lagrange Multiplier test is designed to inform you whether model fit would be substantially improved if a fixed loading were changed to being free (see Chapter 6). In all likelihood, free parameters in a growth curve loading matrix are usually data-driven, and hence are subject to capitalization on chance.

Does this mean you should never free fixed loadings? Probably not. For example, Ghisletta and McArdle (2001) proposed a rationale for modeling a monotonic free slope factor in which successive loadings were free to be estimated, but forced to not decrease with increasing time. The inequality constraints were rationalized to reflect a growth process that does not recede. Also, most alternative structural equation models have a lot of factor loadings that are free to be estimated. With many freed parameters, at some point you may no longer have an interpretable growth curve model, but you will still have an SEM model.

Transforming the Time Scale

Now for more of the bad news that was mentioned earlier. There is an arbitrariness to the factor loading matrix that is used to parameterize growth in the growth curve model. For example, we can recode the slope factor in Table 10.2 so that the sequence of loadings is reversed, that is, {6, 5, 4, 3, 2, 1, 0} instead of {0, 1, 2, 3, 4, 5, 6}. This has the effect that the intercept of the model is now the last variable rather than the first variable, i.e., it becomes a Final Status model instead of an Initial Status model. As long as we reinterpret our results accordingly, there is no problem as the degree of fit to any given data set will be identical. More general transformations are possible, but we need some matrix algebra to write this compactly. You can probably follow the words without the equations.

Suppose we put a full rank transformation matrix *T* and its inverse into equation (10.4). Nothing substantive changes, but we write

$$z = \Lambda T T^{-1} \xi + \varepsilon = \Lambda^* \xi^* + \varepsilon , \qquad (10.7)$$

where $\Lambda^* = \Lambda T$ and the factors have been transformed into $\xi^* = T^{-1}\xi$. It follows that the mean structure can be similarly transformed as

$$\mu = \Lambda^* \mu_\xi^* , \qquad (10.8)$$

where $\mu_\xi^* = T^{-1}\mu_\xi$. It is possible that (10.7) and (10.8), with some particular choice for T that yields some growth-related interpretability to the elements of Λ^*, can provide a more meaningful representation of growth trajectories.

The reproduction of the variable means across time is, after all, identical in (10.5) as it is in (10.8). Note, however, that arbitrary transformations also may not make sense, since the resulting time scale may be uninterpretable.

Unfortunately, if we start with the typical covariance structure M_7 in (10.6), namely, $\Sigma = \Lambda\Phi\Lambda' + \Psi$, the parameters of the covariance structure also change. The new covariance structure is $\Sigma = \Lambda * \Phi * \Lambda' * + \Psi$, where the new covariance matrix of the growth factors is $\Phi^* = T^{-1}\Phi T^{-1\prime}$. This implies that if there were individual differences in growth as shown by the variances of the growth factors (the diagonal elements of Φ), the extent of these individual differences also will change as the variances in $\Phi *$ no longer need to be identical to those in Φ. Similarly, if there was a given correlation, say, between intercept and slope factors under the Initial Status model, in the modified growth curve model the correlation may be completely different. In other words, there is no unique coding of the time scale for growth factors so that the "true" individual differences and correlations between growth factors is obtained. These issues have been discussed in the literature since Rogosa and Willet (1985), e.g., Rovine and Molenaar (1998a) and Mehta and West (2000). Stoel (2003) adds the important point that the correlation of external variables with the growth factors is similarly arbitrary as a result. It depends on the coding of basis functions, that is, on how we choose to parameterize time.

An Orthogonal Polynomial Model

The initial status model is only one way to represent time. Another approach is to use orthogonal polynomials as discussed in books on experimental design (see e.g., Cohen, Cohen, West, & Aiken, 2003, Ch. 6; Maxwell & Delaney, 2004, Ch. 6). These polynomials represent a kind of contrast coding where the codes maintain mutual orthogonality. Due to this orthogonality, modeling with orthogonal polynomials should result in fewer computing problems than with those associated with the initial status model.

Table 10.5 shows the coding for 7 variables on the above 5 factors, taken from Table A.10 of Maxwell and Delaney (2004). Notice that in this factor loading matrix, the sum of cross-products of any two columns equals zero, i.e., this is an orthogonal coding.

Variable or Time	Intercept F1	Slope F2	Quadratic F3	Cubic F4	Quartic F5
V1	1	-3	5	-1	3
V2	1	-2	0	1	-7
V3	1	-1	-3	1	1
V4	1	0	-4	0	6
V5	1	1	-3	-1	1
V6	1	2	0	-1	-7
V7	1	3	5	1	3

Table 10.5

Of course, it is important to understand the meaning of this coding for the interpretation of parameters of the model. Clearly the intercept factor seems the same as previously. In a model with only intercept and slope, by algebra parallel to that in Table 10.3, the model-reproduced mean would reproduce the mean of V4 since at that point the slope factor has no influence on V4. So the intercept can be interpreted as the level of the trend at time indicated by V4, which is sort of the average time. So this model might be called the *Time-Averaged* model. The quadratic and higher order trends now more explicitly code contrasts than in the initial value model. For example, the quadratic factor largely contrasts the time endpoints (V1,V7) with the middle (V3-V5). You might also note that the sum or average of all of the trend factor coefficients, except the intercept, is zero. This means that the factor loading matrix represents a kind of centroid factor analysis, except that the loadings are fixed rather than free. See e.g., Stoolmiller (1995) for more information. Biesanz et al. (2004) provide a good recent discussion of interpretation problems associated with orthogonal polynomial codes.

Trends In Means

Now we can get to an actual EQS model setup and run. Before we get to individual curves, let us look at the SEM approach to modeling trends in means. We do this with the data of Table 10.1, the mouse data. Pan and Fan (2002) proposed to model the means of the mouse data using a second degree polynomial and we will do the same, using the mean structure (10.5) without any covariance structure. That is, we are not interested in any individual differences in trends. This means that we should assume M_1, that is, the saturated covariance structure in (10.6). This is a structure that cannot be rejected, and hence any overall goodness of fit test will test only the mean structure. To implement this using the usual confirmatory factor model M_7 in (10.7), we take the covariance matrix of the factors ξ as null, that is, $\Phi = 0$, and let the residuals ε have variances and correlate freely. It follows that the covariance matrix of the variables is $\Sigma = \Psi$, namely, the covariance matrix of the residuals. The key part of the input for this model is given below. It is duplicated in the file 10mean.eqs.

```
/EQUATIONS
V1 =  + 1F1  + 0F2  +  0F3  + 1E1; !F1 IS THE INTERCEPT
V2 =  + 1F1  + 1F2  +  1F3  + 1E2; !F2 IS THE SLOPE
V3 =  + 1F1  + 2F2  +  4F3  + 1E3; !F3 IS THE QUADRATIC FACTOR
V4 =  + 1F1  + 3F2  +  9F3  + 1E4;
V5 =  + 1F1  + 4F2  + 16F3  + 1E5; !ALL FACTOR LOADINGS ARE FIXED PARAMETERS
V6 =  + 1F1  + 5F2  + 25F3  + 1E6;
V7 =  + 1F1  + 6F2  + 36F3  + 1E7;
F1 = .01*V999;                     !F1 HAS A MEAN TO BE ESTIMATED, WITHOUT ANY RESIDUAL
F2 = .22*V999;                     !F2 HAS A MEAN TO BE ESTIMATED, WITHOUT ANY RESIDUAL
F3 = -.01*V999;                    !F3 HAS A MEAN TO BE ESTIMATED, WITHOUT ANY RESIDUAL
/VARIANCES
V999= 1;
E1 TO E7 = .02*;                   !THE ES CARRY THE VARIANCE
/COVARIANCES                       !AND COVARIANCE INFORMATION
E1 TO E7 =.015*;                   !ON THE VARIABLES
```

The factors F1, F2, and F3 correspond precisely to the intercept, slope, and quadratic factors shown in Table 10.2, and have the fixed coefficients shown there. The only thing new here is that the factors themselves are explained by the V999 variable. It will be remembered from Chapter 8 that regression on a constant is an intercept, and when there is no indirect effect on the dependent variable, it is a mean. So the *V999 parameters are the means of the factors. This is like the *a* and *b* parameters in the much simpler and smaller model of Figure 10.2. Unlike Figure 10.2, here the Es are correlated. The residuals E1…E7 just carry the variance and covariance information of the variables, since the Fs carry no variance information. That is, there is no residual in the equations for F1, F2, or F3. This is unlike Figure 10.2, where there are D1 and D2 residual terms.

The key part of the output is the estimates of the factor means. All are significantly different from zero:

```
F1   =F1  =     .220*V999    ! VALUE NEAR THE SAMPLE MEAN ON V1 (DAY 3) IN TABLE 1
                .005
              40.206@
F2   =F2  =     .187*V999    ! APPROXIMATE INCREMENT IN MEANS FROM Vi TO Vi+1
                .005
              35.729@
F3   =F3  =    -.011*V999    ! A DAMPENING FROM THE LINEAR MODEL ALONE
                .001
             -12.838@
```

The intercept factor F1, with factor mean $\hat{\mu}_{F1} = .22$ does just what it should in an initial value growth model, it estimates almost exactly the sample mean at the initial measurement occasion. See the *a* parameter in Table 10.4. The slope factor F2 mean, $\hat{\mu}_{F2} = .187$, approximates quite well the mean differences from occasion to occasion. For example, from time 1 to time 2, the mean increment in the data of Table 1 is .386 - .211 = .175, the comparable increment from time 2 to time 3 is .551 - .386 = .165, and so on. These empirical values are a bit smaller than the value .187 predicted by the model. This is because the quadratic factor F3 has a negative mean $\hat{\mu}_{F3} = -.011$, which also influences the model-reproduced mean. That is, the negative mean on F3 has the effect that the model-reproduced mean of the variables will be reduced as compared to that reproduced from F1 and F2 alone, especially

at the later measurement occasions. Figure 10.3, which plots the trends of {F1 and F2}, (F1,F2, and F3}, and the data means, tells the story.

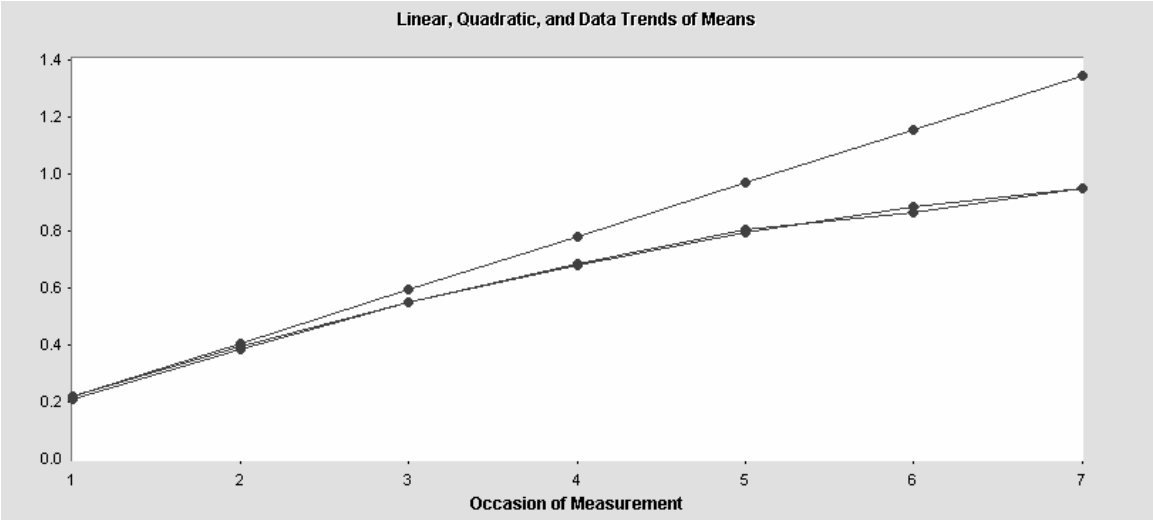


Figure 10.3

The top line is the linear trend obtained by plugging in the estimates $\hat{\mu}_{F1}$ = .22 and $\hat{\mu}_{F2}$ = .187 in the formulae for reproducing the sample means, as shown in the rows of Table 10.3. For example the model-implied mean of V4 is $1\hat{\mu}_{\mathrm{F1}} + 3\hat{\mu}_{\mathrm{F2}}$ = .220 + 3(.187) = .781. Of course, this is not the trend under the actual model used, as we also have the quadratic factor in the model. Adding the quadratic component $9\,\hat{\mu}_{F3}$ = 9(-.011) = -.099 to this (see Table 10.3) drops the model-reproduced trend to .682. The contribution of the quadratic trend grows substantially with later occasions, as the weight attached to these contributions increases from 9 to 16, 25, and 36. The end result is that the overall model-reproduced trend, based on F1, F2 and F3 almost precisely reproduces the actual sample means (bottom two trend lines).

The mean structure, and the overall model, has 4 degrees of freedom (7 sample means minus 3 mean parameters). It fits acceptably $\chi_4^2 = 8.7$, p = .07.

Could we add residuals D1, D2 and D3 to the mean structure equations F1 = *V999, F2 = *V999, and F3 = *V999 in this model, to allow for individual differences in the intercept, slope, and quadratic factors? For example, could we have F1 = *V999 + D1? No, we cannot! As it is, the covariance structure is fully saturated, that is, there are as many variance and covariance parameters for the Es as there are data variances and covariances. The covariance structure has zero degrees of freedom. Adding any more individual difference parameters will make the covariance structure underidentified. If we were determined to allow individual differences in these factors – as one does with typical growth curve models – we would have to have some type of more restricted covariance structure that would leave enough degrees of freedom for these added parameters. This is what the typical factor analytic covariance structure M_7 accomplishes, but it can also be accomplished by other structures.

Note: *Restricted covariance structures are very hard to impose on the mouse* data. The scaling of the mouse data to fractions of a pound has the consequence that the data variances and covariances are very small (even though the correlations are large). For example, the largest variance is .02. Also, these data are based on a very small sample size. As a result, models based on these data are very sensitive to arithmetical accuracy, starting values etc. If you are going to work with these data, you should rescale the variables (e.g., multiply all scores by 10, or by 16 to get ounces instead of pounds). Adjustments such as /TECHNICAL with crit=.000001; tol=.000001; itr=500; also may be needed to get reliable convergence.

Growth or No Growth?

Growth curve models are not the only way to think about modeling changes over time. Many of the structures discussed in time series, such as autoregressive or moving average processes on error variables, can be adapted to SEM (see e.g., van Buuren, 1997; Hamaker, Dolan, & Molenaar, 2003; Molenaar, 1999; Nesselroade & Molenaar, 1999). Using an autoregressive model on variables rather than residuals, and comparing to an intercept only model, in this section we illustrate the point that different model structures focus on different features of the data, and hence can substantially alter what you look at, and thus potentially your conclusions about time-structured data.

Suppose we have 3 repeated measures with covariances and means as follows, using EQS conventions:

	V1	V2	V3
V1	1.000		
V2	.700	1.490	
V3	.420	.894	1.536
V999	1.000	1.000	1.000

The variances and covariances are in the upper triangle, and the means are in the bottom row. We consider two different models for these data. The first is a growth curve model with only an intercept factor (no slope factor), in which we allow no individual differences on the factor and let the residual errors correlate freely. The second model is a saturated mean model with an autoregressive structure, in which the previous variable also influences the next in line variable, that is, V1→V2 and V2→V3. The EQS model setups, with optimal final estimates in the equations, are given in the following.

Intercept Only Unrestricted Covariances	**Autoregressive Uncorrelated Errors**
/EQUATIONS V1 = F1 +E1; V2 = F1 +E2; V3 = F1 +E3; F1 = 1*V999; /VARIANCES E1 TO E3 =*; /COVARIANCES E1 TO E3 =*;	/EQUATIONS V1 = 1*V999 +E1; V2 = .3*V999 + .7*V1 +E2; V3 = .4*V999 + .6*V2 +E3; /VARIANCES E1 TO E3 =*;

First we should recognize that the two models test different features of the data, in fact, nonoverlapping features of the data. The intercept only model evaluates a restricted mean structure, since there are 3 sample means and there is only 1 mean parameter with a saturated covariance structure. The model fits the above data perfectly, with $\chi_2^2 = 0.0$. The autoregressive model evaluates a quite different aspect of the data. Its mean structure is saturated, but the covariance structure has 5 parameters for 6 data points, so there is 1df for a model test. This evaluates only the covariance structure, and also shows that the model fits perfectly, with $\chi_1^2 = 0.0$. Given that we have perfectly "fitting" models, let us interpret the parameter estimates and look at the implications for changes in means.

In the intercept only model, the optimal estimate of the mean of F1 (the F1,V999 parameter) is 1.0. By path tracing (V999→F1→Vi) or computing total effects, this is the model-implied mean for all 3 variables. A conclusion based on this intercept only model is that nothing is happening in means across time. There is no mean growth.

The autoregressive model comes to the same conclusion. The total effects of V999→Vi, which provide the model-reproduced means, all yield 1.0, showing that there is no mean growth. The total effect for V1 is just the direct effect V999→V1, which is 1.0. For V2, the mean of V1 is propagated to V2 via the .7 valued V1→V2 path, so the total effect is 1(.7) + .3 = 1.0. For V3, the total effect is 1(.7)(.6) +.3(.6) + .4 = 1.0.

However, if you focus only on the direct effects, you would be drawing misleading conclusions about the overall trends. The mean increments V999→V2 of .3 and V999→V3 of .4 are conditional mean increments. Stated differently, conditional on V1, an increment in the mean of V2 is required to model the mean of V2. Similarly, conditional on V2, an increment in the mean of V3 is needed. These increments are needed in this model as nonzero values because the repeatedly measured variable is not perfectly stable, leading to regressions in the model-reproduced means that are made up by the added mean increments at T2 and T3. If the V1→V2 path were 1.0 instead, the indirect product V999→V1→V2 would equal 1.0 and be able to perfectly reproduce the mean of V2; and then the V999→V2 path would be zero. A similar rationale holds for V3.

Which model tells the true story? As both reproduce the data equally well, they both do. But the interpretations of the models are quite different.

Mean Structure with Modeled Error

In this section, we use data from an unpublished Italian longitudinal study of prosocial and antisocial behavior in schoolchildren, courtesy of Michele Vecchione and Gian Vittorio Caprara, to illustrate a simple mean structure model with some possible error structures for longitudinal data. In the typical approach to the analysis of means, namely analysis of variance, structures on residuals or covariances is not of interested. But when data are not experimental, standard anova approaches may not be appropriate and SEM provides a flexible alternative (Rovine & Molenaar, 1998b).

Summary statistics on the repeated measures of prosocial behavior from 4th grade through 8th grade of a sample of 364 children is given in Table 10.6. This is only a subset of variables and subjects in the interesting Vecchione data.

		PB_4TH V 1	PB_5TH V 2	PB_6TH V 3	PB_7TH V 4	PB_8TH V 5
PB_4TH	V 1	.128				
PB_5TH	V 2	.069	.123			
PB_6TH	V 3	.058	.069	.125		
PB_7TH	V 4	.059	.061	.068	.128	
PB_8TH	V 5	.043	.055	.056	.071	.135
V999	V999	2.404	2.434	2.437	2.413	2.400

Table 10.6

We would like to consider a simple model for the means and use a modified version of the usual confirmatory factor covariance structure to model the growth.

Actually, there does not seem to be any growth. In the data of Table 10.6, there is no mean change in prosocial behavior from grade 4 to grade 8. This implies that a simple model with an intercept factor and its individual differences may be acceptable. There is a slight tendency for the means to increase and then decrease again, so possibly a quadratic function may be appropriate as well. But for now we use only the level model. Also, it seems that covariances from T_i to T_{i+1} are a bit higher than the rest, and perhaps an autoregressive component of order 1 would help to improve the model. That is, we begin with the combination of a simple growth curve model and an autoregressive model that has recently been made popular (e.g., Bollen & Curran, 2004; Curran & Bollen, 2000; Green & Hershberger, 2000). Bollen and Curran call this an ALT model, an autoregressive latent trajectory model.

```
/EQUATIONS
V1 =   1F1 + E1;
V2 =   1F1 + E2 + *V1;
V3 =   1F1 + E3 + *V2;
V4 =   1F1 + E4 + *V3;
V5 =   1F1 + E5 + *V4;
F1 =   *V999 + D1;
/VARIANCES
E1 TO E5 = *; D1 = *;
```

```
/CONSTRAINTS
(E2,E2)=(E3,E3)=(E4,E4)=(E5,E5);
(V2,V1)=(V3,V2)=(V4,V3)=(V5,V4);
```

F1 is clearly the intercept factor. Its individual differences show up in the variance of D1. What is new here is that in addition to the intercept factor, autoregressive coefficients are added such that a given variable influences its next in line variable, such as V1→V2. We also impose the constraint that all the residual error variances except the first are equal; we do not equate the first, since V1 does not contain an autoregressive predictor as do the other variables. Finally, all the autoregressive coefficients are set equal.

The model does not fit the data. The ML $\chi^2_{15} = 48.1$. Such a large value could be due to inappropriately imposed equality constraints, but all of the LM tests show that the constraints individually and as a group are statistically acceptable. It also could be due to violation of normality, as Mardia's normalized coefficient for the data is 8.1, but the Satorra-Bentler $\chi^2_{15} = 41.7$ and Yuan-Bentler residual based F test (p=.003) also lead to model rejection. Possibly this lack of fit is due to the absence of a quadratic effect, but we will not pursue this here. The model setup for the covariances and means (not the raw data, which is not included) is given in the file 10alt.eqs.

Instead we consider an alternative model in which the autoregressive structure is on the residual error terms (e.g., Browne & du Toit, 1991). What is different conceptually about this model is that the mean parameters are not propagated through the residuals, while in the ALT model, V999→F1 intercept mean has an influence not only through the V999→F1→Vi paths as in the growth curve model, but also through such indirect paths as V999→F1→Vi→Vj. In principle, the model setup is a minor variation on the above with Ei→Ej terms, but EQS does not allow equations with E's as dependent variables (equations can be written only for Vs and Fs). So we adopt the trick of naming the basic residuals Fs rather than Es (starting the numbering with F11 so as not to confuse with F1, our intercept factor), as follows.

```
/EQUATIONS
V1  =   1F1  + F11;      !F11-F15 are residual error terms, not factors
V2  =   1F1  + F12;
V3  =   1F1  + F13;
V4  =   1F1  + F14;
V5  =   1F1  + F15;
F12 =   *F11 + E2;       !Fs are possible dependent variables in EQS, while Es are not
F13 =   *F12 + E3;       !Autoregressive predictions, based on a lag of 1 time interval
F14 =   *F13 + E4;       !New error residuals E2-E5 are introduced
F15 =   *F14 + E5;
 F1 =  *V999 + D1;
/VARIANCES
F11 = *; E2 to E5  = *; D1 = *;
/CONSTRAINTS
(E2,E2)=(E3,E3)=(E4,E4)=(E5,E5);
(F12,F11)=(F13,F12)=(F14,F13)=(F15,F14);
```

This model has the same number of parameters as the ALT model, so its degrees of freedom will be the same. The result is drastically different from that of the ALT model. The ML $\chi^2_{15} = 18.9$, the SB $\chi^2_{15} = 16.5$, and the Yuan-Bentler residual based F test yields a p=.40. This is also an acceptable model by residuals and other fit indices, and the constraints are acceptable within the model. Substantively, the intercept factor mean is 2.416, which is very close to all the sample means; the largest residual is .021. In view of the excellence of the fit, evidently there is no reason to consider a quadratic growth curve factor. The variance of D1 is highly significant, indicating individual differences around the mean that cannot be ignored. The model setup for the covariances and means (not the raw data, which is not included) is given in the file 10errar.eqs.

We complete this section by considering an additional error structure, the moving average structure. A diagram for the model we use is given in Figure 1 of Green and Hershberger (2000). Its key feature is that an error term Ei acts in the usual way on the i^{th} variable Vi, but it also freely predicts the next variable in time, V_{i+1}. In a sense, a residual error here is a common factor for two variables. As before, we impose equality constraints on the residual variances as well the error regression coefficients. The model setup follows.

```
/EQUATIONS
V1 =   1F1 + E1;
V2 =   1F1 + E2 + *E1;      !E1 PREDICTS V1 AND V2, BUT IT'S A FREE PARAMETER HERE
V3 =   1F1 + E3 + *E2;      !E2 PREDICTS V2 AND V3
V4 =   1F1 + E4 + *E3;      !E3 PREDICTS V3 AND V4
V5 =   1F1 + E5 + *E4;      !E4 PREDICTS V4 AND V5
F1 =   *V999 + D1;
/VARIANCES
V999 = 1;
E1 TO E5 = *; D1 = *;
/CONSTRAINTS
(E1,E1)=(E2,E2)=(E3,E3)=(E4,E4)=(E5,E5);
(V2,E1)=(V3,E2)=(V4,E3)=(V5,E4);
```

The model also fits well, with $\chi^2_{16} = 22.1$, a Satorra-Bentler $\chi^2_{16} = 19.1$, and Yuan-Bentler residual F test having a probability of .34. The mean of F1 and variance D1 are almost identical to those of the autoregressive error model. Empirically, it is not possible to choose between the autoregressive and moving average models. The model setup for the covariances and means (not the raw data, which is not included) is given in the file 10mvavg.eqs.

Prediction of Linear Growth

We now return to models of the more basic latent curve type. Curran (2000) developed several models to describe growth in adolescent substance use. We use the model and data provided in his Appendix C, which is a "two factor conditional latent curve model for adolescent alcohol use regressed on age, gender and parent alcoholism diagnosis" (p. 36). The model is shown in Figure 10.4. In the right side are three repeatedly measured alcohol use variables (V4, V5, V6), which are modeled with intercept and slope factors. This part of the model is identical to that of Figure 10.2, except that here the individual differences in intercept (D1) are allowed to be correlated with individual differences in slope (D2).

A model that contains predictors of growth factors is sometimes called a *conditional* latent growth model, as the effects of intercept and slope are conditional on the predictors. In the left are the 3 predictor variables (age, gender, parental status), which are correlated. Since these variables are part of a structured mean model, their means also have to be incorporated into the model. Their means are given by regressions on the constant (V999). As a result, V1-V3 become dependent variables, and their covariances are shown in the covariances of their attached E variables. These 3 variables are predictors of the intercept (F1) and slope (F2) factors, since there are paths going from V1-V3 to F1 and F2. These paths are concerned with whether individual difference in V1-V3 predict individual differences in the intercept and slope factors. However, by path tracing, it is clear that the constant also indirectly affects intercept and slope factors through V1-V3. This implies that the mean intercept and slope is given by the total effect of the constant on F1 and F2, i.e., the sum of direct and indirect effects. As in this model there are no direct effects of the constant V999 on V4-V6, the total effect is just the indirect effect.

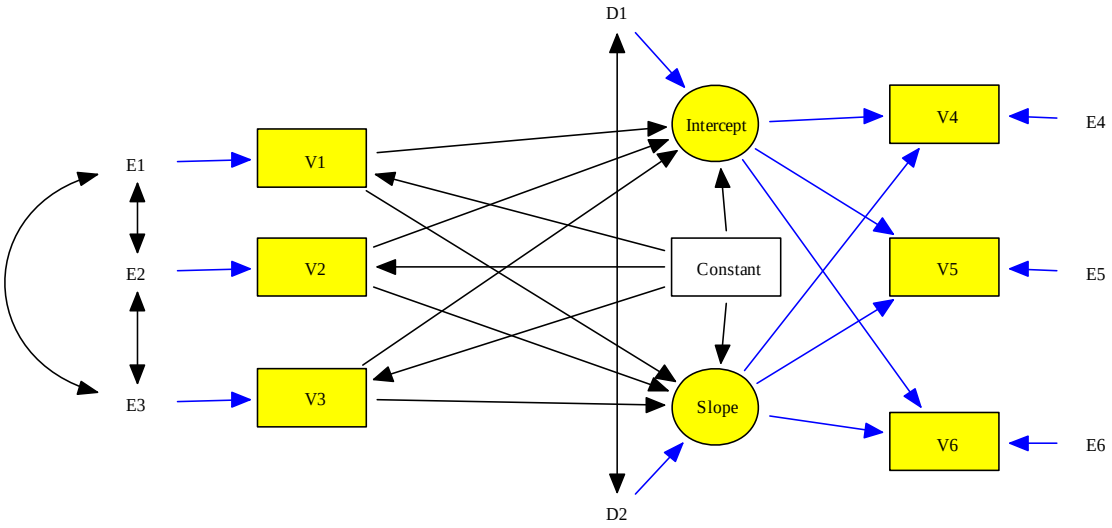


Figure 10.4

The key part of the EQS model setup, duplicated in the file 10curran.eqs, shows that this is a structured mean model with intercept and slope coefficients coded as an initial value model.

```
/SPECIFICATIONS
DATA='C:\EQS61\CURRAN\CURRAN APP C.ESS';
VARIABLES=9; CASES=363;
METHOD=ML; ANALYSIS=MOMENT; MATRIX=COVARIANCE;
/EQUATIONS
V1 =   *V999 + E1;                    !DECOMPOSITION INTO MEANS AND
V2 =   *V999 + E2;                    !DEVIATION FROM MEANS (THE E'S)
V3 =   *V999 + E3;
V4 =   1F1 + 0F2 + E4;                ! F1 IS THE INTERCEPT FACTOR
V5 =   1F1 + 1F2 + E5;                ! F2 IS THE SLOPE FACTOR
V6 =   1F1 + 2F2 + E6;                ! NO "*" ON THE FACTOR LOADINGS
F1 =   *V999 + *V1 + *V2 + *V3 + D1;  ! PREDICT INTERCEPT F1 FROM THE PREDICTORS
F2 =   *V999 + *V1 + *V2 + *V3 + D2;  ! PREDICT SLOPE F2 FROM THE PREDICTORS
/VARIANCES
E1 TO E6 = *;
D1 TO D2 = *;                         ! INDIVIDUAL DIFFERENCES IN INTERCEPT AND SLOPE
/COVARIANCES
E1 TO E3 =*                           ! PREDICTORS ARE CORRELATED
D2, D1 = *;                           ! INDIV DIFFS IN INTERCEPT & SLOPE MAY CORRELATE
```

The sample covariance matrix (top left) and mean vector (bottom line) to be modeled is the following.

	V1	V2	V3	V4	V5	V6
V1	1.968					
V2	-.022	.250				
V3	-.104	.001	.246			
V4	1.238	.019	.164	7.906		
V5	1.750	.110	.331	7.634	15.828	
V6	1.532	.239	.478	6.726	12.917	23.013
V999	12.915	.515	.565	1.364	2.119	3.185

This model fits the data well, with $\chi^2_4 = 5.53$ (p=.24). You will know how to interpret the output of this run, as it follows that of any standard mean and covariance structure model, so we will not provide many details. We will discuss three items from the output, mean growth, the correlation of individual differences in growth, and the effect of the background variables on growth.

In Table 10.4 we had indicated that in an *unconditional* intercept/slope model, the empirical mean differences should be approximately equal across time, as this is what the model demands. Is this true also in the *conditional* model? It

might seem not, as the means of V4-V6 are not just functions of the intercept and slope factors. By path tracing, they also involve the many indirect effects of the constant V999 on the variables (e.g., constant → V1 → intercept → V4). A key observation is that all of the indirect effects operate through the intercept and slope factors. There is no other way to get to V4-V6 except through these growth factors. As a result, whatever the value of the indirect effect might be by the time it gets to F1, the intercept, the subsequent paths to V4-V6 just take that value and multiply it by 1.0. By itself, this of course leads to equal means. But, in addition, whatever the indirect effect might be by the time it gets to F2, the slope, the paths to V4-V6 are multiplied by 0, 1, 2 in the usual way. As a result, equal mean differences will be observed in an intercept/slope model, even when there are predictors of growth. In this instance, the empirical means of V4-V6 are 1.36, 2.12, and 3.18, and the model-reproduced ones are 1.36, 2.24, and 3.13. Even though the empirical means differ by increments of .76 and 1.06, the model-implied ones differ uniformly by .88.

The covariance of D1 and D2 is estimated at -.843. In the standardized solution, this is transformed to a correlation of -.185. Negative correlations between individual differences in intercept and slope are fairly typical, as this implies that those who start high grow less, or conversely, those who start low have more room to grow. But remember that the actual value is not unique, as different codings of time can lead to different correlations.

Substantively, in a model of this sort one is interested in whether the predictor variables significantly influence individual differences in growth. In this model they do. The standardized solution shows influences varying from .023 (which is the only not significant coefficient in the unstandardized solution) to .340. The interpretation of the .340 is that there is a tendency of older kids (higher levels of V1) to have higher initial levels of alcohol use (F1). The interpretation of the .196 is that children whose parents are alcoholics will tend to have higher F2 scores, that is, they will grow faster in alcohol use across time as compared to other children.

```
F1   =F1  =    .340*V1    +  .023*V2    +  .168*V3    +  .000*V999   +  .934 D1
F2   =F2  =    .119*V1    +  .122*V2    +  .196*V3    +  .000*V999   +  .970 D2
```

Alternatives to the Linear Slope Factor

The above examples with slope factors used the usual initial status parameterization. In their fundamental paper, Meredith and Tisak (1990) proposed that spline factors may be useful. A *spline* is a piecewise polynomial function having various sections that are fitted together smoothly. In EQS, Diagrammer does the computation that allows the slope factor to be substituted by a single linear spline factor based on sample means. The essence of this approach is that the initial value coefficients 0, 1, 2, are recoded using the observed means of the variables to create coefficients that empirically stretch or squeeze the otherwise linear slope. As a result, the straight line is replaced by a series of shorter straight lines that are joined. We follow Stoolmiller's (1995, eq. 4.28) suggestion to recode the loadings so that the intercept factor retains its interpretation as initial status. This is done by estimating the loading λ_i (i=1, 2,…, p) on the spline factor by a relative mean growth

$$\hat{\lambda}_i = (\bar{Z}_i - \bar{Z}_1)/(\bar{Z}_2 - \bar{Z}_1)\,, \tag{10.9}$$

using differences in sample means of pairs of variables. Another pair of base variables is chosen if the first two means are equal to within computational accuracy (i.e., $\bar{Z}_1 \cong \bar{Z}_2$). The resulting coefficients are in practice treated as known, though clearly they are sample-dependent. The resulting spline factor based on (10.9) is probably best thought of as a slope/shape factor. This approach can be illustrated with the growth in WISC intelligence scores of children as proposed by Stoolmiller (1995) on data summarized by McArdle and Epstein (1987). The data is included in the file wisc.ess. The model proposed by Stoolmiller is the following.

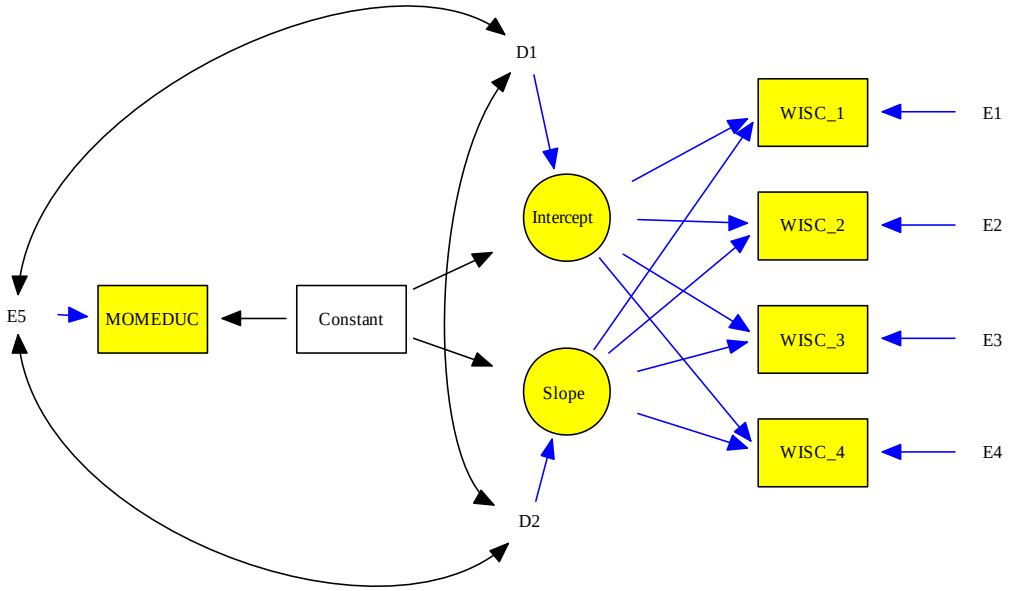


Figure 10.5

From the viewpoint of formal model structure, it seems that Figure 10.5 is pretty much the same as Figure 10.4 above. What makes it different is that the slope factor is really a slope/shape factor with spline coefficients. The model setup is given in the file 10wisc.eqs, a portion of which is copied below.

```
/LABELS
 V1=WISC_1; V2=WISC_2; V3=WISC_3; V4=WISC_4; V5=MOMEDUC;
/EQUATIONS
 V1 =   1F1 +    0F2 + E1;
 V2 =   1F1 +    1F2 + E2;
 V3 =   1F1 + 2.21F2 + E3;
 V4 =   1F1 + 3.67F2 + E4;
 V5 =   *V999 + E5;
 F1 =   *V999 + D1;
 F2 =   *V999 + D2;
/VARIANCES
 E1 TO E5 = *;
 D1 TO D2 = *;
/COVARIANCES
 D1,E5 = *;
 D2,E5 = *;
 D2,D1 = *;
```

Here the slope factor has coefficients 0, 1, 2.21, 3.67 rather than 0, 1, 2, 3 as in the initial status model. The WISC sample means of 18.034, 25.819, 35.255, 46.593 were used to compute the coefficients using (10.9), e.g., $\hat{\lambda}_3$ = (35.255 – 18.034) / (25.819 – 18.034) = 17.221 / 7.785 = 2.21. In the example, this coding proved effective at yielding parameter estimates that reproduced the means and covariances, while the ordinary linear slope model did not. See Stoolmiller for details. Of course, the approach illustrated here is not guaranteed to work in other situations.

There is still another approach. You can decompose the slope factor into two (or more) slope factors whose factor loadings sum to the original coefficients 0, 1, 2, … in Table 10.2. This can be used to model what Raudenbush and Bryk (2002, p. 178) call a *piecewise linear* model, where growth up to some known point is given by the first slope factor and growth after that point is given by the second slope factor. In our mouse example, we might think that one straight line would cover V1-V5, and a different line would cover V5-V7. Then one factor could have coefficients 0, 1, 2, 3, 4, 4, 4 indicating no growth after V5. The other factor would be coded 0, 0, 0, 0, 0, 1, 2 indicating another slope for the last three time points. These pairs of factors could be an alternative to a quadratic factor. Piecewise

nonlinear shapes also can be considered (e.g., Lawrence & Blair, 2003). Seber and Wild (1989, Ch. 9) call models of this type and more general forms as multiphase (here, two phase) and spline regressions.

Cohort-Sequential Growth Modeling

In their latent curve paper, Meredith and Tisak (1990) noted the relevance of their technique to Schaie's (1965) cohort sequential design (see Nesselroade & Baltes, 1979). We illustrate this application with a growth model from Duncan, Duncan, Strycker, Li and Alpert (1999, Chapter 6), based on data these authors obtained from the National Youth Survey (Elliott, 1976). These authors adapted the multiple group missing data modeling approach of Allison (1987) and Muthén, Kaplan and Hollis (1987) to this accelerated longitudinal data design in the context of growth curve modeling. Cohort sequential analyses make it possible to investigate growth trajectories over a large time span even though no sample may have been observed over the entire time span. Information from overlapping measurements of different cohorts is combined under the assumption that all groups come from a single population with no cohort differences. This methodology has been used in a variety of empirical studies (e.g., Duncan, Duncan & Hops, 1994, 1996). Technically, this is an application of multisample mean and covariance structure analysis (Chapter 9) in the context of missing data and growth curve modeling. The methodology of this section also can be implemented by the more general approach to missing data models given in Chapter 12.

In the particular example, we attempt to establish a single mean trend for alcohol use across 5 years (ages 12-16) even though no sample has data on all five years. The first sample has data at ages 12, 13, 14. The second sample has data at ages 13, 14, 15. The third sample has data at ages 14, 15, 16. Duncan and coauthors also compared the results to those obtained from a sample that provided data at all 5 time points, and found the results matched well. This aspect of their work is not shown here.

The input program is formulated as a multi-sample analysis with constraints; it is duplicated in the file 10cohort.eqs. The missing variables in the cohort sequential design are handled by using dummy entries in the sample mean vectors and covariance matrices and by fixing parameters specific to these parts. Variances and covariances involving missing variables have the entries of "0". The means of missing variables have the entries of "1". These numbers are just placeholders to allow one to set up the model in the same way in all groups using the same variables, factors, and notation. The missing means and covariances are not used during estimation. For example, sample 1 contains dummied information for V4 and V5. But since V4 and V5 are not used in the model (equations, variances, and covariances), the dummy entries are ignored.

The model itself has 2 factors in all groups, an intercept (F1) and a slope (F2) factor. Notice that the slope factor has the following loadings in the 3 groups: 0, 1, 2; 1, 2, 3; and 2, 3, and 4; all keyed to the time of measurement of the relevant V. This allows the entire slope range 0, 1, 2, 3, 4 to be covered although no group has data for the entire range. The factors have means (*V999 parameters) and individual differences (variances of D1 and D2) to be estimated. The individual differences are correlated. And there are the usual unique variances.

```
/TITLE
EXAMPLE FROM DUNCAN, DUNCAN, STRYCKER, LI & ALPERT (1999)
COHORT-SEQUENTIAL DESIGN GROUP 1
/SPECIFICATIONS
CAS=248; VAR=5; ME=ML; MA=COV;
ANAL=MOMENT;
GROUPS=3;                                          !THREE COHORTS ARE USED
/LABELS
V1=ALC_T1; V2=ALC_T2; V3=ALC_T3; V4=ALC_T4;
V5=ALC_T5; F1=ALC_INT; F2=ALC_SLP;
/EQUATIONS
V1=F1+0F2+E1;                                      !INTERCEPT AND SLOPE FACTORS
V2=F1+1F2+E2;
V3=F1+2F2+E3;
F1=0*V999+D1;
F2=0*V999+D2;
/VARIANCES
E1 TO E3=*;
D1 TO D2=*;
```

```
/COVARIANCES
D1 TO D2=*;
/MATRIX
0.483
0.303 .899
0.271 0.488 1.258
0.000 0.000 0.000 0.000                              !MISSING DATA HAVE
0.000 0.000 0.000 0.000 0.000                        !DUMMY ENTRIES
/MEANS
0.335 0.774 0.964 1.000 1.000                        !THE VARIABLE 4 AND 5
/END                                                 !HAVE THE ENTRIES OF 1

/TITLE
GROUP 2
/SPECIFICATIONS
CAS=250; VAR=5; ME=ML; MA=COV; ANAL=MOMENT;
/LABELS
V1=ALC_T1; V2=ALC_T2; V3=ALC_T3; V4=ALC_T4;
V5=ALC_T5; F1=ALC_INT; F2=ALC_SLP;
/EQUATIONS
V2=F1+1F2+E2;
V3=F1+2F2+E3;
V4=F1+3F2+E4;
F1=0*V999+D1;
F2=0*V999+D2;
/VARIANCES
E2 TO E4=*;
D1 TO D2=*;
/COVARIANCES
D1 TO D2=*;
/MATRIX
0.000                                               !THE MISSING VARIABLES
0.0 0.759                                           !HAVE DUMMY ENTRIES
0.0 0.400 0.988
0.0 0.412 0.683 1.315
0.0 0.000 0.000 0.000 0.000
/MEANS
1.000 0.764 0.992 1.384 1.000
/END

/TITLE
GROUP 3
/SPECIFICATIONS
CAS=231; VAR=5; ME=ML; MA=COV; ANAL=MOMENT;
/LABELS
V1=ALC_T1; V2=ALC_T2; V3=ALC_T3; V4=ALC_T4;
V5=ALC_T5; F1=ALC_INT; F2=ALC_SLP;
/EQUATIONS
V3=F1+2F2+E3;
V4=F1+3F2+E4;
V5=F1+4F2+E5;
F1=0*V999+D1;
F2=0*V999+D2;
/VARIANCES
E3 TO E5=*;
D1 TO D2=*;
/COVARIANCES
D1 TO D2=*;
/PRINT
EFFECT=YES;COVARIANCE=YES;
/MATRIX
0.000
0.0 0.000
0.0 0.000 1.170
0.0 0.000 0.711 1.371
0.000 0.000 0.662 0.936 1.709
/MEANS
1.000 1.000 0.939 1.541 2.170
/PRINT
FIT=ALL;
```

```
/CONSTRAINTS
(1,D1,D2)=(2,D1,D2)=(3,D1,D2);                    !EQUALITY CONSTRAINTS ARE
(1,D1,D1)=(2,D1,D1)=(3,D1,D1);                    !BASED ON THE ASSUMPTION
(1,D2,D2)=(2,D2,D2)=(3,D2,D2);                    !THAT THREE GROUPS COME FROM
(1,F1,V999)=(2,F1,V999)=(3,F1,V999);              !THE SAME POPULATION
(1,F2,V999)=(2,F2,V999)=(3,F2,V999);
(1,E2,E2)=(2,E2,E2);
(1,E3,E3)=(2,E3,E3)=(3,E3,E3);
(2,E4,E4)=(3,E4,E4);
/END
```

The output showed that the model is largely acceptable and interpretable, though the statistical fit is not perfect.

```
CHI-SQUARE =       65.746 BASED ON      17 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS       .00000
```

Lagrange multiplier tests showed that the model could be substantially improved by dropping two of the equality constraints, and recognizing that some of the variables' means cannot totally be explained by the proposed structure. However, such model adjustments take us beyond the point of the illustration.

The last thing to explain in this chapter is how the combined growth curve is constructed. Using information from either the total effects, or the model-reproduced mean/covariance matrix, the following results are obtained on the reproduced means of the variables.

	Group 1	**Group 2**	**Group 3**
V1=ALC_T1	.310		
V2=ALC_T2	.694	.694	
V3=ALC_T3	1.078	1.078	1.078
V4=ALC_T4		1.463	1.463
V5=ALC_T5			1.847

Each group provides only 3 model-reproduced means, group 1 for V1-V3, group 2 for V2-V4, and group 3 for V3-V5. But, because of the imposed equality constraints, two pieces of information in each group match precisely the corresponding information in another group. As a result, we can conclude that the mean trend in alcohol use under this model from T1 to T5 are: .310, .694, 1.078, l.463, 1.847. Not surprisingly, as this is a *linear* growth curve model, there is a common mean increment of .384 from occasion to occasion.

11. MULTILEVEL METHODS

A typical SEM involves a single data set and one model, either a covariance structure model $\Sigma = \Sigma(\theta)$, or a covariance structure model with an added mean structure $\mu = \mu(\theta)$. Chapters 7 and 9 showed how this basic setup could be extended to study multiple models for data sets arising from several samples. This chapter introduces a methodology for studying multiple models from a single sample. Following some early overviews (e.g., Bock, 1989; Bryk & Raudenbush, 1992), this topic has grown substantially over the past decade. Clearly, if all sample data is summarized in a single mean vector $\bar{x}$ and sample covariance matrix *S* to be modeled, it will be difficult if not impossible to entertain more than a single model. However, if sample data is structured in a special way, and there are indicators of the structure, multiple models can be entertained. We will partition the overall or total model covariance matrix into a between and within part, that is, $\Sigma = \Sigma_B + \Sigma_W$ and allow each part to be expressed as a Bentler-Weeks model. A mean structure may be associated with the between part. Corresponding to these model matrices, of course, are sample estimators that are somewhat complicated to compute. These complications are internal to EQS, however, so that from your point of view there is not much new to accomplishing a multilevel analysis except to use a few additional commands. However, some background is essential so that you can use multilevel models with accuracy and confidence.

Psychological, educational, clinical, and social research often involves data that has a hierarchical, or multilevel, structure. The units from which data have been collected are often nested within a larger social group or context on which data also may be available. For example, data may be collected from individual students within a school, and similarly from many schools, or data may be gathered from patients within a clinic, and also from many clinics. As a result, cluster-sampled data reflect a hierarchical structure. In the study of students and schools, there is the micro-level of students (often called level 1) and the macro-level of schools (level 2). Then Σ_W would represent the covariation at the student level, i.e., individual differences (e.g., in algebra skills) and their correlates, controlling for variation in schools. Σ_B would represent covariation at the school level, i.e., differences among schools (e.g., in teacher qualifications or money available for instruction) and their correlates. The data arising in such setups are often called clustered data, and the effects are called within-cluster and between-cluster effects. The importance of separating these effects in studies of school effectiveness has been long recognized (e.g., Aitkin & Longford, 1986; Burstein, 1980; Morris, 1995), and the parallel importance in other domains has also become clear (e.g., Greenland, 2000b, 2002; Julian, 2001; Leyland & Goldstein, 2001; Mason, Wong, & Entwistle, 1983). In principle, a hierarchical structure may involve more than 2 levels: schools nested within districts, districts nested within states, and so on. This chapter shows how 2-level data can be easily handled by ML (Bentler & Liang, 2002, 2003; Liang & Bentler, 1999a, in press) and Muthén's (1994) approximate ML (MUML) methods. The two-stage approach (Chou, Bentler, & Pentz, 2000) given at the end of the chapter can in principle handle any number of levels. In this release of EQS, multilevel models are restricted to continuous outcome variables.

There are two main problems with ignoring the hierarchical aspects of a data set: (1) a single model matrix Σ may mix two unrelated sources of variance; (2) the statistical analyses may violate assumptions. Obviously, if Σ_B is very small – for example, if schools are homogeneous -- an analysis of Σ and Σ_W would give about the same results. In such a case one can largely ignore school differences. But in some research, say, school effectiveness studies, school differences – however small – are the very focus of an analysis, and Σ_W is just a nuisance source of variance to be controlled. The statistical problem is that, even though it seems that subjects at level 1, say, students, are sampled independently, if they are not independently sampled from a single population (e.g., randomly selected from one master list of students in a state, ignoring school membership) this cannot be true. If sampled only within the school, subjects nested within the same cluster, group, or social context will tend to be more similar than would be expected if they had been sampled from the entire population. Statistical analyses that do not consider the hierarchical structure of the data, or the dependence among observations nested within the same group, can result in erroneous statistical inferences. Valuable recent discussions of these issues and possible solutions include Bryk and Raudenbush (1992), de Leeuw and Kreft (1986, in press), du Toit and du Toit (in press), Goldstein (1986, 2003), Heck and Thomas (2000), Hox (1995, 2002), Jedidi and Ansari (2001), Kreft and de Leeuw (1998), Lee and Song (2001), Longford (1993ab, 2001), Marsh

and Hau (2002), Neuhaus and Kalbfleisch (1998), Rabe-Hesketh, Yang, and Pickles (2001), Raudenbush (1995, 1997), Raudenbush and Bryk (2002), Reise and Duan (2001, 2002), Singer and Willett (2003), Snijders and Bosker (1999), and Stapleton (2002). For a variety of applications, see e.g., Carbonaro and Gamoran (2003), Duncan, Duncan, Alpert et al. (1998), Duncan, Duncan, and Strycker (2002), Duncan, Strycker, Duncan, and Okut (2002), Goldstein, Browne, and Rasbash (2002), Griffith (1999), Kaplan and Elliott (1997), Little, Schnabel, and Baumert (2000), Livert, Rindskopf, Saxe, and Stirratt (2001), Reise (2000), Snijders and Kenny (2000), Stein, Nyamathi, and Bentler (2001), Thompson, Turner, and Warn (2001), van den Oord (2001), van Duijn, van Busschbach, and Snijders (1999), van Yperen and Snijders (2000), Wharton, Rotolo, and Bird (2000), and Yang, Goldstein, Browne, and Woodhouse (2002).

Muthén and Satorra (1995a) refer to the overall analysis of a single Σ as aggregated modeling, and the analysis Σ_W and Σ_B separately as disaggregated modeling. An aggregated analysis of Σ does lead to consistent parameter estimates of a specific overall model even if there are real between and within components, but special adjustments need to be made to compute standard errors and goodness of fit chi-square tests. These adjustments are done automatically in EQS with ME=ML,ROBUST, as described in Chapter 5 and elsewhere in this *Manual*. Nonetheless, such consistent parameter estimates in an overall model cannot separate effects operating at each level. To do this effectively requires modeling both levels, taking into account the lack of independence of observations as part of the modeling procedure. Hierarchical linear models or multilevel models have been recognized as excellent tools for such disaggregated analyses. The basic statistical theory assumes multivariate normality of observations at both level 1 and level 2. However, EQS also provides robust standard errors and test statistics (implemented with the usual ML, ROBUST statement) to handle violations of these assumptions (Yuan & Bentler, 2002bc, 2003). These can be applied even if the normal theory statistics are asymptotically robust (Yuan & Bentler, 2004), a condition that is hard to evaluate in practice and should not be simply assumed to be true.

In the typical multilevel study, there should be many (perhaps hundreds – the more the better) level-2 units such as schools, and there will typically be many level-1 units (perhaps dozens or more) such as students within each level-2 unit. As a result, if *g* is a school index, it is clear that there may be hundreds of within-school covariance matrices Σ_{Wg}. From a statistical point of view, these are too many to be estimated freely, and similarly it is impractical to consider that each might have its own covariance structure $\Sigma_{Wg} = \Sigma_{Wg}(\theta_g)$. To solve this difficulty, the practical assumption is made that these matrices are homogeneous (a test of this assumption is not currently available). Then a single within-group covariance matrix Σ_W can be used to describe within level-1 variation, no matter which level-2 unit is involved. Such simplification is not needed for level-2, as a single covariance matrix Σ_B can be used to describe differences among level-2 units. Thus Σ_W and Σ_B are key unstructured matrices to estimate, and within SEM, as having some further structure that we take to be the Bentler-Weeks structure. Because there are two model matrices, it is *as if* we are dealing with a 2-group multisample model, though of course we do not make the standard statistical assumption of independent groups as discussed in Chapters 7 and 9. In other words, thinking of multilevel modeling as an exercise in multisample analysis can be helpful (e.g., McArdle & Hamagami, 1996). In EQS you will specify basic multilevel models with 2-group setups, one for a within structure, and one for a between structure. Viewed this way, you can approach these models similarly as you approach other SEM models. In particular, latent variable models can be considered at each level (e.g., Goldstein & McDonald, 1988; Lee, 1990; Muthén, 1994; Lee & Poon, 1998). New ways of implementing older approaches as latent variable models continue to be developed (e.g., Rovine, & Molenaar, 2000), while some types of data can be approached from several points of view, e.g., as growth curves or multilevel models (e.g., Chou, Bentler, & Pentz, 1998; Khoo & Muthén, 2000).

Regression vs. SEM Approaches to Multilevel Data

"Multilevel modeling" covers a range of methodologies. *The various approaches are not identical in terms of assumptions, computational details, and types of models that can be handled.* In EQS, as noted above we use the SEM notation and concentrate on within-cluster and between-cluster covariance matrices and between-cluster means, and their latent Bentler-Weeks structures. When developed in a regression framework, multilevel modeling is often known as *hierarchical linear modeling* (HLM) and sometimes as *mixed model regression*. These names

find their way into some of the associated computer programs (e.g., HLM, Raudenbush, Bryk, Cheong, & Congdon, 2000; SAS PROC MIXED, Singer, 1998). In these approaches, an appropriate design along with allowance for dependent residuals can help to determine the contribution of within and between-level predictors on outcomes, and robust standard errors are available to handle violation of assumptions (e.g., Cheong, Fotiu, & Raudenbush, 2001; Maas & Hox, 2004). The sources of variance are sometimes labeled as fixed effects and random effects, and the models themselves may be called random effects (Laird & Ware, 1982) or *random coefficient models* (de Leeuw & Kreft, 1986). The straightforward logic of the regression approaches, as illustrated below, is a real strength. A disadvantage is that latent variables are harder to deal with, as compared to the SEM approach used in EQS. As was noted in Chapter 1, regression-type models can be handled within SEM, and this is similarly true for certain types of multilevel regression models. Recently more general approaches have been developed that combine an even broader range of models, see Skrondal and Rabe-Hesketh (2003, 2004).

The regression-like origin of multilevel models helps to clarify their hierarchical nature. Consider the prediction of algebra skills among high school students from a measure of the extent of time studying. Within a single algebra class – the g^{th} class – for the i^{th} student, we observe his/her algebra test performance Y_{ig} and a measure of the number of hours studying algebra per week X_{ig}. Then we might consider the prediction equation

$$Y_{ig} = \beta_{0g} + \beta_{1g} X_{ig} + e_{ig} . \tag{11.1}$$

If you ignore the class subscript *g*, this is a simple regression model with an intercept β_0 and a slope β_1, and a residual e_i. Estimation of the *fixed β parameters* can be handled easily. The intercept would reflect mean algebra performance, and the slope – expected to be positive in sign -- would represent how strongly increased studying is reflected in increased algebra performance. But in the multilevel setup, we consider that we have many classes, that is g = 1, 2, …G, perhaps G = 100. This means that we actually might have 100 equations of the form (11.1). As a result, we might not expect the intercept and slope parameters to be absolutely identical in each of the 100 classes, and we might consider that they are *random coefficients*, varying across our sample of classes. For example, there are liable to be differences in the intercept β_{0g}, with some classes performing very well on average, and others not so well. Possibly there might be differences in slope β_{1g}, with some classes having been taught effective algebra studying tactics and others not, and some students getting the idea and others not, so that the extent to which studying is translated into algebra performance varies across classes (the old idea of "slopes as outcomes").

Equation (11.1) is a level-1 or student-level equation. We may have some hypotheses about classroom-level variables that might explain larger or smaller values of the random intercept or slope in (11.1). For example, we may believe that the classroom teacher's experience, say Z_g, may explain the overall extent of performance in algebra and perhaps also the effectiveness by which studying is translated into performance. We can express these ideas with the two level-2 equations

$$\beta_{0g} = \gamma_{00} + \gamma_{01} Z_g + u_{0g} \tag{11.2}$$

$$\beta_{1g} = \gamma_{10} + \gamma_{11} Z_g + u_{1g} . \tag{11.3}$$

The γ coefficients are coefficients for regression at level-2, that is, at the classroom level, and u_{0g} and u_{1g} are random level-2 residuals. Equations (11.2)-(11.3) describe the possible effect of teacher experience on the random β coefficients. If teacher experience Z_g has an impact on average classroom algebra performance, γ_{01} would be nonzero (hopefully, positive). Similarly, γ_{11} describes the effect of teacher experience on how well studying is translated into performance. Of course, teacher experience may have no effect, so these coefficients could be zero. In such a situation, aside from random error, $\beta_{0g} \cong \gamma_{00}$ and $\beta_{1g} \cong \gamma_{10}$, so we could interpret γ_{00} as the average intercept (class algebra performance) and γ_{10} as the average slope (effect of studying on performance).

If we plug (11.2)-(11.3) into (11.1), we get the overall model

$$Y_{ig} = \gamma_{00} + \gamma_{10} X_{ig} + \gamma_{01} Z_g + \gamma_{11} X_{ig} Z_g + u_{1g} X_{ig} + u_{0g} + e_{ig} . \tag{11.4}$$

Clearly, this is not a standard regression equation. The residual $u_{1g} X_{ig} + u_{0g} + e_{ig}$ is a combination of the predictor variable X_{ig} and residuals at both levels, and hence this residual is not independent of the Z_g and X_{ig} predictor variables and its variance is not homoscedastic. Also, there is a cross-level interaction term involving the level-1

and level-2 variables $X_{ig}Z_g$. For these reasons, a combined analysis of (11.4) using an ordinary regression program is problematic.

The two-stage HLM approach in EQS will allow you to study systems such as (11.1)-(11.3), but the multilevel SEM approach using ML or MUML will not! In the SEM approach, the variables *Y* and *X* would have variation at both levels 1 and 2 ("within" and "between" classes or clusters) that might be decomposable with an appropriate model, while *Z* would have variation only at level 2 ("between" classes or clusters). But in the SEM approach there is currently no direct way to handle cross-level interactions (although you could externally put the cross-level product $X_{iq}Z_q$ into a data file and proceed from there). This disadvantage is offset by the ability to deal with larger sets of variables and more complicated within- and between-cluster structural models.

Data Format and Structure

Univariate versus Multivariate Layout

To implement a multilevel model, you need a variable that specifies to which cluster or group a given observation belongs. In EQS, this is specified as CLUSTER=V1, for example. If your data set includes the clustering variable along with the variables that you want to decompose into between and within components, the data is ready for the ML and MUML approaches. But sometimes the way data is organized does not obviously show the clustering variable, and it may be necessary to modify a standard data file.

This is especially true with growth curve data, which were previously discussed in Chapter 10. It was noted there that multilevel models can be used to estimate and test models for which growth curve methodology is used. As shown by Chou, Bentler, and Penz (1998), Hox (2000), MacCallum, Kim, Malarkey and Kiecolt-Glaser (1997), Stoel (2003) and others, SEM growth curve modeling and multilevel models can yield identical estimates of relevant parameters and virtually identical standard errors when the two approaches use the same assumptions and data. But multilevel models have an advantage when subjects have different numbers and occasions of measurement. A key difference between these approaches in a basic setup is that, in growth curves, time is specified as some coding of fixed factor loadings, while in multilevel models it is a coded predictor variable. Further discussion of the relations between these approaches is given by Bauer (2003), Singer and Willett (2003), Stoel (2003), and others.

The data matrix used in growth curve analysis is the standard matrix of cases by variables, whose various rows (subjects) are assumed to be independent, and where dependencies are modeled among the column variables that represent the repeatedly measured variables (and other variables). Here is an artificial example of 9 scores for 2 cases:

Case No.	V1=T_{1Y}	V2=T_{2Y}	V3=T_{3Y}	V4=T_{4Y}	V5=T_{1X}	V6=T_{2X}	V7=T_{3X}	V8=T_{4X}	V9
1	20	21	22	23	2.3	2.4	2.8	2.9	110
2	16	18	20	22	3.0	3.0	2.9	3.4	105

V1-V4 is one repeatedly measured variable Y, with implicit measurement time codes 1,2,3,4 that are not part of the data file. V5-V8 is another repeatedly measured variable X, and V9 is some control variable. This is flat-file data, and we might call this format the standard multivariate layout. Singer and Willett (2003) call it a "person-level" data structure.

In the multilevel approach to growth curve data, the data is formatted differently. Below is the data on the same artificial example, now organized to have only 5 variables. (Any Vi in the setup above is not the same Vi below.) V5 is the cluster variable, which here represents the case number of the two subjects, and which helps to define between-subject variability and within-subject variability. In contrast to the setup above, here each case has 4 rows of data (rather than only a single row). V4 is the time coded predictor variable, where both subjects are measured at

four equal intervals (this is not a requirement). There was no time coding variable in the data file setup above, since in growth curve modeling time coding is introduced via the factor loadings. Below, V1 represents the repeatedly measured variable Y, which is a recoding of V1-V4 above. V2 is a recoding of variable X in V5-V8 above. It might represent a time-varying covariate. V3, taken from V9 above, occurs several times in the data matrix for each case below because a person's scores are presumed to be identical at all measurement occasions. For example, it may represent a time-invariant covariate such as IQ that may have individual differences but is presumed not to change within an individual over the interval of measurement.

V1 =Y	V2 =X	V3	V4 Time Code	V5 Case No.
20	2.3	110	1	1
21	2.4	110	2	1
22	2.8	110	3	1
23	2.9	110	4	1
16	3.0	105	1	2
18	3.0	105	2	2
20	2.9	105	3	2
22	3.4	105	4	2

In this data structure, each person has multiple records, specifically 4, one for each repeated measurement. If the data are balanced, each subject in the entire data set will have the same number of records or observations, but this is not needed for a multilevel analysis. In this context, a model for the between-cluster effects is a model for differences across individuals, while a model for the within-cluster effects is a model for effects within individuals. Notice that V3 has no within-cluster variability, and hence can only be used as a between-cluster variable, while in contrast V4 has only within-cluster variability and no between-cluster variability. V5 would be used only to define the clusters, to enable computation of within- and between-cluster covariance matrices. *A similar setup with variables that may vary only at the within-level, or only at the between level, is used in other multilevel setups* (besides growth curves).

In the growth curve context, this modified data setup may be called the univariate representation of the data, since, for example, V1 now contains the repeated measures scores of each individual on the same conceptual variable. Singer and Willett (2003) call this data structure a "person-period" structure.

Contract and Expand Variables

The Windows front end of EQS 6 provides a way to transform data from multivariate to univariate representations, and vice versa. These are obtained the Data menu options. The CONTRACT function takes the multivariate representation and transforms it into the univariate representation appropriate for a multilevel analysis. Suppose we type the above 2-case 9 variable data shown above into a data file 2cases.ess. Then, when you choose Data, and Contract, you get a Contract Variables into Clustered Data dialog box. You specify a set of variables to contract, and EQS provisionally labels the new variable G1. In the above example, we would want to take V1-V4 and contract them to a single variable (V1=Y above, G1 in EQS). Then you choose another set of variables to contract, and provisionally label this G2. In the example, we would want to contract V5-V8 into a new G2. In addition, we tell EQS to create new Cluster and Sequence variables, which are the Case Number (V5) and Time Code (V4) variables in the univariate setup above. The file is given a provisional name. The resulting *.ess file, which you would rename, becomes:

```
     V9          G1          G2        CLUSTER       SEQ
 110.0000    20.0000     2.3000     1.0000     1.0000
 110.0000    21.0000     2.4000     1.0000     2.0000
 110.0000    22.0000     2.8000     1.0000     3.0000
 110.0000    23.0000     2.9000     1.0000     4.0000
 105.0000    16.0000     3.0000     2.0000     1.0000
 105.0000    18.0000     3.0000     2.0000     2.0000
 105.0000    20.0000     2.9000     2.0000     3.0000
 105.0000    22.0000     3.4000     2.0000     4.0000
```

Evidently, this is the same file we constructed above, except that the variables are organized differently. The original variables are given first, followed by the newly contracted variables, followed by cluster and time coding variables.

The EXPAND function does the reverse. It takes a data set in univariate repeated measures format and creates the standard flat file multivariate data set. Typically, you would expand on the cluster variable. But since the time coding sequence variable is in the univariate data file, it also gets expanded into a set of variables. In the example, such variables were not in the original file we started with.

Intraclass Correlations

Evaluating the Necessity of a Multilevel Model

When data are gathered under a cluster sampling design, it may or may not be necessary to analyze the data structure with a multilevel model. It is possible that there are no between-cluster effects. Although we are not necessarily recommending that you always compute intraclass correlations, they can be useful in diagnosing the extent to which a multilevel model may be needed in a given situation. Intraclass correlations are obtainable in the Windows front end of EQS. You can compute them for key variables in a study to indicate the degree of similarity or correlation between subjects within a cluster. It is easier to think of an intraclass correlation coefficient as a ratio of variances: the ratio of between-cluster variance divided by the sum of within- and between-cluster variance on a given variable.

$$ICC = \sigma_B^2 / \left(\sigma_W^2 + \sigma_B^2\right). \tag{11.5}$$

In practice, sample estimates are used in (11.5). If there are essentially no between-cluster effects, these coefficients will be small. If large intraclass correlations are found, the assumption of independent observations is violated. Muthén (1997) states that common values in survey data range from .0 to .5, and when values of .1 or larger are combined with group sizes exceeding 15, the multilevel structure of the data should be modeled. This is because the so-called design effect (deff) formula shows that ordinary standard errors would be seriously underestimated without taking the clustering into account. Since the inflation of sampling variance is given (in the single parameter case) by $V_C/V_{SRS} = 1+(c-1)(ICC)$, where V_C is the variance of an estimator under cluster sampling, V_{SRS} is the variance under simple random sampling, c is the group size and ICC is the intraclass correlation coefficient, clearly the product of large c and large ICC is what is problematic.

Intraclass correlations can be obtained in *EQS for Windows*, by running Analysis/Intraclass. We illustrate this on data kindly provided by Duncan et al. (1998). The data represent substance use reports on children in 435 families. The file contains 1204 cases overall, some from families with 2 children, some from families with 3 children, and so on. We select the alcohol use reports at 4 time points as having between- and within-cluster variances, and use the index variable as the cluster variable. The results show the following distribution of family sizes.

```
INTRACLASS CORRELATION
4   Variables are selected from file   c:\eqs61\examples\duncana.ess

 Number of cases in data file are ...........   1204
 Number of cases used in this analysis are ..   1204

 Cluster Size ......   2   3   4   5   6
 Number of Clusters  198 157  65  13   2
```

Not unexpectly, most families (198 of them) have only 2 children. No families have more than 6 children. On the alcohol use measures across four time points, the between-family variance, relative to the total variance, is more than trivial:

```
Estimated Intraclass Correlations

            ALC_T1    ALC_T2    ALC_T3    ALC_T4
            0.2772    0.2115    0.1798    0.1232
```

At each time point, there are systematic differences between families in how much alcohol they drink. With the largest cluster size of 6 siblings, and the largest intraclass coefficient, $V_C/V_{SRS} = 1+5(.2772) = 2.39$, implying at least a doubling of sampling variance. This suggests a potentially serious consequence of ignoring the cluster structure on ML standard errors. So, unless we use robust corrections to standard errors, a multilevel analysis would be appropriate, and it would be highly desirable if we are trying to distinguish between-family effects from differences among children. Furthermore, a multilevel analysis can, and with a cluster sample design probably should, proceed whether or not intraclass correlations are used as an initial basis for evaluation.

Suppose that all intraclass correlations examined for a given data set are very close to zero. This implies that there are essentially no between cluster or between group differences. In this situation, between cluster variances will be close to zero. The matrix Σ_B would then probably not be full rank, and it may be hard or impossible to fit a structural model to Σ_B. Multilevel estimation and testing may break down. Since $\Sigma = \Sigma_B + \Sigma_W$, if Σ_B is close to zero, $\Sigma \cong \Sigma_W$ and an ordinary SEM model is appropriate. Notice that this single model should be the model you may have considered as the within model of the multilevel setup. The opposite case, where the intraclass correlations are really huge and there is almost no within cluster variance, is quite rare. If it were to occur, a multilevel model may then also be hard to estimate and test, and you should fall back on a single model that amounts to the between cluster model. But, as just stated, the situation of trivial Σ_W will not occur very often.

Model-based Intraclass Correlations

Although we have not yet shown you how to set up a multilevel model, if you do an ML or MUML model, part of the modeling output will be the model-based equivalents to (11.5). That is, for each variable in turn, its diagonal element of $\hat{\Sigma}_B$ gives the numerator in (11.5), and it, along with the corresponding diagonal element of $\hat{\Sigma}_W$, gives the denominator. These are printed out as follows.

MUCH TO BE ADDED HEREAFTER

The basic setup for ML estimation

is a 2 group multiple group model that contains the statement MULTILEVEL=ML in group 1, plus the name of the variable that defines the clusters. If there is a mean structure, it can only be in the Between model (“group 2”).

```
/TITLE
EQS Input Program for Analysis of the School Data Set
Within Model
/SPECIFICATIONS
DATA='SCHOOL.DAT'; CASES=5198; VARIABLES=21; METHOD=ML;
MATRIX=RAW; GROUPS=2; ANALYSIS=MOM; MULTILEVEL=ML; CLUSTER=V19;
/LABELS
V7=Y6; V8=Y7; V9=Y8; V10=Y9; V19=SCHOOL; V20=X3; V21=X4;
F1=FW;
/EQUATIONS
Y6=1FW+E6;
Y7=*FW+E7;
Y8=*FW+E8;
Y9=*FW+E9;
/VARIANCES
FW=*;
E6-E9=*;
/END
/TITLE
Between Model
/LABELS
V7=Y6; V8=Y7; V9=Y8; V10=Y9; V19=SCHOOL; V20=X3; V21=X4;
F1=FB;
/EQUATIONS
Y6=1FB+E6;
Y7=*FB+E7;
Y8=*V999+*FB+E8;
Y9=*FB+E9;
FB=*V999+*X3+*X4+D1;
X3=*V999+E3;
X4=*V999+E4;
/VARIANCES
E6-E9=*;
D1=*;
E3-E4=*;
/COVARIANCES
E3,E4=*;
E8,E4=*;
/TECHNICAL
ITR=200; CON=.000001;
/END
```

12. MISSING DATA

The modeling methods discussed in earlier chapters require that data be complete, i.e., that a dataset contain no missing scores. Yet, studies with no missing responses are rare indeed. In fact, if there is *unit nonresponse*, or the outright failure of a sampled subject to participate in a study, such a case's scores will not appear in the dataset at all. Unit nonresponse is handled by survey statisticians by appropriate weighting of cases in the data file when calculating means and covariances in order to obtain less biased estimates of their population counterparts. If you know these weights – which you have to obtain outside EQS -- you can use case-weighted procedures as described in Chapter 5 to make appropriate corrections. In this chapter we deal with what survey statisticians sometimes call *item nonresponse*, which occurs in any kind of multivariate study (e.g., a survey) in which a subject responds to some items or variables but not to others. This chapter reviews some of the alternative methods available in EQS to appropriately deal with item nonresponse missing data.

Doing nothing is not an option. Already at the point where you import data into EQS, you have to inform the program about your coding for missing scores on given variables. For example, you have to specify that "*" or "8" or some other symbol represents a missing observation and not a score on the variable. Chapter 3 discusses how this is accomplished. Alternatively, once the data are imported, you can inform EQS about your missing data coding. This can be done in the Windows interface via the Data tab, Missing Data Specification. If this information is not adequately given to EQS, computed sample means, covariances, and correlations will have little meaning.

Now suppose that EQS has been made aware that certain entries in a data matrix are actually missing values. Structural modeling relies on means, standard deviations, and correlations as data to be modeled. Yet, ordinary formulae to compute these statistics cannot be applied because such formulae assume a complete, rectangular data matrix. There are a lot of options for dealing with this problem, as discussed in a huge literature (e.g., Allison, 2002; Graham, Cumsille, & Elek-Fisk, 2003; Little & Rubin, 2002; Little & Schenker, 1995; Schafer, 1997; Schafer & Graham, 2002; Wiggins & Sacker, 2002). Some of the options are generally outmoded. Others are recent, from the last decade. Some are so new that they have hardly disseminated beyond the original technical literature (Enders, 2002; Kim & Bentler, 2002; Yuan & Bentler, 2000d). Old as well as new methods will be found in EQS (Bentler, 2001). Using the taxonomy of missing data methods provided by Little and Rubin (2002, p. 19), namely: (1) procedures based on completely recorded units, (2) weighting procedures, (3) imputation-based procedures, and (4) model-based procedures, EQS provides some methods within each type. However, among model-based methods, Bayesian approaches to missing data have not yet been incorporated into EQS (e.g., Song & Lee, 2002c).

Before discussing some of the methods available in EQS to deal with missing data, we should discuss theories of how a data matrix with missing entries might have been generated. This may help guide you in selecting a method to be used to deal with the problem. Then we will discuss the alternatives for missing data handling, such as imputation, which, although relevant primarily to the general statistics sections of EQS that you access through the Windows interface, also could be used in SEM. Only then do we deal with the options that EQS provides for dealing with missing data within structural modeling. These include ML and pairwise methods with robust corrections, as well as multiple group approaches that may apply in specialized situations such as the cohort-sequential design. If you want to get to work running your models without much reading, the main conclusion from this chapter will be that in the /SPECIFICATION section of the model setup, you should declare MISSING=ML, along with the usual METHOD=ML,ROBUST if you have nonnormal data.

Background Concepts and Ideas

Missing Data Mechanisms

In an attempt to provide an understanding of how missing data might be generated, three types of mechanisms have been distinguished in the existing literature: *missing completely at random* (MCAR), *missing at random* (MAR),

and *not missing at random* (NMAR). These are defined technically in terms of probability distributions involving the observed parts of the dataset, the missing parts of the data set, the indicator matrix generated by presence or absence of data, and the parameters of the distributions (Rubin, 1976; Little & Rubin, 2002). Informally we may say that when the presence or absence of data ("missingness") is independent of the values of missing and observed values in the data set, missingness and/or the data is called MCAR; when missingness depends on *observed* values in the data set, it is called MAR; and when missingness depends on *unobserved* values, it is called NMAR. "If income is missing and age is fully observed, then the missing data are MAR if missingness of income depends only on age; however if for subjects of a given age, missingness of income depends on the value of income, then the missing data are not MAR, since missingness depends on the values of a variable that is sometimes missing (namely income itself)" (Little & Schenker, 1995, p. 44). "If values are MCAR or MAR, then the missing data mechanism is said to be *ignorable*, and valid estimates can be obtained without an explicit model of the missing data mechanism. If values are NMAR, then the missing value mechanism is *nonignorable* and a model of it must be incorporated into the estimation process. This can be quite difficult in practice, because the missing data mechanism is rarely known with much certainty" (von Hippel, 2004, p. 160). Nonetheless, ignorability does not require MAR, since "…MAR, together with the assumption of separable parameters, is a sufficient condition for ignorability of the MDM (missing data mechanism) in likelihood based inference. It is, however, not a necessary condition" (Lu & Copas, 2004, p. 755).

Of the statistical methods available in EQS, listwise deletion, pairwise deletion, and the robust missing data methodology are based on an MCAR assumption. When people informally say that data is randomly missing, they really are referring to an MCAR mechanism. That is, nothing systematic exists about the presence or absence of data. This is not the only meaning of MCAR, as missingness also can depend on a systematic mechanism as long as this mechanism is independent of the data. MCAR would be violated if, for example, missingness depends on some features of the observable scores on variables. MAR allows such dependence, and hence is a weaker kind of mechanism. The ML method for dealing with missing data, discussed below, remains valid even if data are not MCAR, providing that they are MAR and the data is multivariate normal. The weaker missing data assumptions thus favor the ML missing data methodology, as long as the stronger multivariate normality assumption is acceptable. If data are not multivariate normal, according to Laird (1988) and Rotnitzky and Wypij (1994), the ML parameter estimates will be inconsistent unless the missing data mechanism is MCAR.

In principle, then, in approaching our modeling analysis, we have to choose between two sets of untestable assumptions (a) data are normal and missingness is MAR or MCAR, or (b) data are nonnormal and missingness is MCAR. If (a) is true, the ML statistics will be well behaved (under the usual conditions, e.g., large sample size). If (b) is true, the robust ML methods will be well behaved (under similar conditions). If neither is true, there may be bias, but it is not possible to say what the bias is or how severe it might be. Of course, an extensive literature on ML with complete data verifies that its test statistic and standard errors are often distorted under non-normality. It is hard to believe that removing some data from the data file would improve the situation, whether the data is removed in a way consistent with MCAR, MAR, or NMAR theory. The consequences of violating MCAR and MAR are not well understood and have recently been the focus of study. Enders (2001b), Enders and Bandalos (2001), and Savalei and Bentler (in press) for example, compared the performance of several methods under MCAR and MAR. Results generally favor an ML methodology over its alternatives in either case, and typically imply relatively better performance under MCAR. But Savalei and Bentler (in press) concluded that the number of patterns of missing data (see below) may be more important than the missingness mechanism, as they surprisingly found better performance on some measures under MAR than MCAR.

In practice, one may wish to choose the weaker MAR assumption. But you should remember that *working with a weaker assumption does not make the assumption true.* Even MAR may be violated, and probably will be violated with real data. Hence ML may still yield biased results even with normal data. Actually, whatever your assumptions on the missing data mechanism may be, the ML parameter estimates will remain the same. Until methods for routine testing of various missing data mechanisms are developed, you simply have to recognize that you are making assumptions on the missing data mechanism. Below we provide a test of normality and, if the data are normal, some tests of MCAR. There seem to be no tests for MAR or NMAR that can generally be applied, even though real data no doubt will contain some missing scores that are MCAR, others that are MAR, and perhaps some that are NMAR.

Whether the data for your model are MAR is to some extent under your control. When even the weaker MAR mechanism is not appropriate for a given data set because the cause of missingness is not well represented in the data to be analyzed, it may make sense to augment the data with auxiliary variables that either may have a good chance to represent the cause of missingness, or that are highly correlated with the variables containing missingness (Collins, Schafer, & Kam, 2001; Enders & Peugh, 2004). This can be accomplished by treating the auxiliary variables as saturated correlates of independent variables in the model (Graham, 2003). Alternatively, you can include those variables in your model. Even then, however, MAR remains an assumption that cannot be tested, though if you are correct in your selection of control variables, you will have minimized potential distorting effects.

Complete Cases: Listwise Deletion

An option available in most statistics programs as well as EQS is to use only those cases in the data file that contain scores on all variables. This is a form of weighting, where each case is given a weight 1 if it has complete data, and 0 if it contains any missing data. Complete-case analysis is often called *listwise deletion*. You can think of the cases arranged in a list, and being selected or nonselected depending on whether even a single score is missing for that case. As a result, a rectangular and complete data matrix is obtained for the selected cases, allowing standard methods of analysis to be used. This is a nice property, especially if only a few cases are eliminated out of a large data file, and if the missing data are MCAR, which is an assumption of the method. For MCAR data, complete case analysis provides an unbiased methodology. That is, if the original sample was unbiased, the reduced sample will remain an unbiased sample of the population.

As recently as 10 years ago, listwise deletion was the most frequently used missing data method in applied psychology (Roth, 1994). This is less true today as problems with the methodology are becoming understood and better alternatives become available. In particular, listwise deletion often eliminates a huge proportion of the sample data, especially as the number of variables increases, so that there may be a substantial loss of precision in estimation, as well as in power. For example, the data file amos17.ess discussed above contains 73 cases. But, since 66 of these cases contain some missing data, listwise deletion would leave only 7 cases in the file to be analyzed! This is clearly an unacceptable loss of information in this situation, and better alternatives should be utilized.

If you choose no specific option for dealing with missing data in a SEM model, EQS will default to listwise deletion.

Available Cases: Pairwise Analysis

Another practical alternative to listwise deletion is to use all the univariate and bivariate information in the data to compute summary statistics. A variable's mean or variance can be calculated from data provided by every case for which scores on that variable are available. This is a default in data description with EQS. A case that provides data for a mean, however, may not be usable when calculating a correlation or covariance if that case is missing the corresponding score on the other variable. A case may contribute bivariate data for a given correlation of, say r_{12}, but it may not be used in the computation of another correlation, say r_{13}, if its score on V3 is missing. As a consequence, the number of subject used to compute the various sample means is not the same. Neither is the number of subjects for computing the various correlations or covariances.

This *pairwise present* method uses all the data available for any given statistic. As a result, Ns in pairwise analysis can substantially exceed the N used in listwise deletion. This is a positive feature. When the missing data mechanism is MCAR, pairwise analysis gives consistent estimates of population parameters. It can perform well (e.g., Brown, 1983). However, the method also has some drawbacks in typical applications: Namely, the lack of a unique sample size that can be used in statistical tests, and possible inconsistencies among various correlations due to their being based on different Ns. This inconsistency is, however, a small sample phenomenon under MCAR. Technically it means that the sample covariance matrix is not positive definite, meaning that some underlying dimensions can have negative variances (a disconcerting concept) or some R^2 values may be outside the 0-1 range.

The varying Ns for the various statistics based on pairwise computations have been considered a serious problem when modeling with pairwise data (Marsh, 1998). This issue has been resolved with the new pairwise SEM methodology in EQS. This methodology is discussed in the SEM part of the chapter below.

EQS' Missing Data Diagnosis

Available via Analysis / Missing Data Analysis, this output provides information about missingness on individual variables and pairs of variables. For example, the amos17.ess file provides the following pairwise missingness information (V1-V3 only shown):

```
              visperc     cubes  lozenges
   visperc        20
     cubes        38        23
  lozenges        38        42        24
```

Thus 20 cases (out of 73) have missing scores for the mean of V1, 23 cases have missing scores in V2, but 38 are missing for computing the correlation of V1,V2. On the positive side, these numbers show that close to half the subjects provide information for correlations, substantially more than are available under listwise deletion.

As part of this output, EQS also provides a simple correlational diagnosis that can help to evaluate the pure randomness of missing information, a special case of MCAR. If we recode the data matrix into a new matrix of "1" and "0" elements depending of whether data was present or absent (and ignoring the actual values of any scores), and then correlate the columns of this matrix, under pure randomness we should observe quite low correlations since missingness or presence of data on one variable should not be predictable by presence or absence of data on another variable. Here

```
Correlations of the Dichotomized Variables
             visperc     cubes  lozenges
   visperc    1.0000
     cubes   -0.0860    1.0000
  lozenges   -0.0376   -0.1608    1.0000
```

These correlations are reassuring, but they do not provide an actual test of MCAR, which is available when modeling (see below).

Imputation

Standard methods in EQS are based on the assumption that the data matrix is rectangular, containing a datum for each subject on each variable. When there are missing values, there are "holes" (missing scores) in this matrix. Imputation can be used to create a full data matrix by substituting a seemingly reasonable estimate for each missing datum. Such an imputed data matrix can then be analyzed by standard procedures. There are at least five possible ways to implement imputation: mean imputation, regression imputation, stochastic regression imputation, hot-deck imputation, and EM imputation. All of these, except hot-deck imputation, are available in the Windows front-end of EQS 6. *As we will note, however, perhaps you should not use imputation at all!* There is a related methodology, multiple imputation, that has fewer limitations, though it is not yet implemented in EQS.

Mean Imputation

One of the most common methods for imputing missing data is, for each variable y, to use the arithmetic mean $\bar{y}$ of existing data to replace a missing datum on that variable. Although researchers on missing data decry the use of this method, it will work acceptably when a very tiny percent, say, less than ½ of 1%, of the data are missing in a completely random way, that is, with MCAR data. With a data matrix of 100 subjects and 10 variables, there are 1000 scores, so this might be an option if fewer than 5 scores are missing. In such a situation, almost anything you do will lead to the same results.

A reason to be cautious about using this method is that mean imputation will reduce the variability of the imputed variable. Imagine replacing scores far from the mean by the mean – variance will be drastically reduced. Similarly,

the correlation between the imputed variable and other variables will be reduced. See, for example, Gleason and Staelin (1975), and Graham, Hofer and MacKinnon (1996). More importantly, there are better methods to use.

Regression Imputation

A linear regression equation is of the form $y = \alpha + X\beta + \varepsilon$. If the i^{th} case on variable *y* contains missing data, the predictor scores x_i' (the i^{th} row of *X*) for that case can be used along with the intercept to obtain the regression estimated score $\hat{y}_i = \hat{\alpha} + x_i'\hat{\beta}$. This will equal the mean imputed value if $x_i'\hat{\beta}$ is zero, but otherwise it allows prediction away from the mean and hence will be a better estimate of the missing score than obtained by mean substitution. The extent to which a predicted score can be above or below the mean depends on the size of the R^2 in predicting *y* from *X*, and hence predictors that are highly related to *y* are desirable to use. There are also other regression-type imputation methods, such as using a person's mean score on similar items, that are used frequently in test construction and, in evaluations, have been found to perform well (e.g., Bernaards & Sijtsma, 1999, 2000). However, EQS does not provide this methodology.

In practice, you can decide which variables to include as predictors. As *X* itself may contain missing scores, the estimate $\hat{\beta}$ must take this into account. In EQS, this is done using listwise deletion (see above).

Stochastic Regression Imputation

Like mean imputation, regression imputation creates estimates that are too close to the mean. The variability of the imputed variable is too low. One way of correcting for this is to add a stochastic or random term to the predicted score of the regression imputation. That is, we impute $\hat{y}_i = \hat{\alpha} + x_i'\hat{\beta} + \hat{\varepsilon}_i$, where $\hat{\alpha}$ and $\hat{\beta}$ are as before, and $\hat{\varepsilon}_i$ is an appropriately chosen random number. One approach is to model the errors as coming from a normal distribution having the required estimated residual variance under the model. Such a procedure seems to work reasonably well in practice (see e.g., Jinn & Sedransk, 1989). This is a good approach under the model, but it is misspecified if the residuals do not have a normal distribution. In EQS, $\hat{\varepsilon}_i$ is a randomly chosen residual from the other cases in the data set. This has the advantage that the actual distribution of the residuals is approximately maintained, even if it is not normal.

Of course, as $\hat{\varepsilon}_i$ is a randomly chosen, it may be positive or negative and may move the regression estimate in the wrong direction relative to the true but unknown y_i. But the imputed variable should have closer to the true variability.

Hot-deck Imputation

Let us consider the existing data set a "hot" deck (dataset), and another similar dataset (say, from an earlier wave of data) as a "cold" deck. Then the hot-deck imputation method uses values of variables of other subjects in the current data set to impute missing values. A cold-deck method would use data from another data file. In hot-deck imputation, donor subject scores could be chosen randomly, or by some type of resemblance procedure. In similar response pattern imputation (Brown, 1994; Gold & Bentler, 2000), the donor case is chosen to maximally resemble the recipient case (with the missing score), using some criterion of similarity on a set of standardized variables for which data are complete. Though this approach can perform quite well, no hot-deck method is available in EQS.

EM Imputation

The EM computational procedure for estimating parameters has a long history (Orchard & Woodbury, 1972; Dempster, Laird, & Rubin, 1977) and has been widely applied to a variety of statistical problems. It has been applied to covariance structures by considering latent variables as missing data (e.g., Kiiveri, 1987). It was only recently imported into normal theory based maximum-likelihood SEM with missing observed data (e.g., Arbuckle, 1996; Jamshidian & Bentler, 1999; Rovine, 1994; Tang & Bentler, 1998). The EM algorithm generates a sequence of parameter estimates by cycling iteratively between an expectation (E) step and a maximization (M) step. In the E-step, the conditional expectation of the complete data log-likelihood is computed given the observed data and the current parameter estimates. In the M-step, the expected log-likelihood obtained in the E-step is maximized to

obtain updated parameter estimates under the given model. The iterative process is stopped when the parameter estimates converge to some pre-established criterion. During this process, missing scores are imputed. The EM-based imputed scores are those obtained at the conclusion of the iterative process.

When the EM algorithm is used with a saturated model on normal variables, the "EM algorithm for this problem can be viewed as an iterative form of regression imputation. In fact the E step effectively imputes the missing values in each case by the predictions from the regression of the missing variables on the observed variables, with coefficients based on current estimates of the parameters. The M step estimates the mean and covariance matrix from the filled-in data, with corrections for the covariance matrix for imputing predicted means" (Little & Schenker, 1995, p. 52). This can be thought of as a sequential process in which variables 1, 2, …p are each imputed in turn, and the procedure cycles through these variables again and again, until convergence. What is interesting about this approach is that it produces estimates of a common mean vector and common covariance matrix for all patterns of incomplete data, even though a given pattern may only contain information about certain means or covariances. Computations for the saturated model are discussed by Graham, Hofer and MacKinnon (1996), Little and Rubin (2002), Rovine (1994), Schafer (1997) and others. (As discussed further below, variants of this algorithm also can be used to obtain estimates of structural model parameters.)

Logically and empirically, this type of iterated improvement seems to yield better estimates of imputed scores than those obtained by the previously described methods. *If you have to impute scores, this is the method we recommend.*

But should you impute? Probably not. If you do imputation, you must be planning to use a complete data method on the resulting full data matrix. But statistics based on this full data matrix do not adequately take into account the fact that you do not, in fact, have a complete data matrix, and still do not adequately reflect variability in the imputed values. If you can, you should use a method that uses all available data, but not any artificially imputed data. The ML and SEM methods in EQS do that, as discussed below.

On the other hand, if you plan to use a method of analysis – perhaps a non-SEM method -- for which adequate missing data procedures do not exist, you may have no option but to analyze an imputed data matrix.

Multiple Imputation

As noted, imputing a given value for a missing score implies greater certainty about the missing observation than is warranted. Rubin (1987) proposed multiple imputation to minimize this problem. In this methodology, rather than use a single imputed value, two or more imputed values are selected from a distribution appropriate to the hypothesized missing data mechanism. The imputed values for each missing observation are ordered, and replacing each missing value by the first such imputed value creates the first complete data set; replacing each missing value by the second such imputed value creates the second complete data set; and so on. Thus this procedure yields two or more complete data sets, each of which can be analyzed by the same traditional procedure. In SEM, this would mean that there are then two or more sets of parameter estimates, model tests, and so on. These several sets of results then are combined in a Bayesian or meta-analytic way to yield an overall summary. See Graham and Hofer (2000), Little and Rubin (2002), Schafer (2001), Schafer and Graham (2002), or Sinharay, Stern, and Russell (2001) for further discussion. Multiple imputation is not yet incorporated into EQS.

Options in General Statistics Modules

We conclude this section by summarizing the types of approaches to missing data that are used in the general statistics front-end of EQS, that is, in methods other than structural modeling. As stated previously, *Descriptive Statistics* analyzes each variable separately, using all available data and excluding missing values only for the given variable. *Frequency Tables* also analyzes each variable separately using all available cases for that variable. "MISSING" becomes one of the rows in the cross-tab table, so that the user can see that e.g., 3 cases out of 188 are missing. There are a large number of procedures that use listwise deletion, i.e. they exclude a case if any variable used in the analysis is missing: *t-Tests, Crosstab, One-way and Two-way ANOVA, Regressions, and Intraclass Correlation. Correlation, Factor Analysis, and Nonparametric Statistics* use listwise deletion by default, but pairwise analysis can be chosen as an option instead. The *General Linear Model ANOVA* excludes any case if a

grouping variable or covariate is missing. A case having the dependent variable missing does not affect the calculations, but the value of the dependent variable is imputed by mean imputation. This is essentially invisible to the user. If the data is also saved, the imputed values will be saved and hence become visible. Finally, *Missing Data Analysis* does the missing data diagnosis, and provides the imputation options discussed above.

Structural Equation Modeling

Missing Data Patterns

EQS prints out missing data diagnostics to help you understand the extent and nature of your missing data. If we consider our dataset as the matrix *Z* with elements z_{ij} containing the score of the i^{th} case on the j^{th} variable, then we can define a new matrix *Y* to represent the presence or absence of data. That is, we have $y_{ij} = 0$ if z_{ij} is missing, and $y_{ij} = 1$ if z_{ij} is observed. A missing data *pattern* is a unique row of this matrix; that is, it is a unique vector of zeros and ones. Then most likely several cases in the data file will have an identical pattern of 1's and 0's. In general, if there are ℓ variables with incomplete data, where $\ell \in \{1,\ldots, p\}$, then there can be up to 2^{ℓ} incomplete data patterns if $\ell < p$, or $\left(2^{\ell} - 1\right)$ incomplete data patterns if $\ell = p$. In practice, there may be far fewer.

In an example data file, amos17.ess, N=73 and p=6. In EQS, you first get summary information about the data file as follows

```
NUMBER OF CASES USED                  =   73
NUMBER OF CASES WITH POSITIVE WEIGHT  =   73
NUMBER OF CASES WITH MISSING DATA     =   66
NUMBER OF MISSING PATTERNS IN THE DATA =   29
```

here showing that of the 73 cases that could be processed, 66 had one or more of the 6 variables missing. Since several cases had the same pattern of missing data, the matrix *Y* of presence/absent data could be reduced to a 29x6 matrix. This is printed out in EQS in the following way.

```
IN THE SUMMARY OF MISSING PATTERNS, M REPRESENTS A MISSING VALUE

                          VARIABLES
    #        #       %
 MISSING   CASES   CASES  123456
 -------   -----   -----
       0       7    9.59
       4       1    1.37   MM MM
       3       1    1.37   MM M
       2       3    4.11   MM
       1       9   12.33    M
       2       4    5.48  M M
       3       2    2.74  M M M
       2       2    2.74    M M
       1       1    1.37      M
       2       4    5.48     MM
       3       1    1.37   M MM
       2       1    1.37   M M
       3       2    2.74  MM M
       2       2    2.74  M  M
       3       1    1.37  M  M M
       4       1    1.37  M  MMM
       3       1    1.37  M  MM
       2       2    2.74  M   M
       3       1    1.37  MM  M
       3       2    2.74  MM   M
       2       2    2.74   M   M
       1       4    5.48       M
       2       1    1.37     M M
       1       4    5.48     M
       2       2    2.74    MM
       1       9   12.33   M
       1       1    1.37  M
       3       1    1.37  M   MM
       2       1    1.37      MM
```

The first row shows that 7 cases – 9.59% of all subjects -- have no missing data. The second row shows that 1 case has 4 missing scores, with missing data on variables 2,3,5,6; alternatively, this case only has data on V1 and V4. And so on. There are two patterns of presence/absence of data that occur most frequently: 9 cases were missing only V3, and 9 cases were missing only V2.

Clearly, the amount and type of data available for different cases varies tremendously. If one were to compute means and covariances for each pattern of missing data, the amount of information available from a given pattern may be very minimal indeed. In fact it is impossible to compute correlations/covariances for patterns that are only exhibited by one subject. All of this poses numerous statistical challenges: evaluating hypotheses about multivariate normality, whether the data could be MCAR, what a common set of means and covariance might be, and so on. Methods of dealing with missing data by treating each pattern of data as a sample in multisample analysis (e.g., Allison, 1987; Muthén, Kaplan, & Hollis, 1987), as was done in the cohort sequential design example of Chapter 10 and will be further illustrated below, simply cannot be used. There are too many patterns, most of which have far too few subjects.

Test of Normality

As normality is a key assumption of the normal theory based maximum likelihood method, it would be desirable to test the assumption that the data are multivariate normally distributed. Since there are typically many patterns of missing data, methods developed for one complete dataset cannot be used. Yuan, Lambert, and Fouladi (in press) developed an extension of the Mardia (1970, 1974) test of multivariate kurtosis that can be applied to missing data. It effectively aggregates information across patterns to get one overall summary. Whenever a model with missing data is run, EQS automatically computes their statistic and prints it out as:

```
                        MULTIVARIATE KURTOSIS
                        ---------------------

YUAN, LAMBERT, & FOULADI'S COEFFICIENT (G2,P) =     -0.7012
NORMALIZED ESTIMATE =                              -0.3961
```

As usual, the normalized estimate is interpretable as a standard normal variate, so if it is outside the -3 to +3 range (roughly), the hypothesis of multivariate normality can be rejected, especially if sample size is large. In small samples, EQS uses an adaptation of the Bonett, Woodward, and Randall (2002) test for multivariate kurtosis. Results of this test are printed as follows (if normality is tenable):

```
BONETT-WOODWARD-RANDALL TEST SHOWS NO SIGNIFICANT EXCESS
KURTOSIS INDICATIVE OF NON-NORMALITY.
```

If you were considering using a normal theory method of analysis and the data is not normal, you should probably use the robust methodology described below.

Tests of MCAR

Assumptions and Use

When data represents a random sample from a multivariate normally distributed population, and data is complete, all the information in the data can be summarized by a single set of sample means $\bar{z}$ and a single covariance matrix *S* which are estimators of the corresponding population parameters μ and Σ. Now let *i* represent a pattern of data. When data are incomplete, there are a lot of possible sample means and covariances since each pattern *i* of missing data could possibly generate an associate sample mean vector $\bar{z}_i$ and covariance matrix S_i (though the latter is not defined if there is only one case for that pattern). These vectors and matrices are of different size, reflecting the particular sets of variables with available data in that pattern. In the above example, with 29 patterns of data, there could be up to 29 separate sets of means and covariance matrices (though with 13 patterns having only a single case, there can be only 16 different sample covariance matrices). If data were MCAR, then all patterns of data would be representative samples from the same population. Then the various means and covariances should be *homogeneous* and can be treated as samples from a single population with only one μ and Σ. In that case it makes sense to evaluate a single mean and covariance structure model $\mu = \mu(\theta)$ and $\Sigma = \Sigma(\theta)$. In principle, the various means can be homogeneous (or not), and separately, the various covariance matrices can be homogeneous (or not). If the covariance matrices are homogeneous, a covariance structure hypothesis $\Sigma = \Sigma(\theta)$ would be appropriate even if the means are not homogenous.

Building on the work of Little (1988) and Tang and Bentler (1998), Kim and Bentler (2002) developed three tests of MCAR *for normally distributed data* that have been incorporated into EQS. The first test, the *GLS test of homogeneity of means*, tests whether the various $\bar{z}_i$ could be considered as samples from a common population with a common mean vector μ. The second test, the *GLS test of homogeneity of covariance matrices*, tests whether the various S_i could be considered as samples from a common population with a common covariance matrix Σ. The third is a combined *GLS test of homogeneity of means and covariance matrices*, that tests both hypotheses simultaneously. Whenever you run MISSING=ML, these tests get computed automatically, and its results are printed in the goodness of fit section of program output as illustrated next.

```
GLS TEST OF HOMOGENEITY OF MEANS
  CHI-SQUARE =      104.103 BASED ON     105 DEGREES OF FREEDOM
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS       .50639

GLS TEST OF HOMOGENEITY OF COVARIANCE MATRICES
  CHI-SQUARE =      168.060 BASED ON     158 DEGREES OF FREEDOM
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS       .27704
    NUMBER OF CASES USED FOR THE ABOVE STATISTIC IS           60
```

```
GLS COMBINED TEST OF HOMOGENEITY OF MEANS/COVARIANCES
  CHI-SQUARE =      272.163 BASED ON     263 DEGREES OF FREEDOM
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS       .33579
```

These tests, especially the combined test, have a large degrees of freedom. This reflects the many data means and covariances from the numerous patterns of data, and the relatively few common means and covariances to be estimated. If the probability associated with a test is large (say, greater than .05), the respective null hypothesis is accepted. You can then conclude that the various patterns of missing data can be considered as samples from a single population, as in this example. This confirmation (non-rejection) of MCAR is reassuring for model building.

If the probability for a test is small, indicating that the corresponding hypothesis is untenable, MCAR can be rejected. In that case, it may or may not make sense to assume that a single structural model would fit your data. If only mean homogeneity is rejected, and the means are irrelevant since you are planning to do a covariance structure model, this may not matter. If covariance homogeneity is rejected, you need to be more cautious. If a model for an unstructured common Σ is misspecified (from the test), it may not make sense to expect much success with a single structural model $\Sigma = \Sigma(\theta)$. Ideally, EQS would provide diagnostics that point to the data pattern(s) that are different from the others, but such diagnostics are still in the research stage and will be added to later releases of EQS. For now, you will have to do your own data exploration.

But there are situations when you can ignore the results of these tests.

Rationale for a single structural model if homogeneity is rejected. *First*, these homogeneity tests currently are based on the assumption of multivariate normal data. Hopefully this assumption will be released in updated versions of EQS. Since the robustness of these tests to violation of normality is not known, you should not assume it. *Second*, the large sample statistical theory requires that there are a reasonably large number of observed cases for each pattern of missing data. In the amos17 example, this condition is not met since about 40% of cases fall in patterns that have only 1 or 2 cases. In such a situation, the statistical theory is liable to be unreliable, especially for the covariance and combined tests. In practice, it may suffice if the vast majority (say, 90%) of cases fall in patterns that have 3 or more cases. *Third*, it has to be recognized that all statistical tests can, by chance, reach the wrong conclusion. Preliminary evidence on the tests' performance is good (the means test performs particularly well; Kim & Bentler, 2002), but further work may find conditions under which they fail. *Fourth*, logically speaking, there exist conditions under which a given structural model form with certain parameters can hold in subpopulations and yet a model of the same form, with modified parameters, also can hold in the total population (e.g., Meredith, 1993, Theorem 8). For example, suppose that the i^{th} pattern is drawn from a subpopulation with covariance matrix $\Sigma_i = \Lambda\Phi_i\Lambda' + \Psi$ and that several patterns have different Φ_i. This means that their covariance matrices Σ_i will not be equal, and if the test for MCAR is working properly, the hypothesis of homogeneity of covariance matrices should be rejected. Yet it remains possible for an overall factor model of the form $\Sigma = \Lambda\Phi\Lambda' + \Psi$ to hold in the overall population! This implies that it would not be a mistake to do a single structural model for all patterns, though doing so would, of course, lose information. We would argue that it is incumbent on the researcher to demonstrate such factorial invariance rather than simply to assume it. Typically this is done by doing a multiple group mean and covariance structure model across the patterns suspected to be different, taking into account missing data. Research on factorial invariance theory, primarily emphasizing complete data, is active (e.g., Lubke, Dolan, Kelderman, & Mellenbergh, 2003; Millsap, 1995, 1997, 1998).

Technical Details

Let $\bar{z}_i$ and S_i be the means and unbiased sample covariance matrix of the p_i observed variable(s) for pattern i (i=1,…, m). Let $\mu_i(\hat{\theta})$ and $\Sigma_i(\hat{\theta})$ be the corresponding estimated mean vector and model covariance matrix. These are a subvector of $\mu(\hat{\theta})$ and submatrix of $\Sigma(\hat{\theta})$, respectively, of the modeled mean vector and covariance matrix of all p variables based on the ML estimator $\hat{\theta}$. The number of cases for pattern i is n_i, with relative proportion given by $c_i = n_i / N$ or $k_i = (n_i - 1)/N$, where $N = \sum_{i=1}^{m} n_i$ is the total number of cases. The number of available sample means is $p_1^* = \sum_{i=1}^{m} p_i$. Let j be any pattern with only one case; then, the number of available sample covariances is

$p_2^* = \sum_{i=1(i\neq j)}^{m} p_i(p_i+1)/2$. The number of available sample means and covariances is $p_3^* = p_1^* + p_2^*$. Based on this, Kim and Bentler (2002, eqs. 4-6) defined the following three tests of homogeneity.

The hypothesis that the population means $\mu_i(\theta)$ are subsets of a single population mean vector $\mu(\theta)$ (that is, $\mu_i(\theta) \subseteq \mu(\theta)$ for all i=1,…,m) is given by the test statistic

$$G_1(\hat{\theta}) = \sum_{i=1}^{m} c_i \left(\left(\bar{z}_i - \mu_i(\hat{\theta}) \right)' \Sigma_i^{-1}(\hat{\theta}) \left(\bar{z}_i - \mu_i(\hat{\theta}) \right) \right). \tag{12.1}$$

This is evaluated against a χ^2 distribution with $df_1 = p_1^* - p$. This is a minor variant of Little's (1988) test.

The hypothesis that the population covariance matrices $\Sigma_i(\theta)$ for the various patterns of incomplete data are subsets of a single population covariance matrix $\Sigma(\theta)$ (that is, $\Sigma_i(\theta) \subseteq \Sigma(\theta)$ for all i=1,…, m) is given by

$$G_2(\hat{\theta}) = \sum_{i=1}^{m} \frac{k_i}{2} tr\left(\left[S_i - \Sigma_i(\hat{\theta}) \right] \Sigma_i^{-1}(\hat{\theta}) \right)^2. \tag{12.2}$$

This is evaluated against a χ^2 distribution with $df_2 = p_2^* - p(p+1)/2$. Note: these df come from Bentler, Kim, and Yuan (2004) and differ from Kim and Bentler (2002) when there are patterns with one case. The latter are omitted from the computations, and as a result, a downward adjustment of df occurs.

The combined test evaluates the simultaneous hypotheses associated with (12.1) and (12.2)

$$G_3(\hat{\theta}) = G_1(\hat{\theta}) + G_2(\hat{\theta}). \tag{12.3}$$

This test has $df_3 = df_1 + df_2$ degrees of freedom, and evaluates in an overall way whether the population means $\mu_i(\theta)$ and covariance matrices $\Sigma_i(\theta)$ are subsets of a single population mean vector $\mu(\theta)$ and covariance matrix $\Sigma(\theta)$ (that is, $\mu_i(\theta) \subseteq \mu(\theta)$ and $\Sigma_i(\theta) \subseteq \Sigma(\theta)$ for all i=1,…, m).

Maximum Likelihood

We are now ready to discuss our recommended method for doing structural modeling with missing data. This is the method of maximum likelihood that you have encountered many times before in this *Manual*. In this context, it is often referred to as *direct maximum likelihood* for reasons we explain below. It is also sometimes called *full information maximum likelihood* (abbreviated FIML), which is meant to be contrasted with "limited information" ML estimators. Limited information estimators are not computed in EQS. Or, it may be called *case-based maximum likelihood.* These expressions are interchangeable. What you have to be clear on, however, is that the functions optimized in normal theory ML estimation with complete data given, e.g., for covariance structures by eq. (5.13), no longer can be applied with missing data since the ordinary sample covariance matrix *S* does not exist. Hence, a different maximum likelihood function is used instead. Nonetheless, all that is obtained is the ML solution, with its associated estimates, test statistic, and standard errors. Relations among the direct and traditional ML approach are summarized in the Appendix to this chapter.

In the context of missing data, you may also hear the ML estimator called the *EM algorithm estimator.* The EM algorithm was mentioned above as one way to get imputed values for missing data, but its wider use is in maximum likelihood estimation. In EQS, the EM algorithm and ML estimation are tied together in the missing data context, that is, we use EM as an iterative way to compute the ML estimator for any model that you specify. If you estimate a structural model, following Gold and Bentler (2000) this may be called structured EM, while if you only estimate a saturated model, this might be called unstructured EM. But EM, which is related to a wider class of optimization methods (e.g., Hunter & Lange, 2004), is mainly an effective computational convenience. In principle, other ways can be used to compute an ML estimator, and, in fact, many structural modeling programs do not use EM to

compute the ML estimator for structured models. However, assuming no problems in optimization, the resulting ML estimator will be identical, no matter how it is obtained.

Separating the EM iterative computations from the model that is being estimated is not a universal practice in the field, however. In standard discussions of missing data, EM is used primarily to describe a method for computing estimates of the saturated mean μ and covariance matrix Σ without any structural model. In EQS you can, of course, obtain the ML estimator of the saturated μ and Σ, but you can also obtain ML estimator $\hat{\theta}$ under structural models $\mu = \mu(\theta)$ and $\Sigma = \Sigma(\theta)$. In another usage, direct ML estimation and EM are treated as different methodologies. For example, EM may denote a two-stage method of estimation that first obtains saturated estimates $\hat{\mu}$ and $\hat{\Sigma}$, and then treats these *as if* they were the ordinary sample mean $\bar{X}$ and covariance matrix *S* as sample input to an ordinary complete data modeling run (Enders & Peugh, 2004). This two stage methodology is not currently available in EQS as it requires special statistical handling. But the theory is known (Yuan & Bentler, 2000d).

There is a long history to the ML approach for handling incomplete data in a variety of saturated modeling contexts, see e.g., Dempster, Laird and Rubin (1977), Little and Rubin (2002), Little and Schenker (1995), Rovine (1994), or Schafer (1997). The direct or casewise ML methodology for incomplete data in structural equation modeling was developed by Finkbeiner (1979) and Lee (1986). It was extended to a wider class of structural equation models by Arbuckle (1996). Jamshidian and Bentler (1999) proposed a different way to do the computations, using an EM algorithm; this is implemented in EQS. The ML approach has become popular for both theoretical as well as empirical reasons. Theoretically, it is based on the weaker MAR assumption while its competitors require the stronger MCAR assumption. Empirically, the method just seems to work better than its alternatives under fairly wide modeling conditions (e.g., Arbuckle, 1996; Enders, 2001b; Enders & Bandalos, 2001; Gold & Bentler, 2000; Gold, Bentler, & Kim, 2003; Graham, Hofer, & MacKinnon, 1996; Muthén, Kaplan, & Hollis, 1987; Savalei & Bentler, in press). In this section, we discuss ML estimation with missing data when the assumption of multivariate normality is reasonable; the next section discusses the associated robust methodology.

Assumptions and Use

Let us first review a key assumption. ML is a normal theory based method, and so, as usual, you should verify that your data is multivariate normally distributed. Procedures for doing this were described above. The Yuan, Lambert and Fouladi (in press) coefficient of multivariate kurtosis should be reasonably small, and, in small samples, the Bonett, Woodward, and Randall (2002) test of normality should not be rejected. If you have nonnormal data, you will want to do robust computations as described in a later section.

The setup for a modeling run with missing data is largely identical to the setup for a modeling run without missing data. To do direct ML, you have to do the following.

In /SPECIFICATION you have to add:
1. MISSING=ML;
2. SE=FISHER; (or, OBSERVED)
3. ANALYSIS=MOMENT;

In /EQUATIONS to every V equation, you have to add the predictor:
4. *V999

The following setup for a 2 factor model illustrates these 4 requirements.

```
/SPECIFICATIONS
DATA='c:\eqs61\examples\amos17.ess';
VARIABLES=6; CASES=73;
METHOD=ML; ANALYSIS=COVARIANCE; MATRIX=RAW;
MISSING=ML; SE=FISHER; ANALYSIS=MOM;
/EQUATIONS
V1 =   *V999 + 1F1 + E1;
V2 =   *V999 + *F1 + E2;
V3 =   *V999 + *F1 + E3;
V4 =   *V999 + 1F2 + E4;
V5 =   *V999 + *F2 + E5;
V6 =   *V999 + *F2 + E6;
```

Let us discuss the four requirements in turn.

MISSING=ML specifies ML rather than pairwise or listwise computations. Pairwise would require the statement MISSING=PAIR (discussed further below) while listwise is a default that is used when there is no MISS statement.

SE=FISHER indicated that Fisher information matrix is used to compute standard errors. The alternative statement would be SE=OBSERVED. With complete data, EQS and most other modeling programs use the *Fisher Information matrix* to compute standard errors. This is a kind of covariance matrix of the parameter estimates based on the assumption that sample size is very large and that the model is correct. As with other covariance matrices, standard deviations -- here, standard errors -- are given as the square roots of the diagonal elements of such a matrix, after appropriate renorming to take the actual sample size into account. With missing data, EQS follows Jamshidian and Bentler (1999) to also provide standard errors based on the *Observed Information matrix*. The observed information is based on the 2nd derivatives of the ML function at the given sample size and allows for discrepancies between model and sample data. As a result, it has been argued the observed information might provide better estimates of standard errors in small samples (Dolan & Molenaar, 1991).

ANALYSIS=MOMENT tells EQS that you are doing *a mean structure model*. That is, data and model involve use of means. You have to do this even if you have no interest in the estimates $\hat{\mu}$. The direct ML function requires estimates of the means of the variables – see eq. (12.5) below – and since the ordinary sample means do not exist with missing data, we cannot use $\bar{z}$ as $\hat{\mu}$. While there exists another ML method that allows only a covariance structure to be analyzed (Yuan & Bentler, 2000d, pp. 178-179), this is not currently programmed in EQS.

***V999 in each V equation** tells EQS to estimate the V's mean. V999 is EQS's name for the constant "1", and, as is fully discussed in Chapter 8, regression on a constant is an intercept. The total effect of V999 on a given V gives the model-reproduced mean of V. When there is no indirect path from V999 to the Vs, as here, the direct effect is the total effect. Hence, the parameter *V999 for a given variable is just that variable's mean.

The above setup is generated quite easily in the Windows version of EQS, using Build_EQS. There is an Advanced Option that shows a Missing Data Handling button. When you click it, you get a dialog box that asks you to choose the method for dealing with missing data. The options are: Complete Cases (listwise deletion); Maximum Likelihood; or the Pairwise Covariance Matrix. To implement ML, you select ML as the missing data method. You also have to choose a method for computing standard errors. The options here are Fisher or Observed. Fisher is the default. With ML chosen, the Structural Mean Analysis option on the main Model Specifications dialog box is automatically checked. Your choices are transferred to the /SPEC part of the model file, and when you click on Equations, you will be cued to use V999s.

The output pretty well follows the usual output of any modeling run, though there are some additions to the output. When you run a missing data ML problem, an initial part of the output is the `MAXIMUM LIKELIHOOD ESTIMATES OF MEANS AND COVARIANCE MATRIX (S) BASED ON THE SATURATED (UNSTRUCTURED) MODEL.` This is used to provide a baseline measure of the log likelihood function (12.4) below, which is used along with the corresponding log likelihood function based on final estimates under the structural model (12.5, below) to get the usual likelihood ratio chi-square test (see 12.6). These statistics are printed out as

```
-2LN(L) BASED ON THE UNSTRUCTURED MODEL =    1893.970
-2LN(L) BASED ON THE   STRUCTURED MODEL =    1905.677

LIKELIHOOD RATIO CHI-SQUARE =       11.707 BASED ON     8 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS      0.16475
```

The log-likelihood values can be ignored, as the χ^2 test is of primary interest. As usual, you should trust the test statistics only if the iterative function has been optimized. The function that is printed out during the iterations is not the log-likelihood, but rather an intermediate function (eqs. 8.27 or 12.35 below, without the N multiplier) that is minimized during the M-step of the EM algorithm. Iterations converge when changes in parameter estimates, E-step expectations, M-step minimizations all result in changes smaller than a fixed criterion (less than .001, but modifiable in /TECH).

Another unique part of the output is called `IMPUTED ESTIMATES OF MEANS AND SAMPLE COVARIANCE MATRIX (S) BASED ON THE STRUCTURED MODEL`. This is an intermediate product of computations and can almost always be ignored. It represents the final expectation (E-step) of the mean and covariance matrix in the EM iterations (see Jamshidian & Bentler, 1999, eqs. (5)-(6). It is used analogously to the usual sample means and covariance matrix as input for optimization (the M-step with 8.27 or 12.35 as just noted) to get the final structural model parameter estimates. The parameter estimates and standard errors are printed out in the standard EQS format. The final model $\hat{\mu} = \mu(\hat{\theta})$ and $\hat{\Sigma} = \Sigma(\hat{\theta})$ can be viewed with the /PRINT command.

Technical Details

ML computations have to be done two times in any structural modeling run: once for the saturated model (with the unrestricted means and covariances as parameters, placed into the vector β) and then again for the structural model of interest (with parameters θ). In complete data problems, the solution for the saturated model depends only on the sample means and covariances and hence is immediately knowable, and similarly the value of the likelihood function for the saturated solution is easily computed without iterations. This is not true in a missing data problem, where the parameters of the saturated model and the value of the likelihood function cannot be known without iterative optimizing computations.

As given by Jamshidian and Bentler (1999, eq. 1) using case-wise notation, for the saturated model we maximize $l(\hat{\beta}) = \sum_{i=1}^{N} l_i(\hat{\beta})$, where

$$l_i(\hat{\beta}) = -\frac{p_i}{2}\log(2\pi) - \frac{1}{2}\{\log\left|\Sigma_i(\hat{\beta})\right| + (z_i - \mu_i(\hat{\beta}))'\Sigma_i^{-1}(\hat{\beta})(z_i - \mu_i(\hat{\beta}))\}. \tag{12.4}$$

Note that here the subscript *i* refers to an individual case, not to a pattern. After this optimization problem is solved for the saturated $\hat{\mu}$ and $\hat{\Sigma}$, we repeat the procedure with the structured model and maximize $l(\hat{\theta}) = \sum_{i=1}^{N} l_i(\hat{\theta})$, where

$$l_i(\hat{\theta}) = -\frac{p_i}{2}\log(2\pi) - \frac{1}{2}\{\log\left|\Sigma_i(\hat{\theta})\right| + (z_i - \mu_i(\hat{\theta}))'\Sigma_i^{-1}(\hat{\theta})(z_i - \mu_i(\hat{\theta}))\}. \tag{12.5}$$

This function $l(\hat{\theta})$ is printed out and, at the minimum, is combined with its corresponding value for the saturated model $l(\hat{\beta})$ in the *likelihood ratio* test statistic (Jamshidian & Bentler, eq. 16) as

$$T_{ML} = -2[l(\hat{\theta}) - l(\hat{\beta})]. \tag{12.6}$$

This also is given as T_2 in Yuan and Bentler (2000d, p. 168), and is used for the robust computations described below.

Although the case-wise functions (12.4) and (12.5) are computed, during its computations EQS also makes use of the usual ML function for complete data, see Jamshidian and Bentler (1999, eq. 4) or the *Manual*'s eq. (8.27). However, the complete data function cannot be used directly since the complete data means $\bar{z}$ and covariance matrix S_b are not available. Jamshidian and Bentler showed how a current E-step, their eqs. (5)-(6), can yield an intermediate $\bar{z}^*$ and S_b^* that can be used as sample matrices to be analyzed during a standard optimization step using complete data procedures. This is the maximization or M-step. These EM steps are repeated until convergence. EQS then prints the final $\bar{z}^*$ and S_b^* (the imputed means and covariances mentioned above) as well as the likelihood ratio test statistic (12.6).

By default, standard errors are given as the square roots of a sample-size corrected version of the inverse of the Fisher information matrix, as is typically done in ML estimation. In the complete data case, the usual form of the information matrix applies, namely $\dot{\sigma}'W\dot{\sigma}$ for a correctly specified weight matrix. See (5.8)-(5.9). The computation is somewhat more complicated in the missing data situation. A computation equivalent to $\dot{\sigma}'W\dot{\sigma}$ is done for every pattern of missing data; then, the final information matrix is a weighted average of the information matrices of the various patterns, with the weights being the number of cases observed for that pattern. See Jamshidian and Bentler (eq. 11) for details. If the option is chosen to compute standard errors from the observed information matrix, the second derivative matrix of the function $l(\hat{\theta})$, the so-called Hessian matrix, is computed and used in place of the information matrix. The second derivatives are obtained numerically using Jamshidian and Bentler's eq. (12)-(13).

Robust Methodology for Nonnormal Data

Research has shown that the normal theory ML estimator $\hat{\theta}$ is quite an excellent estimator even when distributional assumptions are violated. This is true in complete as well as missing data situations. However, this is not true for the associated χ^2 test or the estimates of the standard error of $\hat{\theta}$. Although the scientific details are still being researched, it seems quite certain that in circumstances where the ML method cannot be trusted for complete data, it also should not be trusted when such data is disturbed to yield missing data. Chapter 5 describes a variety of methods that EQS uses to correct the ML χ^2 statistic for violation of normality, and the sandwich matrix approach that is used to obtain more reliable standard error estimates. Yuan and Bentler (2000d) have developed parallel statistics for the incomplete data situation, some of which have been implemented in EQS. Other statistics implied by their work also have been adapted for use in EQS. However, not all available methods are currently implemented; for example, asymptotically distribution free approaches (Yuan & Bentler, 1996, 2000d) are not programmed.

Assumptions and Use

The robust methodology is implemented the same way whether data are complete or contain missing elements. That is, in /SPECIFICATION you state

```
METHOD=ML,ROBUST;
```

In Chapter 5 we noted that with complete data this computed: The Satorra-Bentler scaled chi-square test $\bar{T}$; robust standard errors; and three residual-based test statistics T_{RES}, $T_{\text{YB(RES)}}$, and $T_{\text{F(RES)}}$. With incomplete data, the following parallel computations are obtained:

1. The Yuan-Bentler scaled chi-square test $\bar{T}$ is computed;
2. Robust standard errors are computed; and
3. Three residual-based test statistics T_{RES}, $T_{\text{YB(RES)}}$, and $T_{\text{F(RES)}}$ are computed.

Yuan and Bentler (2000d) provided details on $\bar{T}$, T_{RES}, and the robust standard errors. The modification of T_{RES} to yield $T_{\text{YB(RES)}}$ and $T_{\text{F(RES)}}$ was developed in EQS 6. An illustrative output on the test statistics is the following.

```
GOODNESS OF FIT SUMMARY FOR YUAN-BENTLER CORRECTION BASED ON EXPECTED INFO.

SCALED CHI-SQUARE (YUAN-BENTLER)     =      11.493 ON     8 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS       .17531

RESIDUAL-BASED TEST STATISTIC                        =    12.693
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS       .12286

YUAN-BENTLER RESIDUAL-BASED TEST STATISTIC         =    10.813
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS       .21253

YUAN-BENTLER RESIDUAL-BASED F-STATISTIC     =      1.432
DEGREES OF FREEDOM  =                           8,    65
PROBABILITY VALUE FOR THE F-STATISTIC IS          .20010
```

Yuan and Bentler (2000d) found the scaled chi-square to perform well under a variety of conditions. Similarly, Enders (2001b) found the scaled test to perform quite well, though not necessarily better than a model-based bootstrap approach (Enders, 2002). Although at this time almost no empirical work has been done on the various residual-based test statistics, results with complete data would lead one to expect that Yuan-Bentler residual-based *F*-statistic will be competitive in performance to, and perhaps better than, the scaled chi-square.

The robust standard errors for each individual parameter are printed in the usual way subsequent to the ML estimate, its standard error, and the corresponding z-statistic. They are given in parentheses, and titled `ROBUST STATISTICS IN PARENTHESES.` Enders (2001b, p. 368) concluded that "Both the robust standard errors and naïve bootstrap provided a dramatic improvement over unadjusted ML standard errors…". Note that the robust standard errors will be different when the Fisher information matrix is used in the ML solution as compared to when the observed information matrix is used. The information matrix that you specify for use is the one that is used in the computation of robust standard errors.

Technical Details

As noted above, the Yuan-Bentler scaled χ^2 statistic parallels the Satorra-Bentler scaled χ^2 statistic for the complete data case. As discussed in association with eqs. (5.48)-(5.51), T_{ML} is not necessarily chi-square distributed under non-normality, and the Satorra-Bentler statistic is a simple rescaling of that statistic designed to yield a new statistic whose expected value is correct. The surprising effect is that not only is the mean of the new test statistic more accurate, but so is its tail behavior where decisions to accept or reject models is made. Specifically, the Satorra-Bentler rationale applied to some test statistic T is given in (5.51) as $\bar{T} = T/k$, where $k = tr(\hat{U}\hat{V})/\mathrm{d}$, $U = W - W\dot{\sigma}M\dot{\sigma}'W$ is a residual weight matrix, V is the true asymptotic covariance matrix of the moment data involved, and d is the degrees of freedom. The same principle applies here. That is, Yuan-Bentler (2000d) showed in their Lemma 3.2 that in general T_{ML} is distributed as a mixture of $\chi^2_{(1)}$ variates, from which the scaled statistic (their T_2^* in eq. 20) can be obtained as

$$\bar{T} = T_{ML}/k. \tag{12.7}$$

where T_{ML} is given in (12.6), $k = tr(\hat{U}_\beta \hat{V}_\beta)/(p^* + p - q)$, and U_β and V_β are conceptually defined as above though the specific formulas with missing data are more complicated. The detailed mathematics are given in Yuan and Bentler. Here we give enough of the flavor to help clarify what is going on by explaining the components required for (12.7). First, V_β is the asymptotic covariance matrix of the saturated model parameter estimator $\hat{\beta}$ as defined in (12.4). This is a sandwich estimator of the form (5.21), though with somewhat different constituents. Without constraints on parameters V_β is eq. (5) in Yuan and Bentler. This is

$$V_\beta = A_\beta^{-1} B_\beta A_\beta^{-1} \tag{12.8}$$

where

$$A_\beta = -\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2 l_i(\beta)}{\partial\beta\partial\beta'} \tag{12.9}$$

and

$$B_\beta = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \frac{\partial l_i(\beta)}{\partial\beta}\frac{\partial l_i(\beta)}{\partial\beta'} . \tag{12.10}$$

An estimate of A_β at the given N is taken as the observed information matrix. In EQS, when the user so requests, this can be replaced by its expected value, the Fisher information matrix. When there are constraints on parameters, A_β^{-1} is replaced by a residual matrix of the form given in (5.19), namely, e.g., $A_\beta^{-1} - A_\beta^{-1}\dot{c}'(\dot{c}A_\beta^{-1}\dot{c}')^{-1}\dot{c}A_\beta^{-1}$. B_β represents the covariance (cross-product) matrix of $\partial l_i(\beta)/\partial\beta$, namely, the derivatives of the saturated case-wise log likelihood. This is the part that reflects nonnormality. The residual weight matrix U_β is here defined as

$$U_\beta = A_\beta - A_\beta\dot{\beta}(\dot{\beta}'A_\beta\dot{\beta})^{-1}\dot{\beta}'A_\beta , \tag{12.11}$$

where $\dot{\beta} = \partial\beta(\theta)/\partial\theta$, evaluated at $\hat{\theta}$.

EQS also computes incomplete data versions of the three residual-based test statistics T_{RES}, $T_{\text{YB(RES)}}$, and $T_{\text{F(RES)}}$ discussed in Chapter 5 for complete data. In the current context, T_{RES} is based on comparing the saturated and structural models via Yuan-Bentler's test statistic T_5 (their eq. 21). This is computed as

$$T_{\text{RES}} = \text{n}(\hat{\beta}\text{-}\beta(\hat{\theta}))'\hat{U}_{\text{V}}(\hat{\beta}\text{-}\beta(\hat{\theta})) , \tag{12.12}$$

where

$$U_{\text{V}} = V_\beta^{-1} - V_\beta^{-1}\dot{\beta}\left(\dot{\beta}'V_\beta^{-1}\dot{\beta}\right)^{-1}\dot{\beta}'V_\beta^{-1} . \tag{12.13}$$

Note here that we need to invert V_β, which may be problematic in small samples. The theoretical advantage of T_{RES} is that it is asymptotically χ^2 distributed, while $\bar{T}$ only corrects the mean of T_{ML}. Based on empirical work in the complete data case, one may expect that the sample size needed for T_{RES} to work reliably should be large. The following modified versions of this test, developed here for EQS but named identically to their complete data versions, are liable to perform better in smaller sample sizes. The Yuan-Bentler corrected version of the residual-based test is computed as

$$T_{\text{YB(RES)}} = T_{\text{RES}}/(1 + \text{N}T_{\text{RES}}/\text{n}^2) . \tag{12.14}$$

The denominator in (12.14) is almost always greater than 1.0, and as N goes to infinity, $T_{\text{YB(RES)}}$ and T_{RES} become equivalent. Generally, $T_{\text{YB(RES)}}$ will be smaller than T_{RES}, reducing its inflation, but possibly it may overcorrect. To avoid this, you may want to consider the residual-based F-test

$$T_{\text{F(RES)}} = \{(\text{N} - \text{d})/(\text{nd})\}T_{\text{RES}} . \tag{12.15}$$

As this appears to be the most promising test in complete data situations, it also may be the most reliable test currently available to evaluate models under general conditions. Asymptotically, this test is equivalent to T_{RES}, but in realistic sized samples, it probably takes degrees of freedom of the model and sample size into account in a more appropriate way. Of course, research is needed to evaluate the actual performance of this statistic.

Our description of $\bar{T}$, T_{RES}, and $T_{\text{YB(RES)}}$ is based on the general situation in which means may or may not be structured beyond their base inclusion in models. Assuming there are q parameters in the hypothesized model, the degrees of freedom for these tests is d=p*+p-q. Correspondingly, $T_{\text{F(RES)}}$ is referred to an F-distribution with d and (N-d) degrees of freedom.

Finally, the standard errors of the estimated $\hat{\theta}$ are derived from an expression similar in form to (12.8), namely Y-B's eq. (9). The robust covariance matrix of $\hat{\theta}$ is

$$V_\theta = A_\theta^{-1}B_\theta A_\theta^{-1} , \tag{12.16}$$

where

$$A_\theta = -\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2 l_i(\theta)}{\partial\theta\partial\theta'} \tag{12.17}$$

and

$$B_\theta = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \frac{\partial l_i(\theta)}{\partial\theta}\frac{\partial l_i(\theta)}{\partial\theta'}. \tag{12.18}$$

In EQS, when there are constraints on parameters, A_θ^{-1} is replaced by $A_\theta^{-1} - A_\theta^{-1}\dot{c}'(\dot{c}A_\theta^{-1}\dot{c}')^{-1}\dot{c}A_\theta^{-1}$.

Pairwise-based ML With Robust Corrections

Earlier in this chapter, we discussed the available-case or pairwise present method of computing correlations or covarinaces. The literature states that using the typical pairwise covariance matrix, say S_p in structural modeling poses serious problems, e.g., there are varying sample sizes for the various covariances (e.g., Marsh, 1998). In this section we discuss a method developed by Savalei and Bentler (in press) that resolves these problems with an ML approach involving corrections for misspecifications. While the approach is technically correct, this does not necessarily mean that the method will perform better than the ML,ROBUST method just described. Nonetheless, you should have the option to use it. In this method, unlike ML, you do not need to use ANALYSIS=MOMENT or specify *V999 in equations.

Assumptions and Use

The pairwise matrix S_p is a consistent estimator of the population Σ under an MCAR data mechanism, and hence using this matrix with any optimization method in EQS will give consistent estimates of the model parameters. Here we concentrate on the pairwise-based ML method. That is, we use S_p with the normal theory ML fitting function given in (5.13), but replace the ordinary sample covariance matrix S with S_p. We minimize

$$F = \ln|\Sigma| + tr(S_p\Sigma^{-1}) - \ln|S_p| - \mathrm{p}. \tag{12.19}$$

This function is misspecified for S_p since it is not the unstructured ML estimator, and so we use robust corrections to the ML test statistic and the standard errors. You specify the matrix S_p with a MISSING=PAIR statement in the /SPECIFICATION section, and, as usual, correct the ML statistics with the statement METHOD=ML,ROBUST. Thus a typical line in the /SPEC section is

```
METHOD=ML,ROBUST; ANALYSIS=COVARIANCE; SE=FISHER; MISSING=PAIR;
```

Notice that Fisher and Observed information matrix options for standard errors under ML are still available to you. At the $\hat{\theta}$ that minimizes (12.19), the pairwise ML test statistic is then computed in the usual way as T_{ML}=n $\hat{F}$ (see eq. 5.14). However, this test statistic *cannot be used* as it is misspecified since $\hat{\theta}$ does not maximize the proper likelihood function, even with normal data, and hence T_{ML} is not distributed as a chi-square variate. Instead you should use the robust results. The following is some of the relevant output.

```
GOODNESS OF FIT SUMMARY FOR METHOD = ROBUST

  SCALED (SATORRA-BENTLER TYPE) CHI-SQUARE =      6.7271 ON     8  D.F.
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS       .56634

  RESIDUAL-BASED TEST STATISTIC                      =     6.461
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS       .59573

  YUAN-BENTLER RESIDUAL-BASED TEST STATISTIC         =     5.936
  PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS       .65443
```

```
YUAN-BENTLER RESIDUAL-BASED F-STATISTIC     =        .729
 DEGREES OF FREEDOM  =                          8,    65
 PROBABILITY VALUE FOR THE F-STATISTIC IS          .66530
```

Savalei and Bentler (in press) provided the technical details for the scaled χ^2 test, and showed that it worked reasonably under some conditions – though definitely not as well as the full robust ML approach. Surprisingly, it did much worse with MCAR data than with MAR data. As this relative success also occurred with direct ML, it cannot be due to the pairwise methodology per se, but was probably due to the larger number of patterns of missing data in their MCAR condition.

The three additional residual-based statistics T_{RES}, $T_{\mathrm{YB(RES)}}$, and $T_{\mathrm{F(RES)}}$ were added in EQS to be consistent with the theory for the three residual-based test statistics as discussed above and in Chapter 5. Of course, the formulae were modified to apply to analysis based on S_p.

Standard errors of parameter estimates are the usual sandwich type standard errors, adapted to the current situation. As usual, they are printed out under the ML standard errors and *z*-tests (which should be avoided due to misspecification) with the heading `ROBUST STATISTICS IN PARENTHESES.`

Technical Details

The robust ML pairwise theory developed by Savalei and Bentler (in press) depends on a result of Van Praag, Dijkstra, and Van Velzen (1985), who, in a regression context, determined the asymptotic covariance matrix of S_p under an MCAR missing data mechanism and without any distributional assumptions. Van Praag et al. (their eq. 9) determined the asymptotic covariance between a pair s_{ij}, s_{kl} of pairwise-computed sample variances and covariances. Letting s_p be the vector of lower triangular elements of S_p, they determined

$$\sqrt{n}(s_p - \sigma) \xrightarrow{D} \mathcal{N}(0, V), \tag{12.20}$$

where the matrix V has elements

$$v_{ij,kl} = \frac{p_{ijkl}}{p_{ij}p_{kl}}(\sigma_{ijkl} - \sigma_{ij}\sigma_{kl}), \tag{12.21}$$

where σ_{ijkl} is a fourth order moment with probability p_{ijkl} of being observed, and σ_{ij}, σ_{kl} are covariances with probabilities p_{ij}, p_{kl} of being observed. Their result is a generalization of Browne's (1984) widely known formula for complete cases, as given in (5.6), since it reduces to (5.6) with complete data when $p_{ij} = p_{kl} = p_{ijkl} = 1.0$. In practice, of course, the probabilities and the moments in (12.21) are estimated by sample proportions and moments.

Using this information, and following the application of (12.20) in covariance structure analysis by Gold, Bentler, and Kim (2003), Savalei and Bentler gave the asymptotic covariance matrix of the ML estimator with pairwise covariances as

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}(0, \Omega_{\hat{\theta}}), \tag{12.22}$$

where, without constraints on parameters,

$$\Omega_{\hat{\theta}} = (\dot{\sigma}'W_p\dot{\sigma})^{-1}\dot{\sigma}'W_pVW_p\dot{\sigma}(\dot{\sigma}'W_p\dot{\sigma})^{-1}. \tag{12.23}$$

This robust sandwich type covariance matrix is of the form given for complete data in (5.21), with standard errors obtainable as in (5.23). With constrained estimates, the outer matrix is replaced by the usual residual matrix. Here, W_p is the usual normal theory weight matrix as given in (5.15), but estimated using pairwise data

$$\hat{W}_p = .5D_p'(\hat{\Sigma}_p^{-1} \otimes \hat{\Sigma}_p^{-1})D_p. \tag{12.24}$$

EQS provides robust test statistics $\bar{T}$, T_{RES}, $T_{\text{YB(RES)}}$, and $T_{\text{F(RES)}}$ that correct the ML goodness of fit test for misspecification. Savalei and Bentler (in press) developed the variant of the Satorra-Bentler statistic. They defined it as the usual $\bar{T} = T_{ML} / k$ given in (12.7), where T_{ML}=n $\hat{F}$ is based on the minimum of (12.19), $k = tr(\hat{U}\hat{V})/\text{d}$ for d degrees of freedom, V was defined in (12.20), and U is given (without constraints) as

$$W_p - W_p\dot{\sigma}(\dot{\sigma}'W_p\dot{\sigma})^{-1}\dot{\sigma}'W_p, \tag{12.25}$$

all evaluated at $\hat{\theta}$. (12.25) is an application of (5.49) with W_p replacing W. Savalei and Bentler found this pairwise ML test to perform quite well, though it was a bit inflated as compared to direct casewise ML approach with Yuan-Bentler corrections.

In EQS, we also developed the three additional residual-based test statistics for the pairwise ML method. These are computed as in (5.54)-(5.56), with the use of s_p and S_p replacing the ordinary covariance vector and matrices, and with V as defined in (12.20)-(12.21).

LM and W Tests with ML and Pairwise ML

Chapter 6 provided a detailed discussion of the meaning and use of the Lagrange Multiplier (LM) test for evaluating possible model misspecifications. This test is specified the same way with missing data as with complete data. You should know, however, that in the initial release of EQS 6, both MISSING= ML and MISSING=PAIR will lead the LM test to be based on maximum likelihood computations. If the data are normal, under MISSING=ML, the LM statistics are correct when based on a priori considerations, while for MISSING=PAIR, the LM statistics will not be precisely correct. Nonetheless, the relative seriousness of a given parameter misspecification should be well indicated by the relative size of the LM statistics. In later releases, when you specify ML,ROBUST, the LM test also will be computed using robust estimates of sampling variability of parameter estimates, which will be asymptotically correct for both methods.

The Wald test – also discussed in detail in Chapter 6 – is automatically computed based on the most general distributional assumptions specified. If you use MISSING=ML, the W test will be based on ML estimates of sampling variability of parameter estimates; this will be correct if the data are normal. With both MISSING=ML and MISSING=PAIR, if you specify ML,ROBUST, the W test will be based on robust estimates of sampling variability which are asymptotically correct with both methods. You will see the heading `ROBUST ASYMPTOTIC COVARIANCE MATRIX USED IN THIS WALD TEST.`

Pairwise-based Distribution Free Approach

In a previous section we discussed the use of pairwise present covariances along with ML estimation. In this section, we discuss the use of these covariances with asymptotically distribution free estimation. As you have learned throughout this *Manual*, this methodology should be considered in practice only when sample size is very large.

Assumptions and Use

With MCAR data, the matrix S_p of available case covariances is a consistent estimator of the population Σ. This means that any optimized parameter estimator available in EQS based on S_p will provide consistent estimates. If sample size is large enough, the asymptotically distribution free (ADF) estimates are efficient, that is, they are most precise in the sense of have minimal sampling variability, and the usual AGLS model test statistic is available. You implement this method by stating, in the /SPEC section,

```
METHOD=AGLS; MISSING=PAIR;
```

This instructs the program to implement the method developed by Gold, Bentler, and Kim (2003), called ADFAC (ADF available cases) by them. In addition to the usual AGLS chi-square statistic, corrected versions also are implemented and printed out as illustrated next.

```
CHI-SQUARE =        6.629 BASED ON       8 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS       .57716

YUAN-BENTLER CORRECTED AGLS TEST STATISTIC         =     6.077
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS       .63861

YUAN-BENTLER AGLS F-STATISTIC               =       .748
DEGREES OF FREEDOM  =                           8,    65
PROBABILITY VALUE FOR THE F-STATISTIC IS          .64905
```

In smaller samples, the F-statistic is probably the most reliable one to use in practice. Finally, in addition to the ordinary AGLS standard errors of parameter estimates, the program prints out

```
(YUAN-BENTLER CORRECTED AGLS STATISTICS IN PARENTHESES)
```

The corrected standard error estimates are more appropriate in small samples.

Chapter 5 discussed the AGLS method in the case of complete data, when the ordinary sample covariance matrix *S* is used in an analysis. As it is hard to imagine that a method can perform better when data are missing than when data are complete, you should read the various considerations discussed in Chapter 5 surrounding the use of this method as these still apply when S_p replaces *S*. The main consideration is, of course, the need for very large samples. In their study of this method, Gold et al. concluded that the direct ML approach corrected for nonnormality (discussed above) was generally superior to the ADFAC method. "Thus the main take-home message of these results is that the non-normality corrected EM implementation of ML tends to outperform ADFAC in the most important respects, regardless of missing-data mechanism. Estimates are less biased, and the chi-square test statistic is more trustworthy. The only important respects in which ADFAC outperforms ML involves (a) efficiency of parameter estimates in the robust condition and (b) bias in standard-error estimates with large samples." (Gold, et al., p. 73). This is not exactly a ringing endorsement of the pairwise ADF method by its developers. Further work may find similar or more optimistic results with the corrected test statistics and corrected standard errors, which were not studied by Gold et al.

Technical Details

The statistical theory for the Browne (1982, 1984) and Chamberlain (1982) asymptotically distribution free approach to covariance structures involves minimization of $Q = (s - \sigma(\theta))' W (s - \sigma(\theta))$ given in equation (1.5) of Chapter 1. This hinges on the large sample distribution of sample covariances, as given in eqs. (5.5)-(5.7) of Chapter 5. In the pairwise method, the lower triangular elements of the matrix S_p are vectorized to s_p, whose distribution (12.20) was obtained by van Praag, Dijkstra, and van Velzen (1985). The asymptotic covariance matrix of s_p was given in (12.21). It is simply an adjustment to the normal ADF covariance matrix (5.6) to account for missing data. As a result, Gold, Bentler, and Kim (2003) defined the minimum chi-square function

$$Q = (s_p - \sigma(\theta))' V^{-1} (s_p - \sigma(\theta)) . \qquad (12.26)$$

Using an estimator of V in (12.20)-(12.21), (12.26) is minimized with respect to $\hat{\theta}$. Then this is evaluated at the minimum $\hat{Q}$ to yield the usual AGLS type test statistic

$$T_{\text{AGLS}} = \text{n}\,\hat{Q} . \qquad (12.27)$$

This is referred to the usual (p*-q) df chi-square distribution. Two alternatives to this test, based on prior work described in eqs. (5.2)-(5.3), were developed in EQS 6 as

$$T_{\text{YB}} = T_{\text{AGLS}} / (1 + T_{\text{AGLS}}/\text{n}) \text{ and } T_{\text{F(AGLS)}} = \{(\text{N} - \text{d})/ (\text{nd})\} T_{\text{AGLS}}. \qquad (12.28)$$

These are the Yuan-Bentler corrected AGLS (1997d) and F statistics (1999a). Chapter 5 provides references to their origins and applications. In smaller samples, $T_{F(AGLS)}$ is probably the best statistic among these to use.

As an asymptotically optimal estimator under MCAR, $\hat{\theta}$ in principle has the smallest sampling variability. This property may or may not hold in small to medium sized samples. In large samples, the usual asymptotic covariance matrix given in eqs. (5.9)-(5.10), adjusted to take into account the missing data, holds here. That is, the inverse of an analogue to the information matrix

$$\Delta = (\dot{\sigma}'V^{-1}\dot{\sigma})^{-1} \tag{12.29}$$

gives the asymptotic covariance matrix of $\hat{\theta}$. This is modified to $M = \Delta - \Delta\dot{c}'(\dot{c}\Delta\dot{c}')^{-1}\dot{c}\Delta$ when there are constraints. Then, as in (5.10), $S.E.(\hat{\theta}_i) = \sqrt{\hat{m}_{ii}/n}$. For smaller samples, Yuan and Bentler (1997c) suggested the adjustment M_c = {n/(n – p* - 1)} M with $S.E.(\hat{\theta}_i) = \sqrt{\hat{m}_{c(ii)}/n}$, as described in (5.11)-(5.12). This gives the corrected AGLS *z*-statistics.

Multisample Analysis With Constraints

Before the case-wise or direct ML method became popular for handling normal missing data in a structural equation modeling context, Allison (1987) and Muthén, Kaplan, and Hollis (1987) suggested that the multiple sample option in structural modeling (see Chapter 9) could be used. While this approach is generally superseded with the option MISSING=ML, there are some situations when the multiple group approach might still be interesting. In particular, if the data contain a few predominant patterns of missing data (e.g., some subjects have data only from waves 1 and 2; all others have complete data from waves 1, 2, and 3), the multiple group approach is attractive because it can pinpoint differences among the patterns if they exist and provide consistent ML estimates and tests otherwise. If there are three patterns of missing data, three sets of data are created and a 3-group structural model is used. However, if there are dozens of patterns of missing data, this approach is impractical. Not only may some samples be too small to yield stable results, the multisample approach becomes hard to implement with so many samples.

In the multisample approach, data have to be organized so that samples of subjects are grouped together in accord with their pattern of missing data. This can be done conveniently using BMDP's AM or SPSS's Missing Value Analysis programs, for example. Allison (1987) and Muthén et al. (1987) use dummy variables and factors in the groups with missing data, with pseudo-values replacing the missing means and covariances. Equality constraints across groups are used to assure that the same parameters (means and covariances, or structural modeling parameters) are estimated in both groups when these parameters would be identified if the data were complete (they may not be identified in any single sample), and the process is carried out so that the pseudo-values are fitted exactly. Both means and covariances must be modeled.

An example of the structural modeling approach to missing data is given in the job setup shown below. The data and model are taken from Allison (1987), who used the LISREL program to estimate and test the model. This required using a number of "tricks", such as using dummy variables and parameters, and adjusting the degrees of freedom, to yield the correct missing-data results. These tricks are unnecessary to the theory involved, and serve to confuse the simplicity of the ideas. They are also unnecessary in EQS, in which the model setup is essentially the same as in any multisample analysis with structured means. The critical point in such a setup is that the samples with missing data can contain specifications of equations, intercepts, variances and covariances only for variables that are actually observed, as well as for hypothesized factors and residual variables relevant to those variables.

```
/TITLE
   INCOMPLETE DATA FACTOR MODEL (ALLISON, 1987)
   COMPLETE DATA SUBSAMPLE
/SPECIFICATIONS
    VARIABLES = 4; CASES = 348; ANALYSIS = MOMENT; GROUPS = 2;
/EQUATIONS
    V1 =  17*V999 +   F1        + E1;
    V2 =  17*V999 + 1*F1        + E2;
    V3 =   7*V999        +   F2 + E3;
    V4 =   7*V999        + 1*F2 + E4;
/VARIANCES
    F1 = 117*; F2 = 14*;
    E1 = 94*;  E2 = 47*;  E3 = 2*;  E4 = 1*;
/COVARIANCES
    F1,F2 = 25*;
/MATRIX
 180.90
 126.77 217.56
  23.96  30.20  16.24
  22.86  30.47  14.36  15.13
/MEANS
  16.62  17.39   6.65   6.75
/END
/TITLE
   INCOMPLETE DATA SUBSAMPLE
/SPECIFICATIONS
    VARIABLES = 4; CASES = 1672; ANALYSIS = MOMENT;
/EQUATIONS
    V1 =  17*V999 +   F1        + E1;
    V3 =   7*V999        +   F2 + E3;
/VARIANCES
    F1 = 117*; F2 = 14*;
    E1 = 94*;  E3 = 2*;
/COVARIANCES
    F1,F2 = 25*;
/MATRIX
  217.27
    0     1
   25.57  0  16.16
    0     0   0     1
/MEANS
   16.98  0   6.83  0
/CONSTRAINTS            ! SET COMMAND GENERATES EQUALITY CONSTRAINTS ON
   SET = PFF, PEE, GVV; ! ALL PARAMETERS WHICH ARE FREE IN BOTH GROUPS.
/LMTEST
/PRINT
   COVARIANCE = YES;
/END
```

The model described above is a two-group model in which one sample has complete information on all variables, and the other sample contains a particular pattern of missing data, and thus contains observed data on a subset of the variables. The data originally came from Bielby, Hauser and Featherman (1977), and the specific meaning of the variables in the example can be found in Allison (1987). The model is a simple two-factor model with intercepts for the measured variables (paths from V999 to the measured variables). The data from this sample, based on 348 cases, is complete, that is, all variances and covariances among V1-V4, as well as the means of these variables, are available for analysis. Thus in this first group the model is a rather standard factor analysis model, except that the variable intercepts (which are means in this example) are also being estimated. If this sample were the only one being analyzed, these intercepts would be estimated at the sample means.

The second, much larger, sample, based on 1672 cases, contains incomplete data. Allison states that the data are missing at random. Data are available only on variables V1 and V3, i.e., no data exist for V2 and V4. The covariance matrix and means in the input file are of the same dimension as in the complete data sample, that is, with 4 variables, in order to keep the notation V1 and V3 for the available data (rather than V1 and V2 for two variables, which is what EQS otherwise would assume for two input variables). The entries corresponding to V2 and V4 are completely arbitrary and have no meaning; in fact, by the model setup, EQS will not even read these entries, and

only the data corresponding to V1 and V3 will be read in by the program and analyzed. Since only V1 and V3 have data, equations for only these variables are provided. Variances and covariances are specified for factors and errors given in these equations. The final critical part of the setup lies in the cross-group constraints, which specify that every free parameter in the incomplete data sample is to be estimated at the same value as in the complete data sample.

The model was estimated with EQS, yielding $\chi^2_{(6)} = 7.692$, an acceptable model with probability p = .26. Note that there were 19 sample covariances and means to be analyzed, 20 free parameters in the model setup, and 7 cross-group constraints, yielding 19-20+7=6 degrees of freedom. An advantage of this approach is that all the diagnostics usually available in modeling runs are available. In this example, the Lagrange Multiplier test showed that some cross-pattern constraints can be rejected statistically, implying the possibility that a single model may not be appropriate for both data patterns even under an assumed MAR mechanism.

Cohort Sequential Design

Missing data may arise from a planned data gathering design in which not every subject is tested at every occasion of measurement in a longitudinal study. This enables a huge reduction in time and cost of the study. Good discussions of planned missing data designs can be found in Graham, Hofer and MacKinnon (1996), and Graham, Taylor, and Cumsille (2001). Here we extend the discussion of Chapter 10 on the cohort sequential design (e.g., Nesselroade & Baltes, 1979). In this design, a given cohort is tested on several occasions, another cohort is tested on some overlapping as well as non-overlapping occasions, and yet another cohort can be tested on still other overlapping/non-overlapping set of occasions. This would imply data gathering occasions such as the following Design 1:

Design 1	T1	T2	T3	T4	T5
Cohort x	x1	x2	x3		
Cohort y		y1	y2	y3	
Cohort z			z1	z2	z3

A true longitudinal study would test everyone at T1, T2, …, T5. Here, each of the cohorts x, y, and z are tested 3 times. While such a design would avoid testing everyone across all 5 time points, it would not shorten the length of the overall study. To achieve this, data can be gathered as in Design 2:

Design 2	T1	T2	T3
Cohort x	x1	x2	x3
Cohort y	y1	y2	y3
Cohort z	z1	z2	z3

Now the study is condensed to 3 time periods. But instead of simply having 3 samples tested 3 times, cohort y providing data y1 is assumed to be at the same stage in longitudinal development as cohort x at measurement x2; and cohort z providing z1 is assumed to be at the same stage as cohort x at x3, as well as cohort y at y2. See the columns T2 and T3 in the previous table. As an example, in a study of annual growth during adolescence, at the initial time point, cohort 1 may be a sample of 12-year olds; cohort 2, a sample of 13-year olds; and cohort 3, a sample of 14-year olds. Each is studied for 3 years, cohort 1 at ages 12-15, cohort 2 at ages 13-16, and cohort 3 at ages 14-17. The entire age range 12-17 is covered in only 3 years.

The data gathered under Design 2 are analyzed as if they had been gathered in Design 1, and features of that design are imposed on the analysis. The setup is that there are three groups that contain missing data. The analysis then proceeds as in the Allison example above, a multiple group model with missing data. The baseline model assumes that all cohorts come from a single population with no cohort differences. Examples of this design can be found in studies of adolescent development (Duncan, Duncan & Hops, 1994), deviant behavior (Raudenbush & Chan, 1992), and marriage research (Anderson, 1993). Chapter 10 illustrated an example from the excellent "Accelerated

Designs" chapter in Duncan, Duncan, Strycker, Li, and Alpert (1999). These authors also discuss the interesting question of whether results by such a cohort sequential design can actually match the results that would be observed if one had a true longitudinal design. They had an available longitudinal data set, and show strong similarity in reproduced variable means from the linear growth curve model in the cohort sequential and longitudinal data. Similar results across a six year period also were shown by Duncan, Duncan, and Hops (1996). But they caution: "However, in order to reduce the likelihood of nonconvergence in fitting the cohort-sequential model, cohorts should overlap as much as possible, sample sizes within each cohort should be relatively large, and cohorts should be as comparable as possible (Tonry et al., 1991)" (Duncan, Duncan et al., 1999, p. 82).

Appendix: Direct and Mean/Covariance Structure ML

We provide a short description of how the direct ML function (12.5) (or 12.4) relates to the mean and covariance structure ML function (8.25) and the likelihood ratio test (8.27). The log of the density of a multivariate normal variate is given by (12.5) as

$$l_i(\theta) = -\frac{p_i}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(z_i - \mu_i)'\Sigma_i^{-1}(z_i - \mu_i).$$

Here we suppress the notation that $\mu_i \subseteq \mu(\theta)$ and also $\Sigma_i \subseteq \Sigma(\theta)$. If observations are independent, their joint density is the product of the individual densities, and their logs are additive

$$l(\theta) = \sum_{i=1}^{N} l_i(\theta) = -\frac{Np_i}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\log|\Sigma_i| - \frac{1}{2}\sum_{i=1}^{N}(z_i - \mu_i)'\Sigma_i^{-1}(z_i - \mu_i). \quad (12.30)$$

In complete data we do not need the subscript *i* on the number of variables and population parameters in this equation, as every case has the same number of observed variables and also the same $\mu = \mu(\theta)$ and $\Sigma = \Sigma(\theta)$. As a result, we can write (12.30) as

$$l(\theta) = -\frac{Np}{2}\log(2\pi) - \frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}(z_i - \mu)'\Sigma^{-1}(z_i - \mu). \quad (12.31)$$

The right-most quadratic form can be written as follows

$$\frac{1}{2}\sum_{i=1}^{N}(z_i - \mu)'\Sigma^{-1}(z_i - \mu) = \frac{1}{2}\sum_{i=1}^{N}tr[(z_i - \mu)'\Sigma^{-1}(z_i - \mu)] = \frac{1}{2}\sum_{i=1}^{N}tr[\Sigma^{-1}(z_i - \mu)(z_i - \mu)']$$

$$= \frac{1}{2}tr[\Sigma^{-1}\sum_{i=1}^{N}(z_i - \bar{z} + \bar{z} - \mu)(z_i - \bar{z} + \bar{z} - \mu)'] = \frac{1}{2}tr[\Sigma^{-1}\sum_{i=1}^{N}(z_i - \bar{z})(z_i - \bar{z})'] + \frac{1}{2}tr[\Sigma^{-1}\sum_{i=1}^{N}(\bar{z} - \mu)(\bar{z} - \mu)']$$

$$= \frac{N}{2}tr\Sigma^{-1}S_b + \frac{N}{2}tr[\Sigma^{-1}(\bar{z} - \mu)(\bar{z} - \mu)'] = \frac{N}{2}tr\Sigma^{-1}S + \frac{N}{2}(\bar{z} - \mu)'\Sigma^{-1}(\bar{z} - \mu)$$

In this expression, S_b is the biased (divisor N) sample covariance matrix. In view of this, we can rewrite (12.31) as

$$l(\theta) = -\frac{Np}{2}\log(2\pi) - \frac{N}{2}\log|\Sigma| - \frac{N}{2}\mathrm{tr}\Sigma^{-1}S_b - \frac{N}{2}(\bar{z} - \mu)'\Sigma^{-1}(\bar{z} - \mu). \quad (12.32)$$

Aside from the additive constant (which does not affect optimization) and the sign, this is the mean and covariance structure function (8.25). Maximizing this with respect to θ under our null hypothesis, we obtain $\hat{\mu} = \mu(\hat{\theta})$ and $\hat{\Sigma} = \Sigma(\hat{\theta})$, and the maximized log likelihood can be written as

$$l(\hat{\theta}) = -\frac{Np}{2}\log(2\pi) - \frac{N}{2}\log|\hat{\Sigma}| - \frac{N}{2}\mathrm{tr}\hat{\Sigma}^{-1}S_b - \frac{N}{2}(\bar{z} - \hat{\mu})'\hat{\Sigma}^{-1}(\bar{z} - \hat{\mu}). \quad (12.33)$$

If we similarly maximize (12.32) with respect to the saturated model with parameters β as in (12.4), as is well known the maximum is achieved by $\hat{\mu} = \bar{X}$ and $\hat{\Sigma} = S_b$. Evaluating this we get

$$l(\hat{\beta}) = -\frac{Np}{2}\log(2\pi) - \frac{N}{2}\log|S_b| - \frac{Np}{2}. \tag{12.34}$$

The likelihood ratio χ^2 statistic -2log λ is well known to be

$$\begin{aligned}
&-2\log\lambda = -2\{l(\hat{\theta}) - l(\hat{\beta})\} \\
&= -2\left\{-\frac{Np}{2}\log(2\pi) - \frac{N}{2}\log|\hat{\Sigma}| - \frac{N}{2}\mathrm{tr}\hat{\Sigma}^{-1}S_b - \frac{N}{2}(\bar{z}-\hat{\mu})'\hat{\Sigma}^{-1}(\bar{z}-\hat{\mu})\right\} \\
&-2\left\{\frac{Np}{2}\log(2\pi) + \frac{N}{2}\log|S_b| + \frac{Np}{2}\right\} \\
&= N\log|\hat{\Sigma}| - N\log|S_b| + N\mathrm{tr}\hat{\Sigma}^{-1}S_b + N(\bar{z}-\hat{\mu})'\hat{\Sigma}^{-1}(\bar{z}-\hat{\mu}) - Np.
\end{aligned}$$

Thus, finally we get

$$-2\{l(\hat{\theta}) - l(\hat{\beta})\} = N\left\{\log|\hat{\Sigma}| - \log|S_b| + \mathrm{tr}\hat{\Sigma}^{-1}S_b - p + (\bar{z}-\hat{\mu})'\hat{\Sigma}^{-1}(\bar{z}-\hat{\mu})\right\}. \tag{12.35}$$

This is the LR statistic in (8.27). The special case without structured means is the ML covariance structure function given in (5.13).

13. SIMULATION AND RESAMPLING

In EQS, the SIMULATION section described in Chapter 3 facilitates two different types of simulations. In Monte Carlo simulation, a population is generated from information provided by the user concerning the distributions of variables and the structure of a model, and samples are taken from this population. In resampling, it is presumed that the distribution of the data is a good approximation to the distribution in the population, and hence the empirical data is repeatedly sampled. In both types of simulation, a specific model is fit to each sample, and interesting statistics on that sample are saved. The distribution of those statistics across the many samples provides sampling distributions of statistics to be analyzed and understood. Some relative advantages of the two types of simulation are discussed in Lewis and Orav (1989, pp. 251-252). EQS provides a variety of options for creating interesting populations to be studied by Monte Carlo methods, while its choices of resampling techniques are regular bootstrap, model-based bootstrap, and jackknifing. Examples of these procedures are presented in this chapter.

Monte Carlo simulation is useful when the distribution of actual data can be described by mathematical equations or in testing whether a proposed model fits theoretical data. This simulation methodology has been used for a long time in structural modeling to evaluate whether the abstract properties of statistics hold in practice under various violations of conditions (e.g., Boomsma, 1983; Harlow, 1985; Tanaka, 1984). For example, based on methods for generating multivariate nonnormal populations (Fleishman, 1978; Vale & Maurelli, 1983), Monte Carlo results have provided evidence regarding the robustness or lack of robustness of a chi-square statistic to violation of distributional assumptions. Methods for generating wider classes of simulated populations continue to be developed (e.g., Headrick, 2002; Headrick & Sawilowsky, 1999, 2000; Mattson, 1997; Reinartz, Echambadi, & Chin, 2002; Yuan & Bentler, 1997b). A good recent overview on how to think about Monte Carlo research can be found in Mooney (1997). Paxton, Curran, Bollen, Kirby, and Chen (2001) discuss the design and implementation of Monte Carlo studies in the structural modeling field, and Olsson, Foss, and Breivik (2004) use this methodology to evaluate the adequacy of the noncentral χ^2 distribution under model misspecification. Among various specific applications, Muthén and Muthén (2002) propose its use in determining sample size and power.

Resampling is useful when data-generation would be too time-consuming, even on a high-speed computer, or when the equations that describe the data are not known. Instead of sampling from an artificially created population, sampling is done from the empirical data at hand. The key idea is that variability observed when repeatedly sampling from the data can be a good approximation to the variability that would be observed when sampling from the population. This is an old idea in statistics, as reviewed e.g., by Hall (2003). Today's most popular resampling methodology is the bootstrap (Efron, 1979, 1982, 2000, 2003), which grew out of earlier ideas of cross-validation and the jackknife. It is discussed further below. The bootstrap is probably the most widely used general technique for obtaining statistical results when mathematical statistics cannot or are not desirable to be used due to complexity of the problem. In EQS, for example, you can bootstrap R^2 coefficients and standardized parameters. A special issue of *Statistical Science* (May, 2003) devoted to the Silver Anniversary of the bootstrap discusses the impact of the bootstrap in a variety of fields related to statistics. Among the many excellent articles in this issue, Davison, Hinkley and Young (2003) provide a review of recent developments. The bootstrap was rather quickly adopted in social science (e.g., Stine, 1989) and structural modeling (e.g., Boomsma, 1986; Bollen & Stine, 1993). Reviews are given by Fan (2003b) and Yung and Bentler (1996a). Recent SEM applications include Enders (2002), Nevitt and Hancock (2001), and Yuan, Guarnaccia, and Hayslip (2003). Yung and Bentler (1996a) and Yuan and Hayashi (2003) propose its use in determining power; an approach due to Yuan and Marshall (2004) can also be used.

In this chapter we discuss the basic commands needed to implement a Monte Carlo simulation in EQS. Subsequently, we discuss how resampling methods such as the bootstrap and jackknife are specified in EQS. In the process, we describe output and data files that are created, including estimates and test statistics (ETS) and codebook (CBK) files. Then we turn to some advanced simulation options.

Generating Data for Simulation

The /SIMULATION section allows you to specify how your simulation will be implemented. For a Monte Carlo simulation, you have to specify characteristics of the population. EQS can generate data for all independent variables in the model, or for all variables in a covariance or correlation matrix supplied by the user. We discuss these in turn, and provide a discussion of the simulation itself and the summary statistics that are generated to interpret the simulation. Data that are initially generated will be multivariate normal, and hence you may want to implement transformations to create nonnormal data. These topics are discussed subsequently.

POPULATION=MODEL

We illustrate generating data from a model specification by running the model file below. The model is a 2 factor model with uncorrelated errors. The statement POPULATION=MODEL tells EQS to use the particular parameters that you provide in the model specification to generate a multivariate normal population whose variables are the independent variables in your model, which in turn generate observed variables whose population covariance matrix is the one generated from your parameters. Here, REPLICATIONS=100, informing EQS that 100 samples should be drawn from the population, each with sample size 200 (CASES=200). When this model file is submitted to run, for each sample EQS creates an intermediate file containing the data to be analyzed in a given replication, and then analyzes this data in the usual way (here, using ML). When finished, EQS summarizes the simulation results.

```
/TITLE
 SIMULATION EXAMPLE
 DATA IS GENERATED FROM THE MODEL
/SPECIFICATIONS
 CASES=200; VARIABLES=5; METHOD=ML;     !EACH SAMPLE WILL HAVE N=200, ANALYZED BY ML
/EQUATIONS
 V1 = .8*F1         + E1;
 V2 = .7*F1 - .4*F2      + E2;
 V3 = .9*F1 + .3*F2           + E3;
 V4 = .8*F1 + .4*F2                + E4;
 V5 = .9*F1 - .3*F2                     +E5;
/VARIANCES
 F1 TO F2 = 1.0;
 E1 = .36*; E2 = .51*; E3 = .19*; E4 = .36*; E5 = .19*;
/OUTPUT                                 !SPECIFICATIONS FOR OUTPUT FROM EACH SAMPLE
 PARAMETERS;
 STANDARD ERRORS;                       !PARAMETER ESTIMATES AND SE'S TO BE SAVED
 DATA='EXAMPLE1.ETS';                   !FILE NAME OF SIMULATION OUTPUT
/SIMULATION                             !SPECIFICATIONS FOR NATURE OF SIMULATION
 SEED = 987654321;                      !RANDOM PROCESS STARTS AT AN ARBITRARY PLACE
 REPLICATIONS=100; POPULATION=MODEL;    !NUMBER OF SAMPLES; HOW POPULATION IS GENERATED
/END
```

Corresponding to any input file, such as above, after EQS has run, there will be a corresponding OUT file (in some computer systems, called a LOG file). You will be interested primarily in the very first part and then the last part of the output file. The first part duplicates the model file to be run, and then describes how EQS generated the data as will be shown below. The huge intermediate part of the OUT file provides summary information on each replication, namely, the sample statistics and sample covariance matrix for each replication. Since there are 100 replications in this run, there will be 100 such abbreviated sets of summary statistics. We do not show these here. The really important part of the OUT file is at the end, which provides a statistical summary of what occurred across all 100 replications. That will be illustrated below.

Note, however, that *the OUT file does not contain specific results from the modeling run in each sample* (e.g., the parameter estimates, standard errors, chi-square statistics and so on). That output is placed into the file you

specified in the OUTPUT section (here, EXAMPLE1.ETS). ETS (for Estimates & Test Statistics) is the default file extension for a saved data file, and you provide its main name. These detailed results are put into this external file so that, if you want, you can do your own statistical analysis on this data. However, as noted and is illustrated below, EQS already will provide a simple summary of the simulation results in the OUT file that may suffice.

When POPULATION=MODEL, the variables generated are precisely the independent variables of the model, and in this model, they are the F's and E's. The OUT file gives a summary of univariate characteristics of the generated variables. In the example, note that each variable is generated from a normal distribution, with mean zero. The standard deviations of the error terms are the square roots of the variance estimates, above.

```
SIMULATION DEFINITIONS

                  NUMBER OF REPLICATIONS    : 100
                  SAMPLE DATA GENERATED FROM: MODEL
                  SAMPLE SIZE               :  200
                  DATA IS NORMAL            ? YES
                  DATA TO BE CONTAMINATED   ? NO
                  ORIGINAL SEED             =    987654321.
                  DATA FILE TO BE SAVED     ? NO


UNIVARIATE CHARACTERISTICS OF SIMULATED DATA:

            MEAN    STAND. DEV.  UNIVARIATE TRANSFORMATION
   F1  :   0.0000      1.0000    NORMAL  DISTRIBUTION
   F2  :   0.0000      1.0000    NORMAL  DISTRIBUTION
   E1  :   0.0000      0.6000    NORMAL  DISTRIBUTION
   E2  :   0.0000      0.7141    NORMAL  DISTRIBUTION
   E3  :   0.0000      0.4359    NORMAL  DISTRIBUTION
   E4  :   0.0000      0.6000    NORMAL  DISTRIBUTION
   E5  :   0.0000      0.4359    NORMAL  DISTRIBUTION
```

EQS will write converged parameter values, and their standard errors, plus other default summary information on the file EXAMPLE1.ETS for each replication, but the usual model run output file (OUT file) will contain a detailed description of the contents of EXAMPLE1.ETS. It also will provide the statistical summary of results across the replications as shown below. Of course, the numbered statistics displayed will depend on the model and options you specified. For instance, if ROBUST were added to the model file, several additional statistics would appear below, as well as robust standard errors of parameters. By choosing options in the OUTPUT section, you can get summary statistics for those results that interest you. See details about the OUTPUT section in Chapter 3. As you can see, there is quite a bit of information that can be gleaned from the summary.

As shown below, in 92% of the replications, the model converged with no parameter condition codes. We display the statistics from those replications. The summaries for the other 8% of replications, and for all replications combined, are also provided, but this additional output is not duplicated below. *A complete understanding of the results of a simulation will require understanding why and how replications with no condition codes differ from those with condition codes.*

Some illustrative results are the following. The mean of the tail probability (the first statistic in the summary) is much larger than .05, indicating a good model fit. Next, the fit indices, statistics 4 through 10, are extremely close to 1, and their standard deviations are small. The mean SRMR is very tiny (.0045), indicating an excellent fit on average. Also, the mean values of the parameter standard errors are very close to the standard deviations of the parameter estimates. The latter are the empirical standard errors calculated from the simulation. Notice that for each statistic, the summary information includes the values of that statistic that bound 90% of the empirical results. For example, 90% of the RMSEA values lie in the range 0-.108; the lowest 5% of the cases must equal 0 (to rounding) while the largest 5% of RMSEA values are larger than .108. The distribution of RMSEA is quite skewed. As you can see, a lot of statistical information is described, most of which is given by default.

SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=ML

SUCCESS (CONVERGENCE WITH NO CONDITION CODE) OCCURRED IN
92 REPLICATIONS OUT OF 100 (92.00 PERCENT)

NUMBERED STATISTICS ARE
1. TAIL PROBABILITY FOR MODEL CHI-SQUARE
2. INDEPENDENCE MODEL CHI-SQUARE
3. MODEL CHI-SQUARE
4. LISREL GFI FIT INDEX
5. LISREL AGFI FIT INDEX
6. BOLLEN (IFI) FIT INDEX
7. MCDONALD (MFI) FIT INDEX
8. BENTLER-BONETT NORMED FIT INDEX
9. BENTLER-BONETT NON-NORMED FIT INDEX
10. COMPARATIVE FIT INDEX (CFI)
11. ROOT MEAN-SQUARE RESIDUAL (RMR)
12. STANDARDIZED RMR
13. ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)
14. CRONBACH'S ALPHA
15. GREATEST LOWER BOUND RELIABILITY
16. BENTLER'S DIMENSION-FREE LOWER BOUND RELIABILITY
17. SHAPIRO'S LOWER BOUND RELIABILITY FOR A WEIGHTED COMPOSITE
18. RELIABILITY COEFFICIENT RHO
19. MINIMIZED MODEL FUNCTION VALUE

SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=ML FOR 92 SUCCESSES

STATISTIC	1	2	3	4	5
MEAN	0.5064	568.0553	0.8816	0.9982	0.9737
STANDARD DEV.	0.2662	46.8674	1.2529	0.0025	0.0373
SKEWNESS (G1)	-0.0651	-0.0129	2.8831	-2.9459	-2.9457
KURTOSIS (G2)	-0.9430	-0.2181	10.7702	11.3603	11.3572
LOWER 5%	0.0683	497.3840	0.0056	0.9935	0.9018
UPPER 5%	0.9409	637.8620	3.3228	1.0000	0.9998

STATISTIC	6	7	8	9	10
MEAN	1.0002	1.0003	0.9984	1.0019	0.9993
STANDARD DEV.	0.0023	0.0031	0.0023	0.0232	0.0019
SKEWNESS (G1)	-3.0896	-2.8620	-3.0854	-3.0931	-4.2170
KURTOSIS (G2)	12.5479	10.5946	12.5753	12.5825	21.4807
LOWER 5%	0.9963	0.9942	0.9946	0.9620	0.9962
UPPER 5%	1.0017	1.0025	1.0000	1.0179	1.0000

STATISTIC	11	12	13	14	15
MEAN	0.0049	0.0045	0.0196	0.8727	0.9144
STANDARD DEV.	0.0036	0.0033	0.0393	0.0130	0.0089
SKEWNESS (G1)	1.5243	1.6363	2.1014	-0.0566	-0.2855
KURTOSIS (G2)	3.9289	4.6826	3.8709	-0.6904	0.2160
LOWER 5%	0.0006	0.0005	0.0000	0.8543	0.9006
UPPER 5%	0.0107	0.0096	0.1080	0.8905	0.9290

STATISTIC	16	17	18	19
MEAN	0.9143	0.9190	0.9116	0.0044
STANDARD DEV.	0.0089	0.0669	0.0093	0.0063
SKEWNESS (G1)	-0.2861	-9.2115	-0.3734	2.8830
KURTOSIS (G2)	0.2139	84.2360	0.2098	10.7700
LOWER 5%	0.9005	0.9102	0.8968	0.0000
UPPER 5%	0.9290	0.9381	0.9262	0.0167

PARAMETER ESTIMATES

PARAMETER	E1,E1	E2,E2	E3,E3	E4,E4	E5,E5
MEAN	0.3550	0.5050	0.1875	0.3441	0.1880
STANDARD DEV.	0.0445	0.0926	0.0649	0.0875	0.0742
SKEWNESS (G1)	0.1566	-0.2275	-0.0037	-0.3856	-0.1404
KURTOSIS (G2)	-0.3843	-0.2000	0.4829	-0.0990	-0.6825
LOWER 5%	0.2897	0.3564	0.0917	0.1926	0.0665
UPPER 5%	0.4336	0.6510	0.2985	0.4716	0.2967

```
PARAMETER         V1,F1      V2,F1      V2,F2      V3,F1      V3,F2
MEAN               0.7912     0.6872    -0.3896     0.8847     0.3056
STANDARD DEV.      0.0552     0.0749     0.0874     0.0616     0.0789
SKEWNESS (G1)      0.4007     0.1703    -0.0369    -0.0072     0.2320
KURTOSIS (G2)     -0.2614    -0.4169    -0.4688     0.2947     0.0232
LOWER  5%          0.7153     0.5690    -0.5306     0.7994     0.1833
UPPER  5%          0.8917     0.8073    -0.2598     0.9982     0.4433

PARAMETER         V4,F1      V4,F2      V5,F1      V5,F2
MEAN               0.7882     0.4119     0.8904    -0.2870
STANDARD DEV.      0.0680     0.0921     0.0611     0.0855
SKEWNESS (G1)      0.2343    -0.3012     0.1540    -0.0003
KURTOSIS (G2)     -0.1280    -0.3099    -0.1176    -0.2296
LOWER  5%          0.6796     0.2629     0.7913    -0.4225
UPPER  5%          0.8935     0.5538     0.9885    -0.1516


PARAMETER STANDARD ERRORS

PARAMETER         E1,E1      E2,E2      E3,E3      E4,E4      E5,E5
MEAN               0.0431     0.0968     0.0666     0.0899     0.0772
STANDARD DEV.      0.0044     0.0199     0.0154     0.0228     0.0249
SKEWNESS (G1)      0.4307     1.0639     1.1655     0.9207     1.8723
KURTOSIS (G2)     -0.1671     0.9864     2.4516     0.2312     4.0506
LOWER  5%          0.0369     0.0718     0.0461     0.0592     0.0526
UPPER  5%          0.0506     0.1425     0.0946     0.1376     0.1246

PARAMETER         V1,F1      V2,F1      V2,F2      V3,F1      V3,F2
MEAN               0.0599     0.0768     0.0979     0.0646     0.0896
STANDARD DEV.      0.0027     0.0042     0.0150     0.0030     0.0129
SKEWNESS (G1)      0.3058     0.3211     1.1035     0.0887     0.8729
KURTOSIS (G2)     -0.2380     0.2785     1.0277    -0.8507     0.5096
LOWER  5%          0.0564     0.0701     0.0779     0.0602     0.0736
UPPER  5%          0.0646     0.0843     0.1271     0.0690     0.1145

PARAMETER         V4,F1      V4,F2      V5,F1      V5,F2
MEAN               0.0735     0.0992     0.0643     0.0996
STANDARD DEV.      0.0042     0.0166     0.0028     0.0210
SKEWNESS (G1)     -0.0232     1.1296     0.1147     1.7529
KURTOSIS (G2)     -0.0931     1.3797     0.5631     3.3715
LOWER  5%          0.0667     0.0786     0.0603     0.0779
UPPER  5%          0.0799     0.1292     0.0687     0.1436
```

POPULATION=MATRIX

Data can also be generated from a population covariance matrix that you specify. In such a case, random samples for the simulation will be from a population with the given covariance matrix. We illustrate this method of generating data by running the model file below. In contrast to the previous simulation, here the population covariance matrix Σ is specified. It is included in the data file under /MATRIX, though it could be read in from an external file. The given model is used to specify the model to be used in the analysis. Hence, it is possible for the MATRIX and model to be different, that is, $\Sigma \neq \Sigma(\theta)$ as one would want when studying power.

```
/TITLE
 SIMULATED CONFIRMATORY FACTOR ANALYSIS EXAMPLE
/SPECIFICATIONS
 CASES = 400; VARIABLES = 6; ME = ML; MA = COV;  !SAMPLE SIZES AND ANALYSIS METHODS
/SIMULATION
 POPULATION=MATRIX; REPLICATIONS=100;            !HOW POPULATION IS TO BE GENERATED
/EQUATIONS
 V1 = 0.649*F1      + E1;                        !ANALYSIS MODEL FOR EACH REPLICATION
 V2 = 0.709*F1      + E2;
 V3 = 0.373*F1      + E3;
 V4 =      0.831*F2 + E4;
 V5 =      0.778*F2 + E5;
 V6 =      0.897*F2 + E6;
/VARIANCE
 F1 TO F2 = 1;
 E1 = 0.579*;
 E2 = 0.498*;
 E3 = 0.861*;
 E4 = 0.309*;
 E5 = 0.395*;
 E6 = 0.195*;
/COVARIANCE
 F1,F2 = 0.244*;
/MATRIX                                          !COULD BE SAME OR DIFFERENT FROM MODEL
 1.0000
 0.4598 1.0000
 0.2421 0.2647 1.0000
 0.1317 0.1439 0.0758 1.0000
 0.1232 0.1347 0.0709 0.6469 1.0000
 0.1421 0.1553 0.0818 0.7460 0.6982 1.0000
/OUTPUT
 DATA='EXAMPLE2.ETS';                           !WHERE SIMULATION DETAILS WILL GO
 PARAMETER ESTIMATES;
 STANDARD ERRORS;
/END
```

As in the previous example, the usual OUT file describes how the data are generated. When POPULATION=MATRIX, the variables generated are precisely the V's whose covariance matrix is given. Here again each variable is generated from a normal distribution, but if the population is meant to be nonnormally distributed, either transformations (described in Ch. 3 and below) or the method of Vale and Maurelli (1983) is used to generate data with the given covariance structure.

```
  SIMULATION DEFINITIONS

                  NUMBER OF REPLICATIONS    : 100
                  SAMPLE DATA GENERATED FROM: MATRIX
                  SAMPLE SIZE               :  400
                  DATA IS NORMAL            ? YES
                  DATA TO BE CONTAMINATED   ? NO
                  ORIGINAL SEED             =    123456789.
                  DATA FILE TO BE SAVED     ? NO


  UNIVARIATE CHARACTERISTICS OF SIMULATED DATA:

   VARIABLE            V1         V2         V3         V4         V5         V6

   MEAN              0.0000     0.0000     0.0000     0.0000     0.0000     0.0000
   STAND. DEV.       1.0000     1.0000     1.0000     1.0000     1.0000     1.0000
   SKEWNESS          0.0000     0.0000     0.0000     0.0000     0.0000     0.0000
   KURTOSIS          0.0000     0.0000     0.0000     0.0000     0.0000     0.0000
```

```
 DATA GENERATED BASED ON FOLLOWING CORRELATION MATRIX

                      V1         V2         V3         V4         V5         V6
                     V  1       V  2       V  3       V  4       V  5       V  6
      V1    V  1      1.000
      V2    V  2      0.460      1.000
      V3    V  3      0.242      0.265      1.000
      V4    V  4      0.132      0.144      0.076      1.000
      V5    V  5      0.123      0.135      0.071      0.647      1.000
      V6    V  6      0.142      0.155      0.082      0.746      0.698      1.000
```

For this model, all replications converged. Results are presented for only the first 10 statistics. The format is the same as before. The tail probability and fit indices indicate a good model fit.

```
 SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=ML

 SUCCESS (CONVERGENCE WITH NO CONDITION CODE) OCCURRED IN
   100 REPLICATIONS OUT OF  100 (100.00 PERCENT)

 NUMBERED STATISTICS ARE
   1. TAIL PROBABILITY FOR MODEL CHI-SQUARE
   2. INDEPENDENCE MODEL CHI-SQUARE
   3. MODEL CHI-SQUARE
   4. LISREL    GFI  FIT INDEX
   5. LISREL   AGFI  FIT INDEX
   6. BOLLEN   (IFI) FIT INDEX
   7. MCDONALD (MFI) FIT INDEX
   8. BENTLER-BONETT     NORMED FIT INDEX
   9. BENTLER-BONETT NON-NORMED FIT INDEX
  10. COMPARATIVE FIT INDEX (CFI)
  11. ROOT MEAN-SQUARE RESIDUAL (RMR)
  12. STANDARDIZED RMR
  13. ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)
  14. CRONBACH'S ALPHA
  15. GREATEST LOWER BOUND RELIABILITY
  16. BENTLER'S DIMENSION-FREE LOWER BOUND RELIABILITY
  17. SHAPIRO'S LOWER BOUND RELIABILITY FOR A WEIGHTED COMPOSITE
  18. RELIABILITY COEFFICIENT RHO
  19. MINIMIZED MODEL FUNCTION VALUE

 SUMMARY STATISTICS FOR ALL REPLICATIONS
 ---------------------------------------

   STATISTIC            1          2          3          4          5
   MEAN            0.5463   779.7889     7.3977     0.9939     0.9840
   STANDARD DEV.   0.2835    57.5076     3.6646     0.0030     0.0078
   SKEWNESS (G1)  -0.3027     0.3288     0.9443    -0.8937    -0.8938
   KURTOSIS (G2)  -1.1286    -0.3234     0.5364     0.3803     0.3793
   LOWER  5%       0.0714   699.2500     2.5927     0.9881     0.9687
   UPPER  5%       0.9573   878.4700    14.4210     0.9979     0.9944

   STATISTIC            6          7          8          9         10
   MEAN            1.0009     1.0008     0.9905     1.0016     0.9984
   STANDARD DEV.   0.0047     0.0046     0.0045     0.0089     0.0029
   SKEWNESS (G1)  -0.8483    -0.9331    -0.8181    -0.8448    -2.0589
   KURTOSIS (G2)   0.1879     0.5054     0.1261     0.1862     3.7489
   LOWER  5%       0.9920     0.9920     0.9820     0.9848     0.9919
   UPPER  5%       1.0072     1.0068     0.9965     1.0137     1.0000
```

Generating Nonnormal Data

As noted above, simulation with nonnormal data is often done in order to study the empirical performance of methods that are rationalized on the basis of multivariate normal theory. For example, one may inquire how sensitive a given test statistic is to violation of normality. In EQS, nonnormal data is obtained by transforming or categorizing initially generated normal data. As summarized in Chapter 3, a limited set of nonlinear transformations that result in nonnormal variables is available in EQS. Alternatively, you specify the desired end result in terms of the marginal skew and kurtosis of the variables. Both methods are implemented so as to preserve your original covariance structure. Another way that data can deviate from assumptions is that variables are categorical rather than continuous. Response formats that contain only a few ordinal response options are very typical in the social sciences. Hence, the simulation methodology allows the creation of categorized variables to be used in the simulation. The various options are implemented with TRANSFORMATION and CATEGORIZATION statements.

To illustrate the transformation and categorization of generated data, we add the following lines to the SIMULATION section of the previous example:

```
 TRANSFORMATION = V1:S-0.5,K0.00&
                  V2:S .50,K2.00&
                  V3:LOG        &
                  V4:EXP        &
                  V5:S-0.6,K1.00&
                  V6:S1.00,K1.20;
CATEGORIZATION = V1 - V6: -2.5, -1.5, -0.5, 0.5, 1.5, 2.5;
```

The TRANSFORMATION command defines how each variable is generated. V3 is generated by applying the logarithmic function to random data from a uniform (0,1) distribution. V4 is generated by applying the exponential function to random data from a normal (0,1) distribution. The other V's are generated using the procedure of Vale and Maurelli (1983, eqs. 1-11), to generate multivariate nonnormal distributions having the given correlations, means, variances, skewnesses and kurtoses. Their procedure is an extension of Fleishman (1978). The skewnesses and kurtoses are user-defined, e.g. V1 has a skewness of –0.5 and a kurtosis of 0. The means are all 0 and the standard deviations are all 1. Without a CATEGORIZATION command, the generated data would be continuous and have the population characteristics specified. In samples, the distribution of the variables will differ somewhat from the given specification.

In this setup, since a CAT command is included, after generating the data, categorization is applied to each V in accord with the specified cut points. Specifying 2 cut points generates a 3-point ordinal scale. Here, 6 cut points are specified for each of the 6 variables. This will generate a 7-point ordinal scale, accomplished as follows: Any value less than –2.5 is replaced by 1, any value in the range (-2.5, -1.5) is replaced by 2, etc. Any value exceeding 2.5 is replaced by 7. The usual output file from the run describes how the data are generated:

```
SIMULATION DEFINITIONS

                  NUMBER OF REPLICATIONS    : 100
                  SAMPLE DATA GENERATED FROM: MATRIX
                  SAMPLE SIZE               :  400
                  DATA IS NORMAL            ? NO
                  DATA TO BE CONTAMINATED   ? NO
                  ORIGINAL SEED             =    123456789.
                  DATA FILE TO BE SAVED     ? NO
```

```
 UNIVARIATE CHARACTERISTICS OF SIMULATED DATA:

 VARIABLE            V1        V2        V3        V4        V5        V6

 MEAN              0.0000    0.0000    0.0000    0.0000    0.0000    0.0000
 STAND. DEV.       1.0000    1.0000    1.0000    1.0000    1.0000    1.0000
 SKEWNESS         -0.5000    0.5000    0.0000    0.0000   -0.6000    1.0000
 KURTOSIS          0.0000    2.0000    3.0000    2.0000    0.0000    1.2000


 CATEGORIZATION THRESHOLD

    V1  : -2.5000 -1.5000 -0.5000  0.5000  1.5000  2.5000
    V2  : -2.5000 -1.5000 -0.5000  0.5000  1.5000  2.5000
    V3  : -2.5000 -1.5000 -0.5000  0.5000  1.5000  2.5000
    V4  : -2.5000 -1.5000 -0.5000  0.5000  1.5000  2.5000
    V5  : -2.5000 -1.5000 -0.5000  0.5000  1.5000  2.5000
    V6  : -2.5000 -1.5000 -0.5000  0.5000  1.5000  2.5000
```

The OUT file indicates that this model converged with no condition codes in 77% of the replications. The summary for the successes again shows a good model fit. The results are presented below without comment. However, you should recognize that *you should always understand the reasons for condition codes, and have a well rationalized plan for describing the entire results, not just the converged results.* You will have to study the ETS file in detail to determine precisely what kind of problems the runs had. Some problems are minor, e.g., a variance is held to the non-negative boundary. Others are more serious, e.g., inability to get a converged solution. The latter problem will occur quite frequently if the analysis model differs substantially from the data generation model, and especially when sample size is small. Perhaps you will have to take the sample data from a nonconverged run, to study it in detail in a separate analysis. In general, *when a substantial portion of replications result in condition codes, those replications without condition codes are liable to provide a biased view of the performance of statistics.* The bias may make your statistics look better than they really are. For example, a nonconverged run may occur because a chi-square test statistic could not be computed; if this happens, odds are good that if you were to get convergence, it would be associated with a very large chi-square test.

```
 SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=ML

 SUCCESS (CONVERGENCE WITH NO CONDITION CODE) OCCURRED IN
    77 REPLICATIONS OUT OF  100 ( 77.00 PERCENT)

 NUMBERED STATISTICS ARE
   1. TAIL PROBABILITY FOR MODEL CHI-SQUARE
   2. INDEPENDENCE MODEL CHI-SQUARE
   3. MODEL CHI-SQUARE
   4. LISREL    GFI  FIT INDEX
   5. LISREL   AGFI  FIT INDEX
   6. BOLLEN   (IFI) FIT INDEX
   7. MCDONALD (MFI) FIT INDEX
   8. BENTLER-BONETT     NORMED FIT INDEX
   9. BENTLER-BONETT NON-NORMED FIT INDEX
  10. COMPARATIVE FIT INDEX (CFI)
  11. ROOT MEAN-SQUARE RESIDUAL (RMR)
  12. STANDARDIZED RMR
  13. ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)
  14. CRONBACH'S ALPHA
  15. GREATEST LOWER BOUND RELIABILITY
  16. BENTLER'S DIMENSION-FREE LOWER BOUND RELIABILITY
  17. SHAPIRO'S LOWER BOUND RELIABILITY FOR A WEIGHTED COMPOSITE
  18. RELIABILITY COEFFICIENT RHO
  19. MINIMIZED MODEL FUNCTION VALUE
```

Only selected results are presented in the following, to illustrate the nature of the output.

```
SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=ML          FOR    77 SUCCESSES

  STATISTIC            1          2          3          4          5
  MEAN            0.5250   311.9635     7.5069     0.9938     0.9838
  STANDARD DEV.   0.2717    30.6154     3.3870     0.0027     0.0072
  SKEWNESS (G1)  -0.1607     0.0870     0.5297    -0.4924    -0.4921
  KURTOSIS (G2)  -1.0267     0.0205    -0.3219    -0.3512    -0.3504
  LOWER  5%       0.0934   260.3420     2.4887     0.9887     0.9702
  UPPER  5%       0.9622   358.2095    13.5961     0.9979     0.9946

  STATISTIC            6          7          8          9         10
  MEAN            1.0018     1.0006     0.9758     1.0034     0.9962
  STANDARD DEV.   0.0112     0.0042     0.0109     0.0214     0.0068
  SKEWNESS (G1)  -0.4214    -0.5172    -0.5028    -0.4176    -1.5361
  KURTOSIS (G2)  -0.4988    -0.3283    -0.4994    -0.4981     0.8744
  LOWER  5%       0.9811     0.9930     0.9550     0.9639     0.9807
  UPPER  5%       1.0182     1.0069     0.9915     1.0349     1.0000
  ----

PARAMETER ESTIMATES

  PARAMETER       F2,F1      E1,E1      E2,E2      E3,E3      E4,E4
  MEAN            0.2545     0.5875     0.5699     1.0074     0.1362
  STANDARD DEV.   0.0624     0.1851     0.1892     0.0896     0.0336
  SKEWNESS (G1)  -0.2137    -0.6358    -0.7536     0.1205     0.2717
  KURTOSIS (G2)  -0.5886     0.5156     0.2567     0.6886    -0.2792
  LOWER  5%       0.1560     0.2323     0.1866     0.8872     0.0861
  UPPER  5%       0.3548     0.8201     0.8090     1.1368     0.1870
  ----

  PARAMETER       V4,F2      V5,F2      V6,F2
  MEAN            0.0018     0.8097     0.8141
  STANDARD DEV.   0.0231     0.1031     0.0957
  SKEWNESS (G1)   0.1124     0.2307     0.2016
  KURTOSIS (G2)  -0.0896    -0.4419    -0.6620
  LOWER  5%      -0.0379     0.6490     0.6640
  UPPER  5%       0.0394     0.9796     0.9768


PARAMETER STANDARD ERRORS

  PARAMETER       F2,F1      E1,E1      E2,E2      E3,E3      E4,E4
  MEAN            0.0723     0.1994     0.1761     0.0715     0.0097
  STANDARD DEV.   0.0043     0.1232     0.0855     0.0063     0.0024
  SKEWNESS (G1)   0.5971     2.8122     2.0281     0.1213     0.2700
  KURTOSIS (G2)   0.5437     9.2061     3.7605     0.7145    -0.2839
  LOWER  5%       0.0669     0.1073     0.1034     0.0631     0.0061
  UPPER  5%       0.0802     0.4251     0.3934     0.0806     0.0132
  ----

  PARAMETER       V4,F2      V5,F2      V6,F2
  MEAN            0.0202     0.1183     0.1178
  STANDARD DEV.   0.0027     0.0337     0.0319
  SKEWNESS (G1)   0.0526     1.8041     1.5146
  KURTOSIS (G2)  -0.1132     4.6220     4.0510
  LOWER  5%       0.0161     0.0834     0.0808
  UPPER  5%       0.0240     0.1807     0.1729
```

Generating Contaminated Normal Data

Contaminated normal data are no longer normal. They are symmetrically distributed variables that are members of the class of elliptical distributions (see Chapter 5). To illustrate how to achieve this, we now hypothesize that the data of the first example in this chapter is contaminated. For each independent variable, we assume that 90% of the population is normal, with means zero and variances as specified in the VARIANCE section of the model. The

remaining 10% has mean zero, but the variance is six times larger than the given variance. We add the following line to the SIMULATION section:

```
CONTAMINATION = .1, 6;
```

Also we add ROBUST to the METHOD= command in the SPECIFICATION section. The OUT file indicates how the data are to be contaminated:

```
  SIMULATION DEFINITIONS

                  NUMBER OF REPLICATIONS    : 100
                  SAMPLE DATA GENERATED FROM: MODEL
                  SAMPLE SIZE               :  200
                  DATA IS NORMAL            ? NO
                  DATA TO BE CONTAMINATED   ? YES
                  CONTAMINATION PERCENTILE  =       0.1000
                  CONTAMINATION FACTOR      =       6.0000
                  ORIGINAL SEED             =    987654321.
                  DATA FILE TO BE SAVED     ? NO
```

The OUT file indicates that this model converged in 91% of the replications. Both tail probabilities (the first two statistics) are much larger than .05, and the fit indices (statistics 5-11 and 22-26) are very close to 1, so the fit is quite good. A part of the summary statistics section is duplicated below.

```
 SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=ML   +ROBUST

SUCCESS (CONVERGENCE WITH NO CONDITION CODE) OCCURRED IN
   91 REPLICATIONS OUT OF  100 ( 91.00 PERCENT)

NUMBERED STATISTICS ARE
  1. TAIL PROBABILITY FOR MODEL CHI-SQUARE
  2. TAIL PROBABILITY FOR SATORRA-BENTLER SCALED CHI-SQUARE
  3. INDEPENDENCE MODEL CHI-SQUARE
  4. MODEL CHI-SQUARE
  5. LISREL    GFI  FIT INDEX
  6. LISREL   AGFI  FIT INDEX
  7. BOLLEN   (IFI) FIT INDEX
  8. MCDONALD (MFI) FIT INDEX
  9. BENTLER-BONETT     NORMED FIT INDEX
 10. BENTLER-BONETT NON-NORMED FIT INDEX
 11. COMPARATIVE FIT INDEX (CFI)
 12. ROOT MEAN-SQUARE RESIDUAL (RMR)
 13. STANDARDIZED RMR
 14. ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)
 15. CRONBACH'S ALPHA
 16. GREATEST LOWER BOUND RELIABILITY
 17. BENTLER'S DIMENSION-FREE LOWER BOUND RELIABILITY
 18. SHAPIRO'S LOWER BOUND RELIABILITY FOR A WEIGHTED COMPOSITE
 19. RELIABILITY COEFFICIENT RHO
 20. ROBUST INDEPENDENCE MODEL CHI-SQUARE
 21. SATORRA-BENTLER SCALED CHI-SQUARE
 22. ROBUST  BOLLEN  (IFI) FIT INDEX
 23. ROBUST MCDONALD (MFI) FIT INDEX
 24. ROBUST BENTLER-BONETT     NORMED FIT INDEX
 25. ROBUST BENTLER-BONETT NON-NORMED FIT INDEX
 26. ROBUST COMPARATIVE FIT INDEX
 27. ROBUST ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)
 28. MINIMIZED MODEL FUNCTION VALUE
```

SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=ML +ROBUST FOR 91 SUCCESSES

STATISTIC	1	2	3	4	5
MEAN	0.4918	0.4877	583.8333	0.9781	0.9981
STANDARD DEV.	0.2815	0.2856	51.7194	1.3625	0.0026
SKEWNESS (G1)	-0.0171	-0.0332	-0.1892	3.1072	-2.9927
KURTOSIS (G2)	-1.3169	-1.3353	0.0223	14.0027	13.0585
LOWER 5%	0.0586	0.0525	498.9740	0.0146	0.9931
UPPER 5%	0.9039	0.8993	673.1925	3.5781	1.0000

STATISTIC	6	7	8	9	10
MEAN	0.9712	1.0001	1.0001	0.9983	1.0006
STANDARD DEV.	0.0393	0.0022	0.0034	0.0022	0.0228
SKEWNESS (G1)	-2.9930	-2.7809	-3.0744	-2.7249	-2.7794
KURTOSIS (G2)	13.0625	11.0994	13.7170	10.6703	11.0779
LOWER 5%	0.8966	0.9958	0.9936	0.9942	0.9576
UPPER 5%	0.9996	1.0017	1.0025	1.0000	1.0174

STATISTIC	21	22	23	24	25
MEAN	1.0467	0.9999	0.9999	0.9976	0.9992
STANDARD DEV.	1.5350	0.0034	0.0038	0.0034	0.0349
SKEWNESS (G1)	3.3341	-2.8351	-3.2971	-2.8080	-2.8281
KURTOSIS (G2)	15.5957	10.6421	15.2693	10.1869	10.5661
LOWER 5%	0.0161	0.9934	0.9931	0.9908	0.9320
UPPER 5%	3.7626	1.0025	1.0025	1.0000	1.0256

STATISTIC	26	27	28
MEAN	0.9988	0.0264	0.0049
STANDARD DEV.	0.0029	0.0439	0.0068
SKEWNESS (G1)	-3.8924	1.8794	3.1072
KURTOSIS (G2)	18.0334	3.6407	14.0021
LOWER 5%	0.9932	0.0000	0.0001
UPPER 5%	1.0000	0.1178	0.0180

PARAMETER ESTIMATES

PARAMETER	E1,E1	E2,E2	E3,E3	E4,E4	E5,E5
MEAN	0.3597	0.4660	0.1947	0.3339	0.1988
STANDARD DEV.	0.0451	0.1130	0.0667	0.0982	0.0737
SKEWNESS (G1)	0.0253	-0.8295	-0.2811	-0.6230	-0.3909
KURTOSIS (G2)	-0.4960	0.6336	0.0047	0.2848	0.3678
LOWER 5%	0.2958	0.2616	0.0957	0.1722	0.0658
UPPER 5%	0.4338	0.6019	0.2959	0.4781	0.3145

PARAMETER	V1,F1	V2,F1	V2,F2	V3,F1	V3,F2
MEAN	0.8083	0.7041	-0.4249	0.9003	0.2948
STANDARD DEV.	0.0592	0.0753	0.0987	0.0564	0.0809
SKEWNESS (G1)	0.1942	-0.0295	-0.5246	0.0530	-0.0378
KURTOSIS (G2)	-0.2458	-0.3302	0.2660	0.2355	-0.4664
LOWER 5%	0.7174	0.5884	-0.5732	0.8055	0.1751
UPPER 5%	0.8951	0.8346	-0.2846	0.9856	0.4244

PARAMETER STANDARD ERRORS

PARAMETER	E1,E1	E2,E2	E3,E3	E4,E4	E5,E5
MEAN	0.0435	0.1034	0.0646	0.0891	0.0746
STANDARD DEV.	0.0045	0.0321	0.0161	0.0295	0.0261
SKEWNESS (G1)	0.1413	2.9902	1.3553	2.1536	2.5456
KURTOSIS (G2)	-0.6189	13.5368	2.1254	6.1833	8.5454
LOWER 5%	0.0364	0.0717	0.0463	0.0605	0.0505
UPPER 5%	0.0514	0.1505	0.0997	0.1398	0.1196

```
   PARAMETER         V1,F1      V2,F1      V2,F2      V3,F1      V3,F2
   MEAN               0.0607     0.0770     0.1022     0.0648     0.0878
   STANDARD DEV.      0.0027     0.0039     0.0204     0.0032     0.0139
   SKEWNESS (G1)      0.1322     0.1025     2.3803    -0.1882     1.1991
   KURTOSIS (G2)     -0.0325    -0.4773     8.4113    -0.3955     1.0267
   LOWER  5%          0.0561     0.0705     0.0801     0.0593     0.0727
   UPPER  5%          0.0652     0.0825     0.1434     0.0701     0.1168
   ---

 ROBUST STANDARD ERRORS

   PARAMETER         E1,E1      E2,E2      E3,E3      E4,E4      E5,E5
   MEAN               0.0430     0.1033     0.0634     0.0880     0.0730
   STANDARD DEV.      0.0056     0.0346     0.0167     0.0307     0.0259
   SKEWNESS (G1)      0.4082     3.7040     1.0133     2.3510     2.5440
   KURTOSIS (G2)     -0.0814    20.4991     0.6980     7.3125     8.8472
   LOWER  5%          0.0347     0.0730     0.0442     0.0579     0.0486
   UPPER  5%          0.0536     0.1569     0.0942     0.1449     0.1155

   PARAMETER         V1,F1      V2,F1      V2,F2      V3,F1      V3,F2
   MEAN               0.0601     0.0770     0.1016     0.0639     0.0857
   STANDARD DEV.      0.0053     0.0075     0.0221     0.0064     0.0154
   SKEWNESS (G1)      0.0597     0.0917     2.7213     0.7781     0.8312
   KURTOSIS (G2)      0.1048     1.1064    12.0178     0.3282     0.3232
   LOWER  5%          0.0514     0.0636     0.0763     0.0552     0.0673
   UPPER  5%          0.0700     0.0880     0.1353     0.0752     0.1162
   ---
```

Resampling

As noted in the introduction, resampling from your data can substitute for sampling from an artificial population. In EQS, available resampling methods include the regular (i.e., completely nonparametric) bootstrap, model-based bootstrap, case-robust bootstrap, and jackknife procedures.

Bootstrap

In a given bootstrap sample, we sample independently and repeatedly from the existing data file until the desired number of cases *n* has been selected. Thus these draws are done with replacement. The original data consists of *N* cases and we resample until we have a sample of specified size *n*. The number *n* may be equal to, smaller than, or larger than *N*. Typically, one chooses $n=N$, although theory and empirical work shows that $n<N$ corrects some bootstrap failures. This process is repeated for *B* samples. In each of the *B* samples, EQS will run whatever model you specify and will save the resulting statistics that you specify. At the conclusion, a summary of results is printed, as was illustrated above.

The idea of the bootstrap is simple but useful. An excellent overview is given by Efron and Tibshirani (1993). One of its virtues is that you do not need to specify a population. It is nonparametric. If your sample is a representative sample from some well defined population, and the sample size is large enough, we can expect that the empirical distribution of the bootstrap samples will approach the population distribution. As a result, we can draw inferences about parameters associated with that population, in our case, parameters related to structural models. In practice, since the early work of Boomsma (1983, 1986) the bootstrap method has been shown to usefully approximate the distribution of parameter estimates. Empirical results have matched theoretical expectations when these are clearly known. But an important justification for the bootstrap is that it can be implemented and may give acceptable results on sample statistics that are sufficiently complicated mathematically that their theoretical distribution is not even known.

The empirical sampling distribution of any statistic can be taken as the distribution of that statistic across the repeated bootstrap samples. If the statistic we are interested in estimates some parameter in a model, we thus obtain an approximate empirical standard error estimate for that parameter. In addition to computing an empirical standard error for any statistic of interest, we can also obtain an empirical confidence interval estimate. The average value across bootstrap samples can be used to estimate bias, which indicates the difference between the bootstrap mean and the original estimate.

The bootstrap process can be illustrated as follows, where we consider the model $\Sigma = \Sigma(\theta)$, estimate $\hat{\theta}$ by ML, and use the bootstrap rather than a formula to estimate the standard error of some parameter estimate $\hat{\theta}_i$. Let *Z* be the Nxp matrix of raw scores that we plan to resample. We can sample rows of *Z* independently with replacement until we get one bootstrap sample *Z** (also Nxp). We compute the sample covariance matrix *S** based on this bootstrap sample, fit the same model, and obtain $\hat{\theta}_i^*$. If we repeat this procedure many times, say, 1000 times, we will have 1000 values of $\hat{\theta}_i^*$ and hence we can obtain the empirical (bootstrap) sampling distribution of $\hat{\theta}_i^*$. In particular, we can compute the standard deviation of the $\hat{\theta}_i^*$ and use this as an estimate of $S.E.(\hat{\theta}_i)$. How to do this in practice is illustrated below. Although we do not discuss the details in this chapter, one can also compute additional statistics such as confidence intervals, bias corrections, and so on.

New applications of the bootstrap continue to be developed, e.g., for developing parsimonious predictor models in regression (Austin & Tu, 2004), and hence one may expect continuing new applications to SEM as well. Although the bootstrap is a useful tool for evaluating the quality of statistics, it is not a cure-all. Like the simulation methodology described earlier, it requires a substantial amount of computation. If it is applied to a small and atypical sample from a population, it is hard to see how repeated samples from such a bad sample can lead to correct inferences about the population. Even small to moderate samples can be a problem as illustrated e.g., by Ichikawa and Konishi (1995), who found that the bootstrap consistently over-estimated empirical standard errors at N=150 in a study of the effect of non-normality. See also Yung and Bentler (1994, 1996a). More importantly, this method may fail in some circumstances, as noted next.

Model-Based Bootstraps

In structural equation modeling the usual bootstrapping methods may fail when sampling from the original data (e.g., Beran & Srivastava, 1985; Bollen & Stine, 1993). The regular bootstrap is a "naïve" or completely nonparametric bootstrap. Typically this form of the bootstrap does not work in the expected way with test statistics and fit indices that evaluate overall model fit in SEM. For example, naive bootstrapping of a test statistic such as T_{ML} may not lead to expected results, such as an approximate chi-square variate when that is to be expected. Beran and Srivastava, and Bollen and Stine, suggested a modified bootstrap that works more accurately in this situation. In a model-based bootstrap, the original data is transformed, and the bootstrap is applied to the transformed data. This results in a kind of semi-nonparametric bootstrap (Yung & Bentler, 1996a). EQS provides two versions of this bootstrap, the Bollen-Stine bootstrap and the Enders bootstrap, the former for complete data and the latter for missing data. EQS also provides the Yung-Bentler (1996a) model-based bootstrap for analysis of power. Finally, EQS also provides the Yuan-Hayashi (2003) robust model-based bootstrap, as discussed in a subsequent section.

The basic idea of the model-based bootstrap is that when studying the performance of some test statistic under a given null hypothesis, the "population" from which we sample – really, the data file used -- must meet the null hypothesis. This will probably not happen automatically, as the sample data *Z* may well not represent the particular SEM structure that we have in mind. Thus we have to transform our sample data in such a way that we impose our null hypothesis on the new transformed data. In EQS this is done with the MBB (or MBBOOTSTRAP) statement, which, based on the model specified in the model file, generates the model covariance matrix $\hat{\Sigma}$ and if required the mean vector $\hat{\mu}$, that are then used to modify the resampling space for bootstrapping as given below. Evidence is that this will lead to quite good accuracy in estimating the sampling distribution of test statistics under the null hypothesis (e.g., Nevitt & Hancock, 2001).

Let z_i' be the i^{th} row of Z, that is, the observed data for the i^{th} case, and $\bar{z}$ represent the vector of sample means. We can compute the unbiased sample covariance matrix of Z as the average deviation cross-product

$$S = \left(\sum_{i=1}^{N} (z_i - \bar{z})(z_i - \bar{z})' \right) / (N-1) \tag{13.1}$$

If we resample from Z, we are resampling from a "population" whose covariance matrix is S. This may well not reflect our model. What we would like to do is to resample from a data file that has covariance matrix $\Sigma(\theta)$. Of course, $\Sigma(\theta)$ is not known, so this is not practical. But suppose we fit $\Sigma(\theta)$ to S, and the fitted model is $\Sigma(\hat{\theta})$. What the model-based bootstrap does is to determine a data file Y whose covariance matrix is $\Sigma(\hat{\theta})$, thus limiting the bootstrap resampling space to the null hypothesis. This is accomplished as follows.

We consider the more general case in which there is a mean structure $\hat{\mu} = \mu(\hat{\theta})$ as well as a covariance structure $\hat{\Sigma} = \Sigma(\hat{\theta})$ based on the data matrix Z (Yung & Bentler, 1996a). Then the i^{th} row of the new data matrix Y is computed as the transpose of

$$y_i = \hat{\mu} + \hat{\Sigma}^{1/2} S^{-1/2} (z_i - \bar{z}) . \tag{13.2}$$

Here we use the notation that $A^{1/2}$ is an appropriate square root of a matrix (e.g., a symmetric factorization of A such that $A^{1/2} A^{1/2} = A$, $A^{1/2} A^{-1/2} = I$). Direct computation shows that the sample mean of (13.2) is given by $\hat{\mu}$, and

$$\hat{\Sigma} = \frac{\sum_{i=1}^{N} (y_i - \hat{\mu})(y_i - \hat{\mu})'}{(N-1)} . \tag{13.3}$$

This can be seen by substituting (13.2) in the right part of (13.3), giving

$$\frac{\sum_{i=1}^{N} (\hat{\Sigma}^{1/2} S^{-1/2} (z_i - \bar{z}))(\hat{\Sigma}^{1/2} S^{-1/2} (z_i - \bar{z}))'}{(N-1)} = \hat{\Sigma}^{1/2} S^{-1/2} \left(\frac{\sum_{i=1}^{N} (z_i - \bar{z})(z_i - \bar{z})'}{(N-1)} \right) S^{-1/2} \hat{\Sigma}^{1/2}$$

$$= \hat{\Sigma}^{1/2} S^{-1/2} (S) S^{-1/2} \hat{\Sigma}^{1/2} = \hat{\Sigma}$$

In words, the sample covariance matrix of the scores Z is $\Sigma(\hat{\theta})$. Thus the scores Z meet the null hypothesis, and re-sampling from Z provides a model-based bootstrap.

Now suppose that there is missing data. In that case the procedure (13.2) cannot be used since there are missing entries in the data matrix Z and the sample covariance matrix S does not exist. Enders (2002) showed how to implement the model-based bootstrap in this situation. Let z_i' be the row vector of *observed* data for the i^{th} case. With missing data, this contains only the p_i variables with data for that case, i.e., it is not the entire row of the data matrix. In order to make a model-based transformation of the data, we need to use two sets of model-based matrices based on the ML missing data solution that is fitted to Z. The mean vector $\hat{\mu}_i$ and covariance matrix $\hat{\Sigma}_i$ are p_i subsets of the overall estimated saturated model mean vector $\hat{\mu}$ and covariance matrix $\hat{\Sigma}$. Similarly, the null hypothesized model-based mean vector $\mu_i(\hat{\theta})$ and covariance matrix $\Sigma_i(\hat{\theta})$ are p_i subsets of the overall estimated model mean vector $\mu(\hat{\theta})$ and covariance matrix $\Sigma(\hat{\theta})$. Then the i^{th} case model-based data for bootstrapping is

$$y_i = \mu_i(\hat{\theta}) + \Sigma_i(\hat{\theta})^{1/2} \hat{\Sigma}_i^{-1/2} (z_i - \hat{\mu}_i) . \tag{13.4}$$

Enders' formula reduces to the transformation (13.2) when there are no missing data. In practice, computations are grouped by pattern so that (13.4) can be applied immediately to all cases with the same pattern of missing data. Enders has shown that this procedure works well, though perhaps a bit conservatively in small samples.

POPULATION=MATRIX

The model-based bootstrap essentially implements a POP=MODEL command by default, as the transforming matrices are obtained from an imposed structural model. Yung and Bentler (1996a) proposed using a different covariance matrix Σ_A and mean vector μ_A in the defining equation (13.2) for the purpose of studying power and sample size requirements. These matrices are based on an *alternative* model and involve the transformation

$$y_i = \mu_A + \Sigma_A^{1/2} S^{-1/2}(z_i - \bar{z}) . \tag{13.5}$$

In EQS, this is extended to the case of incomplete data with the parallel equation using appropriate subsets

$$y_i = \mu_{iA} + \Sigma_{iA}^{1/2} \hat{\Sigma}_i^{-1/2}(z_i - \hat{\mu}_i) . \tag{13.6}$$

You implement this approach by specifying POP=MATRIX and then providing the matrix that is used as Σ_A. If you do ANALYSIS=MOMENT and specify /MEANS, these are used in the mean vector μ_A. Resampling is then done from the transformed data (13.5)-(13.6), while your model $\mu = \mu(\theta)$ and $\Sigma = \Sigma(\theta)$ is used for analysis of each resampled replication. If $\Sigma_A \neq \Sigma(\theta)$, you are thus sampling from an incorrect model structure, as is required in the study of power. An excellent discussion of this approach is given by Yuan and Hayashi (2003), who also note its relevance for the study of "close" fit. An advantage of this approach, they note, is that "With the bootstrap approach, there is neither a central chi-square distribution when H_0 holds nor a non-central chi-square distribution when H_1 holds. Instead, the distribution of T is judged by its empirical behaviour based on resampling from observed data" (p. 96). Stated differently, the bootstrap approach to power, as compared to alternative approaches (Satorra & Saris, 1985; Muthén & Muthén, 2002), is asymptotically distribution-free. Finally, we note that Yung and Bentler (1996a) also state that it is possible to use a model-estimated matrix $\hat{\Sigma}_A$ instead of an a priori Σ_A, which in EQS you would have to obtain with a separate modeling run. However, they stress that this approach may not work well.

Yung and Bentler also provide a note of caution. "We stress that the two bootstrap methods proposed here are intended to provide 'estimates' of power because the population distribution has been estimated implicitly from the data instead of being assumed. This approach should be clearly distinguished from the empirical and approximate *calculations* suggested by Satorra and Saris (1985) in Table 7.4. If the multivariate normality assumption can be assumed, then the power calculation given by the empirical method used by Satorra and Saris (1985) is exact (subject to numerical accuracy due to simulation). The approximation method, on the other hand, only gives values of power that are exact for extremely large samples" (pp. 220-221).

Case-robust Bootstrap

The naïve and model-based bootstraps discussed above can be implemented with the huge variety of estimation and analysis techniques available in EQS (see e.g., Chapter 5). This would include case-robust methods implemented with the CROBUST statement in the SPEC section of the model file, thus providing one means of handling heavy tailed distributions, outliers, and influential cases. Since case-robust methods are computationally intensive, and such heavy computations would be applied in every bootstrap sample (e.g., 1000 times), this is not an attractive way to proceed. Yuan and Hayashi (2003) argue for producing case-robustness in a different way that reduces the computations by an order of magnitude. Following Yuan, Chan, and Bentler (2000), they propose that the original data sample should be transformed to a new case-robust sample, and the bootstrap should be applied to the transformed sample using ML estimation. Not only does this limit the heavy computations to a single run rather than the many bootstrap replications, but Yuan and Hayashi also show that T_{ML} is approximately *pivotal* (see discussion below) if a good case-robust transformation is obtained thus assuring that the bootstrap should work effectively if sample size is reasonable. In this section we describe the Yuan-Hayashi bootstrap.

CROBUST=#,#;

When specified in the /SIMULATION section, the CROBUST command requires that either regular (BOOTSTRAP= xx;) or model-based (MBBOOT=xx;) bootstraps have also been specified. It will be remembered that the point of CROBUST is to reduce the impact of observations that are far from the robust mean of the observations in the metric of inverse robust covariance matrix, that is, the robust Mahalanobis distances $d_t = sqrt\{(z_t - \hat{\mu})\hat{\Sigma}^{-1}(z_t - \hat{\mu})\}$. This is achieved in EQS by specifying two tuning constants b_1 and b_2. Chapter 5, eqs. (5.75)-(5.78) shows how these are used to obtain robust estimates of means and covariances. The two default tuning constants # and # are $b_1 = 2$ and $b_2 = 1.25$, and can be obtained by just saying CROBUST. However, Yuan and Hayashi argue that using a Huber-type (Huber, 1981) M-estimator may give better results. Huber-type weights can be obtained with the choice $b_1 = 2$ and $b_2 =$ a very large number. The effect of the CROBUST command is to transform the original data file into a case-robust data file for subsequent bootstrapping. This leaves the analysis method for the null hypothesized method entirely up to you. But, as noted, your transformed data file should be close to normally distributed, and Yuan and Hayashi recommend the use of ML. ML can be done with either the regular or model-based bootstrap.

Case-robust ordinary bootstrap. If the command CROBUST is found in simulation with BOOT, we use the original data to calculate the *saturated* case-robust estimates of unstructured means $\hat{\mu}$ and covariances $\hat{\Sigma}$. Then, using the case weights w_i obtained from the robust computations, the raw data file Z is transformed into a new file Y as follows:

$$y_i = \hat{\mu} + kw_i(z_i - \hat{\mu}) \tag{13.7}$$

where $k = \sqrt{(N-1)/(\sum w_i^2 - 1)}$. The scaling k is used so that the new matrix Y has the property that its mean vector is given by $\hat{\mu}$ and its ordinary covariance matrix (divisor N-1) is $\hat{\Sigma}$. This adjustment is done since, as shown in (5.76), $\sum w_i^2 - 1$ is the sample size equivalent used when computing the case-robust covariance matrix. After this transformation, the regular, or naïve bootstrap is implemented in accord with your specifications. No structural model is imposed on the data Y to be bootstrapped.

Case-robust model-based bootstrap. If the command CROBUST is found in simulation with MBBOOT, we set up your model and estimate it using the case-robust means and covariances computed from the original data matrix Z in accord with (5.75)-(5.78). This yields the *structured* model estimates $\mu(\hat{\theta})$ and $\Sigma(\hat{\theta})$. We also make use of the *saturated* model estimates $\hat{\mu}$ and $\hat{\Sigma}$ as described in association with (13.7), and the corresponding case weights w_i. Now the structured model is imposed on the case-robust data

$$y_i = \mu(\hat{\theta}) + \Sigma(\hat{\theta})^{1/2}\hat{\Sigma}^{-1/2}\{kw_i(z_i - \hat{\mu})\}. \tag{13.8}$$

where the term in braces is precisely the term in (13.7). The new matrix Y will have the property that its mean vector is given by the structured $\mu(\hat{\theta})$ and its ordinary covariance matrix (divisor N-1) is the structured $\Sigma(\hat{\theta})$. Subsequently, ordinary bootstrapping is performed on this new Y matrix. This is the Yuan-Hayashi case-robust extension to the Beran-Srivastava and Bollen-Stine bootstrap.

Some Bootstrap Issues

An important feature of any bootstrap is the ability to estimate sampling distributions without severe distributional assumptions of parametric models such as multivariate normality of the population distribution. But, as noted, the imposition of an SEM model may be required. According to Bollen and Stine (1993), the naïve or completely nonparametric bootstrap should be used for estimating the standard errors, whereas the bootstrap distribution of test statistics such as T_{ML} should be estimated by the model-based bootstrap. The latter is the legitimate bootstrap methodology for SEM model evaluation and selection and for studies of power.

With regard to standard error estimates, the choice among bootstrap methods is not completely clear. Although Bollen and Stine (1993) recommended using the naïve bootstrap for estimating standard errors, Yung and Bentler (1996a) showed that the standard error estimates from the model-based bootstrap were closer to those from traditional estimation methods such as ML-Robust estimation than were those from the naive bootstrap. This issue seems to be not completely settled. It can be argued that the naïve or nonparametric bootstrap is more consistent with the philosophy that structural models are never exactly true.

Another issue is the size of the bootstrap sample. We noted above that the typical choice is to use $n=N$. This is not always the best choice, since $n<N$ or $n \ll N$ does correct some bootstrap failures (e.g., Bickel, Götze & van Zwet, 1997). However, the choice of a smaller n can be quite complicated, and we cannot recommend anything specific. Davison, Hinkley, and Young state "…the optimal choice of the subsample size…depends in a delicate way on the inference being performed. Empirical choice is often difficult…" (2003, p. 149).

We would like to repeat that the bootstrap is not assured to give correct results and a certain amount of caution is needed, especially in small samples. Under the right conditions, it can work very well, but such a result is not guaranteed and should in principle be validated in applications. For example, there is a finding that to get correct critical values with the bootstrap, the statistic being studied should be *pivotal*. It is pivotal if its sampling distribution under the null hypothesis does not depend on unknown parameters, and is asymptotically pivotal if this property holds asymptotically. This is a technical concept that requires evaluation in any given application. "… for more accurate inference we should choose a statistic that is as nearly pivotal as possible" (Yuan & Hayashi, 2003, p. 98). With regard to a test statistic, we might say informally that if we can precisely describe its asymptotic distribution as a χ^2 variate, it is asymptotically pivotal. Under the null hypothesis, T_{ML} under normality or under asymptotic robustness, $\bar{T}$ under elliptical distributions, and T_{AGLS} generally when sample size is very large, are all asymptotically pivotal. See, e.g., Yung and Bentler (1996a) and especially Yuan and Hayashi (2003) for a discussion.

Poor conditions of our sample such as small sample size will work against us. If a model has severe problems when applied to the original sample data Z, one should expect that the same model applied to bootstrap samples also will have severe problems. For example, if a model is not identified, does not converge, or yields boundary estimates such as variances of zero, the same difficulties should be expected in bootstrap samples. Almost all statistics of interest in SEM require converged samples, and hence ideally every bootstrap sample would result in a converged solution. But this is unlikely to occur in reality. "Non-convergence issues exist with resampling and caution is needed for bootstrap inference (Ichikawa & Konishi, 1995). This generally happens when the sample size is not large enough, and especially when a model structure is wrong" (Yuan & Hayashi, 2003, p. 106). For example, if a model hardly fits in the original sample, it is quite likely that naïve bootstraps from that sample also would yield many problematic solutions. Imposing a model structure, as in the model-based bootstrap, will help but may not eliminate the problem of lack of convergence. If the statistic of interest is essentially the same for converged as nonconverged runs, then this is not a serious issue. But if quite different empirical bootstrap distributions are observed, then the converged solutions will give a biased view of the bootstrap results and it is inappropriate to rely only on them. Yet, with such heterogeneity, one also has to question whether the overall combined results across all runs represents an appropriate summary. See Yuan and Hayashi for a further discussion and suggestions on how to evaluate bootstrap results under lack of convergence.

At this time, EQS implements the bootstrap only for the typical data situation in which the observations are independent, in one or multiple groups. If there are dependencies among cases, such as in multilevel modeling (see e.g., Meijer, van der Leeden, & Busing, 1995 for a discussion), the EQS bootstrap can not be used directly. You also will have to do your own modifications if you want to bootstrap special quantities, such as regression residuals, which are computed in the Windows front-end of EQS.

A Regular Bootstrap Example

One of the simplest models that can be run is a bivariate regression model, and this can be used to illustrate the naïve bootstrap. We use the exercise.ess data, which consists of some background variables (age, sex, smoking status) and a measure of pulse rate on 40 subjects before and after exercise. Specifically, we regress V6 on V5. This is a saturated model. However, as explained below, we are suspicious about a standard error estimate and hence decide to bootstrap the parameter estimates.

```
/TITLE
 BOOTSTRAP ON BIVARIATE REGRESSION
/SPECIFICATIONS
 DATA='C:\EQS61\EXAMPLES\EXERCISE.ESS';
 VARIABLES=6; CASES=40;
 METHOD=ML; ANALYSIS=COVARIANCE; MATRIX=RAW;
/LABELS
 V1=ID; V2=SEX; V3=SMOKE; V4=AGE; V5=PULSE_1;
 V6=PULSE_2;
/EQUATIONS
 V6 =   *V5 + E6;
/VARIANCES
 V5 = *;
 E6 = *;
/COVARIANCES
/PRINT
 FIT=ALL;
 TABLE=EQUATION;
/OUTPUT
 PARAMETERS;                  !BOOTSTRAP TO BE DONE ON PARAMETER ESTIMATES
 STANDARD ERRORS;             !ALSO, ON STANDARD ERROR FORMULA ESTIMATES
 DATA='EXERCISE.ETS';         !SUMMARIES OF EACH REPLICATION INTO THIS FILE
/SIMULATION                   !KEY STATEMENT TO SET UP MULTIPLE RUNS
 BOOTSTRAP=40;                !BOOTSTRAP IS SPECIFIED, AND WITH N=4O FOR EACH SAMPLE
 REPLICATIONS=1000;           !NUMBER OF BOOTSTRAP SAMPLES
 SEED=123456789.0;            !NUMBER TO START THE RANDOM NUMBER GENERATOR
/END
```

In the ML run on the bivariate regression in the original sample (not the bootstrap run), the regression results showed the following:

```
PULSE_2 =V6  =     1.386*V5    +  1.000 E6
                    .330
                   4.199@
```

It seems that the regression coefficient is statistically significant. However, in its data description, EQS reported a normalized Mardia coefficient of 13.1, indicative of lack of normality. Also, the program indicated

```
BONETT-WOODWARD-RANDALL TEST SHOWS SIGNIFICANT EXCESS KURTOSIS
   INDICATIVE OF NON-NORMALITY AT A ONE-TAIL .05 LEVEL.
```

This raises the question of whether the standard error, which is based on a normal theory ML formula, can be trusted. Hence, the bootstrap with 1000 replications was run. At the end of the OUT file is a summary as follows.

```
PARAMETER ESTIMATES

   PARAMETER         V5,V5      E6,E6      V6,V5
   MEAN              88.2734   347.5702     1.3754
   STANDARD DEV.     16.2198   205.5547      .5734
   SKEWNESS (G1)       .1876      .3369      .5151
   KURTOSIS (G2)       .0046     -.6235     -.0659
   LOWER  5%         62.0487    89.5272      .5987
   UPPER  5%        115.6821   685.3446     2.4335
```

```
PARAMETER STANDARD ERRORS

   PARAMETER         V5,V5      E6,E6      V6,V5
   MEAN             19.9900    78.7092      .3007
   STANDARD DEV.     3.6731    46.5490      .0885
   SKEWNESS (G1)      .1876      .3369     -.3352
   KURTOSIS (G2)      .0046     -.6235    -1.2336
   LOWER  5%        14.0513    20.2739      .1619
   UPPER  5%        26.1968   155.2000      .4207
```

The average estimate of the regression coefficient across the 1000 replications is 1.3754, with a 90% confidence interval from .5987 to 2.4335 (quite wide, no doubt due to the small sample size). The value 1.386 observed in the original sample is very close to the bootstrap mean and is included in the confidence interval. This is reassuring. With regard to the ML standard error of the regression coefficient, the mean ML S.E. formula value computed across the 1000 replications was .30, not far from the .33 value we had found in our original sample. Also, the confidence interval of .16-.42 contains our original .33. *This means that the ML S.E. formula is quite stable. It does not mean that the formula correctly describes the actual sampling variability of the parameter estimates.* When we look at the standard deviation of the parameter estimate for V6,V5, we find its value is .5734. In other words, the actual bootstrap variability across our 1000 replications was substantially larger than even the upper 5% formula value (.42). This provides some evidence against the accuracy of the ML S.E. formula in this small nonnormal data set. But a new *z*-statistic computed as 1.386/.5734 = 2.42, so the coefficient is still significant when based on the bootstrap estimate of variability.

A Model-Based Bootstrap Example

In their study of the bootstrap, Yuan and Hayashi (2003) studied a 9 variable 3 factor data set on nonnormal cognitive variables. The holza.ess data file is also nonnormal and can be analyzed with a similar 3 factor model. It yields a normalized Mardia coefficient of 18.2, indicating substantial kurtosis. When we evaluate the model below with the ML test statistic, we obtain T_{ML} = 191.4 based on 24 df, implying rejection of the model. Under the null hypothesis, using the χ^2_{24} distribution, only 5% of observations would exceed the critical value of 36.4. We may wonder whether the critical value of 36.4 is appropriate, given the data's violation of normality. We study this issue with a model-based bootstrap. We set up the 3-factor model in the usual way, specifying a model-based bootstrap with 1000 replications. Although our interest lies in the T_{ML} test statistic, which is computed automatically, we also request information on parameter estimates and standard errors.

```
/TITLE
  MODEL-BASED BOOTSTRAP  -- c:\eqs61\examples\holza.ess
/SPECIFICATIONS
  DATA='c:\eqs61\examples\holza.ess';
  VARIABLES=9; CASES=145;
  METHODS=ML;
  MATRIX=RAW;
  ANALYSIS=COVARIANCE;
/LABELS
  V1=V1; V2=V2; V3=V3; V4=V4; V5=V5;
  V6=V6; V7=V7; V8=V8; V9=V9;
/EQUATIONS
  V1 =  + *F1  + 1E1;
  V2 =  + *F1  + 1E2;
  V3 =  + *F1  + 1E3;
  V4 =  + *F2  + 1E4;
  V5 =  + *F2  + 1E5;
  V6 =  + *F2  + 1E6;
  V7 =  + *F3  + 1E7;
  V8 =  + *F3  + 1E8;
  V9 =  + *F3  + 1E9;
/VARIANCES
  F1 TO F3 = 1;
  E1 TO E9 = *;
```

```
/COVARIANCES
  F1,F2=*;
  F2,F3=*;
  F1,F3=*;
/PRINT
  FIT=ALL;
/TECHNICAL
  ITERATIONS=500;
/SIMULATION
  MBBOOTSTRAP = 145;                    !THE MODEL-BASED BOOTSTRAP FROM THE
  REPLICATION = 1000;                   !ORIGINAL DATA WITH 145 OBSERVATIONS.
  SEED = 123456789;                     !PROVIDE WITH A SEED NUMBER.
/OUTPUT
  DATA= 'BOOTRES.ETS';
  PARAMETER ESTIMATES;                  !OUTPUT PARAMETER ESTIMATES AND
  STANDARD ERRORS;                      !STANDARD ERRORS OBTAINED FROM
/END                                    !THE MODEL-BASED BOOTSTRAP SAMPLES.
```

As discussed earlier, EQS segregates the results into those based on successful replications with no error messages of any kind, and those that had boundary conditions, iterative failures or any other problem. To see the details on the nature of these problems in each replication, you can go into the ETS file that was created during the run. Alternatively, you can rerun and create a listing of the entire output, run by run, and examine it. Even though the basic model on the original data had difficulty converging (requiring over 400 iterations), here we find that 85% of the runs had no problems. The following is some selected output.

```
 SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=ML

 SUCCESS (CONVERGENCE WITH NO CONDITION CODE) OCCURRED IN     ! BOOTSTRAP SUCCESS RATE
   852 REPLICATIONS OUT OF 1000 ( 85.20 PERCENT)              ! IS INDICATED

 NUMBERED STATISTICS ARE
   1. TAIL PROBABILITY FOR MODEL CHI-SQUARE
   2. INDEPENDENCE MODEL CHI-SQUARE
   3. MODEL CHI-SQUARE
   4. LISREL    GFI  FIT INDEX
   5. LISREL   AGFI  FIT INDEX
   6. BOLLEN   (IFI) FIT INDEX
   7. MCDONALD (MFI) FIT INDEX
   8. BENTLER-BONETT     NORMED FIT INDEX
   9. BENTLER-BONETT NON-NORMED FIT INDEX
  10. COMPARATIVE FIT INDEX (CFI)
  11. ROOT MEAN-SQUARE RESIDUAL (RMR)
  12. STANDARDIZED RMR
  13. ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)
  14. CRONBACH'S ALPHA
  15. GREATEST LOWER BOUND RELIABILITY
  16. BENTLER'S DIMENSION-FREE LOWER BOUND RELIABILITY
  17. RELIABILITY COEFFICIENT RHO
  18. MINIMIZED MODEL FUNCTION VALUE


 SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=ML   ! SUMMARY STATISTICS FOR THE ABOVE
 FOR   852 SUCCESSES                                ! MEASURES ARE CALCULATED OVER ALL
                                                    ! SUCCESSES

   STATISTIC            1          2          3          4          5
   MEAN            0.2156   397.6179    40.9790     0.9466     0.8999
   STANDARD DEV.   0.2899    56.4425    23.7317     0.0226     0.0424
   SKEWNESS (G1)   1.3039     0.8068     2.7474    -1.1858    -1.1858
   KURTOSIS (G2)   0.4527     1.4501    13.1210     2.5278     2.5279
   LOWER  5%       0.0000   321.7760    16.0178     0.9069     0.8253
   UPPER  5%       0.8874   496.1060    80.0394     0.9764     0.9558
```

```
STATISTIC            6         7         8         9        10
MEAN            0.9585    0.9461    0.8994    0.9358    0.9540
STANDARD DEV.   0.0521    0.0701    0.0460    0.0804    0.0500
SKEWNESS (G1)  -1.5469   -1.9561   -1.5932   -1.5212   -1.9304
KURTOSIS (G2)   4.7649    6.9756    5.0121    4.6230    6.1417
LOWER  5%       0.8703    0.8243    0.8199    0.7987    0.8658
UPPER  5%       1.0247    1.0279    0.9559    1.0383    1.0000

STATISTIC           11        12        13        14        15
MEAN           12.7178    0.0561    0.0579    0.6356    0.8044
STANDARD DEV.   4.1975    0.0133    0.0432    0.0346    0.0308
SKEWNESS (G1)   1.3824    0.8010    0.5219   -0.3450   -0.2306
KURTOSIS (G2)   5.6393    1.9918    0.4697    0.1902   -0.0058
LOWER  5%       7.1891    0.0365    0.0000    0.5770    0.7522
UPPER  5%      20.0722    0.0792    0.1273    0.6889    0.8538

STATISTIC           16        17        18
MEAN            0.8011    0.7842    0.2846
STANDARD DEV.   0.0310    0.0323    0.1648
SKEWNESS (G1)  -0.2648   -0.3413    2.7474
KURTOSIS (G2)   0.0190    0.2879   13.1215
LOWER  5%       0.7493    0.7269    0.1112
UPPER  5%       0.8497    0.8350    0.5558
```

Output similar to the above is given for all replications, the converged replications, and the failed replications. The column labeled 3 provides information on the distribution of the T_{ML} statistic across the 852 converged replications. The mean of this statistic is 40.98, which is substantially larger than the expected value 24 of a χ^2_{24} variate. Clearly the empirical distribution is shifted to the right. However, we are especially interested in the bootstrap critical value at the .05 level, that is, the value of the test statistic beyond which 5% of the replications' T_{ML} fall. With the distribution given by these data, and under the null hypothesis, this value is 80.04. Since the value of T_{ML} in the original sample is 191.4, substantially exceeding 80.04, we reject the proposed 3 factor model.

Results from the 15% of replications with problems (nonconvergence, boundary estimates etc.), shown in part below, reinforces this conclusion. In these samples the mean T_{ML} = 43.09, and the bootstrap critical value beyond which 5% of the replications fall is 72.86.

```
SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=ML          FOR   148 FAILURES

STATISTIC            1         2         3         4         5
MEAN             .1713   383.5416    43.0858    .9425     .8921
STANDARD DEV.    .2593    61.0192    27.4491    .0240     .0451
SKEWNESS (G1)   1.7334     1.1237     4.1613  -1.3527   -1.3526
KURTOSIS (G2)   1.9889     3.8010    24.6709   3.5256    3.5255
LOWER  5%        .0000   294.4880    18.1108    .9056     .8230
UPPER  5%        .7974   478.2400    72.8606    .9728     .9490
```

In order to study these results a bit more carefully, we look at the ETS and CBK files.

ETS and CBK Files

We had specified that the information on results in each of the 1000 bootstrap replications should be saved in the file

```
DATA= 'BOOTRES.ETS';
```

The EQS model file that was run produces a normal OUT file from which the summary of replications (above) was taken. The OUT file also provides a summary of what information has been stored in the ETS file, as illustrated in the following small selection from the output.

```
FOLLOWING TECHNICAL INFORMATION HAS BEEN STORED IN BOOTRES.ETS
  REGULAR EQS OUTPUT WILL NOT BE PRODUCED
  ------------------------------------------------------------------------
          TAIL PROBABILITIES NOT APPLICABLE TO THIS ANALYSIS ARE WRITTEN AS -1.
          OTHER STATISTICS WHICH ARE NOT APPLICABLE ARE WRITTEN AS -9.

     SUMMARY SECTION CONTAINS--

          LINE    1 BEGINNING: ANALYSIS ...

          LINE    2 CONTAINING THESE 11 ELEMENTS OF MODEL STATISTICS:

               ESTIMATION METHOD (LS,GLS,ML,ELS,EGLS,ERLS,AGLS,HKGLS,HKRLS)
               CONDITION CODE (0 FOR NORMAL CONDITION)
               CONVERGENCE (0 FOR MODEL CONVERGED)
               NUMBER OF ITERATIONS FOR CONVERGENCE
               DEGREES OF FREEDOM
               NUMBER OF CONSTRAINTS
               DENOMINATOR DEGREES OF FREEDOM FOR F-TESTS
               DEGREES OF FREEDOM FOR POTENTIAL STRUCTURED MEANS MODEL TEST
               D.F. FOR GLS TEST OF HOMOGENEITY OF MEANS
               D.F. FOR GLS TEST OF HOMOGENEITY OF COVARIANCE MATRICES
               D.F. FOR GLS COMBINED TEST OF HOMOGENEITY OF MEANS/COVAS.

          LINE    3 CONTAINING THESE 10 ELEMENTS OF MODEL STATISTICS:

               TAIL PROBABILITY FOR MODEL CHI-SQUARE
               TAIL PROBABILITY FOR RESIDUAL-BASED TEST STATISTIC
               TAIL PROBABILITY FOR YUAN-BENTLER RESIDUAL-BASED TEST STAT.
               TAIL PROBABILITY FOR YUAN-BENTLER AGLS F-STATISTIC
               TAIL PROBABILITY FOR YUAN-BENTLER RESIDUAL-BASED F-STATISTIC
               TAIL PROBABILITY FOR SATORRA-BENTLER SCALED CHI-SQUARE
               TAIL PROB. FOR YUAN-BENTLER SCALED CHI-SQUARE (CASE ROBUST)
               TAIL PROBABILITY FOR SCALED CHI-SQUARE (YUAN-BENTLER)
               TAIL PROBABILITY FOR GLS TEST OF HOMOGENEITY OF MEANS
               TAIL PROB. FOR GLS TEST OF HOMOGENEITY OF COV. MATRICES
```

If you use Wordpad or Notepad to open the bootres.ets file, you will find the numerical information that corresponds to the index given above. This material begins as follows.

```
ANALYSIS USING METHOD = ML            IN REPLICATION   1
       3       0       0      18      24       0     121       0       0       0       0
      .03653    -1.00000    -1.00000    -1.00000    -1.00000    -1.00000    -1.00000    -1.00000
-1.00000    -1.00000
    -1.00000    -1.00000    -1.00000    -1.00000    -1.00000    -1.00000
  437.62      37.777      .95084      .90782      .96669      .95360      .91367      .94854
.96570      9.2315
  .53442E-01  .63139E-01  .16122E-01  .99652E-01 -9.0000     -9.0000     -9.0000     -9.0000
-9.0000     -9.0000
```

As you can see, aside from the header, the ETS file provides data in a free format that you can import into your favorite analysis package.

However, the ETS file can also be imported directly into EQS, and saved as an ESS file. This is what we recommend. When so imported, the data is organized into rows and columns, with each row representing a replication and the various columns representing the statistical output. The columns have shorthand designations for the various statistics, and in order to understand these, EQS automatically creates a *codebook* or CBK file corresponding to your ETS file. It can be read by ascii editor, for example, Notepad. A part of the bootres.cbk file is given below, showing the shorthand name, the number of columns for each piece of information, and what the information describes. Obviously, the CBK file duplicates the information we already found in the OUT file.

```
            20
             1
METHOD       2 ESTIMATION METHOD (LS,GLS,ML,ELS,EGLS,ERLS,AGLS,HKGLS,HKRLS)
CONDCODE     2 CONDITION CODE (0 FOR NORMAL CONDITION)
CONVRGNC     2 CONVERGENCE (0 FOR MODEL CONVERGED)
ITER         2 NUMBER OF ITERATIONS FOR CONVERGENCE
DF           2 DEGREES OF FREEDOM
CNSTRANT     2 NUMBER OF CONSTRAINTS
DFDENOM      2 DENOMINATOR DEGREES OF FREEDOM FOR F-TESTS
DFSTRMEN     2 DEGREES OF FREEDOM FOR POTENTIAL STRUCTURED MEANS MODEL TEST
DFGLSMEN     2 D.F. FOR GLS TEST OF HOMOGENEITY OF MEANS
DFGLSCOV     2 D.F. FOR GLS TEST OF HOMOGENEITY OF COVARIANCE MATRICES
DFGLSCMB     2 D.F. FOR GLS COMBINED TEST OF HOMOGENEITY OF MEANS/COVAS.
PVAL         3 TAIL PROBABILITY FOR MODEL CHI-SQUARE
TPRESTST     3 TAIL PROBABILITY FOR RESIDUAL-BASED TEST STATISTIC
TPYBRTST     3 TAIL PROBABILITY FOR YUAN-BENTLER RESIDUAL-BASED TEST STAT.
TPYBAGLS     3 TAIL PROBABILITY FOR YUAN-BENTLER AGLS F-STATISTIC
TPYBRESF     3 TAIL PROBABILITY FOR YUAN-BENTLER RESIDUAL-BASED F-STATISTIC
SBPVAL       3 TAIL PROBABILITY FOR SATORRA-BENTLER SCALED CHI-SQUARE
YBPVAL       3 TAIL PROB. FOR YUAN-BENTLER SCALED CHI-SQUARE (CASE ROBUST)
TPSCALYB     3 TAIL PROBABILITY FOR SCALED CHI-SQUARE (YUAN-BENTLER)
TPGLSMEN     3 TAIL PROBABILITY FOR GLS TEST OF HOMOGENEITY OF MEANS
TPGLSCOV     3 TAIL PROB. FOR GLS TEST OF HOMOGENEITY OF COV. MATRICES
TPGLSTOT     4 TAIL PROB. FOR GLS COMBINED TEST OF HOMOGENEITY
TPSTRMEN     4 TAIL PROBABILITY FOR POTENTIAL STRUCTURED MEANS MODEL
TPYBCORA     4 TAIL PROB. FOR YUAN-BENTLER CORRECTED AGLS TEST STATISTIC
TPYBRBTS     4 TAIL PROB. FOR YUAN-BENTLER RESIDUAL-BASED ADF STATISTIC
TPYBCORT     4 TAIL PROB. FOR YUAN-BENTLER CORRECTED RESID.-BASED ADF STAT.
TPYBRBAF     4 TAIL PROB. FOR YUAN-BENTLER RESIDUAL-BASED ADF F-STATISTIC
NULCHI       5 INDEPENDENCE MODEL CHI-SQUARE
CHI          5 MODEL CHI-SQUARE
GFI          5 LISREL    GFI  FIT INDEX
AGFI         5 LISREL   AGFI  FIT INDEX
----
PA1         11 PARAMETER ESTIMATES
PA2         11 PARAMETER ESTIMATES
PA3         11 PARAMETER ESTIMATES
PA4         11 PARAMETER ESTIMATES
----
ST1         14 STANDARD ERRORS
ST2         14 STANDARD ERRORS
ST3         14 STANDARD ERRORS
```

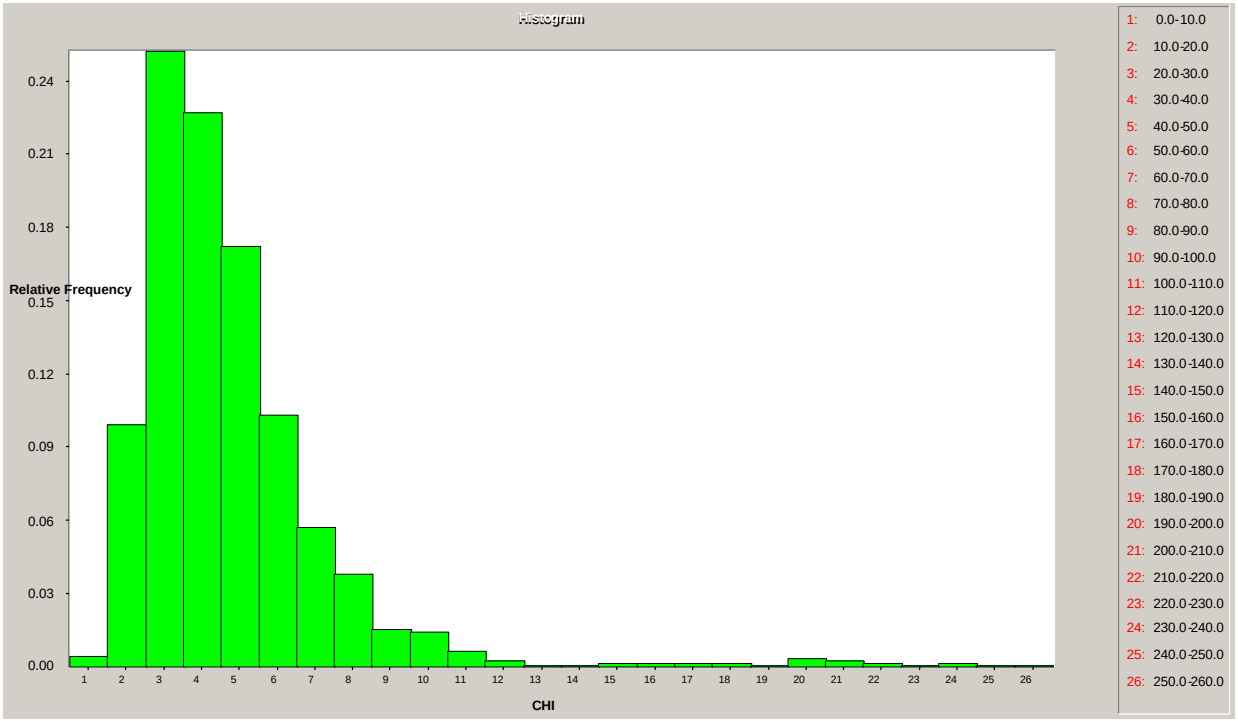
The shorthand name in the left is used as labels for results in the ETS file. To make sense of the specific parameters and standard errors being described sequentially, you should go to the OUT file. It provides a list of these, such as

```
PARAMETERS TO BE PRINTED ARE:

   F2,F1     F3,F1     F3,F2     E1,E1     E2,E2     E3,E3     E4,E4     E5,E5
   E6,E6     E7,E7     E8,E8     E9,E9     V1,F1     V2,F1     V3,F1     V4,F2
   V5,F2     V6,F2     V7,F3     V8,F3     V9,F3
```

To give you the flavor of the distribution of the CHI statistics as seen in the bootres.ets file, we highlighted the CHI column of the bootres.ets file with our pointer, and then copied this column into a word processor where we converted the column to a table. The first part of this table is given next. In this table, each cell gives the T_{ML} value of a particular replication. We show the result for only 42 of the 1000 replications.

37.7770	32.0350	16.9600	31.7440	40.6550	37.8900	55.0130
43.4320	28.9610	31.2450	40.9100	54.9400	21.1110	34.2420
66.5390	80.9500	72.2340	15.9320	29.8610	34.7770	28.3030
51.1900	26.6160	22.1940	42.1390	33.0730	26.0710	21.2940
63.3890	33.1880	54.4190	44.0890	70.4130	36.0470	32.0190
57.9380	36.5910	45.0700	73.7150	28.5910	51.5840	22.6160

Clearly there is a lot of variability. To see this more systematically, we take the CHI column of data in the ETS file and apply a standard plotting tool of EQS. Specifically, we obtain the following histogram of the T_{ML} values. The legend on the right is difficult to read as compressed here, so we mainly note that our original test statistic of 191.4 is at the extreme right of the distribution. Our sample data is quite unlikely to be observed, given that the model is true. This may be reason to reject the model.



The ETS file provides a lot of detailed output. You may need to study it in various ways to truly understand the results of your bootstrap study. Once imported into EQS as an ESS file, it is just a data file that can be studied with the many tools available in EQS.

Standard Errors for Parameters in Standardized Solutions

It is hard to obtain standard errors for the parameter estimates in a standardized solution, because these are not the parameters of the model being estimated. They are transformations of the unstandardized parameters. It is for this reason that most SEM programs do not compute such standard errors. However, it is possible to calculate these standard errors by considering the standardized parameters as functions of the model parameters, and using asymptotic approximations to transform the asymptotic covariance matrix of parameter estimates into that of the standardized parameter estimates. This was done by Jamshidian and Bentler (2000). Their formulae will be implemented soon. In the meantime, you can obtain standard errors for standardized parameters in EQS using the bootstrap. Of course, as always, these are standard error estimates.

The standardized solution is computed in each bootstrap replication, and the standardized parameter estimates are saved. Across the B bootstrap replications, there are thus B estimates of a given standardized parameter, and the standard deviation of that parameter estimate can be taken as the empirical estimate of the standard error of the standardized coefficient. To implement this feature, we add the following line to the OUTPUT section of the bootstrap run:

```
SS;
```

Clearly this can be implemented with the regular bootstrap or with the model-based bootstrap. Here we add this statement to the illustrative model-based bootstrap, and get the results that are partially duplicated below. These are the results printed in the OUT file from the 85% of the runs that were successes. Note that the output includes fixed parameters as well as free parameters estimated by EQS. Remember that some parameters that are free in the unstandardized solution may be fixed in the standardized solution since, e.g., free error variances are transformed into standardized beta weights. As for parameter estimates in the examples above, the mean (standardized) parameter estimates are printed first, followed by their standard deviations. These standard deviations are the empirical bootstrap standard error estimates for the parameters of the standardized solution. So, for example, the standardized factor loading V1,F1 has a bootstrap standard error estimate of .0999.

```
STANDARDIZED SOLUTION

   PARAMETER         F1,F1      F2,F1      F2,F2      F3,F1      F3,F2
   MEAN               1.0000     0.5002     1.0000     0.7350     0.3604
   STANDARD DEV.      0.0000     0.1041     0.0000     0.1007     0.0969
   SKEWNESS (G1)      0.0000     0.0413     0.0000    -0.0868    -0.1455
   KURTOSIS (G2)     -3.0000     0.2946    -3.0000    -0.0941    -0.0451
   LOWER  5%          1.0000     0.3251     1.0000     0.5671     0.2031
   UPPER  5%          1.0000     0.6584     1.0000     0.9050     0.5173

   PARAMETER         F3,F3      E1,E1      E2,E2      E3,E3      E4,E4
   MEAN               1.0000     1.0000     1.0000     1.0000     1.0000
   STANDARD DEV.      0.0000     0.0000     0.0000     0.0000     0.0000
   SKEWNESS (G1)      0.0000     0.0000     0.0000     0.0000     0.0000
   KURTOSIS (G2)     -3.0000    -3.0000    -3.0000    -3.0000    -3.0000
   LOWER  5%          1.0000     1.0000     1.0000     1.0000     1.0000
   UPPER  5%          1.0000     1.0000     1.0000     1.0000     1.0000
   ---

   PARAMETER         V1,F1      V1,E1      V2,F1      V2,E2      V3,F1
   MEAN               0.7010     0.6980     0.3889     0.9167     0.5423
   STANDARD DEV.      0.0999     0.1073     0.0838     0.0363     0.0934
   SKEWNESS (G1)     -0.1232    -1.0266    -0.0446    -0.7641    -0.1545
   KURTOSIS (G2)      0.1170     2.3000     0.1751     0.8675    -0.2652
   LOWER  5%          0.5293     0.4973     0.2532     0.8486     0.3888
   UPPER  5%          0.8676     0.8484     0.5290     0.9674     0.6898

   PARAMETER         V3,E3      V4,F2      V4,E4      V5,F2      V5,E5
   MEAN               0.8327     0.9091     0.3980     0.8168     0.5712
   STANDARD DEV.      0.0614     0.0479     0.1130     0.0454     0.0675
   SKEWNESS (G1)     -0.4734    -0.4796    -0.4396     0.0698    -0.6928
   KURTOSIS (G2)     -0.0149    -0.0153     0.0532    -0.0879     1.2983
   LOWER  5%          0.7240     0.8213     0.1887     0.7409     0.4524
   UPPER  5%          0.9213     0.9820     0.5705     0.8918     0.6716

   PARAMETER         V6,F2      V6,E6      V7,F3      V7,E7      V8,F3
   MEAN               0.6536     0.7530     0.4929     0.8655     0.4662
   STANDARD DEV.      0.0569     0.0500     0.0775     0.0439     0.0758
   SKEWNESS (G1)     -0.0535    -0.4107    -0.2556    -0.3436     0.0754
   KURTOSIS (G2)     -0.0622     0.3182     0.0448    -0.0045     0.1758
   LOWER  5%          0.5609     0.6642     0.3605     0.7884     0.3431
   UPPER  5%          0.7476     0.8279     0.6152     0.9328     0.5918

   PARAMETER         V8,E8      V9,F3      V9,E9
   MEAN               0.8804     0.8782     0.4569
   STANDARD DEV.      0.0414     0.0617     0.1270
   SKEWNESS (G1)     -0.7992    -0.2261    -0.6595
   KURTOSIS (G2)      1.4776    -0.4251     0.1843
   LOWER  5%          0.8061     0.7749     0.2089
   UPPER  5%          0.9393     0.9779     0.6321
```

The Jackknife

We saw above that in the bootstrap, samples of size N are drawn with replacement from the data. In contrast, in the jackknife, samples of size N-1 (or a smaller number) are drawn without replacement. The jackknife was developed before the bootstrap (e.g., Miller, 1974), but it has largely fallen out of favor because in many research contexts the bootstrap provides estimators with less bias and variance. It tends to work best with statistics that are smooth and not too nonlinear functions of the data. But it is a conceptually important procedure that works especially well for certain problems. In particular, it is an effective way to find the contribution of a problematic case to various statistical results. For example, one can measure the contribution of a case to the misfit of a model by comparing the T_{ML} statistic for the entire sample with the various T_{ML} statistics from the leave-one-out samples. If leaving out a particular case results in a huge improvement in model fit, the case is very influential. It is an outlier in the space of model fit. More generally, the ETS file provides information on the contribution of particular cases to standard errors and other statistics of interest.

The jackknife is a systematic sampling method. In one jackknife sample, a given observation is omitted, and the rest of the data is used to run the analysis and compute the statistics of interest. A similar procedure is repeated for the next sample, and typically there are as many samples as there are cases. Specifically, in the first sample, EQS will skip the first observation and use the rest of the data to run a given model. In the second sample, it will skip the second observation, and so on, until the given number of replications provided by user is reached. This is illustrated with the previous holza.ess data file and the corresponding 3 factor model. The key modification is as follows:

```
/SIMULATION
  JACKKNIFE;                      !THE JACKKNIFE RESAMPLING METHOD IS USED.
  REPLICATIONS = 145;             !A TOTAL OF 145 REPLICATIONS ARE DONE.
/OUTPUT
  DATA='JACKRES.ETS';             !OUTPUT PARAMETER ESTIMATES AND STANDARD
  PARAMETER ESTIMATES;            !ERRORS OBTAINED FROM THE JACKKNIFE
  STANDARD ERRORS;                !RESAMPLING METHOD ARE STORED IN JACKRES.ETS FILE
/END
```

As there are 145 cases in the data file, 145 replications are specified. The OUT file from the above run includes the usual summary statistics, the ETS file provides details on the iterative results, and the corresponding CBK file provides the codebook for the ETS file. We do not provide the details, but we illustrate some findings.

It will be remembered that in the original sample, T_{ML} = 191.4 based on 24 df. Studying the column in the jackres.ets file that corresponds to this statistic, we find that most runs give only marginally smaller test statistics. The mean T_{ML} equals 190.3, that is, on average deleting a case does not especially improve or worsen model fit. However, three runs do stand out. These are replications 63, 106, and 142. These have deleted, in turn, case 63, or 106, or 142 from the analysis. In these runs, T_{ML} equals 164.2, 162.2, and 160.1 respectively, i.e., substantially better than the T_{ML} in the original model fit and its jackknife mean of 190.3. Evidently, each of these cases is very influential to the misfit of the model. Probing a bit further, we evaluate the case contributions to kurtosis in the original sample, finding:

```
CASE NUMBERS WITH LARGEST CONTRIBUTION TO NORMALIZED MULTIVARIATE KURTOSIS:
-------------------------------------------------------------------------
CASE NUMBER        24          63          98          106         142
ESTIMATE       308.2982    543.2404    1548.0941    808.2508    656.3097
```

Clearly a possible explanation for why cases 63, 106, and 142 are influential to the χ^2 test is that they are outliers in the data space. They are among the 5 cases with the largest contribution to Mardia's normalized coefficient. However, not all outliers make the fit of the model worse; cases 24 and 98, when deleted from their replication, yield a T_{ML} of 201.3 and 176.4 respectively. Finally, additional study of the ETS file shows additional problems caused by these and other cases. For example, some parameter estimates change radically when case 142 is omitted from the analysis. Consider the error variance E6,E6. Its mean across the jackknife samples is 168.5, with a 90% confidence interval of 167.4 to 170.1. But without case 142 in the analysis, this error variance is estimated at 132.9.

EQS does not implement formal diagnostic measures of influence based on case deletion, such as Cook's (1977) distance.

We have suggested that the jackknife plays a useful role in evaluating case contributions to misfit, parameter estimate variance, and so on. But it should be remembered that a "leave-one-out" methodology cannot predict what would happen if more than one case is deleted from a run due to the phenomenon of masking. For example, although case 63 and case 106 contribute to the misfit of the model, it is not known whether simultaneous deletion of cases {63, 106} would do more damage than either considered alone. To evaluate this would require a "leave-two-out" or more generally "leave-m-out" methodology that is not currently implemented in EQS. To study this issue further, you can delete a suspicious case from the analysis using the DEL command, and then rerun the jackknife on the remaining file. This process may need to be repeated again and again. Recently developed approaches to identification of influential observations, such as Poon and Poon's (2002) methods based on the so-called local influence approach of Cook (1986), remain to be implemented in EQS. In the meantime, you should remember that if you suspect influential cases in your analysis, you can do case-robust estimation in EQS. You can also verify any problems with the case-robust bootstrap of Yuan and Hayashi (2003).

Note on DEL with bootstrap or jackknife. When EQS is instructed to ignore cases in the data file with the DEL command, a reduced file excluding those cases is used for resampling. During the resampling, the printout may refer to cases by case number (e.g., in the CASE CONTRIBUTION statement); those case numbers refer to the case number in the original data file. To avoid any interpretive confusion, it may be better to create a new file that excludes the cases to be ignored, and do resampling from that file.

We now turn to some additional approaches and issues associated with the /SIMULATION section of EQS.

Simulation in Multisample Analyses

Simulation can be performed on multisample models. It does so by generating data in accord with the MODEL or MATRIX specification as given in the first group, and doing the analysis in each of the simulation replications in accord with the usual principles of multisample models. We demonstrate this by adding a SIMULATION section to an example from Chapter 7.

```
/TITLE
2 GROUP EXAMPLE FROM WERTS ET AL 1976 - GROUP 1
1 FACTOR MODEL WITH UNEQUAL FACTOR CORRELATIONS
/SPECIFICATIONS
CASES = 865; VARIABLES = 4; GROUPS = 2;
/MODEL
 (V1,V2) ON F1 = 5*;
 (V3,V4) ON F2 = 5*;
 VAR F1,F2; VAR E1-E4 = 50*;
 COV FF = .5*;
/MATRIX
63.382
70.984 110.237
41.710  52.747 60.584
30.218  37.489 36.392 32.295
/SIMULATION
 POPULATION=MODEL; REPLICATIONS=150; SEED=111111111;
/OUTPUT
 DATA='SIMGROUP.ETS'; PARAMETERS;
/END
/TITLE
2 GROUP EXAMPLE FROM WERTS ET AL 1976  - GROUP 2
/SPECIFICATIONS
CASES = 900; VARIABLES = 4;
```

```
/MODEL
 (V1,V2) ON F1 = 5*;
 (V3,V4) ON F2 = 5*;
 VAR F1,F2; VAR E1-E4 = 50*;
 COV FF = .5*;
/MATRIX
67.898
72.301 107.330
40.549  55.347 63.203
28.976  38.896 39.261 35.403
/CONSTRAINTS
(1,V1,F1)=(2,V1,F1);
(1,V2,F1)=(2,V2,F1);
(1,V3,F2)=(2,V3,F2);
(1,V4,F2)=(2,V4,F2);
(1,F2,F1)=(2,F2,F1);
/END
```

As you see, there is no SIMULATION section in the model for the second group. By default, simulation options from one group carry over to the next group. The following options in the SIMULATION section may appear in the first group only: REPLICATIONS, SEED, DATA_PREFIX, SAVE, and FORMAT.

If MBBOOTSTRAP is used, it must appear in the first group, and it will apply to all groups. Use the statement MBB only (rather than MBB=N), and let EQS determine the sample size. The number of bootstrapped cases is taken to be the sample size of the relevant group, which may differ from one group to the next.

The OUT file includes a summary of the numbered statistics for all groups combined, and parameter estimates for each group separately. In this example, the tail probability and fit indices indicate a good model fit.

```
 SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=ML

 SUCCESS (CONVERGENCE WITH NO CONDITION CODE) OCCURRED IN
   150 REPLICATIONS OUT OF  150 (100.00 PERCENT)

 NUMBERED STATISTICS ARE
   1. TAIL PROBABILITY FOR MODEL CHI-SQUARE
   2. INDEPENDENCE MODEL CHI-SQUARE
   3. MODEL CHI-SQUARE
   4. LISREL    GFI  FIT INDEX
   5. LISREL   AGFI  FIT INDEX
   6. BOLLEN   (IFI) FIT INDEX
   7. MCDONALD (MFI) FIT INDEX
   8. BENTLER-BONETT     NORMED FIT INDEX
   9. BENTLER-BONETT NON-NORMED FIT INDEX
  10. COMPARATIVE FIT INDEX (CFI)
  11. ROOT MEAN-SQUARE RESIDUAL (RMR)
  12. STANDARDIZED RMR
  13. ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)
  14. MINIMIZED MODEL FUNCTION VALUE

 SUMMARY STATISTICS FOR ALL REPLICATIONS
 ---------------------------------------

   STATISTIC            1         2         3         4         5
   MEAN              0.8954   547.1627    2.0354    0.9994    0.9984
   STANDARD DEV.     0.2133    61.8964    2.7091    0.0008    0.0022
   SKEWNESS (G1)    -2.5772     0.0887    2.2469   -2.2356   -2.2354
   KURTOSIS (G2)     6.0068    -0.6536    5.1827    5.1257    5.1229
   LOWER  5%         0.3205   447.0750    0.0938    0.9977    0.9934
   UPPER  5%         1.0000   647.3700    8.1586    1.0000    0.9999
```

```
  STATISTIC            6         7         8         9        10
  MEAN              1.0094    1.0014    0.9963    1.0163    0.9996
  STANDARD DEV.     0.0051    0.0008    0.0047    0.0088    0.0017
  SKEWNESS (G1)    -1.8796   -2.2380   -2.0615   -1.8749   -4.5732
  KURTOSIS (G2)     3.8127    5.1768    4.1909    3.7998   20.3190
  LOWER  5%         0.9978    0.9997    0.9862    0.9962    0.9978
  UPPER  5%         1.0147    1.0020    0.9998    1.0255    1.0000

  STATISTIC           11        12        13        14
  MEAN              0.4961    0.0066    0.0010    0.0012
  STANDARD DEV.     0.2890    0.0038    0.0041    0.0015
  SKEWNESS (G1)     1.0410    1.0221    4.1329    2.2469
  KURTOSIS (G2)     0.6308    0.5364   15.8968    5.1826
  LOWER  5%         0.1658    0.0022    0.0000    0.0001
  UPPER  5%         1.1073    0.0147    0.0095    0.0046


PARAMETER ESTIMATES FOR GROUP  1

  PARAMETER        F2,F1     E1,E1     E2,E2     E3,E3     E4,E4
  MEAN              0.5003   49.3061   49.3006   49.2350   49.8631
  STANDARD DEV.     0.0632    5.9531    5.4238    5.6525    5.8747
  SKEWNESS (G1)    -0.0771   -0.2948   -0.0455   -0.1853   -1.0707
  KURTOSIS (G2)    -0.0824    0.0506   -0.0189    0.1934    3.2048
  LOWER  5%         0.4028   40.3175   41.3712   39.7943   40.1459
  UPPER  5%         0.5906   58.3668   58.7116   58.4242   58.1000

  PARAMETER        V1,F1     V2,F1     V3,F2     V4,F2
  MEAN              5.0206    5.0731    5.0432    5.0009
  STANDARD DEV.     0.5611    0.5343    0.5863    0.5780
  SKEWNESS (G1)     0.5569    0.0981    0.2229    0.5874
  KURTOSIS (G2)     0.1433   -0.3765    0.5483    1.1252
  LOWER  5%         4.2060    4.1259    4.1683    4.1302
  UPPER  5%         5.9574    5.9759    6.1356    6.0274


PARAMETER ESTIMATES FOR GROUP  2

  PARAMETER        F2,F1     E1,E1     E2,E2     E3,E3     E4,E4
  MEAN              0.5003   49.4202   49.3158   49.2663   49.8776
  STANDARD DEV.     0.0632    5.8850    5.2863    5.5528    5.8654
  SKEWNESS (G1)    -0.0771   -0.3023   -0.0899   -0.1823   -1.0646
  KURTOSIS (G2)    -0.0824    0.0417   -0.0160    0.1624    3.2943
  LOWER  5%         0.4028   40.0508   41.3006   39.6479   39.8693
  UPPER  5%         0.5906   58.5761   58.5560   58.4773   57.9840

  PARAMETER        V1,F1     V2,F1     V3,F2     V4,F2
  MEAN              5.0206    5.0731    5.0432    5.0009
  STANDARD DEV.     0.5611    0.5343    0.5863    0.5780
  SKEWNESS (G1)     0.5569    0.0981    0.2229    0.5874
  KURTOSIS (G2)     0.1433   -0.3765    0.5483    1.1252
  LOWER  5%         4.2060    4.1259    4.1683    4.1302
  UPPER  5%         5.9574    5.9759    6.1356    6.0274
```

It is comforting to see that parameters that are constrained, such as (1,V1,F1)=(2,V1,F1), have identical simulation results, i.e., their means, SDs, etc. are identical.

Simulation with Looping

EQS can perform the same analysis on equal-sized subsets of a data set, via the LOOP option in the SPECIFICATION section. This is done with a command such as LOOP=2; which tells EQS to do two runs of whatever model is specified on mutually exclusive subsets of a single data file. In essence the single data file gets divided into groups, and a multiple group analysis is done with the identical model and without any constraints across the groups.

This is implemented as follows. EQS does one run with the specified number of cases, then moves down the file to do a run on the next number of cases, and so on. As a result, LOOP has to be coordinated with the number of cases in the data file such that #loops × #cases ≤ actual size of sample in the data file. So, for example, if there are 50 cases total in the data file, and you want to do LOOP=2, you must set CASES=25 or less. If you were to set CASES=50 when LOOP=2, EQS could do its first run, but would find no data for the second run and would terminate.

This same feature is enabled with simulation, and we illustrate it with the bootstrap. We use the data set poon.ess, which is distributed with the EQS for Windows program. It contains the scores of 200 subjects on eight variables. With LOOP=4, we divide the 200 cases into four subsets of 50 cases each. In each 50-case sample, we run the specified 2 factor model (ignoring that some of the variables are categorical), and do the ordinary bootstrap with 75 replications. The results from all 75× 4=300 replications are aggregated sequentially into the loop.ets file, and are summarized overall.

```
/TITLE
 LOOPING EXAMPLE
/SPECIFICATIONS
 DATA='POON.ESS'; VARIABLES=8; CASES=50; LOOP=4;
 METHODS=GLS;     MATRIX=RAW;  ANALYSIS=ZSCORE;
/EQUATIONS
V1 =  + *F1  + E1;
V2 =  + *F1  + E2;
V3 =  + *F1  + E3;
V4 =  + *F1  + E4;
V5 =  + *F2  + E5;
V6 =  + *F2  + E6;
V7 =  + *F2  + E7;
V8 =  + *F2  + E8;
/VARIANCES
F1-F2 = 1;
E1-E8 = *;
/COVARIANCES
F2 , F1 = *;
/SIMULATION
 REPLICATIONS=75; BOOTSTRAP=50; SEED=987654321;
/OUTPUT
 DATA='LOOP.ETS';
/TECHNICAL
 ITERATIONS=100;
/END
```

As usual the OUT file provides a summary. A small part is duplicated below. We show the results for the 247 successful runs out of the 300 replications that were done.

```
 SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=GLS

 SUCCESS (CONVERGENCE WITH NO CONDITION CODE) OCCURRED IN
   247 REPLICATIONS OUT OF  300 ( 82.33 PERCENT)

 NUMBERED STATISTICS ARE
   1. TAIL PROBABILITY FOR MODEL CHI-SQUARE
   2. INDEPENDENCE MODEL CHI-SQUARE
   3. MODEL CHI-SQUARE
   4. LISREL    GFI  FIT INDEX
   5. LISREL   AGFI  FIT INDEX
   6. BOLLEN   (IFI) FIT INDEX
   7. MCDONALD (MFI) FIT INDEX
   8. BENTLER-BONETT     NORMED FIT INDEX
   9. BENTLER-BONETT NON-NORMED FIT INDEX
  10. COMPARATIVE FIT INDEX (CFI)
  11. ROOT MEAN-SQUARE RESIDUAL (RMR)
  12. STANDARDIZED RMR
  13. ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)
  14. CONFIDENCE INTERVAL FOR RMSEA (LOWER BOUND)
  15. CONFIDENCE INTERVAL FOR RMSEA (UPPER BOUND)
```

```
 16. CRONBACH'S ALPHA
 17. GREATEST LOWER BOUND RELIABILITY
 18. BENTLER'S DIMENSION-FREE LOWER BOUND RELIABILITY
 19. RELIABILITY COEFFICIENT RHO
 20. MINIMIZED MODEL FUNCTION VALUE

SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=GLS          FOR   247 SUCCESSES

  STATISTIC            1         2         3         4         5
  MEAN            0.1341   61.6989   30.1101    0.8464    0.7089
  STANDARD DEV.   0.1703    8.7654    8.2317    0.0420    0.0796
  SKEWNESS (G1)   1.8309    0.6606    0.4973   -0.4973   -0.4973
  KURTOSIS (G2)   3.0777    1.0243   -0.1303   -0.1302   -0.1302
  LOWER  5%       0.0005   48.8200   18.2206    0.7643    0.5534
  UPPER  5%       0.5077   77.0885   46.1959    0.9070    0.8239
  ---

  STATISTIC           16        17        18        19        20
  MEAN            0.8790    0.9564    0.9561    0.9357    0.6145
  STANDARD DEV.   0.0342    0.0110    0.0110    0.0306    0.1680
  SKEWNESS (G1)  -0.9348   -0.6513   -0.6984   -5.2539    0.4973
  KURTOSIS (G2)   0.5255    1.0309    1.0544   43.8948   -0.1303
  LOWER  5%       0.8137    0.9379    0.9379    0.8924    0.3719
  UPPER  5%       0.9224    0.9723    0.9717    0.9618    0.9428
```

Generating Missing Data

You can force some of the generated data to be missing, by specifying MISSING = #; in the SIMULATION section. The number # that you provide specifies the proportion of missing data, e.g. MISSING = .05; will force 5% of the generated data to be missing. In association with the MISSING command, you can also specify the variables that should contain the missing data. You do this by adding VARIABLE = xx,…,zz; (or VAR=), where xx … zz gives a list of variables (e.g., V1, V6, V9) on which you want to have missing values generated. The default (which does not need to be specified) is VAR=ALL; i.e., all dependent and independent V's in the model, excluding V999, will be eligible to contain missing scores.

The mechanism used to generate the missing data is "completely random," that is, it would qualify as a special case of MCAR. If you set VAR=ALL, you will get a huge number of patterns of missing data, most with one to very few observations unless the sample size is really huge. Thus, this option may not represent real data very well. It seems likely that real data almost never will contain such a huge set of patterns of missing scores.

The number of missing values in a simulated data set will be the nearest integer to NV × NC × #, where NV is the number of variables that you specify to have missing data, NC is the number of cases, and # is the proportion of missing data. So if your data file contains 100 cases and 10 variables, and you request 5% missing data, you will get 50 missing scores. If you set VAR=ALL, these 50 scores will be distributed across all 10 variables (and up to 50 cases), generating quite a large number of missing data patterns. If you set VAR=V1, on the other hand, you will have complete data for V2-V10, but five (=1 × 100 × .05) of the 100 cases will have missing scores on V1. [To get the same number 50 of missing scores on only V1, you would need to set # = .5.] Thus there are only two patterns of data: complete, or missing V1 but complete otherwise.

MISSING = # sets the default to MISSING=ML in the SPECIFICATION section. When you run ML, this will be case-wise ML and you should remember that the saturated model also contains mean parameters (one *V999 for each measured variable). Thus your model structure also should account for the variable means, and you should use some kind of a mean structure model. That is, you should have some *V999 parameters (their nature depending on the model) in the equations, and the statement ANALYSIS = MOMENT to tell EQS to estimate these means. If you like, you may override the use of ML, and use MISSING = COMPLETE or PAIR instead. The missing value character (see the SPECIFICATION section) is used when missing values are written onto a file created by the simulation.

In the following example, we simulate a simple model for 4 variables using transformations to achieve an extremely kurtotic distribution of observed data (Mardia's normalized kurtosis going over 100 in some replications). This model is run under the null hypothesis, that is, the model in fact fits the data means and covariance matrix used to generate the data. We also study the distribution of R^2 coefficients for the various equations.

```
/TITLE
 MISSING DATA SIMULATION
/SPECIFICATIONS
 VARIABLES=4; CASES=200; METHODS=ML,ROBUST; MATRIX=COVARIANCE;
 MISSING=ML; ANALYSIS=MOMENT;
/EQUATIONS
V1 =  +  1.0F1  + E1;
V2 =  + 0.6*F1  + E2;
V3 =  + 0.6*F1  + E3;
V4 =  + 0.6*F1  + E4;
F1 =  0.3*V999  + D1;
/VARIANCE
E1-E4 = 0.64*;
D1=1.0*;
/MATRIX
1
0.36 1
0.36 0.36 1
0.36 0.36 0.36 1
/MEANS
0.3 0.3 0.3 0.3
/TECHNICAL
 ITERATIONS=100;
/SIMULATION
 POPULATION = MATRIX;
 REPLICATIONS= 100;                          !IN REAL STUDY, MIGHT USE 1000 INSTEAD
 SEED = 99135473;
 TRANSFORMATION = V1: S3.0, K15.0 &          !VERY HEAVY SKEW AND KURTOSIS
                  V2: S3.0, K15.0 &
                  V3: S3.0, K15.0 &
                  V4: S3.0, K15.0 ;
 MISSING=.025;                                 !PROPORTION OF MISSING DATA ON V1 AND V2
 VARIABLES = V1,V2;                            !LIMITS THE NUMBER OF MISSING DATA PATTERNS
/OUTPUT
    DATA='MISSING1.ETS'; RSQUARE;
/END
```

A portion of the OUT file is displayed below.

```
SUMMARY STATISTICS OF REPLICATIONS FOR METHOD=ML   +ROBUST

 SUCCESS (CONVERGENCE WITH NO CONDITION CODE) OCCURRED IN
   100 REPLICATIONS OUT OF  100 (100.00 PERCENT)

 NUMBERED STATISTICS ARE
   1. TAIL PROBABILITY FOR MODEL CHI-SQUARE
   2. TAIL PROBABILITY FOR RESIDUAL-BASED TEST STATISTIC
   3. TAIL PROBABILITY FOR YUAN-BENTLER RESIDUAL-BASED TEST STAT.
   4. TAIL PROBABILITY FOR YUAN-BENTLER RESIDUAL-BASED F-STATISTIC
   5. TAIL PROBABILITY FOR SCALED CHI-SQUARE (YUAN-BENTLER)
   ---
```

Because the data is heavily nonnormal, we might look at the various tail probabilities observed, that is, statistics #1-#5. The ML statistic #1 slightly over rejects models, with the empirical 5% rejection rate being at the .02 level. The Yuan-Bentler scaled chi-square #5 slightly under rejects models, with the empirical 5% rejection rate being at the .07 level. The various residual-based tests under reject to an even greater extent. The ETS file gives more details.

```
 SUMMARY STATISTICS FOR ALL REPLICATIONS
 ---------------------------------------
   STATISTIC            1         2         3         4         5
   MEAN              .5388     .5412     .5517     .5523     .6009
   STANDARD DEV.     .3171     .2542     .2486     .2503     .2851
```

```
  SKEWNESS (G1)      -.2448     -.1630     -.1779     -.1866     -.4124
  KURTOSIS (G2)     -1.2993    -1.1564    -1.1355    -1.1361     -.9930
  LOWER  5%           .0209      .1356      .1530      .1497      .0702
  UPPER  5%           .9529      .8991      .9007      .9024      .9650
```

The output file also shows that the number of missing patterns in the data is 3.1 on average. The ETS file shows 90% with 3 patterns, and 10% with 4 patterns. If you reran the analysis with the statement VARIABLES = ALL; the corresponding run would have obtained 5.6 missing data patterns on average, with 98% being 5, 6, or 7 patterns. The comparison would lead to greater differences as the number of variables in the model is increased.

```
NUMBER OF MISSING PATTERNS
  MEAN               3.1000
  STANDARD DEV.       .3015
  SKEWNESS (G1)      2.6667
  KURTOSIS (G2)      5.1111
  LOWER  5%          3.0000
  UPPER  5%          4.0000
```

There are five equations in the model, and each generates an R^2 coefficient for the proportion of variance explained in the relevant dependent variable. A summary of these coefficients across the 100 replications is as follows.

```
R-SQUARES
  DEP. VAR.            V1         V2         V3         V4         F1
  MEAN                .3761      .3807      .3631      .3574      .0000
  STANDARD DEV.       .0995      .1146      .0984      .0959      .0000
  SKEWNESS (G1)       .2322      .4733      .2497      .2682      .0000
  KURTOSIS (G2)      -.5261     -.0947     -.5342     -.4773    -3.0000
  LOWER  5%           .2385      .2063      .2092      .2025      .0000
  UPPER  5%           .5655      .5854      .5239      .5256      .0000
```

The population R^2 for each of these equations is .36. The sample R^2s are remarkably close to the population value, all about .36-.38 on average. The sample R^2 for F1 equals 0.0 in every replication, the same as in the population, as there are no predictors of individual differences. The standard deviation for each R^2 could be taken as an estimate of its standard error, while the lower and upper 5% values provide estimates of a 90% confidence interval.

Saving Generated Data

Data that is generated by a simulation can be saved for future use by EQS or other programs. As indicated the description of options in Chapter 3, a save command can be added to the /SIM section as follows.

```
SAVE=CONCATENATE; DATA_PREFIX='FAC'; FORMAT=(6F4.1);
```

The method of saving is to concatenate, i.e., put the various generated data end to end, and the data prefix statement has the effect that EQS will write the file FAC.DAT. If there are 400 cases in a single replication, and 100 replications, the file will have 40000 records, each containing one case generated by the simulation. The 400 cases of the first replication are followed by all 400 for the second, etc. An illustrative 1st ten records of FAC.DAT are given below. Note that numbers are separated by spaces, so that free format can be used to read the data into EQS for Windows or other programs.

```
 4.0 3.0 3.0 4.0 3.0 3.0
 4.0 3.0 5.0 4.0 4.0 5.0
 2.0 3.0 4.0 4.0 4.0 3.0
 5.0 3.0 4.0 4.0 4.0 3.0
 2.0 3.0 5.0 4.0 5.0 4.0
 4.0 4.0 4.0 4.0 3.0 3.0
 3.0 3.0 5.0 4.0 3.0 3.0
 4.0 4.0 1.0 4.0 5.0 6.0
 3.0 4.0 5.0 4.0 4.0 3.0
 3.0 5.0 4.0 4.0 4.0 4.0
```

Simulation with other EQS Options

We close this chapter by noting some additional features and limitations of the simulation option in EQS.

Simulation can be used with the CATEGORICAL option in the SPECIFICATION section, but only for jackknifing or bootstrapping (regular or model-based). For Monte Carlo simulations, data can be grouped by using the CATEGORIZATION option in the SIMULATION section.

Simulation can be used with case weighting. See the WT, WTN, WTE, and CROBUST options in the SPECIFICATION section.

Simulation cannot currently be used in multilevel analyses (MULTILEVEL = HLM or ML or MUML).

14. RELIABILITY AND PRACTICAL FIT

A lot can and should be said about evaluating a model, besides the basic issue of how well a model fares when measured against an overall chi-square goodness of fit test. In fact, in SEM a lot can go wrong, not only in model evaluation, but in all phases of the process. The review of applications by MacCallum and Austin (2000) is a good source for discussion of issues such as generalizabilty of findings, confirmation bias, the role of time, and a variety of model specification, design, and analysis issues that could interfere with gaining optimal knowledge from SEM research. This concluding chapter addresses issues associated with variable quality and model fit. First, by their very nature, latent variable models make use of indicators. The number and quality of indicators of some factors may be inadequate to permit easy detection of misspecifications related to those factors, and hence the measurement status of indicators is critical to draw reasonable conclusions about a model. For example, MacCallum, Widaman, et al. (1999) showed that the levels of communality of variables and overdetermination of factors were key features related to quality of modeling results. Communality is closely related to internal consistency reliability, the first topic of the chapter. The second topic is largely motivated by sample sizes that may be, in practice, "too large." That is, sample size may be so large that power is so high that any reasonable model may be statistically rejected. This has motivated the growth of an industry related to "practical" fit indices. A discussion of these issues – reliability and practical fit indices -- completes this **Manual.**

Reliability

A composite variable is a sum of other variables, for example, V4 = V1+V2+V3. Composites such as V4 are well known, for example, as the total score on a test made up of parts V1, V2, and V3. Another name for a composite score is a scale score, where, in testing or survey research, the scale score may be the sum of the variable or item scores comprising it. Internal consistency reliability coefficients are meant to describe the quality of a composite or scale in terms of its hypothetical common and unique constituents. Suppose for example, that V1 = C1+ U1; V2 = C2+ U2; and V3 = C3+ U3, and one hypothesizes that C1-C3 are uncorrelated with U1-U3, and that the various U's are also uncorrelated. Under this model, for each variable, say V1, $\sigma_{V1}^2 = \sigma_{C1}^2 + \sigma_{U1}^2$ and we may define the reliability of V1 as $\rho_{11}=\sigma_{C1}^2/\sigma_{V1}^2$, the ratio of common to total variance, or as $\rho_{11}=1 - \sigma_{U1}^2/\sigma_{V1}^2$. The smaller the unique variance relative to a variable's variance, the closer to 1.0 is the reliability. More reliable indicators lead to better modeling.

These equations would be in EQS language if we replaced "C" with "F" and "U" with "E", but a C variable may itself be a composite variable such as C1=.5F1 + .8F2, while, as is well known from factor analysis theory, a unique variable U is the sum of specificity and random error. But if C1=C2=C3=F1, a single factor, then this would be a standard EQS factor model since in practice it is difficult to distinguish U and E variables. Although EQS uses the E notation, in modeling practice E variables are really U variables.

Frequently, interest lies in the reliability of a composite such as V4. Under the model V4 = C1+C2+C3 + U1+U2+U3 = C4+U4. So $\rho_{44}=1 - \sigma_{U4}^2/\sigma_{V4}^2$. The reliability of this composite can be higher than the reliability of any of its constituents. For example, suppose that C1=C2=C3=F1 and the variances of U1-U3 are equal. Then the reliability of the *i*th item is the same as the reliability of the first item $\rho_{11} = 1 - \sigma_{U1}^2/(\sigma_{F1}^2+\sigma_{U1}^2)$. Further, with V4=3F1+U4, the variance of U4 is $3\sigma_{U1}^2$, $\sigma_{V4}^2 = (9\sigma_{F1}^2+3\sigma_{U1}^2)$, and hence the reliability of the composite is $\rho_{44} = 1 - 3\sigma_{U1}^2/(9\sigma_{F1}^2+3\sigma_{U1}^2) = 1 - \sigma_{U1}^2/(3\sigma_{F1}^2+\sigma_{U1}^2)$. Thus $\rho_{44} > \rho_{11}$. This is why composites or scales, or averages across components, often are used in practical work to describe subjects. To avoid a specific variable name like V4, let the composite or scale variable be called X, and its internal consistency reliability be ρ_{xx}. In this chapter, "reliability" will mean only internal consistency reliability. Other types of reliability, such as test-retest reliability, are not discussed, and further information on the general topic can be found in McDonald (1999), Nunnally and Bernstein (1994), and Thompson (2003). Lucke (2005) provides an excellent discussion of internal consistency and reliability.

As we shall see, internal consistency reliability may not be synonymous with "true reliability", say ρ^*_{xx}, which must be operationally defined in some way besides internal consistency.

We will define several different composite reliability indices based on a general structural model of the form $\Sigma = \Sigma_C + \Psi$, where Σ_C is the covariance matrix of the C (common) variables, and Ψ is the covariance matrix of the U (unique) variables, typically taken as diagonal. All are specializations of the following general formula (Bentler, 1968, eq. 12; Heise & Bohrnstedt, 1970, eq. 32)

$$\rho_{xx} = \left(1 - \frac{w'\Psi w}{w'\Sigma w}\right), \tag{14.1}$$

where w is a p-length column vector of weights. Here $w'\Psi w$ is a weighted sum of unique variances, and represents the unique variance of the sum. Similarly, $w'\Sigma w$ is a weighted sum of the elements of the model covariance matrix, and represents the variance of the weighted observed sum score. Here we use population defining formulae, but it is possible to substitute the sample covariance matrix S for the model covariance matrix in the above formula. The weights w_i are weights that might be used when computing a composite that gives some variables more influence than others, such as $X = w_1 X_1 + w_2 X_2 + \ldots + w_p X_p$. The weights may be known or unknown. They, along with Σ and Ψ, may need to be estimated by a SEM run, so that $\hat{\rho}_{xx}$ is obtained from $\hat{\Sigma}$, $\hat{\Psi}$, and $\hat{w}$. That is, estimates of population reliability are based on estimates of population parameters of a model, and possibly, a weighting vector. For simplicity, typically we do not use the "hat" when talking about estimates if the context is clear. In typical practice, interest focuses on the equally weighted composite $X = X_1 + X_2 + \ldots + X_p$, so that $w = 1$, a vector with elements $w_i = 1.0$. In the example above, we had $\rho_{44} = 1 - \sigma_{U4}^2/\sigma_{V4}^2$. This formula is a special case of ρ_{xx} with σ_{U4}^2 representing the sum of the diagonal elements of Ψ. Since the variance of a composite is simply the sum of all the elements in the covariance matrix of its parts, σ_{V4}^2, the total variance, is the sum of all the elements in Σ.

Technically speaking, ρ_{xx} is a lower bound to true reliability ρ^*_{xx} (Bentler, 1968, p. 339). Suppose that the unique covariance matrix Ψ is a sum of a random error covariance matrix Ψ_E and the covariance matrix Ψ_S of specific but reliable variables. Since $\Psi = \Psi_E + \Psi_S$, $w'\Psi w = w'\Psi_E w + w'\Psi_S w$, and hence some algebra shows

$$\rho_{xx} + (w'\Psi_S w)/(w'\Sigma w) = \rho^*_{xx},$$

where

$$\rho^*_{xx} = \left(1 - \frac{w'\Psi_E w}{w'\Sigma w}\right), \tag{14.2}$$

Generally, then, $\rho_{xx} \leq \rho^*_{xx}$, but more typically $\rho_{xx} < \rho^*_{xx}$. Equality ($\rho_{xx} = \rho^*_{xx}$) is achieved when unique and error variables are identical, i.e., the specificity covariance matrix Ψ_S is zero. See Bentler & Woodward (1980, p. 249).

Reliability estimates for scales or composites can be discussed without considering a structural model. For example, the well-known Guttman-Cronbach coefficient alpha for the reliability of a composite of p components $X = X_1 + X_2 + \ldots + X_p$ can be computed without reference to any model (although it is usually justified by a model). This has the advantage that α can be universally applied to any data set, and no doubt accounts for the fact it is the single most widely used reliability coefficient in practice (Hogan, Benjamin, & Brezinski, 2000). In the population, it is defined as

$$\alpha = \frac{p}{p-1}\left(1 - \frac{1'D1}{1'\Sigma 1}\right), \tag{14.3}$$

where D is the diagonal of the covariance matrix Σ, and 1 is column vector of unit elements, a summing vector. Thus $1'D1$ is the sum of the variances of the variables, and $1'\Sigma 1$ is the sum of all the elements of the covariance matrix. Typically, α is applied by substituting the sample covariance matrix S in place of Σ, that is, any structure to Σ is irrelevant. Nonetheless, as will be noted below, α is best rationalized when based on a very restricted one-factor

model that is quite unrealistic in real data; in a different approach, Lucke (2005) uses a confirmatory factor model in (14.3). Regardless, when $\Sigma = \Sigma_C + \Psi$ with Ψ being diagonal, $\alpha \leq \rho_{xx} \leq \rho^*_{xx}$, that is, alpha almost always underestimates internal consistency and true reliability (see e.g., Lord & Novick, 1968; Novick & Lewis, 1967). Hence, one motivation for model-based estimates of reliability is to get a better estimate of reliability. However, when correlated errors are present so that Ψ is not diagonal, α also may overestimate reliability -- though it will not do so if the factors that generate correlated errors are specified in Σ_C and not in Ψ. There are many good recent discussions of α, its problems, and its alternatives (e.g., Barchard & Hakstian, 1997; Becker, 2000ab; Bonett, 2003ab; Enders, 2003; Enders & Bandalos, 1999; Feldt & Charter, 2003; Green, 2003; Green & Hershberger, 2000; Hakstian & Barchard, 2000; Komaroff, 1997; Lucke, 2005; Miller, 1995; Raykov, 1997b, 1998b, 2001b; Raykov & Shrout, 2002; Schmidt, Le, & Ileus, 2003; Schmitt, 1996; Shevlin, Miles, Davies & Walker, 2000; Vautier & Jmel, 2003; and Zinbarg et al., 2005).

Although we are talking about lower bounds to reliability, you should know that all the lower-bound coefficients provided in EQS -- as with comparable coefficients obtained elsewhere -- are lower bounds to population reliability only if computed using the population covariance matrix Σ. This is almost never available. Hence, when computed using the sample covariance matrix *S* or a model-based estimate $\hat{\Sigma}$, a supposed lower-bound coefficient can exceed population reliability based on Σ. While there exists a statistical or probabilistic lower bound to population reliability (Woodward & Bentler, 1978), statistical lower bounds are not presently available and will be implemented in EQS in the future.

In this chapter, we assume that the covariance matrix Σ_C has some type of latent variable (LV) structure. The use of model structures for estimating reliability, and obtaining related statistical quantities such as standard errors, has a long history (e.g., Werts, Rock, Linn, & Jöreskog, 1978) and is becoming recognized as important (e.g., Kano & Azuma, 2003; Raykov, 1997a, 1998c, 2001c, 2002, 2004a; Raykov & Shrout, 2002; Steyer, Schmitt, & Eid, 1999; Vautier & Jmel, 2003; Yuan & Bentler, 2002a; Yuan, Guarnaccia, & Hayslip, 2003). In our approach to point estimation, there are four classes of lower bound reliability coefficients ρ_{xx} to be considered:

1. Composite reliability based on a one-factor model, with $\Sigma_C = \Lambda\Lambda'$ (Λ p×1) and w_i =1.0.
2. Maximal weighted reliability based on a one-factor model, with $\Sigma_C = \Lambda\Lambda'$ (Λ p×1) and *w* unknown but chosen to yield the largest ρ_{xx}.
3. Dimension-free composite reliability based on an unknown factor model with an arbitrary number of factors, with $\Sigma_C = \Lambda\Lambda'$ and w_i =1.0. In one variant of this approach, it is also assumed that $\Psi = KK'$, i.e., that the unique or error variances are nonnegative. In another, $\Sigma_C = \Lambda\Lambda'$, $\Psi = KK'$, and *w* is chosen optimally.
4. Composite reliability based on a given LV structural model, with Σ_C having any SEM structure with any number of latent variables and w_i =1.0.

The appropriate roles of these various reliability coefficients is discussed below. Thereafter, their use is illustrated. For all coefficients but #4, it is assumed that the unique or error variables are uncorrelated, that is, Ψ is diagonal. To obtain reliabilities #1 and #2, a one-factor model is estimated using standard EQS methods. To obtain the dimension-free reliabilities #3, computations are based on Jamshidian and Bentler (1998) or Bentler and Woodward (1983) and the optimal value is chosen. Thus, several factor models are computed, and for each model, its parameters generate $\hat{\Psi}$, the covariance matrix of unique variables, and $\hat{\Sigma}$, the model-based covariance matrix. Using these estimates, all of the coefficient types #1-#3 are computed and printed. Coefficient #4, a new coefficient that has not previously been studied, additionally requires the user to specify a model – any latent variable model with additive error variables. A special case of #4 is the confirmatory factor model (see e.g., Lucke, 2005). *When using model-based reliability estimates (especially #1 and #4), you should be sure that the model is consistent with the data.* If the model does not describe the data, model-based reliability is probably not meaningful. Reliability coefficients #3 always are based on a model that fits the data.

Composite Reliability Based On A One-Factor Model

When it is desired to estimate the reliability of $X = X_1 + X_2 + \ldots + X_p$, the unit-weighted sum of a set of components, the typical assumption underlying a composite reliability coefficient is that $\Sigma = \Lambda\Lambda' + \Psi$, where Λ is px1 and w_i =1.0 (e.g., Raykov, 1997a). We will call the 1-factor coefficient *rho*. Actually, a multiple group model that permits factor variances φ to differ across groups is set up in EQS as $\Sigma = F\varphi F' + \Psi$, with $\Lambda = F\varphi^{1/2}$. The elements needed for estimating ρ_{xx} are the sum of the estimated variances of the E terms in the model (to estimate *1'Ψ1*) and the sum of all elements in the model-reproduced covariance matrix (to estimate *1'Σ1*). The printout will indicate:

```
RELIABILITY COEFFICIENT RHO = .875
```

This method actually yields a class of reliability coefficients, as further constraints can be imposed on the model. If it is desired to compute reliability based on a restricted one-factor model, the relevant /CONSTRAINTS must be added to an *.eqs file setup. If you want to compute coefficient α based on a 1-factor model, this should be the model of compound symmetry. This can be set up with a one-factor model with the factor variance free, all factor loadings constrained to be equal, and all error variances set equal. In this case α computed on the model-reproduced covariance matrix $\hat{\Sigma}$ and *rho* will be equal. Note that *rho* for a constrained model will not be the same as *rho* for an unconstrained model.

Maximal Weighted Composite Reliability Based On A One-factor Model

In recent years, it has become popular to recommend computing so-called maximal reliability, the reliability of an optimally weighted composite (Drewes, 2000; Li, 1997; Li, Rosenthal, & Rubin, 1996; Raykov, 2004b). In this situation, the weighted composite is $X = w_1X_1 + w_2X_2 + \ldots + w_pX_p$. Different weights will lead to different composites, and hence the reliability coefficients of these various composites will differ. To get an optimal reliability coefficient – the maximal weighted reliability -- it is proposed that the weights be chosen so that the corresponding ρ_{xx} be as large as possible. This can be done. When the reliabilities of the components are known, i.e., the error variances are known, methods for computing maximal weighted reliability have a very long history (e.g., Thomson, 1940). In EQS, *maximal weighted reliability with known component error variances* can be obtained by using the /RELIABILITY command and setting /CONSTRAINTS to fix the error variances that are known at given values. Computations are based on Green (1950), as given by Bentler (1968, eq. 7). Note that, because EQS is more general than the original publications, some of these error variances also could be estimated.

When the reliabilities of the components are not known – the typical case – *maximal weighted reliability with unknown component error variances*, and the associated weights, can be computed from the solution to the so-called canonical factor model (Bentler, 1968, eqs. 13-15), which is a maximum likelihood solution. Bentler called this alpha-maximized factor analysis, and the coefficient alpha-*O*. "Maximal reliability," as the more recent references call Bentler's alpha-*O*, is a better name. To be more precise, we use the terminology "maximal weighted composite reliability," because there exists a maximal reliability for unweighted or equally weighted composites (see below). *Strictly speaking, maximal weighted reliability requires the use of ME=ML;* but EQS computes the maximal weighted composite reliability for any chosen estimation method. In principle, if Σ_C is multidimensional, additional sets of weights with associated composites and reliabilities also could be computed, but EQS computes only the maximal weighted reliability and the weights that generate it. Results are printed out as follows:

```
MAXIMAL WEIGHTED INTERNAL CONSISTENCY RELIABILITY = .765
MAXIMAL RELIABILITY CAN BE ACHIEVED BY WEIGHTING THE VARIABLES AS FOLLOWS:
 V1   V2   V3
.613 .592 .788
```

Again, a class of maximal weighted reliability coefficients can be obtained. Although maximal weighted reliability has been developed only for unrestricted one-factor solutions, the methodology has been extended in EQS to provide maximal weighted reliability under constraints on parameters as well. *When* /CONSTRAINTS *are part of the model setup with* /RELIABILITY*, the reported maximal weighted reliability coefficient is a* ***restricted*** *maximal reliability coefficient.* It generally will not equal the unrestricted maximal weighted reliability. Also note that if the model specified by the user is not a one-factor model, then 1-factor maximal weighted reliability is not computed or printed.

Note: If maximal weighted reliability is reported, care should be taken to clarify that this does not correspond to the reliability of a typical total score or scale. Such a scale score, in common practice, is a unit-weighted composite of a set of items or components. Maximal weighted reliability is the reliability of a differentially weighted composite, not the reliability of a standard composite. Comparing the maximal weighted reliability and unit-weighted composite reliabilities can give an idea of whether differential weighting is a good idea. If you have no intention of creating composites based on such differential weights, it is probably not a good idea to report this index unless you subscribe to the notion that it represents "construct reliability" as has been argued by Hancock (2001) and Hancock and Mueller (2001). They provided a new approach to maximal weighted reliability and related it to the proportion of variance in a construct optimally explainable by its indicators.

Composite Reliability Based On An Unknown Factor Model

Reliabilities based on a one-factor model do not adequately summarize all the systematic variation when more than one factor underlies Σ. Actually, it is possible to define and compute reliability without making the 1-factor assumption. In fact, in our defining formula for ρ_{xx} an arbitrary number of factors may underlie the covariance matrix Σ of the observed variables. Two obvious possibilities for computing reliability with a multiple factor model are to use an exploratory factor analysis model (see Bentler, 2004, 2005 on how to set this up) or to use some type of confirmatory factor model. In these cases, one must specify the number of factors. We first discuss several methods for computing reliability without any assumption on the number of factors or on possible factor loading patterns. There are several possible approaches: (1) we can assume a factor model with an arbitrary number of factors, with $\Sigma_C = \Lambda\Lambda'$ and $w_i = 1.0$; or (2) we can do (1) with the further assumption that the covariance matrix of the unique variables also can be written as $\Psi = KK'$, thus imposing a nonnegativity constraint on their variances; or (3) we can do (2) with the further restriction that w is chosen optimally. In these approaches, we can *not* specify the number of factors, nor can we put additional constraints on Λ or Ψ. In other words, even if 1-factor reliability coefficients are computed, with or without constraints in one or several groups, the dimension-free coefficients never have additional constraints. *It will be seen that in spite of the fact that there are several possibilities for multidimensional or dimension-free coefficients, the final reliability estimates can be interpreted to represent a unidimensional composite.*

Dimension-free lower bound to reliability

We make the usual assumption that $\Sigma = \Sigma_C + \Psi$, where Σ_C is the covariance matrix of the common variables, and Ψ is the covariance matrix of the unique (or error) variables. We assume that Ψ is diagonal, i.e., that errors are uncorrelated, and that we are interested in the reliability of a unit-weighted composite. Bentler (1972) defined a dimension-free lower bound to reliability of a unit-weighted composite as the smallest possible value of ρ_{xx} consistent with Σ_C being positive semidefinite, that is, that Σ_C has a factor structure with factor loading matrix Λ based on an arbitrary number of factors, i.e., that $\Sigma_C = \Lambda\Lambda'$. He also provided a method to compute this lower bound. In EQS, use the MTFA (minimum trace factor analysis) algorithm of Jamshidian and Bentler (1998) to compute Bentler's lower bound (say, blb), and verify it with Bentler's original method. See also della Riccia and Shapiro (1982), and Shapiro (1982b). The results of an analysis are printed out as follows:

```
BENTLER'S DIMENSION-FREE LOWER BOUND RELIABILITY = .890
```

Bentler (1972) showed that this coefficient is always as large as α, or larger. However, the sample coefficient may overestimate its population counterpart in small samples. Shapiro and ten Berge (2000) provided a bias-correction

to improve this coefficient, and Li and Bentler (2004) provided corrected and resampling estimators. They will be available in future versions of EQS if shown to work well in realistic-sized samples.

Greatest lower bound to reliability

In computing Bentler's dimension-free lower bound, it is possible for some unique variances (diagonal elements of Ψ) to be estimated as zero or negative. Stated differently, a so-called Heywood solution might arise. In this situation, Ψ cannot be the covariance matrix of real-numbered unique or error variables. Thus, when computing blb, it may be desirable to impose the further condition that the elements of Ψ are all nonnegative, i.e., that it would be possible to compute Ψ=KK' (however, since Ψ is diagonal, so is K). This can yield a larger coefficient, the so-called greatest lower bound (glb) to reliability of a unit-weighted composite (Woodhouse & Jackson, 1977). It was shown by ten Berge, Snijders, and Zegers (1981) that the Woodhouse-Jackson method might not find the optimal lower bound. Bentler and Woodward (1980) showed how to compute the greatest lower bound with a small modification of Bentler's (1972) algorithm. See also Bentler and Woodward (1983, 1985) and Shapiro (1985c). In EQS, we use the CMTFA (constrained MTFA) algorithm of Jamshidian and Bentler (1998) and Bentler and Woodward (1983) to do this computation. The results are printed out as follows:

```
GREATEST LOWER BOUND RELIABILITY = .910
```

The greatest lower bound will be identical to the blb when the solution contains no zero or negative unique (error) variances. Otherwise, the glb will be, well, greater. *The larger size of this coefficient may not be desirable in small samples.* The problem is that the glb has an upward sampling bias (ten Berge, 2000; ten Berge & Sočan, 2004). As a result, until a bias-corrected estimator is available, in small to intermediate sample sizes it may be desirable to utilize the less biased dimension-free blb coefficient. Current research by Li and Bentler (2001, 2004) shows that bias-correction of the glb can be achieved through a revised estimator of unique variances and resampling methods. These will be implemented in the future in EQS.

Maximal Dimension-Free Lower Bound To Weighted Composite Reliability

The maximal weighted reliability coefficient based on a 1-factor model (Bentler, 1968) was generalized by Shapiro (1982a) to a factor model with an arbitrary number of factors ($\Sigma_C = \Lambda\Lambda'$), Ψ=KK', and *w* chosen optimally. So, again, interest focuses not on the standard unit-weighted sum of components, but rather, on an optimally weighted sum. Consider the set of possible ρ_{xx} coefficients as Ψ is varied, and consider its minimal value $\min_\Psi(\rho_{xx})$. Shapiro's approach (say, slb) involves selecting *w* (excluding the null vector) so that the value of $\{\min_\Psi(\rho_{xx})\}$ is maximized, i.e., is as large as possible. In EQS, Shapiro's coefficient is computed using the WMTFA (weighted MTFA) algorithm of Jamshidian and Bentler (1998). An illustrative result is the following:

```
SHAPIRO'S LOWER BOUND RELIABILITY FOR A WEIGHTED COMPOSITE = .811
WEIGHTS THAT ACHIEVE SHAPIRO'S LOWER BOUND:
 V1   V2   V3
.822 .779 .245
```

An advantage of the coefficient is that it is scale free, i.e., it has the same value for arbitrarily rescaled covariance matrices. However, as with maximal weighted reliability, it must be recognized that the slb coefficient does not refer to the typical equally weighted composite, and care must be used in reporting the meaning of the index. As with the greatest lower bound, there is liable to be an upward bias in the coefficient that needs further study.

Composite Reliability Based On A Given LV Model

Reliability coefficients can be computed based on your model, which need not be a one-factor model. In fact, it need not be a factor model. This can be done when there are /EQUATIONS or there is a /MODEL paragraph, and

when the model contains F's and E's. This is useful, for example, when the model contains a single factor plus additional correlated errors, but the errors are not intended to be part of the common scores (e.g., Bollen, 1980; Kano & Azuma, 2003; Vautier, 2001). In fact, any model with an arbitrary number of factors also can be run. Whatever such a model might be, its covariance matrix is decomposed as $\Sigma = \Sigma_C + \Psi$, where Ψ contains the variances and covariances of the E's, and Σ_C contains the covariances of everything else. To take a single variable, suppose V1 = .5*F1 + .8*F2 + E1. Then C1 = . 5*F1 + .8*F2 is the variable whose variance is in the first diagonal position of Σ_C, and whose covariances with other similar C variables are in the first row and column of Σ_C. Thus EQS automatically computes reliability of an unweighted composite of the observed variables used in any latent variable model that has additive errors. To illustrate, if one runs the correct two-factor model for the data file *manul7a.ess*, the 2-factor model-based coefficient is:

```
RELIABILITY COEFFIENT RHO = .780
```

Clearly, the labeled printout is the same as when a 1-factor model is run. A special case of *rho* is McDonald's (1970) omega.

Multidimensional Coefficients Are Maximal 1-Factor Reliabilities

As the critiques of coefficient α demonstrate, researchers usually hope that reliability coefficients describe consistency of indicators for one homogeneous dimension, resulting in what might be called a unidimensional coefficient. Yet, as ten Berge and Sočan (2004, p. 613) note, the unidimensional hypothesis "will invariably be rejected when there are more than three test parts." However, it is not necessary to contrast unidimensional vs. multidimensional coefficients, because every multidimensional coefficient implies a particular composite with maximal unidimensional reliability. This idea takes some getting used to!

In the above development, the main assumptions made with regard to dimension-free coefficients are that $\Sigma = \Sigma_C + \Psi$ under the restrictions that Σ_C is a covariance matrix of real-valued (not imaginary) common scores, and that Ψ is diagonal. Thus $\Sigma_C = \Lambda\Lambda'$ where Λ may contain any number (say, *k*) of common factors. Even though there may be *k* factors, Bentler (2004, 2005) showed that the dimension-free coefficients actually can be conceptualized as reliability coefficients for a particular 1-dimensional composite. Specifically, there exists a single factor dimension in any *k*-dimensional factor space that is maximally reliable, and that the *k*-factor and dimension-free coefficients describe this maximally reliable factor dimension. Note: this is under equal weighting of variables and does not refer to maximal weighted reliability of a differentially-weighted composite!

To see this, we start with a factor model such as $x = \Lambda\xi + \varepsilon$ and compute the sum score $X = 1'x = 1'\Lambda\xi + 1'\varepsilon$, where 1 is the unit summing vector. In Bentler's approach, the factor loading matrix (possibly, after rotation) is transformed to $\Lambda = [\lambda \mid \bar{\Lambda}]$, where λ is (*p*x1) and $\bar{\Lambda}$ is (*p*x(*k*-1)) and $1'\bar{\Lambda} = 0$, that is, the *k*-1 columns of $\bar{\Lambda}$ sum to zero. As a result,

$$1'\Lambda = 1'[\lambda \mid \bar{\Lambda}] = [1'\lambda \mid 1'\bar{\Lambda}] = [1'\lambda \mid 0] = [\textstyle\sum_1^p \lambda_i \mid 0] = [\lambda_x \mid 0] \,. \tag{14.4}$$

It follows that the composite can be decomposed into the 1-factor representation

$$X = \lambda_X \xi_\lambda + \varepsilon_X \,, \tag{14.5}$$

where ξ_λ is the factor whose variable loadings are given by λ, and $\varepsilon_X = 1'\varepsilon$. Thus, with $T = \lambda_X \xi_\lambda$ being unidimensional, the true and error scores are uncorrelated, and we obtain the unidimensional internal consistency coefficient

$$\rho_{kk} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\lambda_X^2}{\lambda_X^2 + \sigma_{\varepsilon X}^2} \,. \tag{14.6}$$

This is the *maximal unit-weighted reliability* since the decomposition was proven by Bentler to maximize (14.6). An interesting property of reliability (14.6) is that the remaining factors contribute nothing to reliability, as can be seen by the definition of our *k*-dimensional coefficient

$$\rho_{kk} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{1'\Lambda\Lambda'1}{1'(\Lambda\Lambda'+\Psi)1} = \frac{1'\lambda\lambda'1}{1'(\lambda\lambda'+\Psi)1} = \frac{(\sum_1^p \lambda_i)^2}{(\sum_1^p \lambda_i)^2 + \sum_1^p \psi_i} = 1 - \frac{1'\Psi 1}{1'\Sigma 1}. \qquad (14.7)$$

In other words, any reliability coefficient obtained from a *k*-dimensional EFA, CFA, or general SEM model, or the *k*-dimensional dimension-free and glb coefficients, permits finding the single dimension in the desired multidimensional space that maximizes unit-weighted internal consistency.

To be concrete, the factor loadings on the maximal unit-weighted reliability factor can be computed via $\hat{\lambda} = \{1'(\hat{\Sigma} - \hat{\Psi})1\}^{-1/2}(\hat{\Sigma} - \hat{\Psi})1$, where $\hat{\Sigma}$ is the structural model covariance matrix or just the sample *S* when dealing with dimension-free coefficients. In EQS, $\hat{\lambda}$ is standardized so that the $\hat{\Sigma}$ or *S* have 1.0 diagonals, and the result is printed out as follows:

```
STANDARDIZED FACTOR LOADINGS FOR THE FACTOR THAT GENERATES
  MAXIMAL RELIABILITY FOR THE UNIT-WEIGHT COMPOSITE
  BASED ON THE MODEL (RHO):
     V1        V2        V3        V4        V5        V6
     0.752     0.721     0.743     0.806     0.754     0.762
     V7        V8        V9
     0.461     0.597     0.643
  BASED ON THE GREATEST LOWER BOUND (GLB):
     V1        V2        V3        V4        V5        V6
     0.727     0.743     0.756     0.790     0.772     0.801
     V7        V8        V9
     0.499     0.598     0.635
```

The result (14.5) is relevant to item parceling, in which item composites such as *X* are computed over subsets of items for subsequent inclusion in larger models. Even if the items in a given parcel are multidimensional, it is clear that the composite can be considered to represent the most internally consistent latent dimension for the selected items. However, a different or better selection of items for a given parcel may well produce a composite with greater internal consistency reliability. More generally, how to partition a large set of items into a set of mutually exclusive composites so that the sum of their separate reliabilities is maximal is an unsolved problem in parceling. Most likely, a good approximation to the highest overall reliability will be achieved when the composites mirror the factor structure of the items, that is, if a given composite contains those items with the highest loadings on a given factor. Factor analysis of the composites generated in this fashion, however, will tend to obtain factors that represent the second-order rather than first-order factors in the items. This may not be the goal of parceling! A different procedure is to create composites whose factors more closely mirror the first-order item factors. As you can see, there are important subtleties involved in creating item parcels and we direct you to further reading (Bandalos, 2002; Bandalos & Finney, 2001; Hagtvet, & Nasser, 2004; Hall, Snell, & Foust, 1999; Kim & Hagtvet, 2003; Kishton & Widaman, 1994; Little, Cunningham, Shahar, & Widaman, 2002; Nasser & Takahashi, 2003; Rogers & Schmitt, 2004; Yuan, Bentler, & Kano, 1997).

Composites with Fewer Parts

An Optimal Short Scale Based On Alpha

Since coefficient α is the most widely used coefficient and is routinely available in statistical packages, we want to pursue the observation made above that α actually may be larger for a composite obtained from a subset of variables than for the composite obtained by summing all *p* variables. To determine whether this is possible in a given data set, a new methodology has been implemented in EQS that searches through alternative subset composites of various sizes to obtain an optimal short composite that may yield the highest possible coefficient α . This is printed out for an example as follows.

```
COEFFICIENT ALPHA FOR AN OPTIMAL SHORT SCALE = .915
BASED ON THE FOLLOWING 4 VARIABLES
V1 V3 V8 V15
```

If the list of variables making up the composite is long, only the shorter list of variables excluded from the composite is given, e.g., **`BASED ON 12 VARIABLES, ALL EXCEPT: V5 V9.`** Due to local optima, this methodology is not necessarily guaranteed to find the absolutely best short scale, i.e., the scale with the largest α. However, it should be close, and the proposed short scale will have an α that is at least as large as the one based on all variables. This feature may be useful in item analysis. As in all item analysis, however, you should use theory and substantive knowledge to evaluate whether this short scale is meaningful, since a high alpha can be obtained when a few items have essentially identical wording and hence are more highly interrelated than the pool of items from which they come. It is up to you to be sure that any short scale you develop maintains desired critical content coverage.

An Optimal Short Scale Based On The Greatest Lower Bound

As noted above, it is possible that a total score based on some, but not all, of the variables may have a higher coefficient α than the one based on all variables. In principle, the same phenomenon can occur with the greatest lower bound, although since any covariation is common variance for the glb, the degree of scale shortening may be much less. To find that subset of variables that has the highest glb, iterative glb computations would have to be done for each one of an extremely large number of candidate subsets of variables. This is not practical, and thus EQS uses a simplifying approximation. A backward stepping methodology is used that starts with all variables. At each step, the variable with the largest relative unique variance $\psi_{ii} / \sigma_{ii}^2$ in the current glb decomposition $\Sigma = \Sigma_C + \Psi$ is dropped from the analysis, and a new glb is computed for the reduced composite. If the glb increases, this process is repeated. When the glb no longer increases, the previous set of variables is accepted. The optimal short composite based on fewer variables that is obtained is printed out as

```
GLB RELIABILITY FOR AN OPTIMAL SHORT SCALE  = .945
BASED ON THE FOLLOWING 4 VARIABLES:
V1, V6, V8, V13
```

As before, if the list of variables is long, the shorter list of excluded variables is given instead, e.g., **`BASED ON 12 VARIABLES, ALL EXCEPT: V5, V9.`** Often, no scale shortening is possible.

Some Comments On Choices Of Reliability Coefficients

A large choice of reliability coefficients is made available in EQS so that the research community can study these coefficients further, and come to reasoned conclusions about which coefficient might be best under what circumstances. In this section we provide several further perspectives on these coefficients. The first point we want to make is that care has to be taken with regard to claims made about any reliability coefficient. The actual size of a coefficient often depends on the design of a study, in particular, what variables are involved in the defining coefficients (Cronbach, 1988). A composite made up of 9 items could be looked at in several ways: a sum across 9 items or a sum across three 3-item subtests, for example. Although the composite is the same, reliability coefficients computed from the 9 variables, or the 3 subtest variables, may differ. More generally, composites made up of more components may or may not have larger reliability coefficients. They will tend to be larger, but need not be, because the assumptions underlying the theory may not be correct. Below we describe procedures in EQS to find *optimal short scales* with fewer parts that may have larger reliabilities. The following example illustrates this point further. It is based on the first 3 variables from *manul7a.ess* and also on all 6 variables from this file. The first 3 variables are unidimensional, but the set of 6 variables is not unidimensional, and the 2-factor CFA model with correlated factors has {V1,V2,V3} loading on F1 and {V4,V5,V6} loading on F2. It is seen from the 2nd and 3rd columns that α and 1-factor *rho* are larger for the 3-item composite than for the 6-item composite. On the other hand, the dimension-free coefficients are not fooled this way: they are larger for the 6-item composite. Of course, the 2-factor CFA model cannot be set up for three variables.

No. of vars	Coeff. alpha	1-factor rho	2-factor CFA rho	Bentler's lower bound	Greatest lower bound	1-factor maximal weighted rho	Shapiro's optimal weighted reliability
3	.724	.744	--	.744	.744	.781	.768
6	.682	.700	.780	.794	.794	.791	.824

In addition, there are some other noticeable points. In the 3 variable set, coefficients for 1-factor and x-factor dimension-free models have the same value (.744). Since there are no negative error variances, the blb and glb are equal in both data sets. The set of 6 variables is not unidimensional, so there will be a difference between the 1-factor and 2-factor coefficients, with the 2-factor CFA coefficients typically being larger if the model is correctly specified. The blb and glb are slightly larger than the 2-factor CFA model-based estimate. It is also obvious that all the factor-based coefficients are larger than the Guttman-Cronbach α. It is known that *α is best rationalized with a very restricted one-factor model with all loadings equal and all error variances equal*, even though it seems computationally to be model-free. If one were to compute a *restricted* 1-factor model for the 6 variables with all loadings and error variances equal, both *rho* and the 1-factor maximal weighted *rho* will equal .682, the value for α. We also note in the table that the maximal weighted 1-factor and x-factor coefficients, based on differentially-weighted variables, are greater than their equally weighted counterparts.

It is remarkable that after 2-3 decades, there is very little practical experience with factor-based coefficients. No doubt this is in part because they are not easily available. In deciding on their use, the first practical issue to consider is whether interest lies in unit-weighted composites or optimally weighted composites. Almost always it will be with unit-weighted composites. Then there is a decision to be made on whether a model is to be imposed on the data. If it is, a 1-factor model is the natural place to start, as reliability is most easily interpreted when based on a 1-factor model. However, if variables are not unidimensional, a coefficient based on an assumption of unidimensionality will not accurately describe the internal consistency of a composite. A reliability estimate based on any specified model, called *rho* in EQS, is based on the $\hat{\Sigma}$ that results from the estimation. If $\hat{\Sigma}$ is far from Σ, coefficients based on $\hat{\Sigma}$ will not be adequate to describe reliability based on Σ. Suppose that $\Sigma = \Lambda_1\Lambda_1' + \Lambda_2\Lambda_2' + \Psi$. Then the model-based reliability should equal $1'(\Lambda_1\Lambda_1' + \Lambda_2\Lambda_2')1/1'\Sigma 1$. But a one-factor model reliability estimate will equal $1'(\hat{\Lambda}\hat{\Lambda}')1/1'\hat{\Sigma}1$, where $\hat{\Lambda}$ is a one-factor model, $\hat{\Lambda}\hat{\Lambda}'$ reproduces some unknown mixture of $\Lambda_1\Lambda_1' + \Lambda_2\Lambda_2'$, and the model-based $\hat{\Sigma}$ will be a biased estimator of Σ. Clearly, a reliability coefficient for any fitting latent variable model is much more likely to represent the true relative magnitude of systematic variance. Thus if a user's model is a particular 6-factor latent variable model, and it fits the data, then the corresponding model-based reliability probably should be reported, as it incorporates the various factor sources of variance. As shown in (14.5), *rho* can always be conceived as representing a particular unidimensional composite, where the dimension is chosen to maximize *rho* (Bentler, 2004, 2005). However, this is not the only way to think about reliability, and if focus is on a particular CFA or EFA structure, it would be possible to decompose a multifactor internal consistency coefficient into additive component internal consistencies as is recommended in generalizability theory (e.g., Brennan, 2001; Shavelson & Webb, 1991). For example, $1'(\Lambda_1\Lambda_1' + \Lambda_2\Lambda_2')1/1'\Sigma 1$ can be decomposed into $1'(\Lambda_1\Lambda_1')1/1'\Sigma 1$ plus $1'(\Lambda_2\Lambda_2')1/1'\Sigma 1$. This is not currently done in EQS. A discussion of such decompositions is given in Lucke (2005).

In the absence of a strong model, or if there is controversy about a model (for example, several different models may fit about equally well), a dimension-free coefficient makes the most sense. The blb coefficient is probably safer than the glb, though with a large number of variables these will be essentially identical.

Differential Weighting Of Variables

Although optimal weighting of variables is only rarely of interest, it can be done. When $\hat{w} \neq 1$, there are two options for reliability in EQS. The maximal weighted 1-factor reliability based on variable weighting and Shapiro's dimension-free coefficients are competitive. The choice should depend on what one believes about the data, and on

what one plans to do with the coefficient. If a 1-factor model clearly does not fit, again one can question use of a 1-factor based maximal weighted reliability coefficient. Conceivably one might actually intend to weight the variables such that they would achieve this maximal weighted 1-dimensional reliability, or one could subscribe to the "construct reliability" interpretation of this coefficient. For example, in the two-factor illustration above $\hat{w}'(\hat{\Lambda}\hat{\Lambda}')\hat{w}/\hat{w}'\hat{\Sigma}\hat{w}$ based on one factor may still be a meaningful ratio, since the numerator is some type of true score variance estimate and the corresponding total variance $\hat{w}'\hat{\Sigma}\hat{w}$ is still a consistent estimate of the weighted composite variance $w'\Sigma w$. But the reasoning is tortured, and Shapiro's coefficient is the more natural one to use when a 1-factor model does not fit the data.

Finally, the computational methods in EQS for estimating reliability are not the only ones that can be used. Older algebraic methods (e.g., Jackson & Agunwamba, 1977) or specialized statistical approaches (e.g., Fuller, 1987) also could be used, but these are not implemented in EQS as they are not consistent with the SEM tradition.

Reliability Printouts With Any Model

EQS automatically computes and prints several reliability coefficients. These are:

1. coefficient alpha
2. coefficient alpha for an optimal short scale
3. coefficient *rho* based on the user's LV model
4. the greatest lower bound
5. the glb reliability for an optimal short scale
6. Bentler's dimension-free lower bound
7. Shapiro's lower bound for a weighted composite, and
8. maximal weighted 1-factor reliability.

All of the model-based coefficients, including the dimension-free ones, require an iterative procedure to have converged; if convergence is not achieved, the coefficient may not be printed. The last one in the list, #8, is computed *only if the model is a 1-factor model.* These computations are undertaken within a wide variety of modeling contexts wherever an appropriate covariance matrix can be found, e.g., in multiple group models these computations are done in each of several groups.

/RELIABILITY Without Model Details

One way to get 1-factor reliabilities is to use the /RELIABILITY paragraph without any /EQU, /VAR, and /COV, and without /MODEL. With the statement /RELIABILITY and a specification of which variables are involved, EQS will generate the appropriate one-factor structure. That is, EQS will set up a model and compute a 1-factor ρ_{xx} *when the model file contains NO equations, NO variances, and NO covariances,* and the appropriate scale items are given. Since ρ_{xx} is the reliability of a particular composite, it is necessary to specify which variables are intended to be the components of the sum *X*. The user needs to specify the following statements:

```
/RELIABILITY
   SCALE = V1 TO V4;
```

where it is desired to find the reliability of the scale total score V1+V2+V3+V4. The effect is that, internally, EQS generates (1) a one-factor model and computes composite and maximal weighted reliability, and also (2) computes the dimension-free coefficients discussed above. Although several computations are undertaken, the user sees only the following setup, which is used to estimate the one-factor model:

```
/EQUATIONS

    V1 =  F1 + E1;
    V2 = *F1 + E2;
    V3 = *F1 + E3;
    V4 = *F1 + E4;
/VARIANCES
    F1 = *; E1-E4 = *;
```

The standard output from an EQS run on this model will give the optimal parameter estimates that are used to compute the 1-factor reliability coefficients. Notice that the setup fixes the first loading and frees the variance of the factor. If the factor loadings above are arranged in a vector F, and the variance of F1 is φ, $\Lambda=F\varphi^{1/2}$. We repeat that the resulting 1-factor reliability coefficient may not be meaningful if the 1-factor model does not fit the data.

This command can be used with /CONSTRAINTS. In a one group situation, this allows estimation of restricted one-factor models, for example, a model with equal factor loadings, or equal error variances, or both. In a multiple group setup, it allows for specifying cross-group constraints so that the factor models are in some way equivalent across groups. The reliability coefficients computed under such constraints have different meanings. See e.g., Jöreskog (1971ab) or Lord and Novick (1968). *However, constraints do not affect the computation of any dimension-free reliability coefficients*, whether in one or multiple groups.

Statistical Tests And Fit Indices

As can be seen in the continuing heated discussion on SEMNET, in recent years there has been a lot of controversy about the role of "fit indices" in model evaluation, where a fit index provides evidence on model fit in a standardized continuous range (typically, 0-1, where "1" means perfect fit, though for some indices "0" means perfect fit). It is hard to characterize the many viewpoints about fit indices, but two endpoint perspectives are that fit indices provide irrelevant and unnecessary information, distracting the user from finding problems in the model shown up by the significance test, versus the idea that no significance test can adequately describe the extent to which a model approximates the sample means and covariances. From the latter viewpoint, accepted here, fit indices provide important supplementary information about a model (e.g., Schermelleh-Engel, Moosbrugger, & Müller, 2003).

Consider first a traditional statistical viewpoint. EQS gives you a variety of potential tests T (e.g., T_{ML}) for evaluating your model. Let us assume that you use the best possible such statistic for your given data and modeling situation, for example, one that in principle has an asymptotic χ^2 distribution. If your data meet the conditions for this statistic (e.g., your model is fully a priori, data represent an appropriate sample from a well-defined population, distributional assumptions are appropriate, sample size is near "asymptotic", observations are independent etc.), then the cutoff value based on the selected alpha level tells you whether to accept or reject the null hypothesis $\Sigma = \Sigma(\theta)$. And with a large sample size, presumably you have plenty of power to reject the null. From this viewpoint, there is not much else to say: either your model is consistent with the data, or it is not. Hence there is no need for other indices; they are irrelevant, and a focus on fit indices represents a distraction from the need to admit and possibly find and correct important model misspecifications.

Of course, when running our favorite model we do not really do classical null hypothesis testing, as in the standard experimental design situation, since we want to accept rather than reject the null hypothesis. And we may be building and modifying models rather than testing them. Yet the same accept/reject framework is invoked. No doubt the ability to formally test a model with correlational data, as initially developed by Jöreskog (1969) for confirmatory factor analysis, helped to generate enthusiasm for SEM and permitted it to be contrasted with older-fashioned exploratory factor analysis that generally used no testing machinery (see e.g., Bentler, 1986b). In recent years, even Jöreskog (1993) distinguished between strictly confirmatory, model comparison, and model generating uses of SEM, recognizing that formal testing of models is only a small part of the SEM enterprise. In fact, null hypothesis testing in the broader context has had a long and controversial literature (see, e.g., Sterne & Smith, 2001)

and is also contradictory to the current trend questioning the importance of significance tests (e.g., Harlow, Mulaik, & Steiger, 1997; Schmidt & Hunter, 2002; Thompson, 2002; Wainer & Robinson, 2003). The broadest alternatives in scientific research are meta-analysis and Bayesian methods (e.g., Howard, Maxwell, & Fleming, 2000). The former requires results from multiple data sets that are not typically available in a SEM study, though when available, meta-analysis usually increases statistical power (Cohn & Becker, 2003). Interesting SEM approaches to meta-analysis are possible (e.g., Becker & Schram, 1994; Cheung & Chan, 2005; Furlow & Beretvas, 2005; Viswesvaran & Ones, 1995). The latter, Bayesian inference, rejects classical frequentist thinking that forms the basis for almost all tests used in SEM. It proposes to combine prior information about the model and its parameters with information provided in the data to generate posterior distributions that are used as a basis of inference (see e.g., Rupp, Dey, & Zumbo, 2004). Although recently extensively developed for application to SEM (e.g., Ansari, Jedidi, & Dube, 2002; Arminger & Muthén, 1998; Hayashi & Yuan, 2003; Hoshino, 2001; Jedidi & Ansari, 2001; Lee & Shi, 2000; Lee & Song, 2002, 2003, 2004; Lee & Zhu, 2002; Scheines, Hoijtink, & Boomsma, 1999; Song & Lee, 2001, 2002abc, 2004), it is sufficiently difficult to implement that standard SEM packages do not yet offer its possibilities. In SEM, both prior and posterior distributions are in multivariate form – and in realistic data analyses, of very large dimensionality (orders of magnitude larger than the illustrative applications in the published literature) – requiring more difficult and lengthier computations. The payoff is better handling of difficult problems, such as small sample sizes (Lee & Song, 2004). We expect that as the statistical literature increasingly emphasizes Bayesian inference, as statistical education for the social and behavioral sciences catches up, and as computers continue to increase in capabilities, some Bayesian SEM implementations surely will become available to practitioners in a user-friendly way.

Actually, naïve reliance on significance tests also is problematic within the classical hypothesis testing framework. First of all, the chosen probability level (e.g., reject the model if $p<.05$) is arbitrary. Why not .10, or .01, or .073? Second, as noted earlier, the conditions for a test statistic to be precisely χ^2 distributed will rarely be met exactly, and hence what is printed out as a precise p-value is actually a rather crude and error-prone approximation to what this probability would be under ideal conditions. Third, strong arguments can be made that the null hypothesis typically tested with T is the wrong hypothesis. It has been suggested that the hypothesis $\Sigma = \Sigma(\theta)$ will essentially never be precisely correct (e.g., Bentler & Bonett, 1980; Browne & Cudeck, 1993; de Leeuw, 1988). In discussing models generally, MacCallum (2003, p. 113) notes "All of these models, in their attempt to provide a parsimonious representation of psychological phenomena, are wrong to some degree and are thus implausible if taken literally." Such points of view imply that the classical use of T in a hypothesis testing way is inappropriate, even if the assumptions underlying the test are met and the model test truly represents a confirmatory test of an *a priori* hypothesis. After all, if the null hypothesis really is incorrect, then as sample size increases, the test statistic does not really evaluate a null hypothesis but rather represents a measure of power to reject the null hypothesis. Stated differently, since the typical test statistic is a direct function of sample size, that is, $T = (N-1)F_{min}$, for any given discrepancy F_{min} between data and model, when the model is wrong, increasing N increases the probability of its rejection. Actually, power to reject a model can be very high in not-so-large samples. For example, when unique variances in a latent variable model are small, the test T may have high power for model rejection even though correlational residuals are very small on average (Browne, MacCallum, Kim, Anderson, & Glaser, 2002).

A modified approach is to specify null hypotheses differently, and more usefully. Serlin and Lapsley (1985, 1993) suggested a "good-enough" principle and recommended that tests of *close fit* replace tests of exact fit. In SEM, MacCallum, Browne, and Sugawara (1996) developed a such a test of close fit $\Sigma \approx \Sigma(\theta)$ as a substitute for the exact fit null hypothesis, and recently Yuan, Hayashi, and Bentler (2005), Li (2005), and Ogasawara (2005) developed the distribution of T under alternative rather than the null hypothesis for application in situations where the model structure may be severely incorrect. These are useful developments that will play out in the next few years (see, e.g., Steiger, 2004 on close fit in the context of analysis of variance). Yet, it seems likely that even *exact* tests of hypotheses regarding *approximate* fit (assuming that they are reliable) will be subject to criticism since one person's close fit may be another person's terrible fit.

In summary, it does not make sense to rely on T as the sole measure of model fit. This was recognized a long time ago by Bentler and Bonett (1980). They developed transformations of T designed to map it into a more interpretable 0-1 range. In this, they built upon work by Tucker and Lewis's (1973) on a "reliability" index (which actually has nothing to do with reliability as defined above) for the exploratory factor analysis model estimated by ML, extending this to be applicable to general structured models and fit functions, and developing other indices of fit. A typical name for a transformed statistic is a goodness-of-fit index (e.g., Jöreskog & Sörbom, 1988, p. 43; Bentler,

1983a, p. 507), sometimes abbreviated as GFI (although GFI is typically the name of a specific index, see below). The rationale for such indices was summarized by Bentler (1990a, p. 238): "Acceptance or rejection of the null hypothesis via a test based on *T* may be inappropriate or incomplete in model evaluation for several reasons: 1. Some basic assumptions underlying *T* may be false, and the distribution of the statistic may not be robust to violation of these assumptions; 2. No specific model $\Sigma(\theta)$ may be assumed to exist in the population, and *T* is intended to provide a summary regarding closeness of $\hat{\Sigma}$ to *S*, but not necessarily a test of $\Sigma = \Sigma(\theta)$; 3. In small samples, *T* may not be chi-square distributed; hence the probability values used to evaluate the null hypothesis may not be correct; 4. In large samples any *a priori* hypothesis $\Sigma = \Sigma(\theta)$, although only trivially false, may be rejected."

Fit indices are statistics that, in most cases, estimate some population measure of fit. There has been progress since Bentler (1990a, p. 245) noted that "Essentially nothing is known about the theoretical sampling distribution of the various estimators." Early work on RMSEA (see below) was unpublished (Steiger & Lind, 1980), but has since become well-known (Browne & Cudeck, 1993). More recently, the approximate large sample distribution of a variety of fit indices under correct and misspecified models has been given by Ogasawara (2001ab, 2003, 2005). Treating fit indices as statistics allows statistical goodness-of-fit testing, establishment of confidence intervals, and so on. To a large extent statistical evaluation duplicates information available via *T*, though it should be remembered that when a given *T* is questionable, fit indices based on *T* also may be. Nonetheless, fit indices can perform better than the tests *T* from which they were derived (Lei & Lomax, 2005).

Even though the field appears to have a generally positive viewpoint regarding fit indices, in an outstanding recent paper – really, a "must-read" paper -- Yuan (2005) provided a serious challenge to their value. It has been known for a long time, and superbly verified by Yuan, that fit indices are not just a measure of model fit, but also of choice of *T* (most of which, he says, do not behave as advertised) as well as uncontrolled factors such sample size and the distribution of the data. "Thus, cutoff values for fit indices, confidence intervals for model fit/misfit, and power analysis based on fit indices are open to question" (p. 142). He does support the use of fit indices for comparing different models when based on the same data and test statistic. We feel that a wider use is warranted, though always with caution. As part of this caution, in this chapter we will refrain from recommending particular cutoff values for acceptable models, and refer the reader to the original literature and its controversies on this topic (e.g., Hu & Bentler, 1998, 1999; Marsh, Hau, & Wen, 2004).

The R^2 Analogy and Cutoff Values

Before going into details about fit indices, it is important to point out that these generally are built upon statistics *T*. If you completely reject *T* as meaningful, there is not much point to studying functions of *T* that map it into a more interpretable range. It is still incumbent on you to select the *T* that can be best justified for your particular data. And furthermore, an interpretable and acceptable fit index should not blind you from recognizing that something is still amiss in your model if *T* is large compared to df.

If you have not previously encountered fit indices, but are knowledgeable about regression and its R^2, it may make sense to give an analogy between the two. As in SEM, the most important focus in regression is on the relative size and importance of its parameters (β coefficients in regression). But an important ancillary piece of information is the R^2, which summarizes the proportion of variance accounted for in a dependent variable that is optimally explained by its predictors. Larger values of R^2 imply better predictability, and if prediction is the goal, models with higher R^2 are clearly better. This does not mean that R^2 is a perfect indicator of model quality, nor that there is a precise minimum value (such as .9) that always accompanies an acceptable model. Similar reservations will hold for SEM fit indices.

In fact, it is important to be careful about substantive as well as statistical issues when interpreting an R^2. For example, if nonsense variables are included as predictors, the model is not scientifically valuable regardless of the size of the R^2. Also, if predictors are added to a model in a post-hoc fashion (e.g., by forward search procedures analogous to the LM test), a small-sample based regression model may not cross-validate. Further, when N is small compared to p, an adjusted R^2 provides a better estimate of the population value than does the unadjusted R^2. Considered statistically, the distribution of R^2 depends on the characteristics of the variables, their distributions, the

degree to which assumptions are met, and so on. In spite of all such problems, R^2 does provide some useful information, especially when contrasting several models meant to explain the same phenomenon and data.

It is in this spirit that we present fit indices. These have many of the same advantages and disadvantages as R^2, but instead of a single index, there are a lot of competing ones. In regression, there is typically one method of estimation, but in SEM, there are many, and their quality impinges on fit indices (Yuan, 2005). Further, the field cannot agree on the single best measure of model quality. Given any particular index, when comparing two meaningful models, the one with the better fit index is probably better. And then, as with R^2, capitalization on chance reduces its meaning, and a model with nonsense parameters is not meaningful regardless of the size of the index. Ideally, the index also has a good population estimator rationale. And so on.

The analogy between fit indices and R^2 is not perfect. The most obvious point is that R^2 describes a property of a single equation model, while SEM is a multi-equation system for which the various R^2s may or may not be important. A fit index is not an overall summary of the several R^2s, since the degree to which a SEM model fits is unrelated to the relative size of R^2s in its equation system. Also, R^2 does not describe the quality of the regression model as a SEM model, since regression is a saturated model that will fit any data set, while fit indices are meant to describe the SEM model's ability to explain the sample covariances (and possibly means). Finally, one rarely hears complaints in regression that sample size is so large that one should ignore some statistical result (e.g., on significance of regression coefficients). Yet, excessive power is a complaint that one hears in SEM with regard to *T* (this is rarely a complaint when testing individual parameters) -- unless, of course, the researcher is testing someone else's model and hopes to reject it!

Three Types of Fit Indices

There is a huge catalog of indices. An SEM program may print up to twenty. This diversity is scientifically problematic, since it implies lack of consensus about their quality and allows researchers to select the most optimistic one to report. EQS prints out about 10 indices, depending on the circumstances, and we will limit our discussion to these. (You get all the indices with the /PRINT command, and writing FIT=ALL;) This is still a far greater number than you should use. The many indices can be roughly grouped into three types that provide somewhat different views of the quality of fit. We will discuss each type and recommend that you use one index of each type to summarize your fit; in a later section, we discuss why more than one fit index may be useful. This is consistent with the recommendation of Beauducel and Wittman (2005), and would provide a fairly well-rounded view of global fit that allows you to communicate with readers who prefer one or another index. Within each type, we give one or two recommendations. We also review the concepts of a *noncentrality parameter* and the associated *population fit function*, which is relevant to most of the fit indices.

Before proceeding, it should be noted that there is a long history of classification schemes regarding fit indices (e.g., Bollen, 1989ab; Gerbing & Anderson, 1993; Hu & Bentler, 1995, 1998; Marsh, Balla, & McDonald, 1988; Marsh & Balla, 1994; Tanaka, 1987, 1993). For example, Tanaka noted that indices vary with regard to being: population versus sample based, concerned with simplicity vs. complexity, normed vs. nonnormed, absolute vs. relative, estimation method free vs. estimation method specific, and sample size independent vs. sample size dependent. The several classifications are too complex to review in detail in this chapter; please see the literature. In his technical development of the distribution of fit indices, Ogasawara (2001b) considers the indices that he studied to fall naturally into two groups. These correspond to measures of absolute and incremental fit that we describe below. We begin with an index of a different type.

Averages of Standardized Residuals

In structural modeling we hope that a structured model $\Sigma(\theta)$ is a good approximation to, or identical to, the population covariance matrix Σ, in other words, the population residual $\Sigma - \Sigma(\theta)$ should be small. Of course in practice, we have neither of these matrices. We have the sample covariance matrix *S* and its corresponding model implied covariance matrix $\hat{\Sigma} = \Sigma(\hat{\theta})$. Test statistics aside, how could we find a nice interpretable summary of how large the residuals $S - \hat{\Sigma}$ are? Some type of average across all the elements in this matrix, ignoring the sign of the

discrepancy, seems to make sense. In EQS, two such averages are printed out subsequent to the RESIDUAL COVARIANCE MATRIX (see below). A more well-known measure, the root mean squared residual (Jöreskog & Sörbom, 1981), is defined as

$$\text{RMR} = \sqrt{2\sum_{i=1}^{p}\sum_{j=1}^{i}(s_{ij}-\hat{\sigma}_{ij})^2 / \mathrm{p}(\mathrm{p}+1)}\,. \tag{14.8}$$

Typically this index is not very useful because different variables can have very different variances, which also affects the size of covariances, hence the size of the residuals, and in turn the sheer magnitude of (14.8). It is usually possible to scale the variables by a diagonal matrix *D*, thus modifying the covariance matrix via $\Sigma \to D\Sigma D$, with the effect that the model matrix is similarly scaled while key features of the model are retained. However, when this is done, (14.8) takes on a different value for the same degree of lack of fit. It would be desirable to have a scale-invariant measure corresponding to (14.8) that does not depend on an arbitrary choice for *D*. The rescaling $\Sigma \to D\Sigma D = P$, transforming the covariance matrix into a correlation matrix *P*, accomplishes this. In practice, we do not have *P*. We use the *D* that takes *S* into the sample correlation matrix *R*, and transform the model-implied matrix similarly, giving the discrepancy matrix $R - D\hat{\Sigma}D$ (Bentler, 1995) which is printed out in EQS as the STANDARDIZED RESIDUAL MATRIX. Computing (14.8) on this matrix gives the standardized RMR

$$\text{SRMR} = \sqrt{2\sum_{i=1}^{p}\sum_{j=1}^{i}\{(s_{ij}-\hat{\sigma}_{ij})/(s_{ii}s_{jj})^{\frac{1}{2}}\}^2 / \mathrm{p}(\mathrm{p}+1)}\,, \tag{14.9}$$

where s_{ii} and s_{jj} are the sample variances of variables i and j. In EQS, (14.8)-(14.9) are printed out with other fit indices as

```
ROOT MEAN-SQUARE RESIDUAL (RMR)     =     1.240
STANDARDIZED RMR                    =      .044
```

The meaning of the SRMR here is that the correlations are reproduced to within .044, on average.

Two similar measures have been available in EQS for a long time. They measure the arithmetic average of absolute residuals (rather than a root mean square), and describe either all elements in the matrix or only the off-diagonal elements

```
            AVERAGE ABSOLUTE STANDARDIZED RESIDUALS     =         .0236
AVERAGE OFF-DIAGONAL ABSOLUTE STANDARDIZED RESIDUALS    =         .0331
```

The average absolute standardized residuals (AASR) and SRMR are comparable in size, and provide the same information in covariance structure models. In mean structure models, however, RMR and SRMR provide summaries only of the fit of the model-implied to sample covariances and correlations, respectively, while AASR for the covariance/mean matrix and the corresponding standardized matrix are also based on the mean (V999) parameters. In computing these averages, the (V999,V999) entry of the matrix is ignored, since it is always zero. In the case of multiple group covariance structures, (14.8) and (14.9) are based of the covariances/correlations of all groups, and the printed indices summarize fit over all the groups. AASR statistics are always provided for each group separately.

Hu and Bentler (1998) proposed that SRMR was especially useful in detecting model misspecification, but not everyone agrees with this. While Fan and Sivo (2005) replicated Hu and Bentler's results, they studied other conditions and concluded that SRMR is not sensitive to all types of misspecification. Similarly, Yuan (2005) does not recommend SRMR because other fit indices make better use of residuals. But in our view, it is an especially intuitively interpretable index. We endorse Hu and Bentler (1995, p. 98) on the value of standardized residuals: "If the discrepancy between the observed correlations and the model-reproduced correlations are (sic) very small, clearly the model is good at accounting for the correlations no matter what the χ^2 or [other] fit indexes seem to imply. For example, if the average of the absolute values of the discrepancy between observed and reproduced correlations is .02, it is simply a fact that the model explains the correlations to within an average error of .02. This is true whether the correlations are large or small and whether the χ^2 test is large or small. If the largest discrepancy

between observed and reproduced correlations, among all the correlations, is also small, say .10, the model is only marginally wrong for some variables. Of course if the largest discrepancy is quite large, say .40, clearly the model is not explaining some of the correlations well at all. We suggest that this descriptive information always accompany reports of model fit, to round out the more popularly used χ^2 and [other] fit indexes." This type of information will be informative to readers who understand correlations but have no further knowledge about fit indices, noncentrality parameters, and related topics.

Noncentrality Parameters and Population Lack of Fit

While SRMR and AASR do not depend on the particular fit function used to estimate the model parameters, other fit indices depend on them, specifically on a population parameter called the noncentrality parameter. Instead of simply measuring the discrepancy $\Sigma - \Sigma(\theta)$, it is measured in the metric of the estimation method used to obtain $\hat{\theta}$. In fact, it is hard to think about $\Sigma - \Sigma(\theta)$ without asking how that θ was determined. In general, we cannot specify an a priori $\Sigma(\theta)$; we have to minimize some function to obtain θ, even in the population.

Consider a thought experiment. If we somehow knew the population covariance matrix Σ, we could use it in our estimation method in place of *S*, to obtain $\hat{\theta}$ in the population. That is, we could input Σ into EQS and run our model on this "data." As output, we would obtain the "estimator" $\hat{\theta}$. Let us use the designation θ^0 for $\hat{\theta}$ in the population, as it is not a sample-based estimate. It is obviously a population parameter, and for real clarity, let us add a subscript for the particular estimation method used, such as θ^0_{ML}. In this *Manual* we described a lot of possible test statistics *T*, for example, the ADF statistic T_{AGLS}=n$\hat{Q}$ in (5.1) or the maximum likelihood statistic T_{ML}=n$\hat{F}$ in (5.14). In addition to the θ^0, given our input sample size, the output of the program also gives us the test statistic associated with that estimation method. We can designate that population test statistic *T* as T^0, and again, for greater clarity, we will add another subscript regarding the estimation method, such as T^0_{AGLS}. Then, for example, replacing *s* by σ in minimizing function (1.5) and using the population weight matrix W_0 with our sample size n, we could compute the test statistic

$$T^0_{AGLS} = \text{n}Q^0_{AGLS} = \text{n}(\sigma - \sigma(\theta^0_{AGLS}))' W_0 (\sigma - \sigma(\theta^0_{AGLS})) \; . \qquad (14.10)$$

Since there are no sample quantities in (14.10), this is actually a population parameter that takes on nonnegative values. The name of the parameter T^0_{AGLS} is the *noncentrality parameter*, sometimes abbreviated ncp (Browne, 1984; Satorra, 1989). To be precise, it is the noncentrality parameter associated with a particular model, a particular fit function, and a particular sample size. Change the model, or the fit function, or the n, and we may get a different noncentrality parameter. If our model is precisely correct in the population, (14.10) will be zero because the population discrepancy $\sigma - \sigma(\theta^0_{AGLS})$ is zero. Of course, this is equivalent to saying that $\Sigma - \Sigma(\theta^0_{AGLS}) = 0$ and also that the population fit function Q^0_{AGLS}=0.

In addition, if the original T_{AGLS} actually is distributed as a chi-square variate, then the parameter (14.10) is called the noncentrality parameter of the chi-square distribution. In fact, there are a lot of χ^2 distributions. We already know that these depend on the degrees of freedom d. Now we find that they also depend on the noncentrality parameter. As a result, we really should write $\chi^2(\text{d}, T^0_{AGLS})$. Chi-square distributions with zero noncentrality ($T^0_{AGLS} = 0$) are called central χ^2 distributions. If the noncentrality parameter is nonzero, they are called non-central χ^2 distributions. The typical model probability statements for test statistics *T* described in this *Manual* are computed under the null hypothesis that the model is correct, i.e., they are based on the central χ^2 distribution, whose tables you no doubt encountered in your basic statistics course.

The noncentrality parameter (14.10) thus can be used as a measure of misfit. Larger ncp's mean a greater degree of misfit. Unfortunately, by itself this proposal has at least two drawbacks. The first is that the size of the noncentrality parameter is directly related to sample size n=(N-1). Double the sample size, and the ncp doubles, even though the discrepancy $\sigma - \sigma(\theta^0_{AGLS})$ between our model and the actual population covariance matrix does not

change. Since noncentrality is related to power to reject the null hypothesis (see, e.g., Dolan, van der Sluis, & Grasman, 2005; Kim, 2005; Lei & Dunbar, 2004; Satorra, 2003; Satorra & Saris, 1985), this is a good thing. Larger n, more power. But this is not good for evaluating the degree of fit in some way uninfluenced by sample size as we can do, for example, with SRMR. Several ways around this problem have been proposed in the fit literature and are discussed below. The second obvious problem is that the ncp depends on the fit function used. Suppose for example, we want to evaluate $\Sigma - \Sigma(\theta)$ in terms of the ML function (5.14). With the same n, the associated noncentrality parameter is

$$T_{ML}^0 = \mathrm{n}F_{ML}^0 = \mathrm{n}\left\{\ln|\Sigma(\theta_{ML}^0)| + tr(\Sigma\Sigma(\theta_{ML}^0)^{-1}) - \ln|\Sigma| - \mathrm{p}\right\}, \tag{14.11}$$

where, as before, the actual value of θ_{ML}^0 is not known ahead of time and must be computed by using the population covariance matrix as data in (5.14) and minimizing that function. At the minimum, we have the ncp (14.11). But with nonzero ncp, it is not necessarily true that $\theta_{ML}^0 = \theta_{AGLS}^0$ nor, as a consequence, that $T_{ML}^0 = T_{AGLS}^0$. Unfortunately this means that different fit functions may well have different ncp's, and hence, may provide differently sized fit indices. There is no way around this problem as far as we can see. Hence, it is critical that you use a well-justified estimation method.

There is an additional problem associated with the use of the noncentral χ^2 distribution to describe the distribution of a test statistics *T*, and hence to evaluate fit based on the associated ncp. Under current methodologies, it is assumed that the large sample distribution of *T* is central χ^2 when the model is correct and can be well approximated by a noncentral χ^2 when the model is misspecified. Yuan (2005) notes that the distribution of many test statistics do not behave as predicted by the theory and are influenced by irrelevant factors other than the degree of misspecification: "...using (noncentral) chi-square distributions of *T* to describe the properties of a fit index…is inappropriate" (p. 130). Further, if the misspecification is great, *T* is not noncentral χ^2, nor even as a mixture of 1-df χ^2 variates (e.g., Bentler, 1994; Satorra & Bentler, 1994), but normally distributed (see e.g., Ogasawara, 2005; Shapiro, 1983; Yuan, Hayashi, & Bentler, 2005). And some fit functions are better than others in being described by the noncentral χ^2 (Olsson, Foss, & Breivik, 2004). Thus the uncritical use of ncp's from the noncentral χ^2 to define fit indices is not necessarily the best possible practice (Bentler, 1990a; Yuan, 2005). Yet, there is some support for the adequacy of the noncentral χ^2 distribution in SEM (Curran, Bollen, Paxton, Kirby, & Chen, 2002), especially when the model is only mildly misspecified, and, since indices based on it predominate, in spite of their limitations we recommend their use when based on the best possible choice of *T*.

Let us write the noncentral chi-square distribution as $\chi^2(\mathrm{d}, T^0)$, with noncentrality parameter $T^0 = T_{AGLS}^0$ or T_{ML}^0 or whatever the appropriate function is. Then the important known fit indices are different functions of the noncentrality parameter. Because the noncentrality parameter is a function of sample size, we may consider the definition

$$T^0 = n\tau\ , \tag{14.12}$$

where τ (tau) is the value of the population fit function or discrepancy function at the minimum, such as $\tau = F_{ML}^0$ or F_{AGLS}^0. So τ is a *standardized noncentrality measure of population misfit* or population discrepancy that eliminates the effect of sample size. Although intuitive, we cannot use T^0, or even the population misfit τ, as useful fit indices. This is because both of these measures increase with increasing misspecification and have no natural upper bound (however, below we describe the comparative fit index approach to defining a useful upper bound). We need to define a fit index as a meaningful and interpretable function of T^0 or τ. If various such fit indices were linear functions of T^0 or τ, then those fit indices could be mapped precisely into each other, and cutoff values could be determined that represent identical degrees of misfit across indices. However, *τ depends on the fit function used, and fit indices used in SEM are nonlinear functions of τ and possibly other parameters, and hence it is rarely possible to map one fit index precisely into another fit index and to define exactly equivalent cutoff values indicating good fit.* Furthermore, the relation among fit indices may differ for different values of p, the number of

variables. E.g., Kenny and McCoach (2003) report that two relative fit indices (NNFI, CFI, see below) sometimes decline in correctly specified models when p increases, while an absolute fit index (RMSEA, see below) improves under those circumstances. Also, among the earliest results on the study of fit indices (e.g., Bollen, 1986; Cudeck & Henly, 1991; LaDu & Tanaka, 1989) is that some are unnecessarily sensitive to sample size variations; this remains true (e.g., Fan, Thompson, & Wang, 1999; Hu & Bentler, 1998). Not all statistics T are equally good in measuring model misspecification (e.g., Fan, Thompson, & Wang, 1999; Olsson, Foss, Troye, & Howell, 2000). As a result, you should not expect conclusions based on one fit index to precisely mirror the conclusions based on another fit index (Weng & Cheng, 1997; Schmukle & Hardt, 2005). "There can never be a best coefficient for assessing fit…any more than there is a single best automobile" (Steiger, 1990, p. 179). Since there is no consensus about the best possible mapping of a noncentrality parameter T^0 into a fit index, along with Kenny and McCoach, we recommend the use of more than one index. However, as we will note below, several indices asymptotically represent the identical mapping, and hence they can be used relatively interchangeably.

In practice, fit indices require estimates of the noncentrality parameter, and these estimates should behave well under the null hypothesis and when the model is misspecified. Even with the same fit function and population fit index definitions, different estimators will yield different fit indices. The basic estimator of T^0 is the observed test statistic minus the degrees of freedom, (T-d), which can be negative. The estimator recommended for practice is $\hat{T}^0 = \max\{(T - \mathrm{d}), 0\}$, which dominates the maximum likelihood estimator with squared error as the loss function (Saxena & Alam, 1982). It has been noted that these estimators may overestimate T^0, and hence may make models appear worse than they really are. The bootstrap has been recommended to obtain a less biased estimator (Raykov, 2005).

A related issue is that a clear interpretation of T^0 implies that, when the model is correct, $T^0 = 0$. This requires that the test statistic should behave appropriately when the model is correct (e.g., to yield E(T) = d), as well as when the model is misspecified (e.g., to yield E(T) = d + T^0). Unfortunately, tests often misbehave (Yuan, 2005), and thus fit indices derived from them are subject to distortion. While population misfit as defined in (14.12) is an appropriate concept whether or not the distribution of $\hat{T}$ is actually noncentral χ^2, if it is not so distributed, better estimators of T^0 or τ may be available. The best general alternative approach, based on the bootstrap, is that of Yuan and Marshall (2004). It has not yet been incorporated into EQS, and until it is, current fit indices have to be used with care.

Finally, even though the best current rationale for fit indices is based on a noncentrality parameter T^0 and the associated population misfit τ, fit indices were developed historically before the ncp rationale became evident (Bentler, 1990a; McDonald, 1989; McDonald & Marsh, 1990; Steiger, 1989). In particular, the normed and non-normed fit indices (Bentler & Bonett, 1980; Tucker & Lewis, 1973) were developed without explicit consideration of noncentrality. Bentler (1990a) provided the population parameters corresponding to these indices, defined via noncentrality of the χ^2 distribution.

Absolute Fit

The most popular measure of absolute fit is Steiger's root mean square error of approximation, typically denoted as RMSEA (Browne & Cudeck, 1993; Steiger, 1989, 2000). It is defined as

$$RMSEA^0 = \sqrt{\tau / \mathrm{d})}\,, \tag{14.13}$$

that is, *the square root of population misfit per degree of freedom*. Thus for the same degree of misfit, models with larger degrees of freedom will have smaller values of RMSEA. Since smaller values indicate better fit, RMSEA favors parsimonious models. In practice, an estimator of (14.13) is computed as

$$RMSEA = \sqrt{\max\{(T - \mathrm{d}), 0\}/(\mathrm{nd})}\,. \tag{14.14}$$

Values close to zero are desired. It is often said that values of .05 or less indicate close fit of a model, values of .08 or less indicate a reasonable error of approximation, and .10 or larger indicate bad fit (e.g., Browne & Cudeck, 1993). McCallum, Browne, and Sugawara (1996) consider values between .08 and .10 to represent mediocre fit. Hu and Bentler (1998) reported that with ML, values of .06 or less indicated good fit; they also reported that this

cutoff value overrejects true models at small sample sizes. Complicating matters even more, Breivik and Olsson (2001) and Kenny and McCoach (2003) found that RMSEA decreases as the number of variables increases, "regardless of the type of misspecification and regardless of the degree of misspecification" (Kenny & McCoach, 2003, p. 347), with the latter also noting that in their study "the degree of decline is rather substantial" and "as p approaches infinity, the RMSEA goes to zero."

A useful feature is the confidence interval available for RMSEA (Browne & Cudeck, 1993), based again on the assumption that the noncentral χ^2 distribution accurately describes any misspecification. EQS computes only the 90% confidence interval for the index (14.14) as

```
ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)    =      .071
90% CONFIDENCE INTERVAL OF RMSEA  (        .037,        .103)
```

The confidence interval becomes narrower as sample size increases. An excellently fitting model would have the lower point of the confidence interval be zero. As noted by Browne and Cudeck (p. 145), in that case the standard test of model fit using T "would not reject the null hypothesis at the 5% level. Thus the confidence interval contains the information provided by the corresponding goodness-of-fit test." The upper end of the interval reminds us that fit may not be quite as good as the point estimate implies. Although derived from asymptotic considerations, Curran, Bollen, Chen, Paxton, and Kirby (2003) found that with normal data and ML estimation, under their conditions the confidence intervals were accurate for samples of size 200 or greater. Measurement of differences between RMSEAs of two models is discussed by Browne and du Toit (1992), and confidence intervals are discussed by Raykov (2001d).

Actually, the decline in size of the RMSEA with increasing number of variables p described by Kenny and McCoach (2003) had already been noted by Steiger (1998) in another context. He was not considering the situation of increasing p (or number of covariances) in a single sample, but rather a parallel situation arising from multisample analysis. Suppose there are m samples and a model with d degrees of freedom is fit. In this situation, the number of variables to be modeled (hence the number of covariances) increases substantially with additional samples, and RMSEA as defined by (14.13)-(14.14) decreases with the number of samples. Steiger redefined RMSEA to avoid this problem.

A typical multisample test statistic is given by (7.5) $T = \mathrm{n}_1\hat{Q}_1 + \mathrm{n}_2\hat{Q}_2 + ... + \mathrm{n}_m\hat{Q}_m$ where the separate $\hat{Q}_g$ (see 7.4) represent the chosen fit function (AGLS, ML, etc.) in each of the m samples. The population noncentrality parameter is given by

$$T^0 = \mathrm{n}_1 Q_1^0 + \mathrm{n}_2 Q_2^0 + ... + \mathrm{n}_m Q_m^0 . \qquad (14.15)$$

As usual, we need to standardize, so we take standardized noncentrality or population misfit as $\tau = T^0 / \mathrm{n}$ where $\mathrm{n} = N - G = \mathrm{n}_1 + \mathrm{n}_2 + ... + \mathrm{n}_\mathrm{m}$. As a result, the population RMSEA defined in (14.13), and its estimator (14.14), can be directly applied. However, Steiger (1998) redefined the RMSEA as the *the square root of population misfit per **average** degrees of freedom.* That is, in place of d in (14.12), he uses d/m. A bit of algebra shows that this multiple group (MG) coefficient can be written as

$$RMSEA_{MG}^0 = \sqrt{\mathrm{m}}\sqrt{\tau / \mathrm{d})} = \sqrt{\mathrm{m}}\,RMSEA^0 . \qquad (14.16)$$

A special case occurs when m=1, and the two indices (14.13) and (14.16) are the same. EQS now accepts the definition (14.16) and, with multiple groups, replaces (14.14) with

$$RMSEA_{MG} = \sqrt{\mathrm{m}}\sqrt{\max\{(T - \mathrm{d}), 0\}/(\mathrm{nd})} . \qquad (14.17)$$

Dudgeon (2004) developed the implications of this modification for other indices and applications.

An important feature of Steiger's RMSEA is that it can be applied to models with mean structures as well as to covariance structures alone. The defining formulae given above apply directly. When there are mean structures, EQS prints an approximate RMSEA for the covariance structure part of the model, and the ordinary mean structure RMSEA.

Another useful index based on noncentrality is McDonald's (1989) fit index, defined as $\exp(-.5\tau)$, with estimator

$$\text{MFI} = \exp\{\text{-.5(T-d)/n}\}. \tag{14.18}$$

This index typically ranges between zero and one with larger values indicating better fit. It may exceed 1.0 as is illustrated in the following EQS printout.

```
MCDONALD'S  (MFI) FIT INDEX           =     1.023
```

In this case, fit is interpreted to be essentially perfect. It typically behaves well, like other noncentrality-based indices, but can be affected by irrelevant sources of variance (Hutchinson & Olmos, 1998). With n being the total sample size across all groups, MFI is directly applicable to multiple groups. It also applies to mean structures. An illustrative printout of MFI and RMSEA in a mean structure model is the following.

```
FIT INDICES (BASED ON COVARIANCE MATRIX AND MEANS)
  ----------
  MCDONALD'S        (MFI) FIT INDEX   =     0.944
  ROOT MEAN-SQUARE ERROR OF APPROXIMATION (RMSEA)    =     0.099
  90% CONFIDENCE INTERVAL OF RMSEA  (       0.000,         0.188)
```

Some additional widely used measures of fit are Jöreskog and Sörbom's (1981) goodness of fit and adjusted goodness of fit indices

$$\begin{aligned} GFI_{ML} &= 1-tr\{(\hat{\Sigma}^{-1}S-I)^2\}/tr(\hat{\Sigma}^{-1}S)^2 \\ GFI_{GLS} &= 1 - tr(S^{-1}\hat{\Sigma}-I)^2/p \\ AGFI &= 1-\{p(p+1)/2d\}(1-GFI) \end{aligned} \quad . \tag{14.19}$$

These are printed out in EQS as follows.

```
JORESKOG-SORBOM'S  GFI FIT INDEX           =      .973
JORESKOG-SORBOM'S AGFI FIT INDEX           =      .866
```

Notice that the GFI index is computed in a slightly different way in ML versus GLS estimation. The adjusted GFI index, AGFI, adjusts for degrees of freedom in a similar way as does adjusted R^2. GFI is an estimator of $\gamma = p/\{p+2\tau\}$ as was noted by Steiger (1989, 1990) and Maiti and Mukherjee (1991). The natural approximation $\hat{\gamma} = p/\{p+2(T-d)/n\}$ seems to perform better than GFI, but it is not implemented in EQS. In a multiple group context, GFI in (14.19) is computed for each group and the final GFI is given as a weighted sum of these GFI indices, using weights $n_g/(N-G)$. The correction that takes GFI into AGFI, of course, reflects the sum of p* across groups, where p* = p(p+1)/2.

When the ADF or AGLS method is used, EQS computes two additional R^2-like indices as proposed by Bentler (1983, eqs. 3.5-3.8). The basic index depends on the fit function (5.1) as well as the value of the fit function when Σ is the null matrix, and the adjusted index corrects for degrees of freedom as does AGFI. That is,

$$\begin{aligned} f &= 1-\frac{\hat{Q}}{s'Ws} = 1-\frac{(s-\sigma(\hat{\theta}))'W(s-\sigma(\hat{\theta}))}{s'Ws} \\ \bar{f} &= 1-\{p(p+1)/2d\}(1-f) \end{aligned} \quad . \tag{14.20}$$

In fact, these are generalizations of the GFI indices (14.19). That is, when W takes on its normal theory specialization under ML and GLS, then (14.20) reduces to the GFI indices in (14.19). In the population, f can be written as $f^0 = 1-\tau/\sigma^{0\prime}W_{pop}\sigma^0$. See Tanaka and Huba (1985) for additional discussion. In EQS, the indices (14.20) are printed out as follows.

```
AGLS FIT INDEX                     =      .894
AGLS ADJUSTED FIT INDEX            =      .812
```

In the case of multiple groups $f = 1 - \frac{T_{AGLS}}{\sum n_g (s_g' W_g s_g)}$, where T_{AGLS} is the test statistic given in (7.5), and $\bar{f}$ is computed accordingly based on the sum of p* data points.

In concluding our discussion of absolute fit indices, it should be noted that their actual range in practice can be severely truncated. Those that presumably vary between 0 and 1 do not really do so in any meaningful sense. For example, if we evaluate the uncorrelated variables model by ML on the manul7.ess data, we find that MFI = .617, GFI = .696, and AGFI = .575. Although not printed out by EQS, $\hat{\gamma} = .753$. Similarly, when evaluating this model by AGLS, f = .844, $\bar{f}$ = .781. These fit index results imply that the model – in which no variables have any effect on any other variables -- is really not too bad. In contrast, SRMR = .250 and RMSEA = .256 for ML, and SRMR = .372 and RMSEA = .220 for AGLS, indicating a terrible fit. It is for this reason that we recommend SRMR, and among absolute fit indices, RMSEA. As you will see, the 0-1 indices discussed next, based on relative fit, are not fooled. They take on their worst possible value of 0.0 in this situation.

Nested Models and Relative Fit

Just as we had to discuss noncentrality and population misfit before we could meaningfully discuss absolute fit indices, we need to discuss nested models in order to make sense of measures of relative or incremental fit. We noted above that the ncp T^0 as well as population misfit τ increase with greater misspecification, so we could imagine a series of worse and worse models with greater and greater misspecification. Is there no endpoint to how bad a model can be? Is there no natural way to describe the worst possible model that we might consider, i.e., what the largest possible values of T^0 or τ might be? Stated differently, can we define a $\Sigma(\theta)$ that is maximally distant from Σ in the metric of the fit function? Bentler and Bonett (1980), building on the ideas of Tucker and Lewis (1973), proposed using the model Σ_D of uncorrelated variables as the worst possible model. In such a model, there are no latent variables, nor any other kind of influences of any variables on any other variables. As you will see, their approach allows us to define the fit of the current model relative to the fit of this worst possible model. Before turning to the details, let us consider not just these two models, but rather a whole continuum of models.

In evaluating comparative model fit, it is helpful to focus on more than one pair of models. Typically, there will exist a series of nested models

$$\mathrm{M}_i, \ldots, \mathrm{M}_j, \ldots, \mathrm{M}_k, \ldots, \mathrm{M}_s \tag{14.21}$$

beginning with the most restricted model M_i that one might consider, and extending to the least restricted model. The least restricted model is always the saturated model M_s. M_s contains as many parameters in θ as there are non-redundant elements in Σ, and hence this model, with no falsifiable structural hypothesis, always fits perfectly. The models in (14.21) are assumed to be nested, so that a more restricted model is obtained by imposing constraints on a more general model. For example, M_j may be obtained from M_k by fixing one or more free parameters in M_k to some *a priori* value, often zero. We may write this nesting structure as

$$\mathrm{M}_i \subseteq \mathrm{M}_j \subseteq \mathrm{M}_k \subseteq \mathrm{M}_s \; . \tag{14.22}$$

The sequence of models with structure (14.22) does not necessarily include all possible models that might be considered in a given situation. Models of different types, for example, a path model with measured variables and a confirmatory factor model with two factors, are not nested, and hence cannot be lined up as in (14.22). When we do have a sequence of nested models, Bentler and Bonett (1980) proposed that in covariance structure analysis, typically the most restricted model M_i that one might consider in practice is that of uncorrelated measured variables with diagonal covariance matrix Σ_D. M_i is sometimes called a *null model*, indicating no mutual influences among variables, or a *baseline model*. If the measured variables that generate Σ are multivariate normally distributed, for which uncorrelatedness is identical to independence, we could call M_i the *independence model* (hence the subscript *i*). This model is not necessarily the most restricted model that should ever be considered, but in typical practice it is likely to be the most restricted model that would reasonably be considered. A more restricted model than the independence model is a model with equal variances and no covariances; another is a model containing no free parameters. While such models could possibly be appropriate baseline models in some situations, these models

would rarely describe real data and are thus not recommended as baseline models for routine use. EQS typically takes the uncorrelated variables model as its baseline model. As noted by Bentler and Bonett, when considering models in a nested sequence (14.22), it may sometimes be necessary to use a more restricted baseline model than the model of uncorrelated variables. For example, if model M_k is a confirmatory factor model and contains equality restrictions for the error variances, then the model that allows Σ_D to have all free parameters in its diagonal, the uncorrelated variables model, is not an appropriate baseline model as it is not nested in this particular M_k. An appropriate baseline model also should maintain those equality restrictions for the variances. Procedures and issues in the use of more restricted null models are spelled out in an excellent article by Widaman and Thompson (2003). They note, for example, that in some situations there may be more than one appropriate null model. EQS analyzes models, and, for many model types, automatically adjusts the baseline model so as to maintain model nesting. In such a case, EQS adds a statement such as

```
INDEPENDENCE MODEL HAS BEEN MODIFIED TO INCLUDE  4
CONSTRAINTS FROM THE SPECIFIED MODEL.
```

to the information about the fit of the independence model, and reminds you, when describing relative fit indices that these are

```
BASED ON MODIFIED INDEPENDENCE MODEL
```

If you are worried about whether EQS handled the nesting appropriately, you should set up your own baseline model and use its results in your own fit index computations.

Other proposals for baseline models, usually of a less restricted form, have been made. Sobel and Bohrnstedt (1985) recommended the use of a model that would include estimates of parameters such as factor loadings that have been verified in previous research. Rigdon (1996) considered the baseline model of independence to be problematic, and proposed (1998b) to use an equal-correlation model as a more meaningful baseline model; this proposal generated debate (Marsh, 1998b; Rigdon, 1998c). In the absence of consensus for a better general baseline model, these alternatives are not currently implemented in EQS.

The reason for considering nested models (14.22) is that, for any given choice of fit function (such as T_{ML}), we can now order the test statistics corresponding to the various models

$$T_i \geq ,\ldots,\geq T_j \geq ,\ldots,\geq T_k \geq ,\ldots,\geq T_s = 0 . \tag{14.23}$$

More restricted models can never yield smaller fit functions (if they do, there is a convergence problem!). We can similarly order the corresponding degrees of freedom $d_i,\ldots,d_j,\ldots,d_k,\ldots,d_s$ as $d_i > d_j > d_k > d_s$. Here T_i is the "chi-square value" based on d_i degrees of freedom obtained by fitting model M_i to S; T_j and d_j are the corresponding values obtained for model M_j; T_k and d_k correspond to M_k; and T_s and d_s correspond to M_s. The saturated model M_s, not necessarily unique, has the characteristic that $T_s = 0$ and $d_s = 0$. In the sequence of tests (14.23), the independence model has the worst fit, intermediate models have intermediate degrees of fit, and the saturated model has a perfect fit. These models and test statistics are based on parameter vectors $\theta_i,\ldots,\theta_j,\ldots,\theta_k,\ldots,\theta_s$, the model matrices $\Sigma_i = \Sigma(\theta_i),\ldots,\Sigma_j,\ldots,\Sigma_k,\ldots,\Sigma_s$ as well as their estimated values $\hat{\Sigma}_i,\ldots,\hat{\Sigma}_j,\ldots,\hat{\Sigma}_k,\ldots,\hat{\Sigma}_s$. In covariance structure analysis, under the model of uncorrelated variables, often $\hat{\Sigma}_i = \hat{\Sigma}_D = S_D$ = diag(S), but no matter what the estimator, no covariances are accounted for by the model. Intermediate model matrices $\hat{\Sigma}_j$ and $\hat{\Sigma}_t$ account for the off-diagonal elements of S, the covariances, to varying degrees. At the other extreme, $\hat{\Sigma}_s = S$, so that the model perfectly explains all the variances and covariances.

When the various models we consider are nested, as in (14.22), the population noncentrality parameters are ordered as in (14.23). That is,

$$T_i^0 \geq ,\ldots,\geq T_j^0 \geq ,\ldots,\geq T_k^0 \geq ,\ldots,\geq T_s^0 = 0 . \tag{14.24}$$

Similarly, the values of the population fit functions or discrepancy functions also are ordered

$$\tau_i \geq ,\ldots,\geq \tau_j \geq ,\ldots,\geq \tau_k \geq ,\ldots,\geq \tau_s = 0 \,. \qquad (14.25)$$

These orderings provide the basis for defining incremental or comparative fit indices.

Comparative Fit

Incremental or relative fit indices evaluate the adequacy of a particular model M_k in relation to the end point models M_i and M_s in the sequence of nested models (14.2). We would like to have an index whose values for different models line up on a continuum from zero to one. Thus we can imagine the line

0.0			1.0	(14.26)
M_i	M_j	M_k	M_s	

with the placement of the independence model at 0, the saturated model at 1, and various possible alternatives somewhere along the line. The usual model chi-square test evaluates whether the distance from M_k to M_s is significantly larger than zero; rejection of the null hypothesis says that the gap between these two models is not due to chance. But even if the distance M_k - M_s is "statistically" significant, implying technical model rejection, it is possible that that distance is actually quite small. Stated differently, our model M_k, although not perfect, may be so good that we have substantially explained the covariances among the variables as would be reflected in a quite large distance M_i - M_k as compared to the length of the entire line M_i - M_s. If the distance $M_i - M_k$ is very large (say, over .9), and hence the remaining distance M_k - M_s quite small, we can conclude that our model is quite good at accounting for the variances and covariances, even if it is not perfect as indicated by the significance test. And we would like our fit index to reflect how well we are doing.

An obvious additional question involves the comparison of two models, say M_j and M_k.. We know that if we add restrictions to M_k, the fit will get worse. But how much worse? Perhaps the distance to M_j is very small, so that the decrement in fit is quite tiny when referred to the 0-1 scale in (14.26). Thus the difference between incremental fit indices provides another way to interpret a χ^2 difference test. It is possible that the χ^2 test may have high power to reject the null hypothesis that the gap between M_j and M_k. is zero. But the gap may actually be very small in terms of relative loss of fit. See Bentler (1990a) for further details, and on proposals for how to use Wald and Lagrange Multiplier tests in this context.

Although our preferred comparative fit indices have a noncentrality rationale, the idea of indices defined on fit functions is easiest to explain. Bentler and Bonett (1980) used the test statistics T in (14.23) to define the normed fit index (NFI). Since T_i is the worst possible value of T, and our model of interest yields T_k, they proposed to measure the distance from T_i to T_k relative to the starting point T_i. The defined

$$NFI = \frac{T_i - T_k}{T_i} = 1 - \frac{T_k}{T_i}. \qquad (14.27)$$

which equals 0 when $T_k = T_i$, equals 1.0 when $T_k = 0$, and is in the (0-1) range otherwise, with higher values indicating better fit. Because of the strict 0-1 range, they called this index the normed fit index. Notice that this is a descriptive index that depends on the fit function used, but there is no requirement that the distribution of T be described by the χ^2 distribution or any other distribution. This is an advantage of the index. A disadvantage of NFI, as has been known for a long time (e.g., Bearden, Sharma, & Teel, 1982), is that it is affected by sample size. In small samples, it may not reach 1.0 when the model is correct. This can occur because the expected value of T_k may be greater than 0, for example, when T_k is a χ^2_d variate, $E(T_k) = d_k$.

James, Mulaik and Brett (1982, p. 155) suggested multiplying (14.27) by d_k/d_i to yield an index to reflect model parsimony. This recommendation is not followed in EQS, because we feel that parsimony is best treated as an issue that should be evaluated separately from sheer model fit. Nonetheless, there are virtues to parsimony, and we recommend that you read about parsimony in such sources as Mulaik, James, Van Alstine, Bennett, Lind, and Stillwell (1989), Bentler and Mooijaart (1989), and Cheung and Rensvold (2001). For example, increasing the

number of valid restrictions in a model (e.g., moving from M_k to M_j) increases precision of parameter estimates, i.e., yields smaller standard errors. Parsimony is also an issue for information-based statistics, described below.

Bentler and Bonett (1980) also recognized that degrees of freedom of the baseline model and model of interest may be helpful in defining model fit. They proposed the modified index

$$NNFI = \frac{T_i - d_i d_k^{-1} T_k}{T_i - d_i} \tag{14.28}$$

called the non-normed fit index. They built this index on one developed by Tucker and Lewis (1973) for evaluating the "reliability" of exploratory factor analysis models estimated by maximum likelihood (hence the index is sometimes called the Tucker-Lewis index). The degrees of freedom adjustment in the index was designed to improve its performance near 1.0, not necessarily to permit the index to reflect other model features such as parsimony. Yet it does have a leaning toward preferring parsimonious models; see below. When $E(T_k) = d_k$, then NNFI = 1.0, thus obviating a major difficulty with NFI. However, NNFI can fall outside the (0,1) range. It will be negative when $d_i d_k^{-1} T_k > T_i$, since usually $T_i \gg d_i$. It will exceed 1.0 when $T_k < d_k$. In fact, the index can be anomalously small, especially in small samples, implying a terrible fit when other indices suggest an acceptable model fit (Anderson & Gerbing, 1984). The variance of NNFI is, in sampling studies, substantially larger than the variance of NFI. This is a negative feature.

NNFI has the major advantage of reflecting model fit very well at all sample sizes (for older research, see e.g., Anderson & Gerbing, 1984; Marsh et al., 1988; Wheaton, 1987; for some recent research, Fan & Sivo, 2005; Kenny & McCoach, 2003; Marsh, Hau, & Wen, 2004). A modification relating to sample size was proposed by Bollen (1986), but it did not solve the major problem of variability in the index. This problem was addressed by Bollen (1989a). He defined the incremental fit index

$$IFI = \frac{T_i - T_k}{T_i - d_k}, \tag{14.29}$$

and showed that it behaved like NNFI in a sampling study, but had a smaller sampling variance. This index is discussed further below.

Having presented the NFI (and a few variants of it) based on the sequence among sample test statistics (14.23), we will now discuss fit indices based on the corresponding population counterpart sequences (14.24)-(14.25), that is, using population noncentrality or population misfit. Specifically, let us substitute the population test statistics or fit functions in place of the sample test statistics in (14.27). This yields the population comparative fit index (Bentler, 1990a)

$$CFI^0 = \frac{T_i^0 - T_k^0}{T_i^0} = \frac{\tau_i - \tau_k}{\tau_i} = 1 - \frac{\tau_k}{\tau_i}. \tag{14.30}$$

Then an interesting question is whether the NFI (14.27) is the best possible estimator of (14.30). Based on previously described estimates of noncentrality, the answer is no. If we use the noncentrality parameter estimator $\hat{T}^0 = \max\{(T - d), 0\}$, we obtain the sample Comparative Fit Index

$$CFI = \frac{\hat{T}_i^0 - \hat{T}_k^0}{\hat{T}_i^0} = 1 - \frac{\hat{T}_k^0}{\hat{T}_i^0} = 1 - \frac{\max\{(T_k - d_k), 0\}}{\max\{(T_i - d_i), (T_k - d_k), 0\}}. \tag{14.31}$$

Notice that the denominator of the subtractive term in (14.31) could be just $\max\{(T_i - d_i), 0\}$. This has been replaced by $\max\{(T_i - d_i), (T_k - d_k), 0\}$ to deal with the unlikely possibility that the estimated noncentrality of model M_k is larger than the estimated noncentrality of model M_i. The estimator is consistent with our nesting assumptions (14.24)-(14.25) that do not allow such a reversal in the population. As a result, $0 \le CFI \le 1$. In addition to (14.31), Bentler (1990a) noted that it would be possible to simply use $\hat{T}^0 = (T - d)$ as an estimated noncentrality parameter. He called the resulting index the Fit Index. It is given by

$$FI = 1 - \frac{T_k - d_k}{T_i - d_i} \ . \qquad (14.32)$$

McDonald and Marsh (1990) developed (14.32) at the about same time and called it the relative noncentrality index (RNI). EQS does not print this index as it can be outside the 0-1 range. An illustrative printout of these various indices is the following.

```
BENTLER-BONETT     NORMED FIT INDEX =     0.736
BENTLER-BONETT NON-NORMED FIT INDEX =     0.662
COMPARATIVE FIT INDEX (CFI)         =     0.820
BOLLEN'S          (IFI) FIT INDEX   =     0.843
```

Below we show that under a wide variety of modeling situations, these indices (especially the last 3) correlate in the high .9's. Yet, as you see in this real example, this does not prevent them from taking on quite different values when a model is misspecified.

It is clear from (14.30) that CFI^0 is invariant to a rescaling of the noncentrality parameters by a nonzero constant. It is equivalently defined in terms of population misfit, which has no sample size term in it. If for some odd reason it were necessary to compare baseline and model of interest using different sample sizes, this would have to be done with estimates of population misfit in (24.30).

With AGLS, CFI is computed not only for the ordinary AGLS chi-square test statistic, but also for the Yuan-Bentler corrected AGLS test statistic (for which the independence model chi-square test is not printed). This is printed out as follows.

```
AGLS CORRECTED COMPARATIVE FIT INDEX=      .970
```

All of these relative fit indices have identical definitions in multiple group contexts. They are not appropriate for mean structures, as noted below.

Relations Among Fit Indices

We noted that CFI and FI are estimators of population misfit. It turns out that NFI, NNFI, and IFI can be seen to estimate population misfit as well (Bentler, 1990a; Bollen, 1989a). As sample size increases indefinitely, NFI and IFI have the same asymptotic definition as CFI, given in (14.30). As a result, in large samples, their values will be essentially identical. However, NFI does have a downward bias in small samples as noted earlier. On the other hand, NNFI has a different probability limit

$$NNFI^0 = 1 - \frac{d_i \tau_k}{d_k \tau_i} = 1 - \frac{\tau_k / d_k}{\tau_i / d_i} . \qquad (14.33)$$

Clearly, NNFI can be interpreted as estimating the relative reduction in noncentrality per degree of freedom, i.e., it does have a parsimony rationale. Unlike CFI, it does not just measure relative population misfit.

FI in (14.32) is linearly related to NNFI in (14.28) (Bentler, 1990a). Specifically,

$$FI = \alpha + \beta(NNFI) \qquad (14.34)$$

where $\beta = d_k / d_i$ and $\alpha = 1 - \beta$. This implies that cutoff values for these two indices *can* be translated into each other. This translation shows that for parsimonious models, FI (and hence CFI if 0<FI<1) and NNFI will be close, but for non-parsimonious models, NNFI will be a good amount below FI. For example, in a model with 30 variables and 1 factor, if FI = .9, NNFI = .894, almost no difference; but with 4 exploratory factors, if FI = .9, NNFI = .853. This explains the downward bias in NNFI observed by Anderson and Gerbing (1984). FI also has a smaller standard error or sampling variance than NNFI, since $\text{var}(FI) = \beta^2 \text{var}(NNFI)$ and $0 < \beta < 1$.

It also may be interesting to relate two indices of different types. We will consider a special case for some insight. For a conventional covariance structure model $\Sigma(\theta)$, if for any parameter vector θ and positive constant α there

exists a parameter vector θ^* such that $\Sigma(\theta^*) = \alpha\Sigma(\theta)$, then $\Sigma(\theta)$ is called "invariant under a constant scaling factor" or ICSF. Most models studied in practice are ICSF (Browne, 1984). As noted by Browne (1974), ICSF models estimated by ML have the property $tr(S\Sigma(\theta)^{-1}) = \mathrm{p}$. In such a case, the population ML fit function $F^0 = \ln|\Sigma(\theta)| + tr(\Sigma\Sigma(\theta)^{-1}) - \ln|\Sigma| - \mathrm{p}$ simplifies to $F^0 = \ln|\Sigma(\theta)| - \ln|\Sigma|$. As a result, (14.30) becomes

$$CFI^0_{ML,ICSF} = 1 - \frac{\tau_k}{\tau_i} = 1 - \frac{\ln|\Sigma(\theta)| - \ln|\Sigma|}{\ln|\Sigma_D| - \ln|\Sigma|} = \frac{\ln|\Sigma_D| - \ln|\Sigma(\theta)|}{\ln|\Sigma_D| - \ln|\Sigma|}, \tag{14.35}$$

where Σ_D is any baseline model. Kenny and McCoach (2003) studied these types of models and further supposed that the population covariance matrix Σ is standardized, that is, $\Sigma = P$, the population correlation matrix. In such a case, with the baseline model of independent variables $\Sigma_D = I$, and furthermore $|\Sigma_D| = 1.0$ and $\ln|\Sigma_D| = 0$. In this situation, (14.35) simplifies to

$$CFI^0_{ML,ICSF} = \frac{\ln|P(\theta)|}{\ln|P|}. \tag{14.36}$$

This is Kenny and McCoach's eq. (10). Notice that under these conditions, the CFI seems to no longer depend on the independence model, and it looks like an absolute fit index. This allows us to do a little algebra to relate the population RMSEA to CFI in. Under ISCF,

$$RMSEA^0_{ML,ISCF} = \sqrt{\frac{\ln|P(\theta)| - \ln|P|}{d_k}}. \tag{14.37}$$

given by Kenny McCoach's (eq. 12) (their equation contains a typo). Thus CFI is a ratio of log-determinants, while RMSEA involves their difference, their degrees of freedom, and a square root operation. As noted by Yuan (2005), some of the good behavior of RMSEA may be due to its use of this square root (which of course, could be added to (14.36) to define a new index). It follows from (14.36) and (14.37) that these indices are related nonlinearly in this setup by

$$CFI^0_{ML,ICSF} = 1 + \left(\frac{d_k}{\ln|P|}\right)(RMSEA^0_{ML,ISCF})^2, \tag{14.38}$$

which depends on the correlations among the variables in P as well as on d_k. In view of this relation, clearly there can be no cutoff for a good model as judged by RMSEA (say, .05) that can map into any comparable cutoff for CFI that is independent of the data and model.

Rationale for Choice Among Fit Indices

In many fields, if one has multiple indicators of some target construct, a natural procedure is to obtain an optimal estimate of that construct. Usually this is done by obtaining a factor score estimate or simply summing up indicators to obtain a composite score that is better than any one of its parts (see Reliability section above). And so it would be possible to create an overall measure of fit based on a variety of specific indices, thus not requiring a choice of "the best one." However, this has not yet been done. Then how can one make a rational choice of index?

Of course, it is possible that fit indices are multidimensional, that is, they provide somewhat different information, and hence using one index, or even a composite across several indices, may not be the best way to proceed. Hu and Bentler (1999) suggested using cutoff values on two indices simultaneously, but their recommendation has not been further studied (although their approach has been criticized, see Marsh, Hau & Wen, 2004). More generally, the latent dimensionality of fit indices has hardly been studied in the field of SEM research. That is, considering the empirical performance of fit indices under the types of conditions where such indices are liable to be consulted, e.g., with various model types and sizes, estimation methods, sample sizes, correctly specified and misspecified models, degrees of violation of assumptions, etc., to what extent are these indices correlated and what is their latent

structure? Bentler (1990a) found close to 1.0 correlations among some indices under correct and incorrect specification, but it seems that Hu and Bentler (1998) were the first to present detailed data on the correlations among various indices across a variety of conditions. They found that for ML, correlations among 15 indices ranged from .29 to 1.0, with many correlations of .999. The range was .44 to 1.0 for GLS, and .20 to 1.0 for AGLS. Doing a principal components analysis of each of the matrices in their Table 3, the first unrotated component accounts for 81%, 77%, and 67%, respectively, of the variance in the variables. Factor analysis would yield a similar result, since the communalities of these variables are close to 1. Clearly a huge general factor exists, and on this basis, perhaps only one index, such as one with the highest loading on this factor, would be sufficient. Similar results for ML can be extracted from Table 3 in Fan and Sivo (2005). Under one condition, the first principal component accounted for 84% of the variance; under another condition, 89%. Beauducel and Wittman (2005) reported 57% and 43% of variance accounted for by the first component. These results are consistent with Ogasawara's (2001b) theoretical results. He showed that under certain conditions (multivariate normality, slight misspecification, particular models) many fit indices correlate .97 or above. This is remarkably high, given that many of the indices are related nonlinearly to each other.

However large the first component may be, in our opinion this is not the whole story. For example, in Hu and Bentler's Table 3, even though there are many correlations in the high .99's for many indices with ML, GLS, and AGLS, in each of the tables there are some quite small correlations. In fact, some of our favorite indices correlated not too highly among themselves. For example, SRMR correlated only -.40, -.59, -.64 with CFI, and .45, .63, .58 with RMSEA; the latter two correlated -.95, -.94, -.67 (under ML, GLS, AGLS estimation). As a result, we decided to subject their 3 matrices to a principal components analysis with 3 orthogonally rotated components. This is not the place to report detailed results, but the following general results were observed:

- IFI, FI, CFI, and NNFI (even with its parsimony adjustment) always had the largest loadings (.95-1.0) on one of the factors. RMSEA (~.9) and MFI (~.97) also loaded on this factor for ML and GLS, but with AGLS, RMSEA and MFI formed a separate factor with some other indices not discussed here.
- SRMR with ML, GLS loaded highly on a separate factor with another index not discussed here. With AGLS, it loaded on a separate factor with GFI and AGFI.
- NFI, GFI, AGFI, and some other indices not discussed here formed a separate factor for ML and GLS. These are indices affected by sample size variation (Hu & Bentler, 1998; Fan & Sivo, 2005).

Very similar results were obtained by component analysis of Fan and Sivo's ML simulation correlations given in their Table 3. In one of their conditions similar to Hu and Bentler's, the first component accounted for 84% of the variance, and virtually identical results to those summarized above were obtained. In a condition more dissimilar to those of Hu and Bentler, the first component accounted for 89% of the variance, but again the loading pattern can be described by the summary above.

You may use these results to inform your choice of what index or indices to report. In addition to our general recommendation above to report an absolute fit index, a relative fit index, and SRMR, these results further indicate that several of the incremental fit indices can be used interchangeably. But remember that high correlations do not imply that the same cutoff values for different indices yield identical decision rules about model quality. For example, (14.34) showed that using the same cutoff for FI and NNFI does not make sense, and (14.38) showed that often there is no simple way to relate CFI and RMSEA.

Fit Indices with Structured Means Models

A structured means model uses the sample covariance matrix S and the vector of sample means $\bar{x}$ to obtain the corresponding model matrices $\Sigma(\hat{\theta})$ and $\hat{\mu}$, based on some estimation method. In EQS, only RMSEA and MFI are computed as fit indices on a structured means model in the standard way. See the examples between (14.14) and (14.18). Although the f index in (14.20) could be defined to include means as was allowed by Bentler (1983a), we do not do so, and all other fit indices reported measure only the closeness of S and $\Sigma(\hat{\theta})$. They ignore information on the fit of the

mean structure. You are reminded about this with the heading FIT INDICES (BASED ON COVARIANCE MATRIX ONLY, NOT THE MEANS) that precedes the covariance measures of fit.

The reason this is done in EQS is that there is no natural null or baseline model for the mean structure, needed in the computation of relative fit indices. When considering the sequence of nested models along our 0-1 line given in (14.26), the baseline model Σ_D of uncorrelated variables seems to be a logical choice to represent the worst possible model $\Sigma(\theta)$ for Σ that one might entertain in practice. But when considering a mean structure $\mu(\theta)$ for μ, what is the worst possible $\mu(\theta)$? The possibility $\mu = 0$ seems to be unnatural, as measures of lack of fit would then be heavily influenced by those variables that happen to have large means. In our opinion, no natural worst possible mean structure model can be defined. Hence there seems to be no meaningful nested sequence of mean structure models leading to a unit-length line of fit parallel to (14.26). In particular, we cannot define the left endpoint of that line with a meaningful T_i or T_i^0 to be used in computation of relative fit indices for mean structures.

A fit index based only on the covariance structure has to ignore $\hat{\mu}$ and provide a meaningful measure of fit for the covariance structure. However, the test statistic T that is computed, as well as the associated degrees of freedom, depend on both the fit of $\Sigma(\hat{\theta})$ to S as well was of $\hat{\mu}$ to $\bar{x}$. The estimation method provides no specific quantities for the covariance part of the model alone. In EQS, these are approximated as follows:

1. The fit function, or χ^2 test, for the covariance structure alone is based on the standard fit function for the method involved, using S and $\Sigma(\hat{\theta})$ from the structured means model. This yields T_h, which we will designate as T_C.
2. The df for the covariance structure alone, d_C, is computed from d_{SM}, the df for the overall model, as $d_C = d_{SM} - p + m$, where p is the number of sample means and m is the number of independent mean parameters. The latter is obtained in EQS as the number of paths from V999 to another variable minus the number of independent constraints on such paths.
3. The model $\hat{\Sigma}_D = \hat{\Sigma}_i$ of uncorrelated variables is obtained as usual, as is the associated degrees of freedom d_i.
4. The quantities in #1-#3 are used in the usual definitions of various fit indices.

You should recognize that although this procedure yields fit indices for the covariance structure, these computed components cannot be used to provide actual statistical tests on the covariance structure. This is because, in reality, the covariance and mean components of fit and df may not be partitionable into independent parts. In fact, providing a unique definition of the df for the covariance structure may not be possible. For example, there may be constraints among mean and covariance parameters that are not taken into account in computing d_C. However, the above methodology should provide fit indices that reasonably well reflect the quality of the covariance part of a structured means model.

Compute Your Own Fit Index

You should not forget that the various fit indices printed in EQS may not cover your special situation. First, a number of additional indices exist in the literature that are not automatically computed by EQS. Second, even within the class of printed indices, only a few selected chi-square test statistics are used in the computation of these indices. For example, fit indices for ML are based on the ML fit function, but no fit indices are computed for the equally valid normal theory reweighted least squares function. Or, under robust estimation, fit indices based on the Satorra-Bentler scaled chi-square are computed, but none are computed for the Yuan-Bentler residual-based chi-square statistic. You will need to do your own computations using the formulae provided above. With absolute fit indices, you only need to use the printout from your computer run. In other cases, you may need to set up the independence model to obtain the comparable baseline test statistic. This model can be easily specified, e.g., for 10 variables:

```
/EQUATIONS
V1=E1; V2=E2;...; V10=E10;
/VARIANCES
E1 TO E10 = *;
```

But remember that the independence model may need to be adjusted if it is not nested in the model of interest.

Global Versus Local Fit

You should remember that, like the χ^2 test itself, fit indices are concerned with the fit of the entire model to all the covariances and possibly means. For any degree of global fit except near perfect fit, models can have local misspecifications. Global fit indices cannot inform you about such misspecifications. You need to use other tools to locate these. Two main tools are the Lagrange Multiplier tests and the standardized residuals. The LM test is probably the most informative; see Chapter 6. Even if the SRMR is very small, indicating, say, an average residual of .03, it is still possible for specific correlations to be badly reproduced. After printing the Standardized Residual Matrix, EQS prints a list such as:

```
LARGEST STANDARDIZED RESIDUALS:

 NO.    PARAMETER   ESTIMATE
 ---    ---------   --------
  1     V6,  V2        -.243
  2     V5,  V2        -.219
  3     V6,  V1         .047
```

In this example, V2 is especially implicated in contributing to misfit. Ideally, based on your theory, you could determine why something like this might occur. Possibly the LM test associated with V2 and its precursors might be informative, so you could consider taking corrective action. For a further discussion and additional diagnostics, albeit for a narrow range of models, see McDonald (2004).

AIC and CAIC

Although we do not consider information criteria as fit indices, for completion we mention Akaike's (1987) information criterion (AIC) and Bozdogan's (1987) consistent version of this statistic (CAIC). These indices are only useful when comparing a series of models, not necessarily nested, and not empirically modified, and one wants to choose a model that balances adequacy of model fit with model parsimony. Model fit as measured by a test statistic *T* always improves as more parameters are added, but this increase in fit may be bought at the expense of too many parameters or too much model complexity. In a situation with many different models, and ignoring substantive considerations, the model that produces the minimum AIC or CAIC may be considered to be a potentially useful model. Hence if you only have one given model, these indices are probably not meaningful.

EQS computes these statistics for the independence model and the model of interest. An illustrative printout follows.

```
INDEPENDENCE AIC =    33.24793   INDEPENDENCE CAIC =   -10.43241
       MODEL AIC =      .69169          MODEL CAIC =   -22.60449
```

The formulae for AIC and CAIC, respectively are given next.

$$\begin{aligned}
&\text{Independence AIC} = T_i - 2d_i \\
&\text{Model AIC} = T_k - 2d_k. \\
&\text{Independence CAIC} = T_i - (\ln N + 1)d_i \\
&\text{Model CAIC} = T_k - (\ln N + 1)d_k.
\end{aligned} \tag{14.39}$$

The relevance of these statistics to LM and W tests is discussed in Chou and Bentler (1996).

When selecting among a set of candidate models, the values of AIC and CAIC for the independence model are irrelevant. Clearly, for any sized test statistic T_k , models with the larger degrees of freedom will have smaller AIC and CAIC values, and hence are favored by these criteria. A useful general discussions of model selection with AIC and another widely used information criterion, the Bayesian information criterion BIC (Schwarz, 1978; Raftery, 1995), as well as other methods, is given in the Special Issue on Model Selection in *Sociological Methods & Research* (especially, Burnham & Anderson, 2004; Kuha, 2004; Weakliem, 2004). Wicherts and Dolan (2004) make the important point that accurate model selection by information criteria when some models contain mean structures requires that all models to be compared contain means, even if those means are fully saturated.

15. REFERENCES

Abrahamson, A. C., Baker, L. A., & Caspi, A. (2002). Rebellious teens? Genetic and environmental influences on the social attitudes of adolescents. *Journal of Personality and Social Psychology, 83*, 1392-1408.

Abrams, D., Viki, G. T., Masser, B., & Bohner, G. (2003). Perceptions of stranger and acquaintance rape: The role of benevolent and hostile sexism in victim blame and rape proclivity. *Journal of Personality and Social Psychology*, *84*, 111-125.

Accornero, V. H., Morrow, C. E., Bandstra, E. S., Johnson, A. L., & Anthony, J. C. (2002). Behavioral outcome of preschoolers exposed prenatally to cocaine: Role of maternal behavioral health. *Journal of Pediatric Psychology*, *27*, 259-269.

Adams, R. A., Stanczak, D. E., Leutzinger, M. R., Waters, M. D., & Brown T. (2001). The impact of psychological disturbances on immediate memory. *Archives Of Clinical Neuropsychology*, *16*, 605-618.

Agnew, C. R., Thompson, V. D., & Gaines, S. O., Jr. (2000). Incorporating proximal and distal influences on prejudice: Testing a general model across outgroups. *Personality and Social Psychology Bulletin*, *26*, 403-418.

Agresti, A., Booth, J. G., Hobert, J. P., & Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology 2000* (pp. 27-80). Washington, DC: American Sociological Association.

Ahmed, S., & Mosley, W. H. (2002). Simultaneity in the use of maternal-child health care and contraceptives: Evidence from developing countries. *Demography*, *39*, 75-93.

Aigner, D. J., & Goldberger, A. S. (1977). *Latent variables in socioeconomic models*. Amsterdam: North-Holland.

Aigner, D. J., Hsiao, C., Kapteyn, A., & Wansbeek, T. (1984). Latent variable models in econometrics. In Z. Griliches & M. D. Intriligator (Eds.), *Handbook of econometrics*: *Vol. 2* (pp. 1321-1393). Amsterdam: North-Holland.

Aiken, L. S., Stein, J. A., & Bentler, P. M. (1994). Structural equation analyses of clinical sub-population differences and comparative treatment outcomes: Characterizing the daily lives of drug addicts. *Journal of Consulting and Clinical Psychology*, *62*, 488-499.

Aiken, L. S., West, S. G., & Pitts, S. C. (2003). Multiple linear regression. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (pp. 483-507). Hoboken, NJ: Wiley.

Aitchison, J., & Silvey, D. C. (1958). Maximum likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics*, *29*, 813-828.

Aitkin, M. & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A*, *149*, 1-43.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317-332.

Albarracín, D., & Wyer, R. S., Jr. (2001). Elaborative and nonelaborative processing of a behavior-related communication. *Personality and Social Psychology Bulletin, 27*, 691-705.

Allison, P. D. (1987). Estimation of linear models with incomplete data. In C. Clogg (Ed.), *Sociological Methodology* (pp. 71-103). San Francisco: Jossey Bass.

Allison, P. D. (2002). *Missing data.* Thousand Oaks, CA: Sage.

Alterman, A. I., McDermott, P. A., Mulvaney, F. D., Cacciola, J. S., Rutherford, M. J., Searles, J. S., Lynch, K. & Cnaan, A. (2003). Comparison of embedded and isolated administrations of the California Psychological Inventory's socialization subscale. *Psychological Assessment*, *15*, 64-70.

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review*, *40*, 37-47.

Alwin, D. F., & Jackson, D. J. (1981). Applications of simultaneous factor analysis to issues of factorial invariance. In D. Jackson & E. Borgatta (Eds.), *Factor analysis and measurement in sociological research: A multi-dimensional perspective.* Beverly Hills: Sage.

Amemiya, Y. (1985). Instrumental variable estimator for the nonlinear errors-in-variables model. *Journal of Econometrics*, *28*, 273-289.

Amemiya, Y., & Anderson, T. W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *Annals of Statistics*, *18*, 1453-1463.

Amemiya, Y., Fuller, W. A., & Pantula, S. G. (1987). The asymptotic distributions of some estimators for a factor analysis model. *Journal of Multivariate Analysis*, *22*, 51-64.

Anderson, D. A., Burton, D. B., Parker, J. D., & Godding, P. R. (2001). A confirmatory factor analysis of the cognitive capacity screening examination in a clinical sample. *International Journal of Neuroscience*, *111*, 221-233.

Anderson, E. R. (1993). Analyzing change in short-term longitudinal research using cohort-sequential designs. *Journal of Consulting and Clinical Psychology*, *61*, 929-940.

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analyses. *Psychometrika*, *49*, 155-173.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*, 411-423.

Anderson, T. W. (1990). Linear latent variable models and covariance structures. *Journal of Econometrics*, *41*, 91-119.

Anderson, T. W., & Amemiya, Y. (1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics*, *16*, 759-771.

Ansari, A., Jedidi, K., & Dube, L. (2002). Heterogeneous factor analysis models: A Bayesian approach. *Psychometrika*, *67*, 49-78.
Anstey, K. J., & Smith, G. A. (1999). Interrelationships among biological markers of aging, health, activity, acculturation, and cognitive performance in late adulthood. *Psychology and Aging*, *14*, 605-618.
Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*, 1423-1440.
Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.) *Advanced structural equation modeling: Issues and techniques* (pp. 243-277). Mahwah, NJ: Erlbaum.
Arminger, G. (1986). Linear stochastic differential equation models for panel data with unobserved variables. In N. Tuma (Ed.), *Sociological Methodology* (pp. 187-212). San Francisco: Jossey-Bass.
Arminger, G., & Küsters, U. (1988). Latent trait models with indicators of mixed measurement level. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 51-73). New York: Plenum.
Arminger, G., & Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika*, *63*, 271-300.
Arminger, G., & Sobel, M. E. (1990). Pseudo-maximum likelihood estimation of mean and covariance structures with missing data. *Journal of the American Statistical Association, 85*, 195-203.
Arminger, G., Stein, P., & Wittenberg, J. (1999). Mixtures of conditional mean- and covariance-structure models. *Psychometrika, 64*, 475-494.
Arnold, S. F. (1981). *The theory of linear models and multivariate analysis.* New York: Wiley.
Arthur, W., Bennett, W., & Huffcutt, A. I. (2001). *Conducting meta-analysis using SAS.* Mahwah, NJ: Erlbaum.
Ary, D. V., Duncan, T. E., Biglan, A. M., Metzler, C. W., Noell, J. W., & Smolkowski, K. (1999). Development of adolescent problem behavior. *Journal of Abnormal Child Psychology, 27*, 141-150.
Auerbach, S. M., & Pegg, P. O. (2002). Appraisal of desire for control over healthcare: Structure, stability, and relation to health locus of control and to the 'big five' personality traits. *Journal of Health Psychology, 7*, 393-408.
Austin, J. T., & Calderón, R. F. (1996). Theoretical and technical contributions to structural equation modeling: An updated annotated bibliography. *Structural Equation Modeling*, *3*, 105-175.
Austin, J. T., & Wolfle, D. (1991). Annotated bibliography of structural equation modeling: Technical work. *British Journal of Mathematical and Statistical Psychology*, *44*, 93-152.
Austin, P. C., & Tu, J. V. (2004). Bootstrap methods for developing predictive models. *The American Statistician*, *58*, 131-137.
Avants, S. K., Margolin, A., & McKee, S. (2000). A path analysis of cognitive, affective, and behavioral predictors of treatment response in a methadone maintenance program. *Journal of Substance Abuse, 11*, 215-230.
Azzalini, A., & Valle, A. D. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*, 715-726.
Babakus, E., Ferguson, C. E., Jr., & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research, 24*, 222-228.
Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing, 17*, 1-42.
Bagozzi, R. P. (1977). Structural equation models in experimental research. *Journal of Marketing Research*, *14*, 271-284.
Bagozzi, R. P. (1980). *Causal models in marketing.* New York: Wiley.
Bagozzi, R. P., & Yi, Y. (1989). On the use of structural equation models in experimental design. *Journal of Marketing Research, 26*, 271-284.
Bakan, D. (1954). A generalization of Sidman's results on group and individual functions, and a criterion. *Psychological Bulletin, 51*, 63-64.
Balderjahn, I. (1985). The robustness of LISREL unweighted least squares estimation against small sample size in confirmatory factor analysis models. In W. Gaul & M. Schader (Eds.), *Classification as a tool of research* (pp. 3-10). Amsterdam: Elsevier Science Publishers.
Bandalos, D. L. (1997). Assessing sources of error in structural equation modeling: The effects of sample size, reliability, and model misspecification. *Structural Equation Modeling, 4*, 157-174.
Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, *9*, 78-102.
Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269-296). Mahwah, NJ: Erlbaum.
Bandstra, E. S., Morrow, C. E., Anthony, J. C., Accornero, V. H., & Fried, P. A. (2001). Longitudinal investigation of task persistence and sustained attention in children with prenatal cocaine exposure. *Neurotoxicology and Teratology*, *23*, 545-559.
Bandstra, E. S., Morrow, C. E., Anthony, J. C., Churchill, S. S., Chitwood, D. C., Steele, B. W., Ofir, A. Y., & Xue, L. H. (2001). Intrauterine growth of full-term infants: Impact of prenatal cocaine exposure. *Pediatrics*, *108*, 1309-1319.
Barbarotto, R., Capitani, E., & Laiacona, M. (2001). Living musical instruments and inanimate body parts? *Neuropsychologia, 39*, 406-414.
Barbaranelli, C., & Caprara, G. V. (2000). Measuring the Big Five in self-report and other ratings: A multitrait-multimethod study. *European Journal of Psychological Assessment, 16*, 31-43.

Barchard, K. A., & Hakstian, A. R. (1997). The effects of sampling model on inference with coefficient alpha. *Educational & Psychological Measurement, 57*, 893-905.
Barker, D. C., & Carman, C. J. (2000). The spirit of capitalism? Religious doctrine, values, and economic attitude constructs. *Political Behavior, 22*, 1-27.
Barnes, G. E., Murray, R. P., Patton, D., Bentler, P. M., & Anderson, R. E. (2000). *The addiction-prone personality*. New York: Kluwer AcademicPlenum Publishers.
Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.
Barron, K. E., & Harackiewicz, J. M. (2001). Achievement goals and optimal motivation: Testing multiple goal models. *Journal of Personality and Social Psychology, 80*, 706-722.
Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis.* New York: Edward Arnold.
Bartholomew, D. J., & Tzamourani, P. (1999). The goodness of fit of latent trait models in attitude measurement. *Sociological Methods & Research, 27*, 525-546.
Bartlett, D. J., & Palisano, R. J. (2000). A multivariate model of determinants of motor change for children with cerebral palsy. *Physical Therapy, 80*, 598-614.
Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology, 28*, 97-104.
Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology, Statistical Section, 3*, 77-85.
Batista-Foguet, J. M., Coenders, G., & Ferragud, M. A. (2001). Using structural equation models to evaluate the magnitude of measurement error in blood pressure. *Statistics In Medicine, 20*, 2351-2368.
Batson, C. D., Chang, J., Orr, R., & Rowland, J. (2002). Empathy, attitudes, and action: Can feeling for a member of a stigmatized group motivate one to help the group? *Personality and Social Psychology Bulletin, 28*, 1656-1666.
Battin-Pearson, S., Newcomb, M. D., Abbott, R. D., Hill, K. G., Catalano, R. F., & Hawkins, J. D. (2000). Predictors of early high school dropout: A test of five theories. *Journal of Educational Psychology, 92*, 568-582.
Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics, 28*, 135-167.
Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8*, 338-363.
Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9*, 3-29.
Bavelier, D., Tomann, A., Hutton, C., Mitchell, T., Corina, D., Liu, G., & Neville, H. (2000). Visual attention to the periphery is enhanced in congenitally deaf individuals. *Journal of Neuroscience, 20*, U1-U6.
Baydar, N., Greek, A., & Gritz, R. M. (1999). Young mothers' time spent at work and time spent caring for children. *Journal of Family & Economic Issues, 20*, 61-84.
Bearden, W. O., Sharma, S., & Teel, J. E. (1982). Sample size effects on chi-square and other statistics used in evaluating causal models. *Journal of Marketing Research, 19*, 425-430.
Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling, 12*, 41-75.
Becker, B. J., & Schram, C. M. (1994). Examining explanatory models through research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 357-381). New York: Russell Sage.
Becker, G. (2000a). Coefficient alpha: Some terminological ambiguities and related misconceptions. *Psychological Reports, 86*, 365-372.
Becker, G. (2000b). How important is transient error in estimating reliability? Going beyond simulation studies. *Psychological Methods, 5*, 370-379.
Belsky, J., Friedman, S. L., Hsieh, K. H. (2001). Testing a core emotion-regulation prediction: Does early attentional persistence moderate the effect of infant negative emotionality on later development? *Child Development, 72*, 123-133.
Bendersky, M., & Lewis, M. (1999). Prenatal cocaine exposure and neonatal condition. *Infant Behavior & Development, 22*, 353-366.
Bentler, P. M. (1968). Alpha-maximized factor analysis (Alphamax): Its relation to alpha and canonical factor analysis. *Psychometrika, 33*, 335-345.
Bentler, P. M. (1972). A lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research, 1*, 343-357.
Bentler, P. M. (1976). Multistructural statistical model applied to factor analysis. *Multivariate Behavioral Research, 11*, 3-25.
Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology, 31*, 419-456.
Bentler, P. M. (1982). Linear systems with multiple levels and types of latent variables. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp. 101-130). Amsterdam: North-Holland.
Bentler, P. M. (1983a). Some contributions to efficient statistics for structural models: Specification and estimation of moment structures. *Psychometrika, 48*, 493-517.
Bentler, P. M. (1983b). Simultaneous equations as moment structure models: With an introduction to latent variable models. *Journal of Econometrics, 22*, 13-42.
Bentler, P. M. (1985). *Theory and implementation of EQS, A structural equations program*. Los Angeles: BMDP Statistical Software.
Bentler, P. M. (1986a). *Lagrange Multiplier and Wald tests for EQS and EQSPC*. Los Angeles: BMDP Statistical Software.

Bentler, P. M. (1986b). Structural modeling and Psychometrika: An historical perspective on growth and achievements. *Psychometrika, 51*, 35-51.
Bentler, P. M. (1987). Structural modeling and the scientific method: Comments on Freedman's critique. *Journal of Educational Statistics, 12*, 151-157.
Bentler, P. M. (1989). *EQS structural equations program manual.* Los Angeles: BMDP Statistical Software.
Bentler, P. M. (1990a). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.
Bentler, P. M. (1990b). Fit indexes, Lagrange multipliers, constraint changes and incomplete data in structural models. *Multivariate Behavioral Research, 25*, 163-172.
Bentler, P. M. (1990c). Latent variable structural models for separating specific from general effects. In L. Sechrest, E. Perrin, & J. Bunker (Eds.), *Research methodology: Strengthening causal interpretations of nonexperimental data* (pp. 61-83) (DDS Pub. No. 90-3454). Rockville, MD: Department of Health and Human Services.
Bentler, P. M. (1991). Modeling of intervention effects. In C. G. Leukefeld & W. J. Bukoski (Eds.), *Drug abuse prevention intervention research: Methodological issues* (pp. 159-182, NIDA Research Monograph 107). Rockville, MD: National Institute on Drug Abuse.
Bentler, P. M. (1994). A testing method for covariance structure analysis. In T. W. Anderson, K. T. Fang, & I. Olkin (Eds.), *Multivariate analysis and its applications: Vol. 24* (pp. 123-135). Hayward, CA: Institute of Mathematical Statistics.
Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.
Bentler, P. M. (1998). Causal modeling: New interfaces and new statistics. In J. G. Adair, D. Belanger, & K. L. Dion (Eds.), *Advances in psychological science: Social, personal, and cultural aspects, Vol 1* (pp. 353-370). East Sussex, UK: Psychology Press.
Bentler, P. M. (2000). Rites, wrongs, and gold in model testing. *Structural Equation Modeling, 7*, 82-91.
Bentler, P. M. (2001). New approaches to structural equation models with missing data. In R. K. Silbereisen & M. Reitzle (Eds.), *Psychologie 2000* (pp. 65-74). Lengerich, Germany: Pabst.
Bentler, P. M. (2004). *Maximal reliability for unit-weighted composites.* UCLA Statistics Preprint Series No. 405.
Bentler, P. M., & Berkane, M. (1985). Developments in the elliptical theory generalization of normal multivariate analysis. *Proceedings of the Social Statistics Section, American Statistical Association*, 291-295.
Bentler, P. M., & Berkane, M. (1986). The greatest lower bound to the elliptical theory kurtosis parameter. *Biometrika, 73*, 240-241.
Bentler, P. M., Berkane, M., & Kano, Y. (1991). Covariance structure analysis under a simple kurtosis model. In E. M. Keramidas (Ed.), *Computing science and statistics* (pp. 463-465). Fairfax Station, VA: Interface Foundation of North America.
Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606.
Bentler, P. M., & Chou, C.-P. (1986, April). *Statistics for parameter expansion and contraction in structural models.* Paper presented at American Educational Research Association meeting, San Francisco.
Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*, 78-117.
Bentler, P. M., & Chou, C.-P. (1989, July). *Mean and covariance structure models: A distribution-free approach*. Paper presented at Psychometric Society meetings, Los Angeles.
Bentler, P. M., & Chou, C.-P. (1990). Model search with TETRADII and EQS. *Sociological Methodology, 19*, 67-79.
Bentler, P. M., & Chou, C. –P. (1993). Some new covariance structure model improvement statistics. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 235-255). Newbury Park, CA: Sage.
Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Krishnaiah (Ed.), *Multivariate analysis VI* (pp. 9-42). Amsterdam: North-Holland.
Bentler, P. M., and Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology, 47*, 541-570.
Bentler, P. M., & Houck, E. L. (1996). Structural equation models of multiple sclerosis disease status. *The International Test Commission Newsletter, 6*, 11-13.
Bentler, P. M., & Jamshidian, M. (1994). Gramian matrices in covariance structure models. *Applied Psychological Measurement, 18*, 79-94.
Bentler, P. M. & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research, 25*, 67-74.
Bentler, P. M., Kim, K., & Yuan, K. –H. (2004). *Testing homogeneity of covariances with infrequent missing data patterns.* In preparation.
Bentler, P. M., & Lee, S.-Y. (1979). A statistical development of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology, 32*, 87-104.
Bentler, P. M., & Lee, S.-Y. (1983). Covariance structures under polynomial constraints: Applications to correlation and alpha-type structural models. *Journal of Educational Statistics, 8*, 207-222.
Bentler, P. M., Lee, S.-Y., & Weng, L.-J. (1987). Multiple population covariance structure analysis under arbitrary distribution theory. *Communications in Statistics - Theory, 16*, 1951-1964.
Bentler, P. M., & Liang, J. (2002). Two-level mean and covariance structures: Maximum likelihood via an EM algorithm. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 53-70). Mahwah, NJ: Erlbaum.

Bentler, P. M., & Liang, J. (2003). Simultaneous mean and covariance structure analysis for two level structural equation models on EQS. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J. J. Meulman (Eds.), *New developments in psychometrics* (pp. 123-132). Tokyo: Springer-Verlag.

Bentler, P. M., & Mooijaart, A. (1989). Choice of structural model via parsimony. *Psychological Bulletin, 106*, 315-317.

Bentler, P. M., Newcomb, M. D., & Zimmerman, M. A. (2002). Cigarette use and drug use progression: Growth trajectory and lagged effect hypotheses. In D. B. Kandel (Ed.), *Stages and pathways of drug involvement* (pp. 223-253). New York: Cambridge.

Bentler, P. M., Poon, W.-Y., & Lee, S.-Y. (1988). Generalized multimode latent variable models: Implementation by standard programs. *Computational Statistics & Data Analysis, 7*, 107-118.

Bentler, P. M., & Raykov, T. (2000). On measures of explained variance in nonrecursive structural equation models. *Journal of Applied Psychology, 85*, 125-131.

Bentler, P. M., & Satorra, A. (2000). Hierarchical regression without phantom factors. *Structural Equation Modeling, 7*, 287-291.

Bentler, P. M., & Speckart, G. (1979). Models of attitude-behavior relations. *Psychological Review, 86*, 452-464.

Bentler, P. M., & Speckart, G. (1981). Attitudes "cause" behaviors: A structural equation analysis. *Journal of Personality and Social Psychology, 40*, 226-238.

Bentler, P. M., & Weeks, D. G. (1979). Interrelations among models for the analysis of moment structures. *Multivariate Behavioral Research, 14*, 169-185.

Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika, 45*, 289-308.

Bentler, P. M., & Weeks, D. G. (1982). Multivariate analysis with latent variables. In P. R. Krishnaiah & L. Kanal (Eds.), *Handbook of statistics: Vol. 2* (pp. 747-771). Amsterdam: North-Holland.

Bentler, P. M., & Weeks, D. G. (1985). Some comments on structural equation models. *British Journal of Mathematical and Statistical Psychology, 38*, 120-121.

Bentler, P. M., & Woodward, J. A. (1978). A head start re-evaluation: Positive effects are not yet demonstrable. *Evaluation Quarterly, 2*, 493-510.

Bentler, P. M., & Woodward, J. A. (1980). Inequalities among lower bounds to reliability: With applications to test construction and factor analysis. *Psychometrika, 45*, 249-267.

Bentler, P. M., & Woodward, J. A. (1983). The greatest lower bound to reliability. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement: A Festschrift for Frederic M. Lord* (pp. 237-253). Hillsdale, NJ: Erlbaum.

Bentler, P. M., & Woodward, J. A. (1985). On the greatest lower bound to reliability. *Psychometrika, 50*, 245-246.

Bentler, P. M., & Wu, E. J. C. (2002). *EQS 6 for Windows user's guide.* Encino, CA: Multivariate Software.

Bentler, P. M., Wu, E. J. C., & Houck, E. L. (1996). Graphische Umgebung für Kausalmodelle. In E. Erdfelder, R. Mausfeld, T. Meiser, & G. Rudinger (Eds.), *Handbuch Quantitative Methoden* (pp. 269-278). Weinheim: Psychologie Verlags Union.

Bentler, P. M., & Xie, J. (2000). Corrections to test statistics in principal Hessian directions. *Statistics & Probability Letters, 47*, 381-389.

Bentler, P. M., & Yuan, K. –H. (1996). Test of linear trend in eigenvalues of a covariance matrix with application to data analysis. *British Journal of Mathematical and Statistical Psychology, 49*, 299-312.

Bentler, P. M., & Yuan, K. –H. (1997). Optimal conditionally unbiased equivariant factor score estimators. In M. Berkane (Ed.), *Latent variable modeling with applications to causality* (pp. 259-281). New York: Springer-Verlag.

Bentler, P. M., & Yuan, K. –H. (1998). Tests for linear trend in the smallest eigenvalues of the correlation matrix. *Psychometrika, 63*, 131-144.

Bentler, P. M., & Yuan, K. –H. (1999a). Structural equation models. In S. Kotz, C. B. Read, & D. L. Banks (Eds.), *Encyclopedia of statistical sciences, Update volume 3* (pp. 716-721). New York: Wiley.

Bentler, P. M., & Yuan, K. –H. (1999b). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research, 34*, 181-197.

Bentler, P. M., & Yuan, K. –H. (2000). On adding a mean structure to a covariance structure model. *Educational & Psychological Measurement, 60*, 326-339.

Beran, R. (1979). Testing for ellipsoidal symmetry of a multivariate density. *The Annals of Statistics, 7*, 150-162.

Beran, R., & Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *Annals of Statistics, 13*, 95-115.

Berkane, M., & Bentler, P. M. (1986). Moments of elliptically distributed random variates. *Statistics & Probability Letters, 4*, 333-335.

Berkane, M., & Bentler, P. M. (1987a). Characterizing parameters of multivariate elliptical distributions. *Communications in Statistics-Simulation, 16*, 193-198.

Berkane, M., & Bentler, P. M. (1987b). Distribution of kurtoses, with estimators and tests of homogeneity of kurtosis. *Statistics & Probability Letters, 5*, 201-207.

Berkane, M., & Bentler, P. M. (1988). Estimation of contamination parameters and identification of outliers in multivariate data. *Sociological Methods & Research, 17*, 55-64.

Berkane, M., & Bentler, P. M. (1990). Mardia's coefficient of kurtosis in elliptical populations. *Acta Mathematicae Applicatae Sinica, 6*, 289-294.

Berkane, M., & Bentler, P. M. (1993). The geometry of mean or covariance structure models in multivariate normal distributions: A unified approach. In C. M. Cuadras & C. R. Rao (Eds.), *Multivariate analysis: Future directions 2* (pp.153-169). Amsterdam: Elsevier Science.

Berkane, M., & Bentler, P. M. (1997). Bias and mean square error of the maximum likelihood estimators of the parameters of the intraclass correlation model. In M. Berkane (Ed.), *Latent variable modeling with applications to causality* (pp. 133-147). New York: Springer-Verlag.

Berkane, M., Kano, Y., & Bentler, P. M. (1994). Pseudo maximum likelihood estimation and elliptical distribution theory in a misspecified model. *Computational Statistics and Data Analysis*, *18*, 255-267.

Berkane, M., Oden, K., & Bentler, P. M. (1997). Geodesic estimation in elliptical distributions. *Journal of Multivariate Analysis, 63*, 35-46.

Bernaards, C. A., & Sijtsma, K. (1999). Factor analysis of multidimensional polytomous item response data suffering from ignorable item nonresponse. *Multivariate Behavioral Research, 34*, 277-313.

Bernaards, C. A., & Sijtsma, K. (2000). Influence of imputation and EM methods on factor analysis when item nonresponse in questionnaire data is nonignorable. *Multivariate Behavioral Research, 35*, 321-364.

Berntsson, L. T., Kohler, L., & Gustafsson, J. E. (2001). Psychosomatic complaints in schoolchildren: A Nordic comparison. *Scandinavian Journal of Public Health, 29*, 44-54.

Beunen, G., Thomis, M., Maes, H. H., Loos, R., Malina, R. M., Claessens, A. L., & Vlietinck, R. (2000). Genetic variance of adolescent growth in stature. *Annals of Human Biology, 27*, 173-186.

Berkson, J. (1980). Minimum chi-square, not maximum likelihood! *The Annals of Statistics, 8*, 457-487.

Bickel, P. J., Götze, F., & van Zwet, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statistica Sinica, 7*, 1-32.

Bielby, W. T., Hauser, R. M., & Featherman, D. L. (1977). Response errors of black and nonblack males in models of the intergenerational transmission of socioeconomic status. *American Journal of Sociology, 82*, 1242-1288.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9*, 30-52.

Biesanz, J. C., West, S. G., & Kwok, O.-M. (2003). Personality over time: Methodological approaches to the study of short-term and long-term development and change. *Journal of Personality, 71*, 905-942.

Birch, L. L, & Fisher, J. O. (2000). Mothers' child-feeding practices influence daughters' eating and weight. *American Journal of Clinical Nutrition, 71*,1054-1061.

Blue, C. L., Wilbur, J., & Marston-Scott, M. (2001). Exercise among blue-collar workers: Application of the theory of planned behavior. *Research in Nursing & Health, 24*, 481-493.

Bock, R. D. (1989). *Multilevel analysis of educational data*. San Diego: Academic Press.

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12*, 261-280.

Boeschen, L. E., Koss, M. P., Figueredo, A. J., & Coan, J. A. (2001). Experiential avoidance and post-traumatic stress disorder: A cognitive mediational model of rape recovery. *Journal of Aggression, Maltreatment & Trauma, 4*, 211-245.

Boker, S. M. (2001). Differential structural equation modeling of intraindividual variability. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 5-27). Washington DC: American Psychological Association.

Bollen, K. A. (1980). Issues in the comparative measurement of political democracy. *American Sociological Review, 45*, 370-390.

Bollen, K. A. (1986). Sample size and Bentler and Bonett's nonnormed fit index. *Psychometrika, 51*, 375-377.

Bollen, K. A. (1987a). Total, direct, and indirect effects in structural equation models. In C. Clogg (Ed.), *Sociological methodology* (pp. 37-69). San Francisco: Jossey Bass.

Bollen, K. A. (1987b). Outliers and improper solutions: A confirmatory factor analysis example. *Sociological Methods & Research, 15*, 375-384.

Bollen, K. A. (1989a). A new incremental fit index for general structural equation models. *Sociological Methods & Research, 17*, 303-316.

Bollen, K. A. (1989b). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A. (1995). Structural equation models that are nonlinear in latent variables: A least squares estimator. *Sociological Methodology, 25*, 223-251.

Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika, 61*, 109-121.

Bollen, K. A. (2000). Modeling strategies: In search of the holy grail. *Structural Equation Modeling, 7*, 74-81.

Bollen, K. A. (2001). Two-stage least squares and latent variable models: Simultaneous estimation and robustness to misspecifications. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 119-138). Lincolnwood, IL: Scientific Software International.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53*, 605-634.

Bollen, K. A., & Barb, K. H. (1981). Pearson's r and coarsely categorized measures. *American Sociological Review, 46*, 232-239.

Bollen, K. A., & Biesanz, J. C. (2002). A note on a two-stage least squares estimator for higher-order factor analyses. *Sociological Methods & Research, 30*, 568-579.

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods & Research, 32*, 336-383.

Bollen, K. A., & Jöreskog, K. G. (1985). Uniqueness does not imply identification. A note on confirmatory factor analysis. *Sociological Methods & Research, 14*, 155-163.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110,* 305-314.
Bollen, K. A., & Liang, J. (1988). Some properties of Hoelter's CN. *Sociological Methods & Research, 16*, 492-503.
Bollen, K. A., & Paxton, P. (1998). Two-stage least squares estimation of interaction effects. *Structural Equation Modeling, 5,* 267-293.
Bollen, K. A., & Stine, R. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In K. A. Bollen and J. S. Long (Eds.) *Testing structural equation models* (pp. 111-135). Newbury Park, CA: Sage.
Bollen, K. A., & Ting, K. (2000). A tetrad test for causal indicators. *Psychological Methods, 5,* 3-22.
Bond, R., & Saunders, P. (1999). Routes of success: Influences on the occupational attainment of young British males. *British Journal of Sociology, 50*, 217-249.
Bonett, D. G. (2003a). Sample size requirements for comparing two alpha coefficients. *Applied Psychological Measurement, 27,* 72-74.
Bonett, D. G. (2003b). Sample size requirement for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics, 27*, 335-340.
Bonett, D. G., Woodward, J. A., & Randall, R. L. (2002). Estimating p-values for Mardia's coefficients of multivariate skewness and kurtosis. *Computational Statistics, 17*, 117-122.
Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality.* Ph.D. Thesis, University of Groningen.
Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika, 52*, 345-370.
Boomsma, A. (1986). On the use of bootstrap and jackknife in covariance structure analysis. *COMPSTAT, 7*, 205-210.
Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling, 7,* 461-483.
Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 139-168). Lincolnwood, IL: Scientific Software International.
Borkenau, P., Riemann, R., Angleitner, A. & Spinath, F. M. (2001). Genetic and environmental influences on observed personality: Evidence from the German observational study of adult twins. *Journal of Personality and Social Psychology, 80*, 655-668.
Bornstein, M. H., & Haynes, O. M. (1998). Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Development, 69*, 654-671.
Bornstein, M. H., Haynes, M. O., & Painter, K. M. (1998). Sources of child vocabulary competence: A multivariate model. *Journal of Child Language, 25*, 367-393.
Botha, J. D., Shapiro, A., & Steiger, J. H. (1988). Uniform indices-of-fit for factor analysis models. *Multivariate Behavioral Research, 23*, 443-450.
Bowden, S. C., Dodds, B., Whelan, G., Long, C., Dudgeon, P., Ritter, A., & Clifford, C. (1997). Confirmatory factor analysis of the Wechsler Memory Scale-Revised in a sample of clients with alcohol dependency. *Journal of Clinical and Experimental Neuropsychology, 19*, 755-762.
Bozdogan, H. (1987). Model selection and Akaike's information criteria (AIC): The general theory and its analytical extensions. *Psychometrika, 52*, 345-370.
Breivik, E., & Olsson, U. H. (2001). Adding variables to improve fit: The effect of model size on fit assessment in LISREL. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 169-194). Lincolnwood, IL: Scientific Software International.
Brennan, R. L. (2001). *Generalizability theory.* New York: Springer-Verlag.
Brito, C., & Pearl, J. (2002). A new identification condition for recursive models with correlated errors. *Structural Equation Modeling, 9,* 459-474.
Brook, J. S., Whiteman, M., Finch, S. J., & Cohen, P. (2000). Longitudinally foretelling drug use in the late twenties: Adolescent personality and social-environmental antecedents. *Journal of Genetic Psychology, 161,* 37-51.
Brook, J. S., Whiteman, M., Finch, S. J., Morojele, N. K., & Cohen, P. (2000). Individual latent growth curves in the development of marijuana use from childhood to young adulthood. *Journal of Behavioral Medicine, 23,* 451-464.
Brook, J. S., Finch, S. J., Whiteman, M., & Brook, D. W. (2002). Drug use and neurobehavioral, respiratory, and cognitive problems: Precursors and mediators. *Journal of Adolescent Health, 30* , 433-441.
Brown, C. H. (1983). Asymptotic comparison of missing data procedures for estimating factor loadings. *Psychometrika, 48,* 269-291.
Brown, R. L. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling, 1,* 287-316.
Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal, 8*, 1-24.
Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72-141). London: Cambridge University Press.
Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of co-variance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62-83.
Browne, M. W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrika, 74*, 375-384.
Browne, M. W. (2000). Psychometrics. *Journal of the American Statistical Association, 95*, 661-665.

Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.) *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185-250). New York: Plenum.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Browne, M. W., & du Toit, S. H. C. (1991). Models for learning data. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 47-68). Washington DC: American Psychological Association.

Browne, M. W., & du Toit, S. H. C. (1992). Automated fitting of nonstandard models. *Multivariate Behavioral Research*, *27*, 269-300.

Browne, M. W., MacCallum, R. C., Kim, C. –T., Anderson, B., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods*, *7*, 403-421.

Browne, M. W., & Shapiro, A. (1986). The asymptotic covariance matrix of sample correlation coefficients under general conditions. *Linear Algebra and Its Applications*, *82*, 169-176.

Browne, M. W., & Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology*, *41*, 193-208.

Bryan, A., Fisher, J. D., & Fisher, W. A. (2002). Tests of the mediational role of preparatory safer sexual behavior in the context of the theory of planned behavior. *Health Psychology*, *21*, 71-80.

Bryant, A. L., Schulenberg, J., Bachman, J. G., O'Malley, P. M., & Johnston, L. D. (2000). Understanding the links among school misbehavior, academic achievement, and cigarette use: A national panel study of adolescents. *Prevention Science*, *1*, 71-87.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: Sage.

Budaev, S. V. (1999). Sex differences in the Big Five personality factors: Testing an evolutionary hypothesis. *Personality & Individual Differences*, *26*, 801-813.

Bullmore, E., Horwitz, B., Honey, G., Brammer, M., Williams, S., & Sharma, T. (2000). How good is good enough in path analysis of fMRI data? *Neuroimage*, *11*, 289-301.

Burgess, P. W., Veitch, E., de Lacy Costello, A., & Shallice, T. (2000). The cognitive and neuroanatomical correlates of multitasking. *Neuropsychologia*, *38*, 848-863.

Burke, R. J. (2001). Job stress, work satisfaction and physician militancy. *Stress and Health*, *17*, 263-271.

Burkholder, G. J., & Harlow, L. L. (1996). Using structural equation modeling techniques to evaluate HIV risk models. *Structural Equation Modeling*, *3*, 348-368.

Burkholder, G. J., & Harlow, L. L. (2003). An illustration of a longitudinal cross-lagged design for larger structural equation models. *Structural Equation Modeling*, *10*, 465-486.

Burnham, K. P., & Anderson, D. R. (2004). Multimodal inference: Understanding AIC and BIC I model selection. *Sociological Methods & Research*, *33*, 261-304.

Burns, D. D., & Spangler, D. L. (2000). Does psychotherapy homework lead to improvements in depression in cognitive-behavioral therapy or does improvement lead to increased homework compliance? *Journal of Consulting and Clinical Psychology*, *68*, 46-56.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. In D. C. Berliner (Ed.), *Review of research in education* (pp. 158-223). Washington, DC: American Educational Research Association.

Buse, A. (1982). The likelihood ratio, Wald and Lagrange multiplier tests: An expository note. *American Statistician*, *36*, 153-157.

Bushman, B. J., & Wells, G. L. (2001). Narrative impressions of literature: The availability bias and the corrective properties of meta-analytic approaches. *Personality and Social Psychology Bulletin*, *27*, 1123-1130.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQSWindows*. Thousand Oaks, CA: Sage.

Byrne, B. M. (1999). The nomological network of teacher burnout: A literature review and empirically validated model. In R. Vandenberghe, A. M. Huberman, (Eds.). *Understanding and preventing teacher burnout: A sourcebook of international research and practice* (pp. 15-37). Cambridge University Press: New York.

Byrne, B. M., & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology*, *30*, 555-574.

Byrne, B. M., & Crombie, G. (2003). Modeling and testing change: An introduction to the latent growth curve model. *Understanding Statistics*, *2*, 177-203.

Byrne, B. M., & Goffin, R. D. (1993). Modeling MTMM data from additive and multiplicative covariance structures: An audit of construct validity. *Multivariate Behavioral Research*, *28*, 67-96.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, *34*, 155-175.

Calantone, R. J., & Zhao, Y. S. (2001). Joint ventures in China: A comparative study of Japanese, Korean, and US partners. *Journal of International Marketing*, *9*, 1-23.

Cambanis, S., Huang, S., & Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis, 11*, 368-385.
Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York: Guilford.
Campbell, J. C., & Weber, N. (2000). An empirical test of a self-care model of women's responses to battering. *Nursing Science Quarterly, 13*, 45-53.
Campbell, L., Simpson, J. A., Kashy, D. A., & Fletcher, G. J. O. (2001). Ideal standards, the self, and flexibility of ideals in close relationships. *Personality and Social Psychology Bulletin, 27*, 447-462.
Campbell, N. A. (1980). Robust procedures in multivariate analysis I: Robust covariance estimation. *Applied Statistics, 29*, 231-237.
Caprara, G. V., Scabini, E., Barbaranelli, C., Pastorelli, C., Regalia, C., & Bandura, A. (1998). Impact of adolescents' perceived self-regulatory efficacy on familial communication and antisocial conduct. *European Psychologist, 3*, 125-132.
Carbonaro, W. J., & Gamoran, A. (2002). The production of achievement inequality in high school English. *American Educational Research Journal, 39*, 801-827.
Cardona, P., Lawrence, B. S., & Bentler, P. M. (2004). The influence of social and work exchange relationships on organizational citizenship behavior. *Group & Organization Management, 29*, 219-247.
Carlstedt, B. (2001). Differentiation of cognitive abilities as a function of general intelligence: A latent variable approach. *Multivariate Behavioral Research, 36*, 589-609.
Carmelli, D., Bliwise, D. L., Swan, G. E., & Reed, T. (2001). A genetic analysis of the Epworth Sleepiness Scale in 1560 World War II male veteran twins in the NAS-NRC Twin Registry. *Journal of Sleep Research, 10*, 53-58.
Carmelli, D., Fabsitz, R. R., Swan, G. E., Reed, T., Miller, B., & Wolf, P. A. (2000). Contribution of genetic and environmental influences to ankle-brachial blood pressure index in the NHLBI Twin Study. *American Journal of Epidemiology, 151*, 452-458.
Carrig, M. M., Wirth, R. J., & Curran, P. J. (2004). A SAS macro for estimating and visualizing individual growth curves. *Structural Equation Modeling, 11*, 132-149.
Carroll, R. J., Ruppert, D., & Stefanski, L. A. (1995). *Measurement error in nonlinear models*. New York: Chapman & Hall.
Causgrove, D. J. (2000). Goal orientations, perceptions of the motivational climate, and perceived competence of children with movement difficulties. *Adapted Physical Activity Quarterly, 17*, 1-19.
Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics, 18*, 5-46.
Chambless, D. L., Bryan, A. D., Aiken, L. S., Steketee, G., & Hooley, J. M. (1999). The structure of expressed emotion: A three-construct representation. *Psychological Assessment, 11*, 67-76.
Chan, D. (2000). Detection of differential item functioning on the Kirton Adaption-Innovation Inventory using multiple-group mean and covariance structure analyses. *Multivariate Behavioral Research, 35*, 169-199.
Chan, D., Ramey, S., Ramey, C., & Schmitt, N. (2000). Modeling intraindividual changes in children's social skills at home and at school: A multivariate latent growth approach to understanding between-settings differences in children's social skill development. *Multivariate Behavioral Research, 35*, 365-396.
Chan, R. Y. K. (2001). Determinants of Chinese consumers' green purchase behavior. *Psychology and Marketing, 18*, 389-413.
Chan, W., & Bentler, P. M. (1993). The covariance structure analysis of ipsative data. *Sociological Methods & Research, 22*, 214-247.
Chan, W., & Bentler, P. M. (1996). Covariance structure analysis of partially additive ipsative data using restricted maximum likelihood estimation. *Multivariate Behavioral Research, 31*, 289-312.
Chan, W., & Bentler, P. M. (1998). Covariance structure analysis of ordinal ipsative data. *Psychometrika, 63*, 369-399.
Chan, W., Yung, Y. –F., & Bentler, P. M. (1995). A note on using an unbiased weight matrix in the ADF test statistic. *Multivariate Behavioral Research, 30*, 453-459.
Chan, W., Yung, Y. –F., Bentler, P. M., & Tang, M. –L. (1998). Tests of independence for ordinal data using bootstrap. *Educational &Psychological Measurement, 58*, 221-240.
Chatzisarantis, N. L. D., Hagger, M. S., Biddle, S. J. H., & Karageorghis, C. (2002). The cognitive processes by which perceived locus of causality predicts participation in physical activity. *Journal of Health Psychology, 7*, 685-699.
Chen, E., Zeltzer, L. K., Bentler, P. M., Byrne, J., Nicholson, H. S., Meadows, A. T., Mills, J., Haupt, R., Fears, T. R., & Robison, L. L. (1998). Pathways linking treatment intensity and psychosocial outcomes among adult survivors of childhood leukemia. *Journal of Health Psychology, 3*, 23-38.
Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research, 29*, 468-508.
Chen, F. F., & Kenrick, D. T. (2002). Repulsion or attraction? Group membership and assumed attitude similarity. *Journal of Personality and Social Psychology, 83*, 111-125.
Chen, H. Y., & Little, R. (1999). A test of missing completely at random for generalised estimating equation with missing data. *Biometrika, 86*, 1-13.
Chen, Z.-Y., & Kaplan, H. B. (1997). The impact of family structure during adolescence on deviance in early adulthood. *Deviant Behavior, 18*, 365-391.
Cheng, T. –C., & Victoria-Feser, M. –P. (2002). High-breakdown estimation of multivariate mean and covariance with missing observations. *British Journal of Mathematical and Statistical Psychology, 55*, 317-335.
Cheong, Y. F., Fotiu, R. P., & Raudenbush, S. W. (2001). Efficiency and robustness of alternative estimators for two- and three-level models: The case of NAEP. *Journal of Educational and Behavioral Statistics, 26*, 411-429.

Cheong, Y. F., & Raudenbush, S. W. (2000). Measurement and structural models for children's problem behaviors. *Psychological Methods, 5*, 477-495.
Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*, 1-27.
Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology, 31*, 187-212.
Cheung, G. W., & Rensvold, R. B. (2001). The effects of model parsimony and sampling error on the fit of structural equation models. *Organizational Research Methods, 4*, 236-264.
Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.
Cheung, M. W. –L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods, 10*, 40-64.
Chin, W. W. (1998). The partial least squares approach for structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295-336). Mahwah, NJ: Erlbaum.
Chin, W. W., & Newsted, P. R. (1999). Structural equation modeling analysis with small samples using partial least squares. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 307-341). Thousand Oaks, CA: Sage.
Chou, C.-P., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, Lagrange Multiplier, and Wald tests. *Multivariate Behavioral Research, 25*, 115-136.
Chou, C. –P., & Bentler, P. M. (1993). Invariant standardized estimated parameter change for model modification in covariance structure analysis. *Multivariate Behavioral Research, 28*, 97-110.
Chou, C. –P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 37-55). Thousand Oaks, CA: Sage.
Chou, C. –P., & Bentler, P. M. (1996). Application of AIC to Wald and Lagrange multiplier tests in covariance structure analysis. *Multivariate Behavioral Research, 31*, 351-370.
Chou, C. –P., & Bentler, P. M. (2002). Model modification in structural equation modeling by imposing constraints. *Computational Statistics & Data Analysis, 41*, 271-287.
Chou, C. –P., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis. *Structural Equation Modeling, 5*, 247-266.
Chou, C. –P., Bentler, P. M., & Pentz, M. A. (2000). A two-stage approach to multilevel structural equation models: Application to longitudinal data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 33-49). Mahwah, NJ: Erlbaum.
Chou, C.-P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology, 44*, 347-357.
Christensen, H., Jorm, A. F., Mackinnon, A. J., Korten, A. E., Jacomb, P. A., Henderson, A. S., & Rodgers, B. (1999). Age differences in depression and anxiety symptoms: A structural equation modelling analysis of data from a general population sample. *Psychological Medicine, 29*, 325-339.
Christensen, T. C., Wood, J. V., & Barrett, L. F. (2003). Remembering everyday experience through the prism of self-esteem. *Personality and Social Psychology Bulletin, 29*, 51-62.
Chuang, H. Y., Schwartz, J., Gonzales-Cossio, T., Lugo, M. C., Palazuelos, E., Aro, A., Hu, H., & Hernandez-Avila, M. (2001). Interrelations of lead levels in bone, venous blood, and umbilical cord blood with exogenous lead exposure through maternal plasma lead in peripartum women. *Environmental Health Perspectives, 109*, 527-532.
Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18*, 115-126.
Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science, 3*, 186-190.
Cobb, R. J., Davila, J., & Bradbury, T. N. (2001). Attachment security and marital satisfaction: The role of positive perceptions and social support. *Personality and Social Psychology Bulletin, 27*, 1131-1143.
Coenders, G., & Saris, W. E. (2000). Testing additive and multiplicative MTMM models. *Structural Equation Modeling, 7*, 219-251.
Coenders, G., Satorra, A., & Saris, W. E. (1997). Alternative approaches to structural modeling of ordinal data: A Monte Carlo study. *Structural Equation Modeling, 4*, 261-282.
Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regressioncorrelation analysis for the behavioral sciences* (3rd Ed.). Mahwah, NJ: Erlbaum.
Cohen, P., Kasen, S., Brook, J. S., & Hartmark, C. (1998). Behavior patterns of young children and their offspring: A two-generation study. *Developmental Psychology, 34*, 1202-1208.
Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods, 8*, 243-253.
Colder, C. R., Campbell, R. T., Ruel, E., Richardson, J. L., & Flay, B. R. (2002). A finite mixture model of growth trajectories of adolescent alcohol use: Predictors and consequences. *Journal of Consulting and Clinical Psychology, 70*, 976-985.
Colder, C. R., Mott, J., Levy, S., & Flay, B. (2000). The relation of perceived neighborhood danger to childhood aggression: A test of mediating mechanisms. *American Journal of Community Psychology, 28*, 83-103.

Collins, J. M., & Bagozzi, R. P. (1999). Testing the equivalence of the socialization factor structure for criminals and noncriminals. *Journal of Personality Assessment, 72*, 68-73.
Collins, L. M., Cliff, N., McCormick, D. J., & Zatkin, J. L. (1986). Factor recovery via binary data sets: A simulation. *Multivariate Behavioral Research, 21*, 377-391.
Collins, L. M., Schafer, J. L., & Kam, C.–M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*, 330-351.
Comrey, A. L. (1985). A method for removing outliers to improve factor analytic results. *Multivariate Behavioral Research, 20*, 273-281.
Conant, L. L. (1998). Psychological variables associated with pain perceptions among individuals with chronic spinal cord injury pain. *Journal of Clinical Psychology in Medical Settings, 5*, 71-90.
Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics, 19*, 15-18.
Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B, 48*, 133-169.
Cooke, D. J., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment, 13*, 171-188.
Corral-Verdugo, V., & Figueredo, A. J. (1999). Convergent and divergent validity of three measures of conservation behavior: The multitrait-multimethod approach. *Environment & Behavior, 31*, 805-820.
Corten, I. W., Saris, W. E., Coenders, G., van der Veld, W., Aalberts, C. E., & Kornelis, C. (2002). Fit of different models for multitrait-multimethod experiments. *Structural Equation Modeling, 9*, 213-232.
Cox, B. J., Enns, M. W., & Clara, I. P. (2002). The multidimensional structure of perfectionism in clinically depressed and college student samples. *Psychological Assessment, 14*, 365-373.
Cox, D. R., & Wermuth, N. (1996). *Multivariate dependencies: Models, analysis and interpretation.* London: Chapman & Hall.
Crawford, T. N., Cohen, P., & Brook, J. S. (2001a). Dramatic-erratic personality disorder symptoms: I. Continuity from early adolescence into adulthood. *Journal of Personality Disorders, 15*, 319-335.
Crawford, T. N., Cohen, P., & Brook, J. S. (2001b). Dramatic-erratic personality disorder symptoms: II. Developmental pathways from early adolescence into adulthood. *Journal of Personality Disorders, 15*, 336-350.
Crawford, T. N., Cohen, P., Midlarsky, E., & Brook, J. S. (2001). Internalizing symptoms in adolescents: Gender differences in vulnerability to parental distress and discord. *Journal of Research on Adolescence, 11*, 95-118.
Cronbach, L. J. (1988). Internal consistency of tests: Analyses old and new. *Psychometrika, 53*, 63-70.
Crowley, S. L., Fan, X. (1997). Structural equation modeling: Basic concepts and applications in personality assessment research. *Journal of Personality Assessment, 68*, 508-531.
Cruz, J., Joiner, T. E., Jr., Johnson, J. G., Heisler, L. K., Spitzer, R. L., & Pettit, J. W. (2000). Self-defeating personality disorder reconsidered. *Journal of Personality Disorders, 14*, 64-71.
Cua, K. O., Junttila, M. A., & Schroeder, R. G. (2002). A perceptual measure of the degree of development of proprietary equipment. *Structural Equation Modeling, 9*, 579-598.
Cudeck, R. (1988). Multiplicative models and MTMM matrices. *Journal of Educational Statistics, 13*, 131-147.
Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105*, 317-327.
Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 18*, 147-157.
Cudeck, R., & du Toit, S. H. C. (2002). A version of quadratic regression with interpretable parameters. *Multivariate Behavioral Research, 37*, 501-519.
Cudeck, R., du Toit, S., & Sörbom, D. (Eds.) (2001). *Structural equation modeling: Present and future.* Lincolnwood, IL: Scientific Software International.
Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structure analysis and the "problem" of sample size. *Psychological Bulletin, 109*, 512-519.
Cudeck, R., & Klebe, K. J. (2002). Multiphase mixed-effects models for repeated measures data. *Psychological Methods, 7*, 41-63.
Cumsille, P. E., Sayer, A. G., & Graham, J. W. (2000). Perceived exposure to peer and adult drinking as predictors of growth in positive alcohol expectancies during adolescence. *Journal of Consulting & Clinical Psychology, 68*, 531-536.
Curkovic, S., Melnyk, S. A., Handfield, R. B., & Calantone, R. (2000). Investigating the linkage between total quality management and environmentally responsible manufacturing. *IEEE Transactions on Engineering Management, 47*, 444-464.
Curran, P. J. (1994). *The robustness of confirmatory factor analysis to model misspecification and violations of normality.* Ph.D. dissertation, Arizona State University.
Curran, P. J. (2000). A latent curve framework for the study of developmental trajectories in adolescent substance use. In J. R. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods for new questions* (pp. 1-42). Mahwah, NJ: Erlbaum.
Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 107-135). Washington, DC: American Psychological Association.
Curran, P. J., Bollen, K. A., Chen, F., Paxton, P., & Kirby, J. B. (2003). Finite sampling properties of the point estimates and confidence intervals of the RMSEA. *Sociological Methods & Research, 32*, 208-252.

Curran, P. J., Bollen, K. A., Paxton, P., Kirby, J., & Chen, F. (2002). The noncentral chi-square distribution in misspecified structural equation models: Finite sample results from a Monte Carlo simulation. *Multivariate Behavioral Research, 37*, 1-36.

Curran, P. J., & Hussong, A. M. (2002). Structural equation modeling of repeated measures data: Latent curve analysis. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Methods and applications* (pp. 59-85). Mahwah, NJ: Erlbaum.

Curran, P. J., & Muthén, B. O. (1999). The application of latent curve analysis to testing developmental theories in intervention research. *American Journal of Community Psychology, 27*, 567-595.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16-29.

Daly, D. L, Bath, K. E., & Nesselroade, J. R. (1974). On the confounding of inter- and intraindividual variability in examining change patterns. *Journal of Clinical Psychology, 30*, 33-36.

Damphousse, K. R., Kaplan, H. B. (1998). Intervening processes between adolescent drug use and psychological distress: An examination of the self-medication hypothesis. *Social Behavior & Personality, 26*, 115-130.

Davis, P. J., Reeves, J. L., Hastie, B. A., Graff-Radford, S. B., & Naliboff, B. D. (2000). Depression determines illness conviction and pain impact: A structural equation modeling analysis. *Pain Medicine, 1*, 238-246.

Davison, A. C., Hinkley, D. V., & Young, G. A. (2003). Recent developments in bootstrap methodology. *Statistical Science, 18*, 141-157.

Deci, E. L., Ryan, R. M., Gagné, M., Leone, D. R., Usunov, J., & Kornazheva, B. P. (2001). Need satisfaction, motivation, and well-being in the work organizations of a former eastern bloc country: A cross-cultural study of self-determination. *Personality and Social Psychology Bulletin, 27*, 930-942.

de Jonge, J., Mulder, M. J., & Nijhuis, F. J. (1999). The incorporation of different demand concepts in the job demand-control model: Effects on health care professionals. *Social Science and Medicine, 48*,1149-1160.

de Leeuw, J. (1983). Models and methods for the analysis of correlation coefficients. *Journal of Econometrics, 22*, 113-137.

de Leeuw, J. (1988). Model selection in multinomial experiments. In T. K. Dijkstra (Ed.), *On model uncertainty and its statistical implications* (pp. 118-138). Berlin: Springer.

de Leeuw, J., & Kreft, I. G. G. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics, 11*, 57-85.

de Leeuw, J., & Kreft, I. G. G. (Eds.) (in press). *Handbook of quantitative multilevel analysis*. Boston: Kluwer Academic.

Della-Maggiore, V., Sekuler, A. B., Grady, C. L., Bennett, P. J., Sekuler, R., & McIntosh, A. R. (2000). Corticolimbic interactions associated with performance on a short-term memory task are modified by age. *Journal of Neuroscience, 20*, 8410-8416.

della Riccia, G., & Shapiro, A. (1982). Minimum rank and minimum trace of covariance matrices. *Psychometrika, 47*, 443-448.

DeMier, R. L., Hynan, M. T., Hatfield, R. F., Varner, M. W., Harris, H. B., & Manniello, R. L. (2000). A measurement model of perinatal stressors: Identifying risk for postnatal emotional distress in mothers of high-risk infants. *Journal of Clinical Psychology, 56*, 89-100.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, Ser. B, *39*, 1-38.

Deng, L.–Y., & Lin, D. K. J. (2000). Random number generation for the new century. *American Statistician, 54*, 145-150.

Devlin, S. J., Gnanadesikan, G., & Kettenring, J. R. (1981). Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association, 76*, 354-362.

Diggle, P. J., Liang, K. –Y., & Zeger, S. L. (1994). *Analysis of longitudinal data*. Oxford: Clarendon.

Dijkstra, T. (1981). *Latent variables in linear stochastic models*. Ph.D. Thesis, University of Groningen.

Dijkstra, T. (1983). Some comments on maximum likelihood and partial least squares methods. *Journal of Econometrics, 22*, 67-90.

Dijkstra, T. K. (1990). Some properties of estimated scale invariant covariance structures. *Psychometrika, 55*, 327-336 (erratum, 1991, *56*, 169).

Dijkstra, T. K. (1992). On statistical inference with parameter estimates on the boundary of the parameter space. *British Journal of Mathematical & Statistical Psychology, 45*, 289-309.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis: Methods and applications*. New York: Wiley.

Dillon, W. R., Kumar, A., & Mulani, N. (1987). Offending estimates in covariance structure analysis: Comments on the causes of and solutions to Heywood cases. *Psychological Bulletin, 101*, 126-135.

Dishion, T. J., & Patterson, G. R. (1999). Model building in developmental psychopathology: A pragmatic approach to understanding and intervention. *Journal of Clinical Child Psychology, 28*, 502-512.

Dishman, R. K., Motl, R. W., Saunders, R. P., Dowda, M., Felton, G., Ward, D. S., & Pate, R. R. (2002). Factorial invariance and latent mean structure of questionnaires measuring social-cognitive determinants of physical activity among black and white adolescent girls. *Preventive Medicine, 34*, 100-108.

Dixon, W. J. (Ed.) (1988). *BMDP statistical software manual, Vols. 1 & 2*. Berkeley, CA: University of California Press.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47*, 309-326.

Dolan, C. V., & Molenaar, P. C. M. (1991). A comparison of four methods of calculating standard errors of maximum-likelihood estimates in the analysis of covariance structure. *British Journal of Mathematical and Statistical Psychology, 44*, 359-368.
Dolan, C. V., & Molenaar, P. C. M. (1994). Testing specific hypotheses concerning latent group differences in multi-group covariance structure analysis with structured means. *Multivariate Behavioral Research, 29*, 203-222.
Dolan, C. V., Molenaar, P. C., & Boomsma, D. I. (1991). Simultaneous genetic analysis of longitudinal means and covariance structure in the simplex model using twin data. *Behavior Genetics, 21*, 49-65.
Dolan, C. V., & van der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika, 63*, 227-253.
Dolan, C., van der Sluis, S., & Grasman, R. (2005). A note on normal theory power calculation in SEM with data missing completely at random. *Structural Equation Modeling, 12*, 245-262.
Dormann, C., & Zapf, D. (1999). Social support, social stressors at work, and depressive symptoms: Testing for main and moderating effects with structural equations in a three-wave longitudinal study. *Journal of Applied Psychology, 84*, 874-884.
Doyle, A. B., Markiewicz, D., Brendgen, M., Lieberman, M., & Voss, K. (2000). Child attachment security and self-concept: Associations with mother and father attachment style and marital quality. *Merrill-Palmer Quarterly, 46*, 514-539.
Drewes, D. W. (2000). Beyond the Spearman-Brown: A structural approach to maximal reliability. *Psychological Methods, 5*, 214-227.
Duckitt, J., Wagner, C., du Plessis, H., & Birum, I. (2002). The psychological bases of ideology and prejudice: Testing a dual process model. *Journal of Personality and Social Psychology, 83*, 75-93.
Dudgeon, P. (2004). A note on extending Steiger's (1998) multiple sample RMSEA adjustment to other noncentrality parameter-based statistics. *Structural Equation Modeling, 11*, 305-319.
Dudgeon, P., Mackinnon, A. J., Bell, R. C., & McGorry, P. D. (2001). Failure to confirm the 3 domains model of psychosis. *Archives of General Psychiatry, 58*, 94-96.
Dumenci, L. (2000). Multitrait-multimethod analysis. In E. A. H. Tinsley & S. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 583-611). San Diego: Academic.
Duncan, S. C., Duncan, T. E., Biglan, A., & Ary, D. (1998). Contributions of the social context to the development of adolescent substance use: A multivariate latent growth modeling approach. *Drug & Alcohol Dependence, 50*, 57-71.
Duncan, S. C., Duncan, T. E., & Hops, H. (1994). The effect of family cohesiveness and peer encouragement on the development of adolescent alcohol use: A cohort-sequential approach to the analysis of longitudinal data. *Journal of Studies on Alcohol, 55*, 588-599.
Duncan, S. C., Duncan, T. E., & Hops, H. (1996). The role of parents and older siblings in predicting adolescent substance use: Modeling development via structural equation latent growth methodology. *Journal of Family Psychology, 10*, 158-172.
Duncan, S. C., Duncan, T. E., & Hops, H. (1998). Progressions of alcohol, cigarette, and marijuana use in adolescence. *Journal of Behavioral Medicine, 21*, 375-388.
Duncan, S. C., Duncan, T. E., & Strycker, L. A. (2002). A multilevel analysis of neighborhood context and youth alcohol and drug problems. *Prevention Science, 3*, 125-133.
Duncan, S. C., Duncan, T. E., Strycker, L. A., & Chaumeton, N. R. (2002). Relations between youth antisocial and prosocial activities. *Journal of Behavioral Medicine, 25*, 425-438.
Duncan, S. C., Strycker, L. A., & Duncan, T. E. (1999). Exploring associations in developmental trends of adolescent substance use and risky sexual behavior in a high-risk population. *Journal of Behavioral Medicine, 22*, 21-34.
Duncan, S. C., Strycker, L. A., Duncan, T. E., & Okut, H. (2002). A multilevel contextual model of family conflict and deviance. *Journal of Psychopathology & Behavioral Assessment, 24*, 169-175.
Duncan, T. E., & Duncan, S. C. (1995). Modeling the processes of development via latent variable growth curve methodology. *Structural Equation Modeling, 2*, 187-213.
Duncan, T. E., Duncan, S. C., Alpert, A., Hops, H., Stoolmiller, M., & Muthén, B. (1998). Latent variable modeling of longitudinal and multilevel substance use data. *Multivariate Behavioral Research, 32*, 275-318.
Duncan, T. E., Duncan, S. C., & Stoolmiller, M. (1994). Modeling developmental processes using latent growth curve structural equation methodology. *Applied Psychological Measurement, 18*, 343-354.
Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Erlbaum.
Dunkley, D. M., Zuroff, D. C., & Blankstein, K. R. (2003). Self-critical perfectionism and daily affect: Dispositional and situational influences on stress and coping. *Journal of Personality and Social Psychology, 84*, 234-252.
Dunn, G. (1999). The problem of measurement error in modelling the effect of compliance in a randomized trial. *Statistics in Medicine, 18*, 2863-2877.
Dunn, G., Everitt, B., & Pickles, A. (1993). *Modelling covariances and latent variables using EQS*. London: Chapman & Hall.
du Toit, S. H. C., & Browne, M. W. (2001). The covariance structure of a vector ARMA time series. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 279-314). Lincolnwood, IL: Scientific Software International.
du Toit, S. H. C., & Cudeck, R. (2001). The analysis of nonlinear random coefficient regression models with LISREL using constraints. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 259-278). Lincolnwood, IL: Scientific Software International.
du Toit, S. H. C., & du Toit, M. (in press). Multilevel structural equation modeling. In J. de Leeuw & I. G. G. Kreft (Eds.), *Handbook of quantitative multilevel analysis*. Boston: Kluwer Academic.

Dwyer, J. H. (1992). Differential equation models for longitudinal data. In J. H. Dwyer, M. Feinlieb, P. Lippert, & H. Hoffmeister (Eds.), *Statistical models for longitudinal studies of health* (pp. 71-98). New York: Oxford.
Ebreo, A., Linn, N., & Vining, J. (1996). The impact of procedural justice on opinions of public policy: Solid waste management as an example. *Journal of Applied Social Psychology, 26*, 1259-1285.
Edwards, D. (1995). *Introduction to graphical modeling.* New York: Springer Verlag.
Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*, 155-174.
Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7*, 1-26.
Efron, B. (1982). *The jackknife, bootstrap and other resampling plans.* Philadelphia: SIAM.
Efron, B. (2000). The bootstrap and modern statistics. *Journal of the American Statistical Association, 95*, 1293-1296.
Efron, B. (2003). Second thoughts on the bootstrap. *Statistical Science, 18*, 135-140.
Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. London: Chapman & Hall.
Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology, 84*, 5-17.
Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika, 65*, 241-261.
Elliot, A. J., & McGregor, H. A. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*, 501-519.
Elliott, D. S. (1976). *National youth survey.* Boulder, CO: ICPSR.
Ellis, D., & Allaire, J. C. (1999). Modeling computer interest in older adults: The role of age, education, computer knowledge, and computer anxiety. *Human Factors, 41*, 345-55.
Enders, C. K. (2001a). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling, 8*, 128-141.
Enders, C. K. (2001b). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods, 6*, 352-370.
Enders, C. K. (2002). Applying the Bollen-Stine bootstrap for goodness-of-fit measures to structural equation models with missing data. *Multivariate Behavioral Research, 37*, 359-377.
Enders, C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological Methods, 8*, 322-337.
Enders, C. K., & Bandalos, D. L. (1999). The effects of heterogeneous item distributions on reliability. *Applied Measurement in Education, 12*, 133-150.
Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*, 430-457.
Enders, C. K., & Peugh, J. L. (2004). Using an EM covariance matrix to estimate structural equation models with missing data: Choosing an adjusted sample size to improve the accuracy of inferences. *Structural Equation Modeling, 11*, 1-19.
Engle, R. F. (1984). Wald, likelihood ratio, and Lagrange Multiplier tests in econometrics. In A. Griliches & M. D. Intriligator (Eds.), *Handbook of econometrics* (pp. 776-826). Amsterdam: North-Holland.
Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*, 309-331.
Epstein, J. A., Griffin, K. W., & Botvin, G. J. (2000a). Role of general and specific competence skills in protecting inner-city adolescents from alcohol use. *Journal of Studies on Alcohol, 61*, 379-386.
Epstein, J. A., Griffin, K. W., & Botvin, G. J. (2000b). Competence skills help deter smoking among inner city adolescents. *Tobacco Control, 9*, 33-39.
Epstein, J. A., Griffin, K. W., & Botvin, G. J. (2000c). A model of smoking among inner-city adolescents: The role of personal competence and perceived social benefits of smoking. *Preventive Medicine, 31*, 107-114.
Epstein, J. A., Griffin, K. W., & Botvin, G. J. (2001). Risk taking and refusal assertiveness in a longitudinal model of alcohol use among inner-city adolescents. *Prevention Science, 2*, 193-200.
Epstein, J. A., Griffin, K. W., & Botvin, G. J. (2002). Positive impact of competence skills and psychological wellness in protecting inner-city adolescents from alcohol use. *Prevention Science, 3*, 95-104.
Erickson, D. J., Wolfe, J., King, D. W., King, L. A., & Sharkansky, E. J. (2001). Posttraumatic stress disorder and depression symptomatology of Gulf War veterans: A prospective analysis. *Journal of Consulting and Clinical Psychology, 69*, 41-49.
Escudero, A., Iriondo, J. M., Olano, J. M., Rubio, A., & Somolinos, R. C. (2000). Factors affecting establishment of a gypsophyte: The case of Lepidium subulatum (Brassicaceae). *American Journal of Botany, 87*, 861-871.
Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review, 9*, 3-25.
Etezadi-Amoli, A., & McDonald, R. P. (1983). A second generation of nonlinear factor analysis. *Psychometrika, 48*, 315-342.
Everitt, B. S. (1984). *An introduction to latent variable models*. London: Chapman & Hall.
Evans, M. A. (1999). An analysis and cross-validation of the role of information during downsizing. *Military Psychology, 11*, 345-364.
Fabrigar, L. R., Visser, P. S., & Browne, M. W. (1997). Conceptual and methodological issues in testing the circumplex structure of data in personality and social psychology. *Personality and Social Psychology Review, 1*, 184-203.
Fan, X. (2003a). Power of latent growth modeling for detecting group differences in linear growth trajectory parameters. *Structural Equation Modeling, 10*, 380-400.

Fan, X. (2003b). Using commonly available software for bootstrapping in both substantive and measurement analyses. *Educational and Psychological Measurement, 63,* 24-50.
Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling*, *12*, 343-367.
Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, *6*, 56-83.
Fang, C. Y., Manne, S. L., & Paper, S. J. (2001). Functional impairment, marital quality, and patient psychological distress as predictors of psychological distress among cancer patients' spouses. *Health Psychology*, *20*, 452-457.
Fang, K. –T., & Anderson, T. W. (1990). *Statistical inference in elliptically contoured and related distributions.* New York: Allerton Press.
Fang, K. –T., Kotz, S., & Ng, K. W. (1990). *Symmetric multivariate and related distributions*. New York: Chapman and Hall.
Fang, K. –T., Li, R., Zhu, L. (1998). A projection NT-type test of elliptical symmetry based on the skewness and kurtosis measures. *Acta Mathematica. Applicatae Sinica* (English Ser.), *14*, 314-323.
Fang, K. –T., Yuan, K. –H., & Bentler, P. M. (1992). Applications of sets of points uniformly distributed on a sphere to testing multinormality and robust estimation. In Z. –P. Jiang, S. –J. Yan, P. Cheng, & R. Wu, (Eds.), *Probability and statistics* (pp. 56-73). Singapore: World Scientific.
Faulbaum, F., & Bentler, P. M. (1994). Causal modeling: Some trends and perspectives. In I. Borg & P. P. Mohler (Eds.), *Trends and perspectives in empirical social research* (pp. 224-249). Berlin: Walter de Gruyter.
Fawcett, S. E., & Myers, M. B. (2001). Product and employee development in advanced manufacturing: Implementation and impact. *International Journal of Production Research*, *39*, 65-79.
Federico, C. M., & Sidanius, J. (2002). Racism, ideology, and affirmative action revisited: The antecedents and consequences of "principled objections" to affirmative action. *Journal of Personality and Social Psychology, 82*, 488-502.
Feldman, P. J., Dunkel-Schetter, C., Sandman, C. A., & Wadhwa, P. D. (2000). Maternal social support predicts birth weight and fetal growth in human pregnancy. *Psychosomatic Medicine*, *62*, 715-725.
Feldt, L. S., & Charter, R. A. (2003). Estimating the reliability of a test split into two parts of equal or unequal length. *Psychological Methods, 8*, 102-109.
Felix-Ortiz, M., & Newcomb, M. D. (1999). Vulnerability for drug use among Latino adolescents. *Journal of Community Psychology, 27*, 257-280.
Fenlon, M. R., Sherriff, M., & Walter, J. D. (2000). An investigation of factors influencing patients' use of new complete dentures using structural equation modelling techniques. *Community Dentistry and Oral Epidemiology*, *28*,133-140.
Ferguson, T. S. (1958). A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities. *Annals of Mathematical Statistics, 29*, 1046-1062.
Ferrando, P. J., & Lorenzo-Seva, U. (2001). Fitting the normal ogive factor analytic model to scores on tests. *Multivariate Behavioral Research, 36*, 445-469.
Ferrara, F. F. (1999). Validation of the child sex abuse attitude scale through confirmatory factor analysis. *Structural Equation Modeling, 6*, 99-112.
Ferrer, E., & McArdle, J. J. (2003). Alternative structural models for multivariate longitudinal data. *Structural Equation Modeling, 10*, 493-524.
Figueredo, A.J., Brooks, A.J., Leff, H.S., & Sechrest, L. (2000). A meta-analytic approach to growth curve analysis. *Psychological Reports, 87*, 441-465.
Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling*, *4*, 87-107.
Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, *44*, 409-420.
Fleishman, A. I. (1978). A method of simulating non-normal distributions. *Psychometrika*, *43*, 521-532.
Fontana, A., Litz, B., & Rosenheck, R. (2000). Impact of combat and sexual harassment on the severity of posttraumatic stress disorder among men and women peacekeepers in Somalia. *Journal of Nervous and Mental Disease*, *188*, 163-169.
Foster, M. D. (2000). Utilization of global attributions in recognizing and responding to gender discrimination among college women. *Current Psychology: Developmental, Learning, Personality, Social*, *19*, 57-69.
Fouladi, R. T. (1999, April). *Model fit in covariance structure analysis under small sample conditions: Modified maximum likelihood and asymptotically distribution free generalized least squares procedures.* Paper presented at American Educational Research Association meetings, Montreal, Canada.
Fouladi, R. T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Structural Equation Modeling*, *7*, 356-410.
Fouladi, R., & Steiger, J. H. (1999). Tests of identity correlation structure. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 167-194). Thousand Oaks, CA: Sage.
Fox, J. (1980). Effect analysis in structural equation models: Extensions and simplified methods of computation. *Sociological Methods & Research, 9*, 3-28.
Fox, J. (1985). Effect analysis in structural-equation models II: Calculation of specific indirect effects. *Sociological Methods & Research, 14*, 81-95.
Frank, E., & Brandstätter, V. (2002). Approach versus avoidance: Different types of commitment in intimate relationships. *Journal of Personality and Social Psychology, 82*, 208-221.
Frane, J. W. (1977). A note on checking tolerance in matrix inversion and regression. *Technometrics, 19*, 513-514.

Fraser, C., & McDonald, R. P. (1988). NOHARM: Least squares item factor analysis. *Multivariate Behavioral Research*, *23*, 267-269.
Freedman, D. A. (1987). As others see us: A case study in path analysis. *Journal of Educational Statistics*, *12*, 101-128. (With discussion)
Freund, A. M., & Baltes, P. B. (2002). Life-management strategies of selection, optimization, and compensation: Measurement by self-report and construct validity. *Journal of Personality and Social Psychology, 82*, 642-662.
Frosch, D. L., Stein, J. A., & Shoptaw, S. (2002). Using latent-variable models to analyze smoking cessation clinical trial data: An example among the methadone maintained. *Experimental and Clinical Psychopharmacology*, *10*, 258-267.
Fuller, W. A. (1987). *Measurement error models*. New York: Wiley.
Furlow, C. F., & Beretvas, S. N. (2005). Meta-analytic methods of pooling correlation matrices for structural equation modeling under different patterns of missing data. *Psychological Methods, 10*, 227-254.
Galaif, E. R., Newcomb, M. D., & Carmona, J. V. (2001). Prospective relationships between drug problems and work adjustment in a community sample of adults. *Journal of Applied Psychology, 86*, 337-350.
Galaif, E. R., Nyamathi, A. M., & Stein, J. A.(1999). Psychosocial predictors of current drug use, drug problems, and physical drug dependence in homeless women. *Addictive Behaviors, 24*, 801-814.
Gallini, J. K., & Casteel, J. F. (1987). An inquiry into the effects of outliers on estimates of a structural equation model of basic skills assessment. In P. Cuttance & R. Ecob (Eds.), *Structural modeling by example* (pp. 189-201). New York: Cambridge.
Gallini, J. K., & Mandeville, G. K. (1984). An investigation of the effect of sample size and specification error on the fit of structural equation models. *Journal of Experimental Education, 53*, 9-19.
Garst, H. (2000). *Longitudinal research using structural equation modeling applied in studies of determinants of psychological well-being and personal initiative in East Germany after the unification*. Unpublished doctoral dissertation, University of Amsterdam.
Gayán, J., & Olson, R. K. (1999). Reading disability: Evidence for a genetic etiology. *European Child and Adolescent Psychiatry, 8*, 52-55.
Geiger, D., & Heckerman, D. (2002). Parameter priors for directed acyclic graphical models and the characterization of several probability distributions. *Annals of Statistics, 30*, 1412-1440.
Gerbing, D. W., & Anderson J. C. (1985). The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. *Multivariate Behavioral Research*, *20*, 255-271.
Gerbing, D. W., & Anderson, J. C. (1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. *Psychometrika, 52*, 99-111.
Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40-65). Newbury Park, CA: Sage.
Gerpott, T. J., Rams, W., & Schindler, A. (2001). Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. *Telecommunications Policy, 25*, 249-269.
Geurts, S. A., Schaufeli, W. B., & Rutte, C. G. (1999). Absenteeism, turnover intention and inequity in the employment relationship. *Work & Stress, 13*, 253-267.
Geweke, J. F., & Singleton, K. J. (1980). Interpreting the likelihood ratio statistic in factor models when the sample size is small. *Journal of the American Statistical Association*, *75*, 133-137.
Ghisletta, P., & McArdle, J. J. (2001). Latent growth curve analyses of the development of height. *Structural Equation Modeling, 8*, 531-555.
Gibbons, D. E., & Weingart, L. R. (2001). Can I do it? Will I try? Personal efficacy, assigned goals, and performance norms as motivators of individual performance. *Journal of Applied Social Psychology*, *31*, 624-648.
Gil, A. G., Wagner, E. F., & Vega, W. A. (2000). Acculturation, familism and alcohol use among Latino adolescent males: Longitudinal relations. *Journal of Community Psychology*, *28*, 443-458.
Gill, R. D., & Robins, J. M. (2001). Causal inference for complex longitudinal data: The continuous case. *Annals of Statistics, 29*, 1785-1811.
Gleason, T. C. & Staelin, R. (1975). A proposal for handling missing data. *Psychometrika*, *40*, 229-252.
Glymour, C., Scheines, R., Spirtes, P., & Kelly, K. (1987). *Discovering causal structure*. Orlando: Harcourt, Brace, Jovanovich.
Gold, M. S., & Bentler, P. M. (2000). Treatments of missing data: A Monte Carlo comparison of resemblance-based hot-deck imputation, iterative stochastic regression imputation, and expectation-maximization. *Structural Equation Modeling*, *7*, 319-355.
Gold, M. S., Bentler, P. M., & Kim, K. H. (2003). A comparison of maximum-likelihood and asymptotically distribution-free methods of treating incomplete nonnormal data. *Structural Equation Modeling, 10*, 47-79.
Goldberger, A. S., & Duncan, O. D. (Eds.) (1973). *Structural equation models in the social sciences*. New York: Seminar.
Goldenberg, C., Gallimore, R., Reese, L., & Garnier, H. (2001). Cause or effect? A longitudinal study of immigrant Latino parents' aspirations and expectations, and their children's school performance. *American Educational Research Journal, 38*, 547-582.
Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika, 73*, 43-56.
Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Edward Arnold.
Goldstein, H., Browne, W., & Rasbash, J. (2002). Multilevel modeling of medical data. *Statistics in Medicine, 21*, 3291-3315.
Goldstein, H. & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, *53*, 435-467.

Goldstein, R. B., Prescott, C. A., & Kendler, K. S. (2001). Genetic and environmental factors in conduct problems and adult antisocial behavior among adult female twins. *Journal of Nervous & Mental Diseases, 189*, 201-209.
Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58*, 80-92.
Golob, T. F., & Regan, A. C. (2001). Impacts of highway congestion on freight operations: Perceptions of trucking industry managers. *Transportation Research, 35*, 577-599.
Gomez, J. M., & Zamora, R. (2000). Spatial variation in the selective scenarios of Hormathophylla spinosa (Cruciferae). *American Naturalist, 155*, 657-668.
Gomez, R., Burns, G. L., Walsh, J. A., & de Moura, M. A. (2003). A multitrait-multisource confirmatory factor analytic approach to the construct validity of ADHD rating scales. *Psychological Assessment, 15*, 3-16.
Gonzalez, R., & Griffin, D. (2001). Testing parameters in structural equation modeling: Every "one" matters. *Psychological Methods, 6*, 258-269.
Goodyear, R. K., Newcomb, M. D., & Allison, R. D. (2000). Predictors of Latino men's paternity in teen pregnancy: Test of a mediational model of childhood experiences, gender role attitudes, and behaviors. *Journal of Counseling Psychology, 47*, 116-128.
Gorsuch, R. L. (2003). Factor analysis. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (pp. 143-164). Hoboken, NJ: Wiley.
Graff, J., & Schmidt, P. (1982). A general model for decomposition of effects. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation, Part I* (pp. 131-148). Amsterdam: North-Holland.
Graham, J. M., Guthrie, A. C., & Thompson, B. (2003). Consequences of not interpreting structure coefficients in published CFA research: A reminder. *Structural Equation Modeling, 10*, 142-153.
Graham, J. W. (2003). Adding missing-data relevant variables to FIML-based structural equation models. *Structural Equation Modeling, 10*, 80-100.
Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). Methods for handling missing data. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (pp. 87-114). Hoboken, NJ: Wiley.
Graham, J. W., & Hofer, S. M. (2000). Multiple imputation in multivariate research. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 201-218). Mahwah, NJ: Erlbaum.
Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research, 31*, 197-218.
Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing-data designs in analysis of change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 335-353). Washington, DC: American Psychological Association.
Granzin, K. L., & Olsen, J. E. (1998). Americans' choice of domestic over foreign products: A matter of helping behavior? *Journal of Business Research, 43*, 39-53.
Grayson, D. A., Mackinnon, A., Jorm, A. F., Creasey, H., & Broe, G. A. (2000). Item bias in the center for epidemiologic studies depression scale: Effects of physical disorders and disability in an elderly community sample. *Journals of Gerontology Series B, 55*, 273-282.
Green, B. F., Jr. (1950). A note on the calculation of weights for maximum battery reliability. *Psychometrika, 15*, 57-61.
Green, S. B. (2003). A coefficient alpha for test-retest data. *Psychological Methods, 8*, 88-101.
Green, S. B., & Babyak, M. A. (1997). Control of Type I error with multiple tests of constraints in structural equation modeling. *Multivariate Behavioral Research, 32*, 39-52.
Green, S. B., & Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling, 7*, 251-270.
Green, S. B., Thompson, M. S., & Babyak, M. A. (1998). A Monte Carlo investigation of methods for controlling Type I errors with specification searches in structural equation modeling. *Multivariate Behavioral Research, 33*, 365-383.
Green, S. B., Thompson, M. S., & Poirier, J. (1999). Exploratory analyses to improve model fit: Errors due to misspecification and strategy to reduce their occurrence. *Structural Equation Modeling, 6*, 113-126.
Green, S. B., Thompson, M. S., & Poirier, J. (2001). An adjusted Bonferroni method for elimination of parameters in specification addition searches. *Structural Equation Modeling, 8*, 18-39.
Greene, V. L. (1977). An algorithm for total and indirect causal effects. *Political Methodology, 44*, 369-381.
Greenland, S. (2000a). Causal analysis in the health sciences. *Journal of the American Statistical Association, 95*, 286-289.
Greenland, S. (2000b). Principles of multilevel modeling. *International Journal of Epidemiology, 29*, 158-167.
Greenland, S. (2002). A review of multilevel theory for ecologic analyses. *Statistics in Medicine, 21*, 389-395.
Greenland, S., Schesselman, J. J., & Criqui, M. H. (1986). The fallacy of employing standardized regression coefficients and correlations as measures of effect. *American Journal of Epidemiology, 123*, 203-208.
Grella, C. E., Hser, Y. I., Joshi, V., & Anglin, M. D. (1999). Patient histories, retention, and outcome models for younger and older adults in DATOS. *Drug and Alcohol Dependence, 57*, 151-166.
Griffin, K. W., Botvin, G. J., Scheier, L. M., Epstein, J. A., & Doyle, M. M. (2002). Personal competence skills, distress, and well-being as determinants of substance use in a predominantly minority urban adolescent sample. *Prevention Science, 3*, 23-42.

Griffin, K. W., Epstein, J. A., Botvin, G. J., Spoth, R. L. (2001). Social competence and substance use among rural youth: Mediating role of social benefit expectancies of use. *Journal of Youth & Adolescence, 30*, 485-498.
Griffin, K. W., Scheier, L. M., Botvin, G. J., & Diaz, T. (2000). Ethnic and gender differences in psychosocial risk, protection, and adolescent alcohol use. *Prevention Science, 1*, 199-212.
Griffin, K. W., Scheier, L. M., Botvin, G. J., & Diaz, T. (2001). Protective role of personal competence skills in adolescent substance use: Psychological well-being as a mediating factor. *Psychology of Addictive Behaviors, 15*, 194-203.
Griffith, J. (1999). School climate as "social order" and "social action": A multi-level analysis of public elementary school student perceptions. *Social Psychology of Education, 2*, 339-369.
Grote, N. K., & Clark, M. S. (2001). Perceiving unfairness in the family: Cause or consequence of marital distress? *Journal of Personality and Social Psychology, 80*, 281-293.
Grube, J. W., & Voas, R. B. Predicting underage drinking and driving behaviors. *Addiction, 91*, 1843-1857.
Gump, B. B., Matthews, K. A., & Raeikkoenen, K. (1999). Modeling relationships among socioeconomic status, hostility, cardiovascular reactivity, and left ventricular mass in African American and White children. *Health Psychology, 18*, 140-150.
Guo, G., & Harris, K. M. (2000). The mechanisms mediating the effects of poverty on children's intellectual development. *Demography, 37*, 431-447.
Gupta, A. K., & Varga, T. (1993). *Elliptically contoured models in statistics*. Dordrecht: Kluwer Academic.
Gutman, L. M., & Eccles, J. S. (1999). Financial strain, parenting behaviors, and adolescents' achievement: Testing model equivalence between African American and European American single- and two-parent families. *Child Development, 70*, 1464-1476.
Gwaltney, C. J., Shiffman, S., Norman, G. J., Paty, J. A., Kassel, J. D., Gnys, M., Hickox, M., Waters, A., & Balabanis, M. (2001). Does smoking abstinence self-efficacy vary across situations? Identifying context-specificity within the relapse situation efficacy questionnaire. *Journal of Consulting and Clinical Psychology, 69*, 516-527.
Hagemann, D., Naumann, E., Thayer, J. F., & Bartussek, D. (2002). Does resting electroencephalograph asymmetry reflect a trait? An application of latent state-trait theory. *Journal of Personality and Social Psychology, 82*, 619-641.
Hagtvet, K. A., & Nasser, F. M. (2004). How well do item parcels represent conceptually defined latent constructs? A two-facet approach. *Structural Equation Modeling, 11*, 168-193.
Hakeberg M., Hagglin C., Berggren U., & Carlsson, S. G. (2001). Structural relationships of dental anxiety, mood, and general anxiety. *Acta Odontologica Scandinavica, 59*, 99-103.
Hakstian, A. R., & Barchard, K. A. (2000). Toward more robust inferential procedures for coefficient alpha under sampling of both subjects and conditions. *Multivariate Behavioral Research, 35*, 427-456.
Hakstian, A. R., Zidek, J. V., & McDonald, R. P. (1977). Best univocal estimates of orthogonal factors. *Psychometrika, 42*, 627-630.
Haley, R. W., Luk, G. D., & Petty, F. (2001). Use of structural equation modeling to test the construct validity of a case definition of Gulf War syndrome: Invariance over developmental and validation samples, service branches and publicity. *Psychiatry Research, 102*, 175-200.
Hall, P. (2003). A short prehistory of the bootstrap. *Statistical Science, 18*, 158-167.
Hall, R. J., Snell, A. F., & Foust, M. S. (1999). Item parcelling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods, 2*, 233-256.
Hamagami, F., & McArdle, J. J. (2001). Advanced studies of individual differences linear dynamic models for longitudinal data analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 203-246). Mahwah, NJ: Erlbaum.
Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2003). ARMA-based SEM when the number of time points T exceeds the number of cases N: Raw data maximum likelihood. *Structural Equation Modeling, 10*, 352-379.
Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association, 69*, 383-393.
Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. New York: Wiley.
Hancock, G. R. (1999). A sequential Scheffe-type respecification procedure for controlling Type I error in exploratory structural equation model modification. *Structural Equation Modeling, 6*, 158-168.
Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika, 66*, 373-388.
Hancock, G. R., Kuo, W. –L., & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling, 8*, 470-489.
Hancock, G. R., Lawrence, F. R., & Nevitt, J. (2000). Type I error and power of latent mean methods and MANOVA in factorially invariant and noninvariant latent variable systems. *Structural Equation Modeling, 7*, 534-556.
Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 195-216). Lincolnwood, IL: Scientific Software International.
Harford, T. C., & Muthén, B. O. (2001). The dimensionality of alcohol abuse and dependence: A multivariate analysis of DSM-IV symptom items in the National Longitudinal Survey of Youth. *Journal of Studies on Alcohol, 62*, 150-157.

Harlow, L. L. (1985). *Behavior of some elliptical theory estimators with nonnormal data in a covariance structures framework: A Monte Carlo study.* Ph.D. Thesis, University of California, Los Angeles.
Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.
Harned, M. S. (2000). Harassed bodies - An examination of the relationships among women's experiences of sexual harassment, body image, and eating disturbances. *Psychology of Women Quarterly, 24*, 336-348.
Harned, M. S., & Fitzgerald, L. F. (2002). Understanding a link between sexual harassment and eating disorder symptoms: A mediational analysis. *Journal of Consulting and Clinical Psychology, 70*, 1170-1181.
Hasegawa, M., Mishima, M., Matsumoto, I., Sasaki, T., Kimura, T., Baba, Y., Senami, K., Kanemura, K., Takano, O., Shibata, T. (2001). Confirming the theoretical structure of the Japanese version of the McGill Pain Questionnaire in chronic pain. *Pain Medicine, 2*, 52-59.
Haugland, S., Wold, B., Stevenson, J., Aaroe, L. E., & Woynarowska, B. (2001). Subjective health complaints in adolescence - A cross-national comparison of prevalence and dimensionality. *European Journal of Public Health, 11*, 4-10.
Hayashi, K., & Bentler, P. M. (2000a). On the relations among regular, equal unique variances, and image factor analysis models. *Psychometrika, 65*, 59-72.
Hayashi, K., & Bentler, P. M. (2000b). The asymptotic covariance matrix of maximum-likelihood estimates in factor analysis: The case of nearly singular matrix of estimates of unique variances. *Linear Algebra and Its Applications, 321*, 153-173.
Hayashi, K., & Yuan, K.–H. (2003). Robust Bayesian factor analysis. *Structural Equation Modeling,* 10, 525-533.
Hayduk, L. A. (1987). *Structural equation modelling with LISREL: Essentials and advances.* Baltimore: Johns Hopkins.
Hayduk, L. A. (1996). *LISREL issues, debates, and strategies.* Baltimore: Johns Hopkins.
Hayduk, L. A., & Glaser, D. N. (2000). Jiving the four-step, waltzing around factor analysis, and other serious fun. *Structural Equation Modeling, 7,* 1-35, 111-123.
Hays, R. D., Widaman, K. F., DiMatteo, M. R., & Stacy, A. W. (1987). Structural equation models of current drug use: Are appropriate models so simple(x)? *Journal of Personality and Social Psychology, 52*, 134-144.
Headrick, T. C. (2002). Fast fifth-order polynomial transforms for generating univariate and multivariate nonnormal distributions. *Computational Statistics & Data Analysis, 40*, 685-711.
Headrick, T. C., & Sawilowsky, S. S. (1999). Simulating correlated multivariate nonnormal distributions: Extending the Fleishman power method, *Psychometrika, 64*, 25-35.
Headrick, T. C. & Sawilowsky, S. S. (2000). Weighted simplex procedures for determining boundary points and constants for the univariate and multivariate power methods. *Journal of Educational and Behavioral Statistics, 25*, 417-436.
Heck, R. H., & Thomas, S. L. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Erlbaum.
Hedecker, D., & Gibbons, R. (1997). Application of random effects pattern mixture models for missing data in social sciences. *Psychological Methods, 2*, 64-78.
Hedecker, D., & Rose, J. S. (2000). The natural history of smoking: A pattern-mixture random-effects regression model. In J. R. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods for new questions* (pp. 79-112). Mahwah, NJ: Erlbaum.
Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods, 6*, 203-217.
Heermann, E. F. (1963). Univocal or orthogonal estimators of orthogonal factors. *Psychometrika, 32*, 363-379.
Heise, D. R., & Bohrnstedt, G. W. (1970). Validity, invalidity, and reliability. In E. F. Borgatta (Ed.), *Sociological methodology 1970* (pp. 104-129). San Francisco: Jossey-Bass.
Helson, R., & Srivastava, S. (2002). Creative and wise people: Similarities, differences, and how they develop. *Personality and Social Psychology Bulletin, 28*, 1430-1440.
Hemenover, S. H. (2001). Self-reported processing bias and naturally occurring mood: Mediators between personality and stress appraisals. *Personality and Social Psychology Bulletin, 27*, 387-394.
Henly, S. J. (1993). Robustness of some estimators for the analysis of covariance structures. *British Journal of Mathematical & Statistical Psychology, 46*,313-338.
Hernán, M. Á., Brumback, B., & Robins, J. M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association, 96*, 440-448.
Hernández, A., & González-Romá, V. (2002). Analysis of multitrait-multioccasion data: Additive versus multiplicative models. *Multivariate Behavioral Research, 37*, 59-87.
Herrenkohl, T. I., Huang, B., Kosterman, R., Hawkins, J. D., Catalano, R. F., & Smith, B. H. (2001). A comparison of the social development processes leading to violent behavior in late adolescence for childhood initiators and adolescent initiators of violence. *Journal of Research in Crime and Delinquency, 38*, 45-63.
Hershberger, S. L. (1994). The specification of equivalent models before the collection of data. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis* (pp. 68-108). Thousand Oaks, CA: Sage.
Hershberger, S. L. (1998). Dynamic factor analysis. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 217-250). Mahwah, NJ: Erlbaum.
Hershberger, S. L. (2003). The growth of structural equation modeling: 1994-2001. *Structural Equation Modeling, 10*, 35-46.
Hershberger, S. L., Molenaar, P. C. M., & Corneal, S. E. (1996). A hierarchy of univariate and multivariate time series models. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 159-194). Mahwah, NJ: Erlbaum.
Herting, J. R. (2002). Evaluating and rejecting true mediation models: A cautionary note. *Prevention Science, 3*, 285-289.

Herting, J. R., & Costner, H. L. (2000). Another perspective on "the proper number of factors" and the appropriate number of steps. *Structural Equation Modeling, 7*, 92-110.
Heuer, L., Penrod, S., Hafer, C. L., & Cohn, I. (2002). The role of resource and relational concerns for procedural justice. *Personality and Social Psychology Bulletin, 28*, 1468-1482.
Hines, M., Chiu, L., McAdams, L. A., Bentler, P. M., & Lipcamon, J. (1992). Cognition and the corpus callosum: Verbal fluency, visuospatial ability and language lateralization related to midsagittal surface areas of callosal subregions. *Behavioral Neuroscience, 106*, 3-14.
Hoaglin, D. C., Mosteller, F. & Tukey, J. W. (1983). *Understanding robust and exploratory data analysis*. New York: Wiley.
Hogan, T. P., Benjamin, A., & Brezinsky, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement, 60*, 523-531.
Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*, 945-970.
Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equations models. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 449-484). Washington, DC: American Sociological Association.
Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology, 65*, 599-610.
Holmbeck, G. N. (2002). Post-hoc probing of significant moderational and mediational effects in studies of pediatric populations. *Journal of Pediatric Psychology, 27*, 87-96.
Holsen, I., Kraft, P., & Røysamb, E. (2001). The relationship between body image and depressed mood in adolescence: A 5-year longitudinal panel study. *Journal of Health Psychology, 6*, 613-627.
Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General, 128*, 3-31.
Hoogland, J. J. (1999). *The robustness of estimation methods for covariance structure analysis*. Unpublished Ph.D. dissertation, Rijksuniversiteit Groningen.
Horwitz, B., Friston, K. J., & Taylor, J. G. (2000). Neural modeling and functional brain imaging: An overview. *Neural Networks, 13*, 829-846.
Hoshino, T. (2001). Bayesian inference for finite mixtures in confirmatory factor analysis. *Behaviormetrika, 28*, 37-63.
Howard, G. S., Maxwell, S. E., & Fleming, K. J. (2000). The proof of the pudding: An illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. *Psychological Methods, 5*, 315-332.
Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.
Hox, J. J. (2000). Multilevel analysis of grouped and longitudinal data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*. Mahwah, NJ: Erlbaum.
Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.
Hox, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling, 8*, 157-174.
Hoyle, R. H. (Ed.) (1995). *Structural equation modeling*. Thousand Oaks, CA: Sage.
Hoyle, R. H. (Ed.) (1999). *Statistical strategies for small sample research*. Thousand Oaks, CA: Sage.
Hoyle, R. H. (2000). Confirmatory factor analysis. In H. E. A. Tinsely & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 465-497). New York: Academic Press.
Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195-222). Thousand Oaks, CA: Sage.
Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 158-176). Thousand Oaks, CA: Sage.
Hsu, P. L. (1949). The limiting distribution of functions of sample means and applications to testing hypotheses. *Proceedings of the First Berkeley Symposium in Mathematical Statistics and Probability*, pp. 359-402. Reprinted in Collected Papers, New York: Springer-Verlag, 1983.
Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Issues, concepts, and applications* (pp. 76-99). Newbury Park, CA: Sage.
Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structural equation modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.
Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*, 351-362.
Huang, B., Kosterman, R., Catalano, R. F., Hawkins, J. D., & Abbott, R. D. (2001). Modeling mediation in the etiology of violent behavior in adolescence: A test of the social development model. *Criminology, 39*, 75-107.
Huba, G. J., & Harlow, L. L. (1987). Robust structural equation models: Implications for developmental psychology. *Child Development, 58*, 147-166.
Huber, P. J. (1977). Robust covariances. In S. S. Gupta & D. S. Moore (Eds.), *Statistical decision theory and related topics: Vol. 2* (pp. 165-191). New York: Academic Press.
Huber, P. J. (1981). *Robust statistics*. New York: Wiley.

Hunter, D. R., & Lange, K. (2004). A tutorial on MM algorithms. *American Statistician, 58,* 30-37.
Hunter, J. A., Jr., & Figueredo, A. J. (1999). Factors associated with treatment compliance in a population of juvenile sexual offenders. *Sexual Abuse, 11,* 49-67.
Hunter, J. A., Jr., & Figueredo, A. J. (2000). The influence of personality and history of sexual victimization in the prediction of juvenile perpetrated child molestation. *Behavior Modification, 24,* 241-263.
Hutchinson, S. R., & Olmos, A. (1998). Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Structural Equation Modeling, 5,* 344-364.
Hyman, J. B., & Leibowitz, S. G. (2001). JSEM: A framework for identifying and evaluating indicators. *Environmental Monitoring and Assessment, 66,* 207-232.
Ichikawa, M. & Konishi, S. (1995). Application of the bootstrap methods in factor analysis. *Psychometrika, 60,* 77-93.
Iyer, A., Leach, C. W., & Crosby, F. J. (2003). White guilt and racial compensation: The benefits and limits of self-focus. *Personality and Social Psychology Bulletin, 29,* 117-129.
Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation modeling. *Psychological Bulletin, 117,* 348-357.
Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: A Monte Carlo investigation. *Structural Equation Modeling, 8,* 205-223.
Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling, 10,* 128-141.
Jackson, K. M., & Aiken, L. S. (2000). A psychosocial model of sun protection and sunbathing in young women: The impact of health beliefs, attitudes, norms, and self-efficacy for sun protection. *Health Psychology, 19,* 469-478.
Jackson, P. H., & Agunwamba, C. C. (1977). Lower bounds for the reliability of total scores on a test composed of nonhomogeneous items: I. Algebraic lower bounds. *Psychometrika, 42,* 567-578.
Jacobsen, H. P., Herskind, A. M., Nielsen, B. W., & Husby, S. (2001). IgE in unselected like-sexed monozygotic and dizygotic twins at birth and at 6 to 9 years of age: High but dissimilar genetic influence on IgE levels. *Journal of Allergy and Clinical Immunology, 107,* 659-663.
James, G. M., & Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association, 98,* 397-408.
James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data.* Beverly Hills, CA: Sage.
Jamshidian, M., & Bentler, P. M. (1993). A modified Newton method for constrained estimation in covariance structure analysis. *Computational Statistics & Data Analysis, 15,* 133-146.
Jamshidian, M., & Bentler, P. M. (1998). A quasi-Newton method for minimum trace factor analysis. *Journal of Statistical Computation and Simulation, 62,* 73-89.
Jamshidian, M., & Bentler, P. M. (1999). ML estimation of mean and covariance structures with missing data using complete data routines. *Journal of Educational and Behavioral Statistics, 24,* 21-41.
Jamshidian, M., & Bentler, P. M. (2000). Improved standard errors of standardized parameters in covariance structure models: Implications for construct explication. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment,* (pp. 73-94). Dordrecht, Netherlands: Kluwer Academic.
Jamshidian, M., & Jennrich, R. I. (2000). Standard errors for EM estimation. *Journal of the Royal Statistical Society,* Series B, *62,* 257-270.
Jedidi, K., & Ansari, A. (2001). Bayesian structural equation models for multilevel data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 129-158). Mahwah, NJ: Erlbaum.
Jedidi, K., Jagpal, H. S., & DeSarbo, W. (1997). STEMM: A general finite mixture structural equation model. *Journal of Classification, 14,* 23-50.
Jedidi, K., Ramaswamy, V., DeSarbo, W. W., & Wedel, M. (1996). On estimating finite mixtures of multivariate regression and simultaneous equation models. *Structural Equation Modeling, 3,* 266-289.
Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimators. *Annals of Mathematical Statistics, 40,* 633-643.
Jennrich, R. I. (1970). An asymptotic chi-square test for the equality of two correlation matrices. *Journal of the American Statistical Association, 65,* 904-912.
Jennrich, R. I. (1995). *An introduction to computational statistics*. Englewood Cliffs, NJ: Prentice Hall.
Jennrich, R. I., & Sampson, P. F. (1968). Application of stepwise regression to nonlinear estimation. *Technometrics, 10,* 63-72.
Jinn, J. H., & Sedransk, J. (1989). Effect on secondary data analysis of common imputation methods. *Sociological Methodology, 19,* 213-241.
Jo, B., & Muthén, B. O. (2001). Modeling of intervention effects with noncompliance: A latent variable approach for randomized trials. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 57-88). Mahwah, NJ: Erlbaum.
Johnson, D. R., & Elliott, L. A. (1998). Sampling design effects: Do they affect the analyses of data from the National Survey of Families and Households. *Journal of Marriage and the Family, 60,* 993-1001.
Johnson, J. D., & Lecci, L. (2003). Assessing anti-white attitudes and predicting perceived racism: The Johnson-Lecci scale. *Personality and Social Psychology Bulletin, 29,* 299-312.
Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183-202.

Jöreskog, K. G. (1971a). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409-426.
Jöreskog, K. G. (1971b). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109-133.
Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85-112). New York: Seminar.
Jöreskog, K. G. (1977). Structural equation models in the social sciences: Specification, estimation and testing. In P. R. Krishnaiah (Ed.), *Applications of statistics* (pp. 265-287). Amsterdam: North-Holland.
Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika, 43*, 443-477.
Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.) *Testing structural equation models* (pp. 294-316). Newbury Park, CA: Sage.
Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika, 59*, 381-389.
Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*, 631-639.
Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research, 36*, 347-387.
Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relationships by the method of maximum likelihood*. Chicago: National Educational Resources.
Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications.* Chicago: SPSS.
Jöreskog, K. G., & Sörbom, D. (1994). *LISREL 8 user's reference guide*. Mooresville, IN: Scientific Software.
Jöreskog, K. G., & Wold, H. (1982). The ML and PLS techniques for modeling with latent variables: Historical and comparative aspects. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp. 263-270). Amsterdam: North- Holland.
Jorm, A. F., Christensen, H., Medway, J., Korten, A. E., Jacomb, P. A., & Rodgers, B. (2000). Public belief systems about the helpfulness of interventions for depression: Associations with history of depression and professional help-seeking. *Social Psychiatry and Psychiatric Epidemiology, 35*, 211-219.
Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-sample designs. *Psychological Methods, 6*, 115-134.
Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct? *Journal of Personality and Social Psychology, 83*, 693-710.
Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling, 8*, 325-352.
Kagee, A., & Dixon, D. N. (2000). Worldview and health promoting behavior: A causal model. *Journal of Behavioral Medicine, 23*, 163-179.
Kano, Y. (1992). Robust statistics for test-of-independence and related structural models. *Statistics & Probability Letters, 15*, 21-26.
Kano, Y. (1994). Consistency property of elliptical probability density functions. *Journal of Multivariate Analysis, 51*, 343-350.
Kano, Y. (1997). Graphical multivariate analysis with AMOS, EQS, and LISREL: A visual approach to covariance structure analysis (in Japanese). Kyoto, Japan: Gendai-Sugakusha.
Kano, Y. (2001). Structural equation modeling for experimental data. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 381-402). Lincolnwood, IL: Scientific Software International.
Kano, Y. (2002). Variable selection for structural models. *Journal of Statistical Planning and Inference, 108*, 173-187.
Kano, Y., & Azuma, Y. (2003). Use of SEM programs to precisely measure scale reliability. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J. J. Meulman (Eds.), *New developments in psychometrics* (pp. 141-148). Tokyo: Springer-Verlag.
Kano, Y., Bentler, P. M., & Mooijaart, A. (1993). Additional information and precision of estimators in multivariate structural models. In K. Matusita, M.L. Puri, & T. Hayakawa (Eds.), *Statistical sciences and data analysis* (pp. 187-196). Zeist: VSP (International Science Publishers).
Kano, Y., Berkane, M., & Bentler, P. M. (1990). Covariance structure analysis with heterogeneous kurtosis parameters. *Biometrika, 77*, 575-585.
Kano, Y., Berkane, M., & Bentler P. M. (1993). Statistical inference based on pseudo-maximum likelihood estimators in elliptical populations. *Journal of the American Statistical Association, 88*, 135-143.
Kano, Y., & Harada, A. (2000). Stepwise variable selection in factor analysis. *Psychometrika, 65*, 7-22.
Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research, 23*, 69-86.
Kaplan, D. (1989). Model modification in covariance structure analysis: Application of the expected parameter change statistic. *Multivariate Behavioral Research, 24*, 285-305.
Kaplan, D. (1991). On the modification and predictive validity of covariance structure models. *Quality & Quantity, 25*, 307-314.
Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions.* Thousand Oaks, CA: Sage.
Kaplan, D., & Elliott, P. R. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling, 4*, 1-24.

Kaplan, D., & Ferguson, A. J. (1999). On the utilization of sample weights in latent variable models. *Structural Equation Modeling, 6,* 305-321.
Kaplan, D., Harik, P., & Hotchkiss, L. (2001). Cross-section estimation of dynamic structural equation models in disequilibrium. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 315-340). Lincolnwood, IL: Scientific Software International.
Kaplan, D., & Wenger, R. N. (1993). Asymptotic independence and separability in covariance structure models: Implications for specification error, power, and model modification. *Multivariate Behavioral Research, 28,* 483-498.
Kaplan, H. B., & Johnson, R. J. (2001). *Social deviance: Testing a general theory.* New York: Kluwer AcademicPlenum.
Kaplan, H. B., & Liu, X. (2001). Adolescent self-rejection and adult social action: A conditional relationship. *Social Science Quarterly, 82,* 701-715.
Kauermann, G., & Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association, 96,* 1387-1396.
Kawamura, M., Tsurumoto, A., Fukuda, S., & Sasahara, H. (2001). Health behaviors and their relation to metabolic control and periodontal status in type 2 diabetic patients: A model tested using a linear structural relations program. *Journal of Periodontology, 72,* 1246-1253.
Kearns, G. S., & Lederer, A. L. (2000). The effect of strategic alignment on the use of IS-based resources for competitive advantage. *Journal of Strategic Information Systems, 9,* 265-293.
Keats, J. A. (1983). Ability measures and theories of cognitive development. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement: A Festschrift for Frederick M. Lord* (pp. 81-101). Hillsdale, NJ: Erlbaum.
Keesling, J. W. (1972). *Maximum likelihood approaches to causal analysis.* Ph.D. thesis, University of Chicago.
Keller, S. D., Ware J. E., Bentler, P. M., Aaronson, N. K., Alonso, J., Apolone, G., Bjorner, J. B., Brazier, J., Bullinger, M., Kaasa, S., Leplege, A., Sullivan, M. & Gandek, B. (1998). Use of structural equation modeling to test the construct validity of the SF-36 Health Survey in ten countries: Results from the IQOLA Project. *Journal of Clinical Epidemiology, 51,* 1179-1188.
Kelley, B. C., Eklund, R. C., & Ritter-Taylor, M. (1999). Stress and burnout among collegiate tennis coaches. *Journal of Sport & Exercise Psychology, 21,* 113-130.
Kempf, D. S., & Smith, R. E. (1998). Consumer processing of product trial and the influence of prior advertising: A structural modeling approach. *Journal of Marketing Research, 35,* 325-338.
Kendler, K. S., Neale, M. C., Sullivan, P., Corey, L. A., Gardner, C. O., & Prescott, C. A. (1999). A population-based twin study in women of smoking initiation and nicotine dependence. *Psychological Medicine, 29,* 299-308.
Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley.
Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 96,* 201-210.
Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112,* 165-172.
Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10,* 333-351.
Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology, 63,* 52-59.
Kenny, D. A., & Zautra, A. (2001). Trait-state models for longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 243-263). Washington, DC: American Psychological Association.
Kessler, R. C., & Greenberg, D. F. (1981). *Linear panel analysis.* New York: Academic.
Kessler, T., & Mummendey, A. (2002). Sequential or parallel processes? A longitudinal field study concerning determinants of identity-management strategies. *Journal of Personality and Social Psychology, 82,* 75-88.
Khoo, S. –T. (2001). Assessing program effects in the presence of treatment-baseline interactions: A latent curve approach. *Psychological Methods, 6,* 234-257.
Khoo, S. –T., & Muthén, B. (2000). Longitudinal data on families: Growth modeling alternatives. In J. S. Rose, L. Chassin, C. C. Preston, & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods for new questions* (pp. 43-78). Mahwah, NJ: Erlbaum.
Kiiveri, H. T. (1987). An incomplete data approach to the analysis of covariance structures. *Psychometrika, 52,* 539-554.
Kim, K. H. (2005). The relation among fit indexes, power, and sample size in structural equation modeling. *Structural Equation Modeling, 12,* 368-390.
Kim, K. H., & Bentler, P. M. (1999). *Tests of means and covariance matrices for multivariate incomplete data.* UCLA Statistics Series preprint #265.
Kim, K. H., & Bentler, P. M. (2002). Tests of homogeneity of means and covariance matrices for multivariate incomplete data. *Psychometrika, 67,* 609-624.
Kim, S., & Hagtvet, K. A. (2003). The impact of misspecified item parceling on representing latent variables in covariance structure modeling: A simulation study. *Structural Equation Modeling, 10,* 101-127.
King, S. (2000). Is expressed emotion cause or effect in the mothers of schizophrenic young adults? *Schizophrenia Research, 45,* 65-78.
Kiriski, L., Dunn, M. G., Mezzich, A. C., & Tarter, R. E. (2001). Impact of parental substance use disorder and child neglect severity on substance use involvement in male offspring. *Prevention Science, 2,* 241-255.

Kirk, K. M., Bailey, J. M., Dunne, M. P., & Martin, N. G. (2000). Measurement models for sexual orientation in a community twin sample. *Behavior Genetics*, *30*, 345-356.
Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational & Psychological Measurement*, *54*, 757-765.
Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65*, 457-474.
Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.
Komaroff, E. (1997). Effect of simultaneous violations of essential tau-equivalence and uncorrelated error on coefficient alpha. *Applied Psychological Measurement, 21*, 337-348.
Kosson, D. S., Cyterski, T. D., Steuerwald, B. L., Neumann, C. S., & Walker-Matthews, S. (2002). The reliability and validity of the psychopathy checklist: Youth Version (PCL:YV) in nonincarcerated adolescent males. *Psychological Assessment*, *14*, 97-109.
Koster, J. T. A. (1999). On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scandinavian Journal of Statistics*, *26*, 413-431.
Koufteros, X. A., Vonderembse, M. A., & Doll, W. J. (2001). Concurrent engineering and its consequences. *Journal of Operations Management, 19*, 97-115.
Koufteros, X. A., Vonderembse, M. A., & Doll, W. J. (2002). Examining the competitive capabilities of manufacturing firms. *Structural Equation Modeling, 9*, 256-282.
Kratt, P., Reynolds, K., & Shewchuk, R. (2000). The role of availability as a moderator of family fruit and vegetable consumption. *Health Education & Behavior, 27*, 471-482.
Krauss, D. A., Sales, B. D., Becker, J. V., & Figueredo, A. J. (2000). Beyond prediction to explanation in risk assessment research: A comparison of two explanatory theories of criminality and recidivism. *International Journal of Law & Psychiatry, 23*, 91-112.
Kreft, I. & de Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage.
Krijnen, W. P. (2002). On the construction of all factors of the model for factor analysis. *Psychometrika, 67*, 161-172.
Krijnen, W. P., Dijkstra, T. K., & Gill, R. D. (1998). Conditions for factor (in)determinacy in factor analysis. *Psychometrika*, *63*, 359-367.
Krijnen, W. P., Wansbeek, T. J., & ten Berge, J. M. F. (1996). Best linear predictors for factor scores. *Communication in Statistics: Theory and Methods*, *25*, 3013-3025.
Kroonenberg, P. M., van Dam, M., van Ijzendoorn, M. H., & Mooijaart, A. (1997). Dynamics of behaviour in the strange situation: A structural equation approach. *British Journal of Psychology*, *88*, 311-332.
Kuha, J. (2004). AIC and BIC: Comparison of assumptions and performance. *Sociological Methods & Research, 33*, 188-229.
Kühnel, S. M. (1988). Testing MANOVA designs with LISREL. *Sociological Methods & Research, 16*, 504-523.
Kulkarni, R. R., Kamal, M. M., & Arjune, D. G. (1999). Improvement of the Mair scoring system using structural equations modeling for classifying the diagnostic adequacy of cytology material from thyroid lesions. *Diagnostic Cytopathology*, *21*, 387-393.
Kuppam, A. R., & Pendyala, R. M. (2001). A structural equations analysis of commuters' activity and travel patterns. *Transportation, 28*, 33-54.
Küsters, U. (1987). *Hierarchische Mittelwert- und Kovarianzstructurmodelle mit nicht-metrischen endogenen Variablen.* Heidelberg: Physica Verlag.
Labouvie, E., & Pinsky, I. (2001). Substance use and driving: The coexistence of risky and safe behaviors. *Addiction*, *96*, 473-484.
LaDu, T. J., & Tanaka, J. S. (1989). The influence of sample size, estimation method, and model specification on goodness-of-fit assessments in structural equation models. *Journal of Applied Psychology, 74*, 625-635.
Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine, 7*, 305–315.
Laird, N. M., & Ware, H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963-974.
Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods, 7*, 228-244.
Lange, K. L., Little, R. J. A., & Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association, 84*, 881-896.
La Peyre, M. K., Mendelssohn, I. A., Reams, M. A., Templet, P. H., & Grace, J. B. (2001). Identifying determinants of nations' wetland management programs using structural equation modeling: An exploratory analysis. *Environmental Management, 27*, 859-868.
Larose, S., Guay, F., & Boivin, M. (2002). Attachment, social support, and loneliness in young adulthood: A test of two models. *Personality and Social Psychology Bulletin, 28*, 684-693.
Larwood, L., Wright, T. A., Desrochers, S., & Dahir, V. (1998). Extending latent role and psychological contract theories to predict intent to turnover and politics in business organizations. *Group & Organization Management*, *23*, 100-123.
Latimer, W. W., Newcomb, M., Winters, K. C., & Stinchfield, R. D. (2000). Adolescent substance abuse treatment outcome: The role of substance abuse problem severity, psychosocial, and treatment factors. *Journal of Consulting & Clinical Psychology*, *68*, 684-696.

Lau, A., & Takeuchi, D. (2001). Cultural factors in help-seeking for child behavior problems: Value orientation, affective responding, and severity appraisals among Chinese-American parents. *Journal of Community Psychology, 29*, 675-692.
Lauritzen, S. L. (1996). *Graphical models.* Oxford: Clarendon.
Lauritzen, S. L. (2001). Causal inference from graphical models. In O. E. Barndorff-Nielsen, D. R. Cox, & C. Klüppelberg (Eds.), *Complex stochastic systems* (pp. 63-107). Boca Raton, FL: Chapman & HallCRC.
Law, K. S., & Wong, C.-S. (1999). Multidimensional constructs in structural equation analysis: An illustration using the job perception and job satisfaction constructs. *Journal of Management*, *25*, 143-160.
Lawrence, F. R., & Blair, C. (2003). Factorial invariance in preventive intervention: Modeling the development of intelligence in low birth weight, preterm infants. *Prevention Science, 4*, 249-261.
Lea, M., Spears, R., de Groot, D. (2001). Knowing me, knowing you: Anonymity effects on social identity processes within groups. *Personality and Social Psychology Bulletin, 27*, 526-537.
Lee, S., & Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research, 25*, 313-334.
Lee, S. –Y. (1985a). Analysis of covariance and correlation structures. *Computational Statistics & Data Analysis, 2*, 279-295.
Lee, S. –Y. (1985b). On testing functional constraints in structural equation models. *Biometrika, 72*, 125-131.
Lee, S. –Y. (1986). Estimation for structural equation models with missing data. *Psychometrika, 51*, 93-99.
Lee, S. –Y. (1990). Multilevel analysis of structural equation models. *Biometrika, 77*, 763-772.
Lee, S. –Y., & Bentler, P. M. (1980). Some asymptotic properties of constrained generalized least squares estimation in covariance structure models. *South African Statistical Journal, 14*, 121-136.
Lee, S. –Y., & Jennrich, R. I. (1979). A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. *Psychometrika*, *44*, 99-113.
Lee, S. –Y., & Leung, K. –M. (1992). Estimation of multivariate polychoric and polyserial correlations with missing observations. *British Journal of Mathematical and Statistical Psychology*, *45*, 225-238.
Lee, S. –Y., & Poon, W. –Y. (1985). Further developments on constrained estimation in the analysis of covariance structures. *The Statistician, 34*, 305-316.
Lee, S. –Y., & Poon, W. –Y. (1998). Analysis of two-level structural equation models via EM type algorithms. *Statistica Sinica, 8,* 749-766.
Lee, S. –Y., Poon, W. –Y., & Bentler, P. M. (1989). Simultaneous analysis of multivariate polytomous variates in several groups. *Psychometrika, 54*, 63-73.
Lee, S. –Y., Poon, W. –Y., & Bentler, P. M. (1990a). Full maximum likelihood analysis of structural equation models with polytomous variables. *Statistics & Probability Letters*, *9*, 91-97.
Lee, S. –Y., Poon, W. –Y., & Bentler, P. M. (1990b). A three-stage estimation procedure for structural equation models with polytomous variables. *Psychometrika*, *55*, 45-51.
Lee, S. –Y., Poon, W. –Y., & Bentler, P. M. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika, 57*, 89-105.
Lee, S. –Y., Poon, W. –Y., & Bentler, P. M. (1994). Covariance and correlation structure analyses with continuous and polytomous variables. In T. W. Anderson, K. -T. Fang, & I. Olkin (Eds.), *Multivariate analysis and its applications: Vol. 24* (pp. 347-358). Hayward CA: Institute of Mathematical Statistics.
Lee, S. –Y., Poon, W. –Y., & Bentler, P. M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology*, *48*, 339-358.
Lee, S. –Y., & Shi, J. –Q. (2000). Bayesian analysis of structural equation model with fixed covariates. *Structural Equation Modeling, 7*, 411-430.
Lee, S. –Y., & Song, X. –Y. (2001). Hypothesis testing and model comparison in two-level structural equation models. *Multivariate Behavioral Research, 36*, 639-655.
Lee, S. –Y., & Song, X. –Y. (2002). Bayesian selection on the number of factors in a factor analysis model. *Behaviormetrika*, *29*, 23-39.
Lee, S. –Y., & Song, X. –Y. (2003). Bayesian model selection for mixtures of structural equation models with an unknown number of components. *British Journal of Mathematical and Statistical Psychology, 56*, 145-165.
Lee, S. –Y., & Song, X. –Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*, *39*, 653-686.
Lee, S. –Y., & Tsui, K. L. (1982). Covariance structure analysis in several populations. *Psychometrika, 47*, 297-308.
Lee, S. –Y., & Xu, L. (2003). Case-deletion diagnostics for factor analysis models with continuous and ordinal categorical data. *Sociological Methods & Research, 31*, 389-419.
Lee, S. –Y., Zhang, W., & Song, X. –Y. (2002). Estimating the covariance function with functional data. *British Journal of Mathematical and Statistical Psychology, 55*, 247-261.
Lee, S. –Y., & Zhu, H.-T. (2000). Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, *53*, 209-232.
Lee, S. –Y., & Zhu, H.-T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika, 67*, 189-210.
Leese, M., & Brown, C. (1997). Outlier detection in psychiatric epidemiology. *Epidemiologia e Psichiatria Sociale, 6*, 151-154.
Lei, M., & Lomax, R. G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. *Structural Equation Modeling, 12*, 1-27.

Lei, P. –W., & Dunbar, S. B. (2004). Effects of score discreteness and estimating alternative model parameters on power estimation methods in structural equation modeling. *Structural Equation Modeling, 11*, 20-44.
Lemery, K. S., Goldsmith, H. H., Klinnert, M. D., & Mrazek, D. A. (1999). Developmental models of infant and childhood temperament. *Developmental Psychology, 35*, 189-204.
Leon, A. C., Solomon, D. A., Mueller, T. I., Turvey, C. L., Endicott, J., & Keller, M. B. (1999). The Range of Impaired Functioning Tool (LIFE-RIFT): A brief measure of functional impairment. *Psychological Medicine, 29*, 869-878.
Leung, Y. P., & Chan, W. (1997). Constrained generalized least squares and asymptotically distribution free estimation in correlation structure analysis. *American Statistical Association: Proceedings of the Section on Government Statistics and Section on Social Statistics*, 422-427.
Levin, P. F. (1999). Test of the Fishbein and Ajzen models as predictors of health care worker's glove use. *Research in Nursing & Health, 22*, 295-307.
Levitz, M., Perlman, M. D., & Madigan, D. (2001). Separation and completeness properties for AMP chain graph markov models. *Annals of Statistics, 29*, 1751-1784.
Lewis, P. A. W., & Orav, E. J. (1989). *Simulation methodology for statisticians, operations analysis, and engineers*. Pacific Grove, CA: Wadsworth.
Leyland, A. H. & Goldstein, H. (Eds.) (2001). *Multilevel modelling of health statistics*. London: Wiley.
Li, F., Duncan, T. E., & Acock, A. (2000). Modeling interaction effects in latent growth curve models. *Structural Equation Modeling, 7*, 497-533.
Li, F., Duncan, T. E., Duncan, S. C., & Acock, A. (2001). Latent growth modeling of longitudinal data: A finite growth mixture modeling approach. *Structural Equation Modeling, 8*, 493-530.
Li, F., Duncan, T. E., Duncan, S. C., McAuley, E., Chaumeton, N. R., & Harmer, P. (2001). Enhancing the psychological well-being of elderly individuals through Tai Chi exercise: A latent growth curve analysis. *Structural Equation Modeling, 8*, 53-83.
Li, F., Duncan, T. E., Duncan, S. C., Yang-Wallentin, F., Acock, A. C., & Hops, H. (2001). In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 173-202). Mahwah, NJ: Erlbaum.
Li, F., Harmer, P., McAuley, E., Fisher, K. J., Duncan, T. E., & Duncan, S. C. (2001). Tai Chi, self-efficacy, and physical function in the elderly. *Prevention Science, 2*, 229-239.
Li, H. (1997). A unifying expression for the maximal reliability of a linear composite. *Psychometrika, 62*, 245-249.
Li, H., Rosenthal, R., & Rubin, D. B. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods, 1*, 98-107.
Li, K. –C., & Shedden, K. (2002). Identification of shared components in large ensembles of time series using dimension reduction. *Journal of the American Statistical Association, 97*, 759-765.
Li, L. (2005). *Normal approximation of the sample RMSEA and its application in mean and covariance structure analysis.* Ph.D. dissertation, UCLA.
Li, L., & Bentler, P. M. (2001). *Resampling approach to greatest lower bound.* Unpublished manuscript.
Li, L., & Bentler, P. M. (2004). *The greatest lower bound to reliability: Corrected and resampling estimators.* Under review.
Liang, J. –J., & Bentler, P. M. (1998). Characterizations of some subclasses of spherical distributions. *Statistics & Probability Letters, 40*, 155-164.
Liang, J. –J., & Bentler, P. M. (1999a). *A new EM algorithm for maximum likelihood estimation in two-level structural equation models with arbitrary sample designs.* UCLA Statistics Series preprint #268.
Liang, J. –J., & Bentler, P. M. (1999b). A t-distribution plot to detect non-multinormality. *Computational Statistics & Data Analysis, 30*, 31-44.
Liang, J. –J., & Bentler, P. M. (2001). *An EM algorithm for clustered data analysis*. Los Angeles: Authors.
Liang, J. –J., & Bentler, P. M. (2004). An EM algorithm for fitting two-level structural equation models. *Psychometrika, 69*, 101-122.
Liang, K. –Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13-22.
Linares, L. O., Heeren, T., Bronfman, E., Zuckerman, B., Augustyn, M., & Tronick, E. (2001). A mediational model for the impact of exposure to community violence on early child behavior problems. *Child Development, 72*, 639-652.
Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*, 1198-1202.
Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd Ed.). New York: Wiley.
Little, R. J. A., & Schenker, N. (1995). Missing data. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.) *Handbook of statistical modeling for the social and behavioral sciences* (pp. 39-76). New York: Plenum.
Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*, 53-76.
Little, T. D. (2000). On the comparability of constructs in cross-cultural research: A critique of Cheung and Rensvold. *Journal of Cross-Cultural Psychology, 31*, 213-219.
Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*, 151-173.

Little, T. D., Linderberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods*, *4*, 192-211.
Little, T. D., Schnabel, K. U., & Baumert, J., (Eds.). (2000). *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*. Mahwah, NJ: Erlbaum.
Livert, D., Rindskopf, D., Saxe, L., & Stirratt, M. (2001). Using multilevel modeling in the evaluation of community-based treatment programs. *Multivariate Behavioral Research*, *36*, 155-183.
Llabre, M. M., Spitzer, S. B., Saab, P. G., & Schneiderman, N. (2001). Piecewise latent growth modeling of systolic blood pressure reactivity and recovery from the cold pressor test. *Psychophysiology*, *38*, 951-960.
Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*, 767-778.
Loas, G., Corcos, M., Stephan, P., Pellet, J., Bizouard, P., Venisse, J. L., Perez-Diaz, F., Guelfi, J. D., Flament, M., & Jeammet, P. (2001). Factorial structure of the 20-item Toronto alexithymia scale: Confirmatory factorial analyses in nonclinical and clinical samples. *Journal of Psychosomatic Research*, *50*, 255-261.
Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (3rd Ed.). Mahwah, NJ: Erlbaum.
Lohmöller, J. –B. (1989). *Latent variable path modeling with partial least squares.* New York: Springer-Verlag.
Lonczak, H. S., Huang, B., Catalano, R. F., Hawkins, J. D., Hill, K. G., Abbott, R. D., Ryan, J. A. M., & Kosterman, R. (2001). The social predictors of adolescent alcohol misuse: A test of the social development model. *Journal of Studies on Alcohol*, *62*, 179-189.
Long, J. D., Feng, D., & Cliff, N. (2003).Ordinal analysis of behavioral data. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (pp. 635-661). Hoboken, NJ: Wiley.
Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, *74*, 817-827.
Longford, N. T. (1993a). Regression analysis of multilevel data with measurement error. *British Journal of Mathematical and Statistical Psychology*, *46*, 301-311.
Longford, N. T. (1993b). *Random coefficient models*. Oxford, England: Clarendon.
Longford, N. T. (2001). Multilevel analysis with messy data. *Statistical Methods in Medical Research*, *10*, 429-444.
Loo, R. (1999). Confirmatory factor analyses of Kolb's Learning Style Inventory (LSI-1985). *British Journal of Educational Psychology*, *69*, 213-219.
Lord, F. M., & Novick, R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Wesley.
Lu, G., & Copas, J. B. (2004). Missing at random, likelihood ignorability and model completeness. *Annals of Statistics*, *32*, 754-765.
Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). Weak measurement invariance with respect to unmeasured variables: An implication of strict factorial invariance. *British Journal of Mathematical and Statistical Psychology*, *56*, 231-248.
Lucke, J. F. (2005). The α and ω of congeneric test theory: An extension of reliability and internal consistency to heterogeneous tests. *Applied Psychological Measurement*, *29*, 65-81.
Luijben, T. C. (1991). Equivalent models in covariance structure analysis. *Psychometrika*, *56*, 653-665.
Luijben, T. C., Boomsma, A., & Molenaar, I. W. (1988). Modification of factor analysis models in covariance structure analysis, A Monte Carlo study. In T. K. Dijkstra (Ed.), *On model uncertainty and its statistical implications* (pp. 70-101). New York: Springer-Verlag.
Maas, C. J. M., & Hox, J. J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics & Data Analysis*, *46*, 427-440.
Maasen, G. H., & Bakker, A. B. (2001). Suppressor variables in path models: Definitions and interpretations. *Sociological Methods & Research*, *30*, 241-270.
MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*, 107-120.
MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In R. H. Hoyle, (Ed.). *Structural equation modeling* (pp. 16-36). Thousand Oaks, CA: Sage.
MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research*, *38*, 113-139.
MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201-226.
MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, *114*, 533-541.
MacCallum, R. C., Browne, M. W., & Preacher, K. J. (2002). Comments on the Meehl-Waller (2002) procedure for appraisal of path analysis models. *Psychological Methods*, *7*, 301-306.
MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130-149.
MacCallum, R.C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent growth curve models. *Multivariate Behavioral Research*, *32*, 215-253.
MacCallum, R. C., Roznowski, M. & Necowitz, L. B. (1992). Model modification in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490-504.

MacCallum, R. C., Tucker, L. R, & Briggs, N. E. (2001). An alternative perspective on parameter estimation in factor analysis and related methods. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 39-57). Lincolnwood, IL: Scientific Software International.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114*, 185-199.

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research, 36*, 611-637.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84-99.

MacCallum, R. C., Zhang, S., Preacher, K. J., Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19-40.

MacDonald, W. L. (1998). The difference between Blacks' and Whites' attitudes toward voluntary euthanasia. *Journal for the Scientific Study of Religion, 37*, 411-426.

MacKinnon, D. P. (2000). Contrasts in multiple mediator models. In J. S. Rose, L. Chassin, C. C. Presson, & S. J. Sherman *Multivariate applications in substance use research: New methods for new questions* (pp. 141-160). Mahwah, NJ: Erlbaum.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of mediation, confounding and suppression effect. *Prevention Science, 1*, 173-181.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variables. *Psychological Methods, 7*, 83-104.

MacKinnon, D. P., Nohre, L., Cheong, J., Stacy, A. W., & Pentz, M. A**.** (2001). Longitudinal relationship between the alcohol warning label and alcohol consumption. *Journal of Studies on Alcohol, 62*, 221-227.

MacKinnon, D. P., Taborga, M. P., Morgan-Lopez, A. A. (2002). Mediation designs for tobacco prevention research. *Drug and Alcohol Dependence, 68 Suppl. 1*, 69-83.

Maffei, C., & Fossati, A. (1999). The structure of DSM-IV borderline personality disorder and its implications for treatment. In J. Derksen, C. Maffei, (Eds.), *Treatment of personality disorders.* New York: Kluwer AcademicPlenum Publishers.

Magill, J. M., & Pressing, J. L. (1997). Asymmetric cognitive clock structures in West African rhythms. *Music Perception, 15*, 189-221.

Magnus, J. R., & Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics* (revised edition). Chichester UK: Wiley.

Maia, J. A. R., Lefevre, J., Claessens, A., Renson, R., Vanreusel, B., & Beunen, G. (2001). Tracking of physical fitness during adolescence: A panel study in boys. *Medicine and Science in Sports and Exercise, 33*, 765-771.

Maiti, S. S., & Mukherjee, B. N. (1991). Two new goodness-of-fit indices for covariance matrices with linear structures. *British Journal of Mathematical and Statistical Psychology*, *44*, 153-180.

Maitland, S. B., Intrieri, R. C., Schaie, K. W., & Willis, S. L. (2000). Gender differences and changes in cognitive abilities across the adult life span. *Aging, Neuropsychology, & Cognition*, *7*, 32-53.

Mak, B. L., & Sockel, H. (2001). A confirmatory factor analysis of IS employee motivation and retention. *Information and Management*, *38*, 265-276.

Malaeb, Z. A., Summers, J. K., & Pugesek, B. H. (2000). Using structural equation modeling to investigate relationships among ecological variables. *Environmental and Ecological Statistics, 7*, 93-111.

Malmberg, L. –E., Wanner, B., Sumra, S., & Little, T. D. (2001). Action-control beliefs and school experiences of Tanzanian primary school students. *Journal of Cross-Cultural Psychology*, *32*, 577-596.

Mandys, F., Dolan, C. V., & Molenaar, P. C. M. (1994). Two aspects of the simplex model: Goodness of fit to linear growth curve structures and the analysis of mean trends. *Journal of Educational and Behavioral Statistics*, *19*, 201-215.

Maner, J. K., Luce, C. L., Neuberg, S. L., Cialdini, R. B., Brown, S., & Sagarin, B. J. (2002). The effects of perspective taking on motivations for helping: Still no evidence for altruism. *Personality and Social Psychology Bulletin*, *28*, 1601-1610.

Manne, S., & Schnoll, R. (2001). Measuring cancer patients' psychological distress and well-being: A factor analytic assessment of the Mental Health Inventory. *Psychological Assessment, 13*, 99-109.

Marcoulides, G. A., & Drezner, Z. (2001). Specification searches in structural equation modeling with a genetic algorithm. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 247-268). Mahwah, NJ: Erlbaum.

Marcoulides, G. A., & Drezner, Z. (2003). Model specification searches using ant colony optimization algorithms. *Structural Equation Modeling, 10*, 154-164.

Marcoulides, G. A., Drezner, Z., & Schumacker, R. E. (1998). Model specification searches in structural equation modeling using Tabu search. *Structural Equation Modeling, 5*, 365-376.

Marcoulides, G. A., & Moustaki, I. (Eds.). (2002). *Latent variable and latent structure models*. Mahwah, NJ: Erlbaum.

Marcoulides, G. A., & Schumacker, R. E. (Eds.). (1996). *Advanced structural equation modeling: Issues and techniques.* Mahwah, NJ: Erlbaum.

Marcoulides, G. A., & Schumacker, R. E. (Eds.). (2001). *New developments and techniques in structural equation modeling.* Mahwah, NJ: Erlbaum.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519-530.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya B, 36*, 115-128.
Marini, M. M., & Singer, B. (1988). Causality in the social sciences. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 347-409). Washington, DC: American Sociological Association.
Markus, K. A. (2002). Statistical equivalence, semantic equivalence, eliminative induction, and the Raykov-Marcoulides proof of infinite equivalence. *Structural Equation Modeling*, *9*, 503-522.
Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Annals of Statistics*, *4*, 51-67.
Marsh, H. W. (1993). Stability of individual differences in multiwave panel studies: Comparison of simplex models and one-factor models. *Journal of Educational Measurement*, *30*, 157-183.
Marsh, H. W. (1998). Pairwise deletion for missing data in structural equation models: Nonpositive definite matrices, parameter estimates, goodness of fit, and adjusted sample sizes. *Structural Equation Modeling, 5*, 22-36.
Marsh, H. W. (1998b). The equal correlation baseline model: Comment and constructive alternatives. *Structural Equation Modeling, 5*, 78-86.
Marsh, H. W. (1999). Cognitive discrepancy models: Actual, ideal, potential, and future self-perspectives of body image. *Social Cognition, 17*, 46-75.
Marsh, H. W., & Balla, J. (1994). Goodness of fit in confirmatory factor analysis: The effect of sample size and model parsimony. *Quality & Quantity*, *28*, 185-217.
Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness-of-fit indices in confirmatory factor analysis: Effects of sample size. *Psychological Bulletin, 103*, 391-411.
Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and revised recommendations. *Educational Psychologist, 34*, 155-167.
Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 177-198). Thousand Oaks, CA: Sage.
Marsh, H. W., & Hau, K. -T. (2002). Multilevel modeling of longitudinal growth and change: Substantive effects or regression toward the mean artifacts? *Multivariate Behavioral Research*, *37*, 245-282.
Marsh, H. W., Hau, K. –T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, *33*, 181-220.
Marsh, H. W., Hau, K. -T., & Kong, C. –K. (2002). Multilevel causal ordering of academic self-concept and achievement: Influence of language of instruction (English compared with Chinese) for Hong Kong students. *American Educational Research Journal, 39*, 727-763.
Marsh, H. W., Hau, K. –T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in over-generalizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 391-410.
Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology*, *92*, 202-228.
Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *American Educational Research Journal*, *35*, 705-738.
Marshall, G. N., Orlando, M., Jaycox, L. H., & Foy, D. W. (2002). Development and validation of a modified version of the peritraumatic dissociative experiences questionnaire. *Psychological Assessment, 14*, 123-134.
Martinez, C. R., Jr., & Forgatch, M. S. (2001). Preventing problems with boys' noncompliance: Effects of a parent training intervention for divorcing mothers. *Journal of Consulting and Clinical Psychology, 69*, 416-428.
Mason, W. A., & Windle, M. (2001). Family, religious, school and peer influences on adolescent alcohol use: A longitudinal study. *Journal of Studies on Alcohol, 62*, 44-53.
Mason, W. M., Wong, G. M., & Entwistle, B. (1983). Contextual analysis through the multilevel linear model. In S. Leinhart (Ed.), *Sociological methodology* (pp. 72-103). San Francisco, CA: Jossey-Bass.
Matsueda, R. L., & Bielby, W. T. (1986). Statistical power in covariance structure models. In N. B. Tuma (Ed.), *Sociological Methodology* (pp. 120-158). San Francisco: Jossey Bass.
Mattson, S. (1997). How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research, 32*, 355-373.
Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective*. Mahwah, NJ: Erlbaum.
Maydeu-Olivares, A. (2001). Multidimensional item response theory modeling of binary data: Large sample properties of NOHARM estimates. *Journal of Educational and Behavioral Statistics, 26*, 51-71.
Mayer, M. J., & Leone, P. E. (1999). A structural analysis of school violence and disruption: Implications for creating safer schools. *Education & Treatment of Children, 22*, 333-356.
McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics, 16*, 163-200.
McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In R. B. Cattell & J. Nesselroade (Eds.), *Handbook of multivariate experimental psychology* (pp. 561-614). New York: Plenum.
McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 341-380). Lincolnwood, IL: Scientific Software International.

McArdle, J. J., & Bell, R. Q. (2000). An introduction to latent growth models for developmental data analysis. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 69-108). Mahwah, NJ: Erlbaum.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*, 110-133.

McArdle, J. J., & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 89-125). Mahwah, NJ: Erlbaum.

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology, 37*, 234-251.

McArdle, J. J., & Nesselroade, J. R. (2003).Growth curve analysis in contemporary psychological research. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (pp. 447-480). Hoboken, NJ: Wiley.

McAuley, E., Blissmer, S., Marquez, D. X., Jerome, G. J., Kramer, A. F., & Katula J. (2000). Social relations, physical activity, and well-being in older adults. *Preventive Medicine, 31*, 608-617.

McCaffery, J. M., Pogue-Geile, M. F., Debski, T. T., & Manuck, S. B. (1999). Genetic and environmental causes of covariation among blood pressure, body mass and serum lipids during young adulthood: A twin study. *Journal of Hypertension, 17*, 1677-1685.

McCreary, D. R., & Sadava, S. W. (2000). Stress, alcohol use and alcohol-related problems: The influence of negative and positive affect in two cohorts of young adults. *Journal of Studies on Alcohol, 61*, 466-474.

McCullough, M. E., Emmons, R. A., & Tsang, J. –A. The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology, 82*, 112-127.

McDonald, R. P. (1967). Nonlinear factor analysis. *Psychometric Monograph*, No. 15.

McDonald, R. P. (1970). The theoretical foundations of common factor analysis, principal factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology, 23*, 1-21.

McDonald, R. P. (1980). A simple comprehensive model for the analysis of covariance structures: Some remarks on applications. *British Journal of Mathematical and Statistical Psychology, 33*, 161-183.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification, 6*, 97-103.

McDonald, R. P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological Methods & Research, 22*, 399-413.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

McDonald, R. P. (2002). What can we learn from the path equations? Identifiability, constraints, equivalence. *Psychometrika, 67*, 225-249.

McDonald, R. P. (2004). The specific analysis of structural equation models. *Multivariate Behavioral Research, 39*, 687-713.

McDonald, R. P., & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical & Statistical Psychology, 42*, 215-232.

McDonald, R. P., & Ho, M. –H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*, 64-82.

McDonald, R. P., & Krane, W. R. (1979). A Monte Carlo study of local identifiability and degrees of freedom in the asymptotic likelihood ratio test. *British Journal of Mathematical and Statistical Psychology, 32*, 121-132.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin, 107*, 247-255.

McIntosh, A. R. (1999). Mapping cognition to the brain through neural interactions. *Memory, 7*, 523-548.

McNulty, J. K., & Karney, B. R. (2002). Expectancy confirmation in appraisals of marital interactions. *Personality and Social Psychology Bulletin, 28*, 764-775.

Meehl, P. E., & Waller, N. G. (2002). The path analysis controversy: A new statistical approach to strong appraisal of verisimilitude. *Psychological Methods, 7*, 283-300.

Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5*, 23-43.

Meijer, E., & Mooijaart, A. (1996). Factor analysis with heteroscedastic errors. *British Journal of Mathematical and Statistical Psychology, 49*, 189-202.

Meijer, E., van der Leeden, R., & Busing, F. M. T. A. (1995). Implementing the bootstrap for multilevel models. *Multilevel Modeling Newsletter, 7*, 7-11.

Mellor, S., & Mathieu, J. E. (1999). A discriminant validity study of aggregate-level constructs and measures of local union formalization, centralization, and innovation. *Journal of Psychology, 133*, 669-683.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543.

Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 203-240). Washington, DC: American Psychological Association.

Meredith, W., & Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika, 57*, 289-311.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*, 107-122.

Merrill, L. L., Thomsen, C. J., Gold, S. R., & Milner, J. S. (2001). Childhood abuse and premilitary sexual assault in male navy recruits. *Journal of Consulting and Clinical Psychology, 69*, 252-261.

Merrill, L. L., Thomsen, C. J., Sinclair, B. B., Gold, S. R., & Milner, J. S. (2001). Predicting the impact of child sexual abuse on women: The role of abuse severity, parental support, and coping strategies. *Journal of Consulting and Clinical Psychology, 69*, 992-1006.

Metsapelto, R. L., Pulkkinen, L., & Poikkeus, A. M. (2001). A search for parenting style: A cross-situational analysis of parental behavior. *Genetic Social and General Psychology Monographs, 127*, 169-192.

Meuleners, L. B., Lee, A. B., Binns, C. W., & Lower, A. (2003). Quality of life for adolescents: Assessing measurement properties using structural equation modeling. *Quality of Life Research, 12*, 283-290.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156-166.

Michailidis, G., & de Leeuw, J. (1998). The Gifi system of descriptive multivariate analysis. *Statistical Science, 13*, 307-336.

Migneault, J. P., Velicer, W. F., Prochaska, J. O., & Stevenson, J. F. (1999). Decisional balance for immoderate drinking in college students. *Substance Use and Misuse, 34*, 1325-1346.

Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling*, *2*, 255-273.

Miller, R. G. (1974). The jackknife – a review. *Biometrika, 61*, 1–15.

Millsap, R. E. (1995). Measurement invariance, predictive invariance, and the duality paradox. *Multivariate Behavioral Research, 30*, 577-605.

Millsap, R. E. (1997). Invariance in measurement and prediction: Their relationship in the single-factor case. *Psychological Methods, 2*, 248-260.

Millsap, R. E. (1998). Group differences in regression intercepts: Implications for factorial invariance. *Multivariate Behavioral Research, 33*, 403-424.

Miltiades, H. B., & Pruchno, R. (2002). The effect of religious coping on caregiving appraisals of mothers of adults with developmental disabilities. *Gerontologist, 42*, 82-91.

Mitchell, W. G., Scheier, L. M., & Baker, S. A. (2000). Adherence to treatment in children with epilepsy: Who follows "doctor's orders"? *Epilepsia, 41*, 1616-1625.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*, 181-202.

Molenaar, P. C. M. (1994). Dynamic latent variable models in developmental psychology. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 155-180). Thousand Oaks, CA: Sage.

Molenaar, P. C. M. (1999). Comment on fitting MA time series by structural equation models. *Psychometrika, 64*, 91-94.

Molenaar, P. C. M., de Gooijer, J. G., & Schmitz, B. (1992). Dynamic factor analysis of nonstationary multivariate time series. *Psychometrika, 57*, 333-349.

Mooijaart, A. (1985a). Factor analysis for non-normal variables. *Psychometrika*, *50*, 323-342.

Mooijaart, A. (1985b). A note on computational efficiency in asymptotically distribution-free correlation models. *British Journal of Mathematical & Statistical Psychology, 38*, 112-115.

Mooijaart, A. (2003). Estimating the statistical power in small samples by empirical distributions. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J. J. Meulman (Eds.), *New developments in psychometrics* (pp. 149-156). Tokyo: Springer-Verlag.

Mooijaart, A., & Bentler, P. M. (1985). The weight matrix in asymptotic distribution-free methods. *British Journal of Mathematical and Statistical Psychology, 38*, 190-196.

Mooijaart, A., & Bentler, P. M. (1986). Random polynomial factor analysis. In E. Diday et al. (Eds.), *Data analysis and informatics IV* (pp. 241-250). Amsterdam: Elsevier Science.

Mooijaart, A., & Bentler, P. M. (1991). Robustness of normal theory statistics in structural equation models. *Statistica Neerlandica, 45*, 159-171.

Mooney, C. Z. (1997). *Monte Carlo simulation*. Thousand Oaks, CA: Sage.

Morokoff, P. J., Quina, K., Harlow, L. L., Whitmire, L., Grimley, D. M., Gibson, P. R., & Burkholder, G. J. (1997). Sexual Assertiveness Scale (SAS) for women: Development and validation. *Journal of Personality and Social Psychology*, *73*, 790-804.

Morris, C. (1995). Hierarchical models for educational data: An overview. *Journal of Educational Statistics, 20*, 190-200.

Mose, S., Budischewski, K. M., Rahn, A. N., Zander-Heinz, A. C., Bormeth, S., & Bottcher, H. D. (2001). Influence of irradiation on therapy-associated psychological distress in breast carcinoma patients. *International Journal of Radiation Oncology Biology Physics, 51*, 1328-1335.

Moskowitz, D. S., & Hershberger, S. L. (2002). *Modeling intraindividual variability with repeated measures data*. Mahwah, NJ: Erlbaum.

Motl, R. W., Dishman, R. K., Saunders, R. P., Dowda, M., Felton, G., Ward, D. S., & Pate, R. R. (2002). Examining social-cognitive determinants of intention and physical activity among black and white adolescent girls using structural equation modeling. *Health Psychology, 21*, 459-467.

Motl, R. W., & DiStefano, C. (2002). Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Structural Equation Modeling, 9*, 562-578.

Moulder, B. C., & Algina, J. (2002). Comparison of methods for estimating and testing latent variable interactions. *Structural Equation Modeling, 9*, 1-19.

Moustaki, I. (2000). A latent variable model for ordinal variables. *Applied Psychological Measurement, 24*, 211-224.

Moustaki, I. (2001). A review of exploratory factor analysis for ordinal categorical data. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 461-480). Lincolnwood, IL: Scientific Software International.

Moustaki, I., & Knott, M. (2000). Generalized latent trait models. *Psychometrika*, *65*, 391-411.

Mueller, M. J., Wetzel, H., Szegedi, A., & Benkert, O. (1999). Three dimensions of depression in patients with acute psychotic disorders: A replication study. *Comprehensive Psychiatry*, *40*, 449-457.
Mueller, R. O. (1996). *Basic principles of structural equation modeling.* New York: Springer Verlag.
Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. New York: Wiley.
Muirhead, R. J., & Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika*, *67*, 31-43.
Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stillwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, *105*, 430-445.
Mulligan, S. (1998). Application of structural equational modeling in occupational therapy research. *American Journal of Occupational Therapy*, *52*, 829-834.
Muntaner, C., Oates, G., & Lynch, J. (1999). Social class and social cohesion: A content validity analysis using a nonrecursive structural equation model. *Annals of the New York Academy of Sciences*, *896*, 409-413.
Muraishi, Y., & Toyoda, H. (1998). Analysis of standardized achievement tests by classical test theory and genetic factor analysis models. *Japanese Journal of Educational Psychology*, *46*, 395-402.
Muraki, E., & Carlson, E. (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement*, *24*, 211-224.
Murray, R. P., Istvan, J. A., Cribbie, R. A., & Barnes, G. E. (2002). Longitudinal analysis of the relationship between changes in smoking and changes in drinking in a community sample: The Winnipeg Health and Drinking Survey. *Health Psychology*, *21*, 237-243.
Murray, S. L., Holmes, J. G., Griffin, D. W., Bellavia, G., & Rose, P. (2001). The mismeasure of love: How self-doubt contaminates relationship beliefs. *Personality and Social Psychology Bulletin*, *27*, 423-436.
Murray, S. L., Holmes, J. G., Bellavia, G., Griffin, D. W., & Dolderman, D. (2002). Kindred spirits? The benefits of egocentrism in close relationships. *Journal of Personality and Social Psychology*, *82*, 563-581.
Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115-132.
Muthén, B. (1987). Response to Freedman's critique of path analysis: Improve credibility by better methodological training. *Journal of Educational Statistics*, *12*, 178-184.
Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 557-585.
Muthén, B. (1993). Goodness of fit test with categorical and other nonnormal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 205-234). Newbury Park: Sage.
Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*, 376-398.
Muthén, B. (1997). Latent variable modeling of longitudinal and multilevel data. In A. E. Raftery (Ed.) *Sociological methodology 1997* (pp. 453-481). Washington, DC: American Sociological Association.
Muthén, B. (2000). Methodological issues in random coefficient growth modeling using a latent variable framework: Applications to the development of heavy drinking ages 18-37. In J. R. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods for new questions* (pp. 113-140). Mahwah, NJ: Erlbaum.
Muthén, B. (2001a). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1-34). Mahwah, NJ: Erlbaum.
Muthén, B. (2001b). Second-generation structural equation modeling with a combination of categorical and continuous latent variables. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 291-322). Washington, DC: American Psychological Association.
Muthén, B. (2002). Beyond *SEM*: General latent variable modeling. *Behaviormetrika*, *29*, 81-117.
Muthén, B., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, *2*, 371-402.
Muthén, B., & Hofacker, C. (1988). Testing the assumptions underlying tetrachoric correlations. *Psychometrika*, *53*, 563-578.
Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171-189.
Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, *45*, 19-30.
Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*, 431-462.
Muthén, B., & Satorra, A. (1995a). Complex sample data in structural equation modeling. In P. V. Marsden (ed.), *Sociological methodology 1995* (pp. 267-316). Cambridge, MA: Blackwell Publishers.
Muthén, B., & Satorra, A. (1995b). Technical aspects of Muthen's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika*, *60*, 489-503.
Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*, 463-469.
Muthén, L. K., & Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*, 599-620.
Muthén, L. K., & Muthén, B. (2003). *Mplus*. Los Angeles, CA: Muthen & Muthen.

Nandakumar, R., Yu, F., Li, H.-H., & Stout, W. (1998). Assessing unidimensionality of polytomous data. *Applied Psychological Measurement, 22*, 99-115.
Nasser, F. & Takahashi, T. (2003). The effect of using item parcels on ad hoc goodness-of-fit indexes in confirmatory factor analysis: An example using Sarason's Reactions to Tests. *Applied Measurement in Education, 16*, 75-97.
Nauck, B. (2001). Intercultural contact and intergenerational transmission in immigrant families. *Journal of Cross-Cultural Psychology, 32*, 159-173.
Neale, M. C. (2000). Individual fit, heterogeneity, and missing data in multigroup structural equation modeling. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 249-267). Mahwah, NJ: Erlbaum.
Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 92-105). Washington, DC: American Psychological Association.
Nesselroade, J. R., & Baltes, P. B. (1979). *Longitudinal research in the study of behavior and development*. New York: Academic.
Nesselroade, J. R., & Molenaar, P. C. M. (1999). Pooling lagged covariance structures based on short, multivariate time series for dynamic factor analysis. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 223-250). Thousand Oaks, CA: Sage.
Neudecker, H., & Wesselman, A. M. (1990). The asymptotic variance matrix of the sample correlation matrix. *Linear Algebra and Its Applications, 127*, 589-599.
Neuhaus, J. M., & Kalbfleisch, J. D. (1998). Between- and within-cluster covariance effects in the analysis of clustered data. *Biometrics, 54*, 638-645.
Neumark-Sztainer, D., Wall, M. M., Story, M., & Perry, C. L. (2003). Correlates of unhealthy weight-control behaviors among adolescents: Implications for prevention programs. *Health Psychology, 22*, 88-98.
Nevitt, J. (2000). *Evaluating small sample approaches for model test statistics in structural equation modeling*. Unpublished Ph.D. dissertation, University of Maryland.
Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling, 8*, 353-377.
Newcomb, M. D. (1994). Drug use and intimate relationships among women and men: Separating specific from general effects in prospective data using structural equations models. *Journal of Consulting and Clinical Psychology, 62*, 463-476.
Newcomb, M. D., & Bentler, P. M. (1988). *Consequences of adolescent drug use: Impact on the lives of young adults.* Beverly Hills: Sage.
Nezafat, R., Shadmehr, R., & Holcomb, H. H. (2001). Long-term adaptation to dynamics of reaching movements: A PET study. *Experimental Brain Research, 140*, 66-76.
NICHD Early Child Care Research Network (2002). Child care structure->process->outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science, 13*, 199-206.
Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*, 241-301.
Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika, 32*, 1-13.
Nuamah, I. F., Cooley, M. E., Fawcett, J., & McCorkle, R. (1999). Testing a theory for health-related quality of life in cancer patients: A structural equation approach. *Research in Nursing & Health, 22*, 231-242.
Nunnally, J., & Bernstein, I. H. (1994). *Psychometric theory* (3rd. ed.). New York: McGraw-Hill.
Nyamathi, A. M., & Stein, J. A. (1997). Assessing the impact of HIV risk reduction counseling in impoverished African American women: A structural equations approach. *AIDS Education & Prevention, 9*, 253-273.
O'Callaghan, F. V., Callan, V. J., & Baglioni, A. (1999). Cigarette use by adolescents: Attitude-behavior relationships. *Substance Use & Misuse, 34*, 455-468.
O'Connor, B. P. (2002). The search for dimensional structure differences between normality and abnormality: A statistical review of published data on personality and psychopathology. *Journal of Personality and Social Psychology, 83*, 962-982.
Ogasawara, H. (2000). Some relationships between factors and components. *Psychometrika, 65*, 167-185.
Ogasawara, H. (2001a). Standard errors of fit indices using residuals in structural equation modeling. *Psychometrika, 66*, 421-436.
Ogasawara, H. (2001b). Approximations to the distributions of fit indices for misspecified structural equation models. *Structural Equation Modeling, 8*, 556-574.
Ogasawara, H. (2002). Asymptotic standard errors of estimated standard errors in structural equation modeling. *British Journal of Mathematical and Statistical Psychology, 55*, 213-229.
Ogasawara, H. (2003). A note on the approximate distributions of fit indexes for misspecified structural equation models. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J. J. Meulman (Eds.), *New developments in psychometrics* (pp. 103-108). Tokyo: Springer-Verlag.
Ogasawara, H. (2005). *Approximations to the distributions of fit indexes under fixed alternatives in nonnormal samples.* Paper presented at International Meeting of the Psychometric Society, Tilburg.
Okabayashi, H., Sugisawa, H., Takanashi, K., Nakatani, Y., & Shibata, H. (1999). The factor structure of coping strategies and their effects on burnout among primary caregivers of impaired elderly persons. *Japanese Journal of Psychology, 69*, 486-493.

Olkin, I., & Siotani, M. (1964). *Asymptotic distribution of functions of a correlation matrix.* Technical Report No. 6, Stanford University.
Olson, J. M., Vernon, P. A., Jang, K. L., & Harris, J. A. (2001). The heritability of attitudes: A study of twins. *Journal of Personality and Social Psychology, 80*, 845-860.
Olsson, U. (1979a). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*, 443-460.
Olsson, U. (1979b). On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research, 14*, 485-500.
Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika, 47*, 337-347.
Olsson, U. H., Foss, T., & Breivik, E. (2004). Two equivalent discrepancy functions for maximum likelihood estimation: Do their test statistics follow a noncentral chi-square distribution under model misspecification? *Sociological Methods & Research, 32*, 453-500.
Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling, 7*, 557-595.
Olzak, L. A., & Wickens, T. D. (1999). Paradigm shifts: New techniques to answer new questions. *Perception, 28*, 1509-1531.
Oort, F. J. (1999). Stochastic three-mode models for mean and covariance structures. *British Journal of Mathematical and Statistical Psychology, 52*, 243-272.
Oort, F. J. (2001). Three-mode models for multivariate longitudinal data. *British Journal of Mathematical and Statistical Psychology, 54*, 49-78.
Orchard, T., & Woodbury, M. A. (1972). A missing information principle: Theory and applications. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, 1*, 697-715.
Orlando, M., & Marshall, G. N. (2002). Differential item functioning in a Spanish translation of the PTSD checklist: Detection and evaluation of impact. *Psychological Assessment, 14*, 50-59.
Orlando, M., Ellickson, P. L., & Jinnett, K. (2001). The temporal relationship between emotional distress and cigarette smoking during adolescence and young adulthood. *Journal of Consulting and Clinical Psychology, 69*, 959-970.
Ostberg, M., & Hagekull, B. (2000). A structural modeling approach to the understanding of parenting stress. *Journal of Clinical Child Psychology, 29*, 615-625.
Oud, J. H. L., & Jansen, R. A. R. G. (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika, 65*, 199-215.
Oxford, M. L., Harachi, T. W., Catalano, R. F., Haggerty, K. P., & Abbott, R. D. (2000). Early elementary school-aged child attachment to parents: A test of theory and implications for interventions. *Prevention Science, 1*, 61-69.
Oyserman, D., Bybee, D., Mowbray, C. T., MacFarlane, P. (2002). Positive parenting among African American mothers with a serious mental illness. *Journal of Marriage and the Family, 64*, 65-77.
Pan, J. –X., & Fang, K. –T. (2002). *Growth curve models and statistical diagnosis.* New York: Springer.
Papke-Shields, K. E., & Malhotra, M. K. (2001). Assessing the impact of the manufacturing executive's role on business performance through strategic alignment. *Journal of Operations Management, 19*, 5-22.
Park, J., Kosterman, R., Hawkins, J. D., Haggerty, K. P., Duncan, T. E., Duncan, S. E., & Spoth, R. (2000). Effects of the "Preparing for the Drug Free Years" curriculum on growth in alcohol use and risk for alcohol use in early adolescence. *Prevention Science, 1*, 125-138.
Patterson, D. M. (2001). Causal effects of regulatory, organizational and personal factors on ethical sensitivity. *Journal of Business Ethics, 30*, 123-159.
Patterson, G. R., DeGarmo, D. S., & Knutson, N. (2000). Hyperactive and antisocial behaviors: Comorbid or two points in the same process? *Development & Psychopathology, 12*, 91-106.
Patterson, G. R., Dishion, T. J., & Yoerger, K. (2000). Adolescent growth in new forms of problem behavior: Macro- and micro-peer dynamics. *Prevention Science, 1*, 3-13.
Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling, 8*, 287-312.
Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: Cambridge University Press.
Pearl, J. (2001). Causal inference in the health sciences: A conceptual introduction. *Health Services and Outcomes Research Methodology.* Technical report R-282, Cognitive Systems Laboratory, Computer Science Department, UCLA.
Peckarsky, B. L., Taylor, B. W., McIntosh, A. R., McPeek, M. A., & Lytle, D.A. (2001). Variation in mayfly size at metamorphosis as a developmental response to risk of predation. *Ecology, 82*, 740-757.
Pedersen, N. L., Steffensson, B., Berg, S., Johansson, B., & McClearn, G. E. (1999). The importance of genetic and environmental effects for self-reported health symptoms: A 30-year follow-up considering survival and selection effects. *Journal of Aging and Health, 11*, 475-493.
Peek, M. K. (2000). Structural equation modeling and rehabilitation research. *American Journal of Physical Medicine and Rehabilitation, 79*, 301-309.
Pepermans, R., & Verleye, G. (1998). A unified Europe? How euro-attitudes relate to psychological differences between countries. *Journal of Economic Psychology, 19*, 681-699.
Pérez-Escamilla, R., Cobas, J. A., Balcazar, H., & Holland, B. M. (1999). Specifying the antecedents of breast-feeding duration in Peru through a structural equation model. *Public Health Nutrition, 2*, 461-467.

Perugini, M., & Bagozzi, R. P. (2001). The role of desires and anticipated emotions in goal-directed behaviours: Broadening and deepening the theory of planned behaviour. *British Journal of Social Psychology*, *40*, 79-98.
Petersson, K. M., Reis, A., Askeloef, S., Castro-Caldas, A., & Ingvar, M. (2000). Language processing modulated by literacy: A network analysis of verbal repetition in literate and illiterate subjects. *Journal of Cognitive Neuroscience*, *12*, 364-382.
Peterzell, D. H., & Teller, D. Y. (2000). Spatial frequency tuned covariance channels for red-green and luminance-modulated gratings: Psychophysical data from human adults. *Vision Research*, *40*, 417-430.
Pillai, V. K., & Gupta, R. (2000). Modeling sexual activity among schoolgirls in Zambia. *Journal of Family Issues, 21*, 170-190.
Ping, R. A., Jr. (1996). Latent variable interaction and quadratic effect estimation: A two-step technique using structural equation analysis. *Psychological Bulletin*, *119*, 166-175.
Ping, R. A., Jr. (1998). EQS and LISREL examples using survey data. In R. E. Schumacker & G. Marcoulides (Eds.). *Interaction and nonlinear effects in structural equation modeling* (pp. 63-100). Mahwah, NJ: Erlbaum.
Plewis, I. (2001). Explanatory models for relating growth processes. *Multivariate Behavioral Research, 36*, 207-225.
Pollitt, E., Durnin, J. V. G. A., Husaini, M., & Jahari, A. (2000). Effects of an energy and micronutrient supplement on growth and development in undernourished children in Indonesia: Methods. *European Journal of Clinical Nutrition*, *54 Suppl. 2*, 16-20.
Pollitt, E., Jahari, A., & Walka, H. (2000). A developmental view of the effects of an energy and micronutrient supplement in undernourished children in Indonesia. *European Journal of Clinical Nutrition*, *54 Suppl. 2*, 107-113.
Poon, W. –Y., & Lee, S. –Y. (1987). Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. *Psychometrika, 52*, 409-430.
Poon, W. –Y., & Lee, S. –Y. (1994). A distribution free approach for analysis of two-level structural equation model. *Computational Statistics & Data Analysis*, *17*, 265-275.
Poon, W. –Y., Lee, S. –Y., Afifi, A. A., & Bentler, P. M. (1990). Analysis of multivariate polytomous variates in several groups via the partition maximum likelihood approach. *Computational Statistics & Data Analysis*, *10*, 17-27.
Poon, W. –Y., Lee, S. –Y., & Bentler, P. M. (1990). Pseudo maximum likelihood estimation of multivariate polychoric and polyserial correlations. *Computational Statistics Quarterly*, *6*, 41-53.
Poon, W. –Y., Lee, S. –Y., Bentler, P. M., & Afifi, A. A. (1993). Covariance structure analysis with polytomous variables in several populations. *Acta Mathematicae Applicatae Sinica*, *9.1*, 63-70.
Poon, W. –Y., Lew, S. –F., & Poon, Y. S. (2000). A local influence approach to identifying multiple multivariate outliers. *British Journal of Mathematical and Statistical Psychology*, *53*, 255-273.
Poon, W. –Y., & Poon, Y. S. (2002). Influential observations in the estimation of mean vector and covariance matrix. *British Journal of Mathematical and Statistical Psychology*, *55*, 177-192.
Poon, W. –Y., Wang, S. –J., & Lee, S. –Y. (1999). Influence analysis of structural equation models with polytomous variables. *Psychometrika, 64*, 461-473.
Posner, S. F., & Baker, L. (2000). Evaluation and extensions of a structural equation modeling approach to the analysis of survival data. *Behavior Genetics*, *30*, 41-50.
Posthuma, D., de Geus, E. J. C., Neale, M. C., Pol, H. E. H., Baare, W. E. C., Kahn, R. S., & Boomsma, D. (2000). Multivariate genetic analysis of brain structure in an extended twin design. *Behavior Genetics*, *30*, 311-319.
Postmes, T., & Branscombe, N. R. (2002). Influence of long-term racial environmental composition on subjective well-being in African Americans. *Journal of Personality and Social Psychology*, *83*, 735-751.
Powell, D. A., & Schafer, W. D. (2001). The robustness of the likelihood ratio chi-square test for structural equation models: A meta-analysis. *Journal of Educational and Behavioral Statistics*, *26*, 105-132.
Power, M., Harper, A., & Bullinger, M. (1999). The World Health Organization WHOQOL-100: Tests of the universality of quality of life in 15 different cultural groups worldwide. *Health Psychology, 18*, 495-505.
Pratt, E. M., Telch, C. F., Labouvie, E. W., Wilson, G. T., & Agras, W. S. (2001). Perfectionism in women with binge eating disorder. *International Journal of Eating Disorders, 29*, 177-186.
Pratt, J. W., & Schlaifer, R. (1984). On the nature and discovery of structure. *Journal of the American Statistical Association*, *79*, 9-21.
Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, *2*, 13-43.
Prescott, C. A., Aggen, S. H., & Kendler, K. S. (2000). Sex-specific genetic influences on the comorbidity of alcoholism and major depression in a population-based sample of US twins. *Archives of General Psychiatry, 57*, 803-811.
Prinstein, M. J., Boergers, J., Spirito, A., Little, T. D., & Grapentine, W. L. (2000). Peer functioning, family dysfunction, and psychological symptoms in a risk factor model for adolescent inpatients' suicidal ideation severity. *Journal of Clinical Child Psychology, 29*, 392-405.
Pugesek, B. H., Tomer, A., & von Eye, A. (Eds.) (2003). *Structural equation modeling: Applications in ecological and evolutionary biology*. New York: Cambridge.
Punnett, L. & van der Beek, A. J. (2000). A comparison of approaches to modeling the relationship between ergonomic exposures and upper extremity disorders. *American Journal of Industrial Medicine, 37*, 645-655.
Purkayastha, S., & Srivastava, M. S. (1995). Asymptotic distributions of some test criteria for the covariance matrix in elliptical distributions under local alternatives. *Journal of Multivariate Analysis, 55*, 165-186.
Quintana, S. M., & Maxwell, S. E. (1999). Implications of recent developments in structural equation modeling for counseling psychology. *Counseling Psychologist*, *27*, 485-527.

Rabe-Hesketh, S., Yang, S., & Pickles A. (2001). Multilevel models for censored and latent responses. *Statistical Methods in Medical Research, 10*, 409-427.
Raftery, A. E. (1995). Bayesian model selection in social research. In A. E. Raftery (Ed.), *Sociological Methodology 1995* (pp. 111-164). Oxford, UK: Blackwell.
Rajah, M. N., McIntosh, A. R., & Grady, C. L. (1999). Frontotemporal interactions in face encoding and recognition. *Cognitive Brain Research, 8*, 259-269.
Ramamurthy, K., Premkumar, G., & Crum, M. R. (1999). Organizational and interorganizational determinants of EDI diffusion and organizational performance: A causal model. *Journal of Organizational Computing and Electronic Commerce, 9*, 253-285.
Ramirez-Valles, J., Zimmerman, M. A., & Newcomb, M. D. (1998). Sexual risk behavior among youth: Modeling the influence of prosocial activities and socioeconomic factors. *Journal of Health & Social Behavior, 39*, 237-253.
Ramsay, J. O., & Silverman, B. W. (1997). *Functional data analysis.* New York: Springer-Verlag.
Ramsay, J. O., & Silverman, B. W. (2002). *Applied functional data analysis*. New York: Springer-Verlag.
Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation. *Proceedings of the Cambridge Philosophical Society, 44*, 50-57.
Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics, 14*, 1-17.
Rapport, M. D., Scanlan, S. W., & Denney, C. B. (1999). Attention-deficithyperactivity disorder and scholastic achievement: A model of dual developmental pathways. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 40*, 1169-1183.
Rascle, N., & Irachabal, S. (2001). Mediators and moderators: Theoretical and methodological implications in stress and health psychology research. *Travail Humain, 64*, 97-118.
Rasmussen, J. L. (1988). Evaluating outlier identification tests: Mahalanobis D squared and Comrey Dk. *Multivariate Behavioral Research, 23*, 189-202.
Raudenbush, S. W. (1995). Maximum likelihood estimation for unbalanced multilevel covariance structure models via the EM algorithm. *British Journal of Mathematical & Statistical Psychology, 48*, 359-370.
Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*, 173-185.
Raudenbush, S. W. (2001a). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology, 52*, 501-525.
Raudenbush, S. W. (2001b). Toward a coherent framework for comparing trajectories of individual change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 33-64). Washington, DC: American Psychological Association.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd Ed.). Thousand Oaks: Sage.
Raudenbush, S. W., Bryk, A., Cheong, Y. F., & Congdon, R. (2000). *HLM 5. Hierarchical linear and nonlinear modeling*. Chicago: Scientific Software International.
Raudenbush, S. W., & Chan, W. (1992). Growth curve analysis in accelerated longitudinal designs. *Journal of Research in Crime and Delinquency, 29*, 387-411.
Raudenbush, S. W., & Sampson, R. J. (1999). Assessing direct and indirect effects in multilevel designs with latent variables. *Sociological Methods & Research, 28*, 123-153.
Raykov, T. (1997a). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*, 173-184.
Raykov, T. (1997b). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32*, 329-353.
Raykov, T. (1998a). Satisfying a simplex structure is simpler than it should be: A latent curve analysis revisit. *Multivariate Behavioral Research, 33*, 343-363.
Raykov, T. (1998b). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement, 22*, 375-385.
Raykov, T. (1998c). A method for obtaining standard errors and confidence intervals of composite reliability for congeneric measures. *Applied Psychological Measurement, 22*, 369-374.
Raykov, T. (2000). Modeling simultaneously individual and group patterns of ability growth or decline. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.). *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 127-146). Mahwah, NJ: Erlbaum.
Raykov, T. (2001a). Testing multivariable covariance structure and means hypotheses via structural equation modeling. *Structural Equation Modeling, 8*, 224-256.
Raykov, T. (2001b). Bias of coefficient α for fixed congeneric measures with correlated errors. *Applied Psychological Measurement, 25*, 69-76.
Raykov, T. (2001c). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical & Statistical Psychology, 54*, 315-323.
Raykov, T. (2001d). Approximate confidence interval for difference in fit of structural equation models. *Structural Equation Modeling, 8*, 458-469.

Raykov, T. (2002). Analytic estimation of standard error and confidence interval for scale reliability. *Multivariate Behavioral Research, 37*, 89-103.
Raykov, T. (2004a). Point and interval estimation of reliability for multiple-component measuring instruments via linear constraint covariance structure modeling. *Structural Equation Modeling, 11*, 342-356.
Raykov, T. (2004b). Estimation of maximal reliability: A note on a covariance structure modelling approach. *British Journal of Mathematical and Statistical Psychology, 57*, 21-27.
Raykov, T. (2005). Bias-corrected estimation of noncentrality parameters of covariance structure models. *Structural Equation Modeling, 12*, 120-129.
Raykov, T., & Marcoulides, G. A. (1999). On desirability of parsimony in structural equation model selection. *Structural Equation Modeling, 6*, 292-300.
Raykov, T., & Marcoulides, G. A. (2000). *A first course in structural equation modeling*. Mahwah, NJ: Erlbaum.
Raykov, T., & Marcoulides, G. A. (2001). Can there be infinitely many models equivalent to a given covariance structure model. *Structural Equation Modeling, 8*, 142-149.
Raykov, T., & Penev, S. (2001). The problem of equivalent structural equation models: An individual residual perspective. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 297-322). Mahwah, NJ: Erlbaum.
Raykov, T., & Shrout, P. E. (2002). Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling, 9*, 195-212.
Reichardt, C. S., & Gollob, H. F. (1986). Satisfying the constraints of causal modeling. In W. M. K. Trochim (Ed.), *Advances in quasi-experimental design and analysis* (pp. 91-107). San Francisco: Jossey-Bass.
Reid, M. J., Webster-Stratton, C., Beauchaine, T. P. (2001). Parent training in Head Start: A comparison of program response among African American, Asian American, Caucasian, and Hispanic mothers. *Prevention Science, 2*, 209-227.
Reinartz, W. J., Echambadi, R., & Chin, W. W. (2002). Generating non-normal data for simulation of structural equation models using Mattson's method. *Multivariate Behavioral Research, 37*, 227-244.
Reise, S. P. (2000). Using multilevel logistic regression to evaluate person-fit in IRT models. *Multivariate Behavioral Research, 35*, 543-568.
Reise, S. P., & Widaman, K. F. (1999). Assessing the fit of measurement models at the individual level: A comparison of item response theory and covariance structure approaches. *Psychological Methods, 4*, 3-21.
Reise, S. P., & Duan, N. (2001). Introduction to special issue on multilevel models. *Multivariate Behavioral Research, 36*, 153.
Reise, S. P., & Duan, N., (Eds.) (2002). *Multilevel modeling: Methodological advances, issues, and applications*. Mahwah, NJ: Erlbaum.
Remington, N. A., Fabrigar, L. R., & Visser, P. S. (2000). Reexamining the circumplex model of affect. *Journal of Personality & Social Psychology, 79*, 286-300.
Rensvold, R. B., & Cheung, G. W. (1999). Identification of influential cases in structural equation models using the jackknife method. *Organizational Research Methods, 2*, 293-308.
Reynolds, K. D., Yaroch, A. L., Franklin, F. A., & Maloy, J. (2002). Testing mediating variables in a school-based nutrition intervention program. *Health Psychology, 21*, 51-60.
Richardson, T., & Spirtes, P. (2002). Ancestral graph Markov models. *Annals of Statistics, 30*, 962-1030.
Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling, 3*, 369-379.
Rigdon, E. E. (1998). Structural equation modeling. In G.A. Marcoulides, (Eds.). *Modern methods for business research* (pp. 251-294). Mahwah, NJ: Erlbaum.
Rigdon, E. E. (1998b). The equal correlation baseline model for comparative fit assessment in structural equation modeling. *Structural Equation Modeling, 5*, 63-77.
Rigdon, E. E. (1998c). The equal correlation baseline model: A reply to Marsh. *Structural Equation Modeling, 5*, 87-94.
Riley, J. L., Robinson, M. E., Wade, J. B., Myers, C. D., & Price, D. D. (2001). Sex differences in negative emotional responses to chronic pain. *Journal of Pain, 2*, 354-359.
Riley, J. L., Wade, J. B., Robinson, M. E., & Price, D. D. (2000). The stages of pain processing across the adult lifespan. *Journal of Pain, 1*, 162-170.
Rimal, R. N. (2001). Longitudinal influences of knowledge and self-efficacy on exercise behavior: Tests of a mutual reinforcement model. *Journal of Health Psychology, 6*, 31-46.
Rindskopf, D. (1983). Parameterizing inequality constraints on unique variances in linear structural models. *Psychometrika, 48*, 73-83.
Rindskopf, D. (1984a). Structural equation models: Empirical identification, Heywood cases, and related problems. *Sociological Methods & Research, 13*, 109-119.
Rindskopf, D. (1984b). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika, 49*, 37-47.
Rindskopf, D. (2001). The next steps in latent variable models for change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 323-332). Washington, DC: American Psychological Association.
Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research, 23*, 51-67.

Rini, C. K., Dunkel-Schetter, C., Wadhwa, P. D., & Sandman, C. A. (1999). Psychological adaptation and birth outcomes: The role of personal resources, stress, and sociocultural context in pregnancy. *Health Psychology, 18*, 333-345.

Robins, J. M. (2001). Data, design, and background knowledge in etiologic inference. *Epidemiology, 12*, 313-320.

Robins, J. M., Hernán, M. Á., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology, 11*, 550-560.

Roesch, S. C. (1999). Modeling stress: A methodological review. *Journal of Behavioral Medicine, 22*, 249-269.

Rogers, W. M., & Schmitt, N. (2004). Parameter recovery and model fit using multidimensional composites: A comparison of four empirical parceling algorithms. *Multivariate Behavioral Research, 39*, 379-412.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 90*, 726-748.

Rogosa, D., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*, 203-228.

Roland, E., & Idsoe, T. (2001). Aggression and bullying. *Aggressive Behavior, 27*, 446-462.

Rose, J. S., Chassin, L., Presson, C. C., & Sherman, S. J. (2000). *Multivariate applications in substance use research: New methods for new questions.* Mahwah, NJ: Erlbaum.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*, 59-82.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach.* New York: Cambridge University Press.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologist. *Personnel Psychology, 47*, 537-560.

Rothenberg, S. J., Khan, F., Manalo, M., Jiang, J., Cuellar, R., Reyes, S., Acosta, S., Jauregui, M., Diaz, M., Sanchez, M., Todd, A. C., & Johnson, C. (2000). Maternal bone lead contribution to blood lead during and after pregnancy. *Environmental Research, 82*, 81-90.

Rotheram-Borus, M. J., Stein, J. A., & Lin, Y. –Y. (2001). Impact of parent death and an intervention on the adjustment of adolescents whose parents have HIVAIDS. *Journal of Consulting and Clinical Psychology, 69*, 763-773.

Rotnitzky, A., & Wypij, D. (1994). A note on the bias of estimators with missing data. *Biometrics, 50*, 1163–1170.

Roussel, P., Durrieu, F., Campoy, E., & El Akremi, A. (2002). *Méthodes d'équations structurelles: Recherche et applications en gestion*. Paris: Economica.

Rovine, M. J. (1994). Latent variables models and missing data analysis. In A. von Eye & C. C. Clogg (Eds.) *Latent variables analysis: Applications for developmental research* (pp. 181-225). Thousand Oaks, CA: Sage.

Rovine, M. J., & Molenaar, P. C. M. (1998a). A nonstandard method for estimating a linear growth model in LISREL. *International Journal of Behavioral Development, 22*, 453-473.

Rovine, M. J., & Molenaar, P. C. M. (1998b). A LISREL model for the analysis of repeated measures with a patterned covariance matrix. *Structural Equation Modeling, 5*, 318-343.

Rovine, M. J., & Molenaar, P. C. M. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research, 35*, 51-88.

Rovine, M. J., & Molenaar, P. C. M. (2001). A structural equations modeling approach to the general linear mixed model. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 67-96). Washington, DC: American Psychological Association.

Rowe, D. C., den Oord, E. J., Stever, C., Giedinghagen, L. N., Gard, J. M., Cleveland, H. H., Gilson, M., Terris, S. T., Mohr, J. H., Sherman, S., Abramowitz, A., & Waldman, I. D. (1999). The DRD2 TaqI polymorphism and symptoms of attention deficit hyperactivity disorder. *Molecular Psychiatry, 4*, 580-586.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581-592.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Rudinger, G., & Rietz, C. (1998). The neglected time dimension? Introducing a longitudinal model testing latent growth curves, stability, and reliability as time bound processes. *Methods of Psychological Research Online, 3*, 109-130.

Rupp, A. A., Dey, D. K., & Zumbo, B. D. (2004). To Bayes or not to Bayes, from whether to when: Applications of Bayesian methodology to modeling. *Structural Equation Modeling, 11*, 424-451.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin, 28*, 1629-1646.

Ryan, A. M., Chan, D., Ployhart, R.E., & Slade, L. A. (1999). Employee attitude surveys in a multinational organization: Considering language and culture in assessing measurement equivalence. *Personnel Psychology, 52*, 37-58.

Sadler, P., & Woody, E. (2003). Is who you are who you're talking to? Interpersonal style and complementarity in mixed-sex interactions. *Journal of Personality and Social Psychology, 84*, 80-96.

Sagarin, B. J., Cialdini, R. B., Rice, W. E., & Serna, S. B. (2002). Dispelling the illusion of invulnerability: The motivations and mechanisms of resistance to persuasion. *Journal of Personality and Social Psychology, 83*, 526-541.

Sager, J. K., Griffeth, R. W., & Hom, P. W. (1998). A comparison of structural models representing turnover cognitions. *Journal of Vocational Behavior, 53*, 254-273.

Salthouse, T. A., Toth, J., Daniels, K., Parks, C., Pak, R., Wolbrette, M., & Hocking, K. J. (2000). Effects of aging on efficiency of task switching in a variant of the trail making test. *Neuropsychology, 14*, 102-111.

Sani, F., & Todman, J. (2002). Should we stay or should we go? A social psychological model of schisms in groups. *Personality and Social Psychology Bulletin, 28*, 1647-1655.
Santos, S. J., Bohon, L. M., & Sanchez-Sosa, J. J. (1998). Childhood family relationships, marital and work conflict, and mental health distress in Mexican immigrants. *Journal of Community Psychology, 26*, 491-508.
Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. In C. Clogg (Ed.), *Sociological methodology 1987* (pp. 105-129). San Francisco: Jossey Bass.
Saris, W. E., & Stronkhurst, L. H. (1984). *Causal modelling in nonexperimental research*. Amsterdam: Sociometric Research Foundation.
Sato, M. (1987). Pragmatic treatment of improper solutions in factor analysis. *Annals of the Institute of Statistical Mathematics, 39*, 443-455.
Sato, M., & Ito, M. (2003). Some cautionary notes on factor loadings estimated by principal component analysis. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J. J. Meulman (Eds.), *New developments in psychometrics* (pp. 117-122). Tokyo: Springer-Verlag.
Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika, 54*, 131-151.
Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology, 22*, 249-278.
Satorra, A. (1993). Multi-sample analysis of moment-structures: Asymptotic validity of inferences based on second-order moments. In I, Haagen, D. J. Bartholomew, & M. Deistler (Eds.), *Statistical modeling and latent variables* (pp. 283-298). Amsterdam: North-Holland.
Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis: A Festschrift for Heinz Neudecker* (pp. 233-247). Dordrecht: Kluwer Academic Publishers.
Satorra, A. (2001). Goodness of fit testing of structural equation models with multiple group data and nonnormality. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 231-256). Lincolnwood, IL: Scientific Software International.
Satorra, A. (2002). Asymptotic robustness in multiple group linear-latent variable models. *Econometric Theory, 18*, 297-312.
Satorra, A. (2003). Power of χ^2 goodness-of-fit tests in structural equation models: The case of non-normal data. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J. J. Meulman (Eds.), *New developments in psychometrics* (pp. 57-68). Tokyo: Springer-Verlag.
Satorra, A., & Bentler, P. M. (1986). Some robustness properties of goodness of fit statistics in covariance structure analysis. *American Statistical Association: Proceedings of the Business & Economic Statistics Section*, 549-554.
Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *Proceedings of the American Statistical Association*, 308-313.
Satorra, A., & Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics & Data Analysis, 10*, 235-249.
Satorra, A., & Bentler, P. M. (1991). Goodness of fit test under IV estimation: Asymptotic robustness of a NT test statistic. In R. Gutierrez & M. J. Valderrama (Eds.), *Proceedings of the Fifth International Symposium on Applied Stochastic Models and Data Analysis* (pp. 555-567). London: World Scientific.
Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.
Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507-514.
Satorra, A., & Neudecker, H. (2003). A matrix equality useful in goodness-of-fit testing of structural equation models. *Journal of Statistical Planning and Inference, 114*, 63-80.
Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50*, 83-90.
Saucier, G. (2002). Orthogonal markers for orthogonal factors: The case of the Big Five. *Journal of Research in Personality, 36*, 1-31.
Saucier, G., & Goldberg, L. R. (2002). Assessing the big 5: Applications of 10 psychometric criteria to the development of marker scales. In B. de Raad & M. Perugini (Eds.), *Big Five Assessment* (pp. 30-54). New York: Hogrefe & Huber.
Savage, C. W., & Ehrlich, P. (Eds.) (1992). *Philosophical and foundational issues in measurement theory*. Hillsdale, NJ: Erlbaum.
Savalei, V., & Bentler, P. M. (in press). A statistically justified pairwise ML method for incomplete nonnormal data: A comparison with direct ML and pairwise ADF. *Structural Equation Modeling*.
Saxe, L., Kadushin, C., Beveridge, A., Livert, D., Tighe, E., Rindskopf, D., Ford, J., & Brodsky, A. (2001). The visibility of illicit drugs: Implications for community-based drug control strategies. *American Journal of Public Health, 91*, 1987-1994.
Saxena, K. M. L., & Alam, K. (1982). Estimation of the non-centrality parameter of a chi-squared distribution. *The Annals of Statistics, 10*, 1012-1016.
Sayer, A. G., & Cumsille, P. E. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 179-200). Washington, DC: American Psychological Association.
Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Schafer, J. L. (2001). Multiple imputation with PAN. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 357-377). Washington, DC: American Psychological Association.
Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177.
Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin, 64*, 92-107.
Schefczyk, M. (2001). Determinants of success of German venture capital investments. *Interfaces, 31*, 43-61.
Scheier, L. M., Botvin, G. J., & Griffin, K. W. (2001). Preventive intervention effects on developmental progression in drug use: Structural equation modeling analyses using longitudinal data. *Prevention Science, 2*, 91-112.
Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika, 64*, 37-52.
Scheines, R., Spirtes, P., Glymour, C., & Meek, C. (1994). *TETRAD II: Tools for discovery*. Hillsdale, NJ: Erlbaum.
Schepers, A., & Arminger, G. (1992). *MECOSA: A program for the analysis of general mean- and covariance structures with non-metric variables, User's guide*. Frauenfeld, Switzerland: SLI-AG.
Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23-74.
Scherpenzeel, A. C., & Saris, W. E. (1997). The validity and reliability of survey questions: A meta-analysis of MTMM studies. *Sociological Methods & Research, 25*, 341-383.
Scheufele, D. A., & Shah, D. V. (2000). Personality strength and social capital: The role of dispositional and informational variables in the production of civic participation. *Communication Research, 27*, 107-131.
Schimmack, U., Radhakrishnan, P., Oishi, S., Dzokoto, V., & Ahadi, S. (2002). Culture, personality, and subjective well-being: Integrating process models of life satisfaction. *Journal of Personality and Social Psychology, 82*, 582-593.
Schmidt, F. L., Le, H., & Ileus, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual-differences constructs. Psychological Methods. 8, 206-224.
Schmidt, F., & Hunter, J. (2002). Are there benefits from NHST? *American Psychologist,* 57, 65-66.
Schmitt, N. (1996). Uses and abuses of coefficient alpha. Psychological Assessment, 8, 350-353.
Schmitt, N. (2002). Do reactions to tests produce changes in the construct measured? *Multivariate Behavioral Research, 37*, 105-126.
Schmukle, S. C., & Hardt, J. (2005). A cautionary note on incremental fit indices reported by LISREL. *Methodology, 1*, 81-85.
Schnoll, R. A., Harlow, L. L., Stolbach, L. L., & Brandt, U. A. (1998). Structural model of the relationships among stage of disease, age, coping, and psychological adjustment in women with breast cancer. *Psycho-Oncology, 7*, 69-77.
Schoenberg, R., & Arminger, G. (1992). LINCS2: A program for linear covariance structure analysis. Kensington, MD: RJS.
Schönemann, P. H., & Steiger, J. H. (1979). Regression component analysis. *British Journal of Mathematical and Statistical Psychology, 29*, 175-189.
Schröder, K. E. E., Hobfoll, S. E., Jackson, A. P., & Lavin, J. (2001). Proximal and distal predictors of AIDS risk behaviors among inner-city African American and European American women. *Journal of Health Psychology, 6*, 169-190.
Schulte-Koerne, G., Deimel, W., Bartling, J., & Remschmidt, H. (1999). The role of phonological awareness, speech perception, and auditory temporal processing for dyslexia. *European Child & Adolescent Psychiatry, 8*, (Suppl. 3), 28-34.
Schumacker, R. E. (2002). Latent variable interaction modeling. *Structural Equation Modeling, 9*, 40-54.
Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Erlbaum.
Schumacker, R. E., & Marcoulides, G. A., (Eds.). (1998). *Interaction and nonlinear effects in structural equation modeling.* Mahwah, NJ: Erlbaum.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464.
Schwarzer, R., & Renner, B. (2000). Social-cognitive predictors of health behavior: Action self-efficacy and coping self-efficacy. *Health Psychology, 19*, 487-495.
Scott, M. A., & Handcock, M. S. (2001). Covariance models for latent structure in longitudinal data. *Sociological Methodology, 31*, 265-303.
Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression*. New York: Wiley.
Seibert, S. E., Kraimer, M. L., & Liden, R. C. (2001). A social capital theory of career success. *Academy of Management Journal, 44*, 219-237.
Seo, T., & Toyama, T. (1996). On the estimation of kurtosis parameter in elliptical distributions. *Journal of the Japan Statistical Society, 26*, 59-68.
Serlin, R. C., & Lapsley, D. K. (1985). Rationality in psychological theory: The good-enough principle. *American Psychologist, 40*, 73-82.
Serlin, R. C., & Lapsley, D. K. (1993). Rational appraisal of psychological research and the good-enough principle. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 129-228). Hillsdale, NJ: Erlbaum.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* New York: Houghton Mifflin.
Shah, J. Y., Friedman, R., & Kruglanski, A. W. (2002). Forgetting all else: On the antecedents and consequences of goal shielding. *Journal of Personality and Social Psychology, 83*, 1261-1280.
Shapiro, A. (1982a). Weighted minimum trace factor analysis. *Psychometrika, 47*, 243-264.

Shapiro, A. (1982b). Rank reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis. *Psychometrika, 47*, 187-199.
Shapiro, A. (1983). Asymptotic distribution theory in the analysis of covariance structures. South *African Statistical Journal, 17*, 33-81.
Shapiro, A. (1985a). Asymptotic equivalence of minimum discrepancy function estimators to GLS estimators. *South African Statistical Journal, 19*, 73-81.
Shapiro, A. (1985b). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika, 72*, 133-144.
Shapiro A. (1985c). A note on the asymptotic-distribution of the greatest lower bound to reliability. *Psychometrika, 50*, 243-244.
Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *Journal of the American Statistical Association, 81*, 142-149.
Shapiro, A. (1987). Robustness properties of the MDF analysis of moment structures. *South African Journal, 21*, 39-62.
Shapiro, A., & Browne, M. W. (1983). On the investigation of local identifiability: A counterexample. *Psychometrika, 48*, 303-304.
Shapiro, A., & Browne, M. W. (1987). Analysis of covariance structures under elliptical distributions. *Journal of the American Statistical Association, 82*, 1092-1097.
Shapiro, A., & Browne, M. W. (1990). On the treatment of correlation structures as covariance structures. *Linear Algebra and Its Applications, 127*, 567-587.
Shapiro, A., & ten Berge, J. M. F. (2000). The asymptotic bias of minimum trace factor analysis, with applications to the greatest lower bound to reliability. *Psychometrika, 65*, 413-425.
Shapiro, A., & ten Berge, J. M. F. (2002). Statistical inference of minimum rank factor analysis. *Psychometrika, 67*, 79-94.
Shapiro, D. L., & Levendosky, A. A. (1999). Adolescent survivors of childhood sexual abuse: the mediating role of attachment style and coping in psychological and interpersonal functioning. *Child Abuse and Neglect, 23*, 1175-1191.
Sharma, S., Durvasula, S., & Dillon, W. R. (1989). Some results on the behavior of alternate covariance structure estimation procedures in the presence of non-normal data. *Journal of Marketing Research, 26*, 214-221.
Shavelson, R., & Webb, N. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.
Sheehan, T. J. (1998). Stress and low birth weight: A structural modeling approach using real life stressors. *Social Science & Medicine, 47*, 1503-1512.
Shen, H., Bentler, P. M., & Comrey, A. L. (1995). A comparison of models of medical school student selection. *Structural Equation Modeling, 2*, 93-100.
Shenkin, S.D., Starr, J. M., Pattie, A., Rush, M. A., Whalley, L. J., & Deary, I. J. (2001). Birth weight and cognitive function at age 11 years: the Scottish Mental Survey 1932. *Archives of Disease in Childhood, 85*, 189-195.
Shevlin, M., Miles, J. N. V., Davies, M. N. O., & Walker, S. (2000). Coefficient alpha: A useful indicator of reliability? *Personality & Individual Differences, 28*, 229-237.
Shifren, K., Park, D. C., Bennett, J. M., & Morrell, R. W. (1999). Do cognitive processes predict mental health in individuals with rheumatoid arthritis? *Journal of Behavioral Medicine, 22*, 529-547.
Shipley, B. (2000a). *Cause and correlation in biology*. New York: Cambridge University Press.
Shipley, B. (2000b). A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling, 7*, 206-218.
Shipley, B. (2002). Start and stop rules for exploratory path analysis. *Structural Equation Modeling, 9*, 554-561.
Shriberg, L. D., Friel-Patti, S., Flipsen, P., Jr., & Brown, R. L.(2000). Otitis media, fluctuant hearing loss, and speech-language outcomes: A preliminary structural equation model. *Journal of Speech, Language, & Hearing Research, 43*, 100-120.
Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422-445.
Siddiqui, O, & Ali, M. W. (1999). Linear structural equation model in analyzing quality-of-life-data from clinical trials. *Journal of Biopharmaceutical Statistics, 9*, 661-681.
Sideridis, G. D., Kaissidis, A., & Padeliadu, S. (1998). Comparison of the theories of reasoned action and planned behaviour**.** *British Journal of Educational Psychology, 68*, 563-580.
Sidman, M. (1952). A note on functional relations obtained from group data. *Psychological Bulletin, 49*, 263-269.
Siegrist, M. (2000). The influence of trust and perceptions of risks and benefits on the acceptance of gene technology. *Risk Analysis, 20*,195-203.
Sieving, R. E., Maruyama, G., Williams, C. L., & Perry, C. L. (2000). Pathways to adolescent alcohol use: Potential mechanisms of parent influence. *Journal of Research on Adolescence, 10*, 489-514.
Silvia, E. S. M., & MacCallum, R. C. (1988). Some factors affecting the success of specification searches in covariance structure modeling. *Multivariate Behavioral Research, 23*, 297-326.
Simonton, D. K. (1991). Latent-variable models of posthumous reputation: A quest for Galton's G. *Journal of Personality and Social Psychology, 60*, 607-619.
Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23*, 323-355.
Singer, J. D., & Willett, J. D. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford.

Singhapakdi, A., Vitell, S. J., & Franke, G. R. (1999). Antecedents, consequences, and mediating effects of perceived moral intensity and personal moral philosophies. *Journal of the Academy of Marketing Science, 27*, 19-36.
Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods, 6*, 317-329.
Sivo, S. A., & Willson, V. L. (2000). Modeling causal error structures in longitudinal panel data: A Monte Carlo study. *Structural Equation Modeling, 7*, 174-205.
Skeem, J. L., Mulvey, E. P., & Grisso, T. (2003). Applicability of traditional and revised models of psychopathy to the Psychopathy Checklist: Screening version. *Psychological Assessment, 15*, 41-55.
Skinner, N., & Brewer, N. (2002). The dynamics of threat and challenge appraisals prior to stressful achievement events. *Journal of Personality and Social Psychology, 83*, 678-692.
Skitka, L. J., & Mullen, E. (2002). Understanding judgments of fairness in a real-world political context: A test of the value protection model of justice reasoning. *Personality and Social Psychology Bulletin, 28*, 1419-1429.
Skrondal, A., & Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika, 68*, 267-287.
Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models.* London: Chapman & Hall.
Slade, M., Leese, M., Taylor, R., & Thornicroft, G. (1999). The association between needs and quality of life in an epidemiologically representative sample of people with psychosis. *Acta Psychiatrica Scandinavica, 100*, 149-157.
Smith, C. S., Tisak, J., Hahn, S. E., & Schmieder, R. A. (1997). The measurement of job control. *Journal of Organizational Behavior, 18*, 225-237.
Smith, M. A., Brown, M. E., Borowsky, S. J., Davis, M. K., Center, B., & Lurie, N. (2001). Measuring the physician perspective on quality of care in health plans. *Evaluation and The Health Professions, 24*, 18-35.
Smith, P., Perrin, S., Yule, W., & Rabe-Hesketh, S. (2001). War exposure and maternal reactions in the psychological adjustment of children from Bosnia-Hercegovina. *Journal of Child Psychology and Psychiatry, 42*, 395-404.
Smither, J. W., Millsap, R. E., Stoffey, R. W., & Reilly, R. R. (1996). An experimental test of the influence of selection procedures on fairness perceptions, attitudes about the organization, and job pursuit intentions. *Journal of Business and Psychology, 10*, 297-318.
Snijders, T. A .B., & Bosker, R. J. (1999). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.
Snijders, T. A. B., & Hagenaars, J. A. (2001). Guest editors' introduction to the Special Issue on Causality at Work. *Sociological Methods & Research, 30*, 3-10.
Snijders, T. A. B., & Kenny, D. A. (2000). The social relations model for family data: A multilevel approach. *Personal Relationships, 6*, 471-486.
Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology* (pp. 290-312). San Francisco: Jossey Bass.
Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. In N. B. Tuma (Ed.), *Sociological Methodology* (pp. 159-186). Washington, DC: American Sociological Association.
Sobel, M. E. (1987). Direct and indirect effects in linear structural equation models. *Sociological Methods & Research, 16*, 155-176.
Sobel, M. E. (1990). Effect analysis and causation in linear structural equation models. *Psychometrika, 55*, 495-515.
Sobel, M. E. (1994). Causal inference in latent variable models. In A. von Eye, C. C. Clogg, (Eds.). *Latent variables analysis: Applications for developmental research* (pp. 3-35). Thousand Oaks, CA: Sage.
Sobel, M. E. (1995). Causal inference in the social and behavioral sciences. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.) *Handbook of statistical modeling for the social and behavioral sciences* (pp. 1-38). New York: Plenum.
Sobel, M. E. (2000). Causal inference in the social sciences. *Journal of the American Statistical Association, 95*, 647-651.
Sobel, M. E., & Bohrnstedt, G. W. (1985). Use of null models in evaluating the fit of covariance structure models. In N. B. Tuma (Ed.), *Sociological methodology* (pp. 152-178). San Francisco: Jossey Bass.
Solberg, E. C., Diener, E., Wirtz, D., Lucas, R. E., & Oishi, S. (2002). Wanting, having, and satisfaction: Examining the role of desire discrepancies in satisfaction with income. *Journal of Personality and Social Psychology, 83*, 725-734.
Song, X. -Y., & Lee, S. -Y. (2001). Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. *British Journal of Mathematical and Statistical Psychology, 54*, 237-263.
Song, X. -Y., & Lee, S. -Y. (2002a). A Bayesian model selection method with applications. *Computational Statistics & Data Analysis, 40*, 539-557.
Song, X. -Y., & Lee, S. -Y. (2002b). A Bayesian approach for multigroup nonlinear factor analysis. *Structural Equation Modeling, 9*, 523-553.
Song, X. -Y., & Lee, S. -Y. (2002c). Analysis of structural equation model with ignorable missing continuous and polytomous data. *Psychometrika, 67*, 261-288.
Song, X. -Y., & Lee, S. -Y. (2004). Bayesian analysis of two-level nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology, 57*, 29-52.
Sörbom, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229-239.

Sörbom, D. (1978). An alternative to the methodology for analysis of covariance. *Psychometrika, 43*, 381-396.
Sörbom, D. (1982). Structural equation models with structured means. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp. 183-195). Amsterdam: North-Holland.
Sörbom, D. (1989). Model modification. *Psychometrika*, *54*, 371-384.
Speckart, G., & Bentler, P. M. (1982). Application of attitude-behavior models to varied content domains. *Academic Psychology Bulletin, 4*, 453-466.
Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search* (2nd Ed.). Cambridge, MA: MIT Press.
Spirtes, P., Scheines, R., & Glymour, C. (1990). Simulation studies of the reliability of computer-aided model specification using the TETRAD II, EQS, and LISREL programs. *Sociological Methods & Research, 19*, 3-66. (With discussion)
Spoth, R., Redmond, C. (2002). Project Family prevention trials based in community-university partnerships: Toward scaled-up preventive interventions. *Prevention Science, 3*, 203-221.
Spoth, R., & Redmond, C. , & Shin, C. (2000). Modeling factors influencing enrollment in family-focused preventive intervention research. *Prevention Science, 1*, 213-225.
Spoth, R., Redmond, C., Shin, C., & Huck, S. (1999). A protective process model of parent-child affective quality and child mastery effects on oppositional behaviors: A test and replication. *Journal of School Psychology, 37*, 49-71.
Srivastava, A., Locke, E. A., & Bartol, K. M. (2001). Money and subjective well-being: It's not the money, it's the motives. *Journal of Personality and Social Psychology, 80*, 959-971.
Stacy, A. W., Bentler, P. M., & Flay B. R. (1994). Attitudes and health behavior in diverse populations: Drunk driving, alcohol use, binge eating, marijuana use, and cigarette use. *Health Psychology, 13*, 73-85.
Stacy, A. W., Newcomb, M. D., & Bentler, P. M. (1995). Expectancy in mediational models of cocaine use. *Personality and Individual Differences, 19*, 655-668.
Stapleton, L. (2002). The incorporation of sample weights into multilevel structural equation models. *Structural Equation Modeling, 9*, 475-502.
Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78-90.
Steiger, J. H. (1989). *EzPATH: A supplementary module forSYSTAT and SYGRAPH.* Evanston, IL: SYSTAT.
Steiger, J. H. (1990). Structural model evaluation and modification: An internal estimation approach. *Multivariate Behavioral Research, 25*, 173-180.
Steiger, J. H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling*, *5*, 411-419.
Steiger, J. H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser. *Structural Equation Modeling, 7*, 149-162.
Steiger, J. H. (2001). Driving fast in reverse: The relationship between software development, theory, and education in structural equation modeling (review of 3 texts). *Journal of the American Statistical Association, 96*, 331-338.
Steiger, J.H. (2002). When constraints interact: A caution about reference variables, identification constraints, and scale dependencies in structural equation modeling. *Psychological Methods, 7*, 210-227.
Steiger, J. H. (2004). Beyond the *F t*est: Effect size, confidence intervals, and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9*, 164-182.
Steiger, J. H., & Hakstian, A. R. (1982). The asymptotic distribution of elements of a correlation matrix: Theory and application. *British Journal of Mathematical & Statistical Psychology, 35*, 208-215.
Steiger, J. H., & Lind, J. M. (1980). *Statistically based tests for the number of common factors*. Paper presented at Psychometric Society, Iowa City, IA.
Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, *50*, 253-264.
Stein, J. A., Newcomb, M. D., & Bentler, P. M. (1996). Initiation and maintenance of tobacco smoking: Changing personality correlates in adolescence and young adulthood. *Journal of Applied Social Psychology, 26*, 160-187.
Stein, J. A., Nyamathi, A., & Bentler, P. M. (2001). Applying multilevel models to AIDS reduction strategies in high-risk populations: Individual and couple level effects. *Under editorial review.*
Stelzl, I. (1986). Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research, 21*, 309-331.
Stephan, W. G., Boniecki, K. A., Ybarra, O., Bettencourt, A., Ervin, K. S., Jackson, L. A., McNatt, P. S., & Renfro, C. L. (2002). The role of threats in the racial attitudes of blacks and whites. *Personality and Social Psychology Bulletin, 28*, 1242-1254.
Sterne, J. A. C., & Smith, G. D. (2001). Sifting the evidence – what's wrong with significance tests? *British Medical Journal, 322*, 226-231.
Stewart, S. H., Taylor, S., Jang K. L., Cox, B. J., Watt, M. C., Fedoroff, I. C., & Borger, S. C. (2001). Causal modeling of relations among learning history, anxiety sensitivity, and panic attacks. *Behaviour Research and Therapy*, *39*, 443-456.
Steyer, R., & Partchev, I. (2001). Latent state-trait modeling with logistic item response models. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 481-520). Lincolnwood, IL: Scientific Software International.
Steyer, R., Partchev, I., & Shanahan, M. J. (2000). In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.). *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 109-126). Mahwah, NJ: Erlbaum.
Steyer, R., Schmitt, M. J., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Psychology, 13*, 389-408.

Steyn, H. S. (1996). The distribution of the covariance matrix for a subset of elliptical distributions with extension to two kurtosis parameters. *Journal of Multivariate Analysis*, *58*, 96-106.
Stine, R. (1989). An introduction to bootstrap methods: Examples and ideas. *Sociological Methods & Research*, *8*, 243-291.
Stoel, R. D. (2003). *Issues in growth curve modeling*. Amsterdam: TT-Publikaties.
Stoel, R. D., & van den Wittenboer, G. (2003). Time dependence of growth parameters in latent growth curve models with time invariant covariates. *Methods of Psychological Research*, *8*, 21-41.
Stoolmiller, M. (1994). Antisocial behavior, delinquent peer association and unsupervised wandering for boys: Growth and change from childhood to early adolescence. *Multivariate Behavioral Research*, *29*, 263-288.
Stoolmiller, M. (1995). Using latent growth curve models to study developmental processes. In J. M. Gottman (Ed.), *The analysis of change* (pp. 103-138). Mahwah, NJ: Erlbaum.
Strauss, M. E., Thompson, P., Adams, N. L., Redline, S., & Burant, C. (2000). Evaluation of a model of attention with confirmatory factor analysis. *Neuropsychology*, *14*, 201-208.
Strenio, J. L. F., Weisberg, H. I. & Bryk, A. S. (1983). Empirical Bayes estimation of individual growth curve parameters and their relationship to covariates. *Biometrics*, *39*, 71-86.
Struthers, C. W., Weiner, B., & Allred, K. (1998). Effects of causal attributions on personnel decisions: A social motivation perspective. *Basic & Applied Social Psychology*, *20*, 155-166.
Sullivan, M.D., Kempen, G.I.J.M., Van Sonderen, E., & Ormel, J. (2000). Models of health-related quality of life in a population of community-dwelling Dutch elderly. *Quality of Life Research*, *9*, 801-810.
Surakka, J., Alanen, E., Aunola, S., & Karppi, S. L. (2001). Isoresistive dynamometer measurement of trunk muscle velocity at different angular phases of flexion and extension. *Clinical Physiology*, *21*, 504-511.
Swanson, E. B., & Dans, E. (2000). System life expectancy and the maintenance effort: Exploring their equilibration. *MIS Quarterly*, *24*, 277-297.
Tait, K., & Purdie, N. (2000). Attitudes toward disability: Teacher education for inclusive environments in an Australian university. *International Journal of Disability, Development & Education*, *47*, 25-38.
Tanaka, J. S. (1984). *Some results on the estimation of covariance structure models*. Ph.D. Thesis, University of California, Los Angeles.
Tanaka, J. S. (1987). "How big is enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development*, *58*, 134-146.
Tanaka, J. S. (1993). Multifaceted conceptions of fit in structure equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
Tanaka, J. S., & Bentler, P. M. (1985). Quasi-likelihood estimation in asymptotically efficient covariance structure models. *American Statistical Association: 1984 Proceedings of the Social Statistics Section*, pp. 658-662.
Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, *38*, 197-201.
Tang, M. –L., & Bentler, P. M. (1997). Maximum likelihood estimation in covariance structure analysis with truncated data. *British Journal of Mathematical and Statistical Psychology*, *50*, 339-349.
Tang, M. –L., & Bentler, P. M. (1998). Theory and method for constrained estimation in structural equation models with incomplete data. *Computational Statistics & Data Analysis*, *27*, 257-270.
Taris, T. W. (1999). The mutual effects between job resources and mental health: A prospective study among Dutch youth. *Genetic, Social, & General Psychology Monographs*, *125*, 433-450.
Taylor, B. J., Graham, J. W., Cumsille, P., & Hansen, W. B. (2000). Modeling prevention program effects on growth in substance use: Analysis of five years of data from the Adolescent Alcohol Prevention Trial. *Prevention Science*, *1*, 183-197.
Taylor, J., McGue, M., Iacono, W. G., & Lykken, D. T. (2000). A behavioral genetic analysis of the relationship between the socialization scale and self-reported delinquency. *Journal of Personality*, *68*, 29-50.
Taylor, J. G., Horwitz, B., Shah, N. J., Fellenz, W. A., Mueller-Gaertner, H. W., & Krause, J. B. (2000). Decomposing memory: Functional assignments and brain traffic in paired associate learning. *Neural Networks*, *13*, 923-940.
Taylor, J. G., Krause, B., Shah, N. J., Horwitz, B., & Mueller-Gaertner, H.W. (2000). On the relation between brain images and brain neural networks. *Human Brain Mapping*, *9*, 165-182.
Teachman, J., Duncan, G. J., Yeung, W. J., & Levy, D. (2001). Covariance structure models for fixed and random effects. *Sociological Methods & Research*, *30*, 271-288.
Teel, J. E., Bearden, W. O., & Sharma, S. (1986). Interpreting LISREL estimates of explained variance in nonrecursive structural equation models. *Journal of Marketing Research*, *23*, 164-168.
ten Berge, J. M. F. (2000). Linking reliability and factor analysis: Recent developments in some classical psychometric problems. In S. E. Hampson (Ed). *Advances in personality psychology: Vol. 1* (pp. 138-156). Philadelphia: Psychology PressTaylor & Francis.
ten Berge, J. M. F. (2004), & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, *69*, 613-626.
ten Berge, J. M. F., Snijders, T. A. B., & Zegers, F. E. (1981). Computational aspects of the greatest lower bound to reliability and constrained minimum trace factor analysis. *Psychometrika*, *46*, 201-213.
Thompson, B. (2000). Ten commandments of structural equation modeling. In L. G. Grimm, P. R. Yarnold, (Eds,). *Reading and understanding more multivariate statistics* (pp. 261-283). Washington, DC: American Psychological Association.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31*, 25-32.
Thompson, B. (Ed.) (2003). *Score reliability: Contemporary thinking on reliability issues*. Thousand Oaks, CA: Sage.
Thompson, J. K., Coovert, M. D., & Stormer, S. M. (1999). Body image, social comparison, and eating disturbance: A covariance structure modeling investigation. *International Journal of Eating Disorders, 26*, 43-51.
Thompson, J. K., & Heinberg, L. J. (1999). The media's influence on body image disturbance and eating disorders: We've reviled them, now can we rehabilitate them? *Journal of Social Issues, 55*, 339-353.
Thompson, J. R. (1999). *Simulation: A modeler's approach*. New York: Wiley.
Thompson, S. G., Turner, R. M., & Warn, D. E. (2001). Multilevel models for meta-analysis, and their application to absolute risk differences. *Statistical Methods in Medical Research, 10*, 375-392.
Thomson, G. H. (1936). Some points of mathematical technique in the factorial analysis of ability. *Journal of Educational Psychology, 27*, 37-54.
Thomson, G. H. (1940). Weighting for battery reliability and prediction. *British Journal of Psychology, 30*, 357-366.
Thurstone, L. L. (1935). *The vectors of mind*. Chicago: University of Chicago Press.
Timm, N. H. (1970). The estimation of variance-covariance and correlation matrices from incomplete data. *Psychometrika, 35*, 417-437.
Tisak, J., & Tisak, M. S. (2000). Permanency and ephemerality of psychological measures with application to organizational commitment. *Psychological Methods, 5*, 175-198.
Tobias, R. D. (1997). *An introduction to partial least squares regression*. Cary, NC: SAS Institute.
Tomas, J. M., Hontangas, P. M., & Oliver, A. (2000). Linear confirmatory factor models to evaluate multitrait-multimethod matrices: The effects of number of indicators and correlation among methods. *Multivariate Behavioral Research, 35*, 469-499.
Tonry, M., Ohlin, L., & Farrington, D. P. (1991). *Human development and criminal behavior*. New York: Springer.
Tormala, Z. L., Petty, R. E., & Briñol, P. (2002). Ease of retrieval effects in persuasion: A self-validation analysis. *Personality and Social Psychology Bulletin, 28*, 1700-1712.
Tremblay, P. F., Gardner, R. C., & Heipel, G. (2000). A model of the relationships among measures of affect, aptitude, and performance in introductory statistics. *Canadian Journal of Behavioural Science, 32*, 40-48.
Trinkoff, A. M., Zhou, Q., Storr, C. L., & Soeken, K. L. (2000). Workplace access, negative proscriptions, job strain, and substance use in registered nurses. *Nursing Research, 49*, 83-90.
Trumbo, C. W. (1999). Heuristic-systematic information processing and risk judgment. *Risk Analysis, 19*, 391-400.
Tu, Q., Vonderembse, M. A., & Ragu-Nathan, T. S. (2001). The impact of time-based manufacturing practices on mass customization and value to customer. *Journal of Operations Management, 19*, 201-217.
Tucker, J. S., Orlando, M., & Ellickson, P. L. (2003). Patterns and correlates of binge drinking trajectories from early adolescence to young adulthood. *Health Psychology, 22*, 79-87.
Tucker, L. R (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika, 23*, 19-23.
Tucker, L. R, & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1-10.
Turbin, M. S., Jessor, R., & Costa, F. M. (2000). Adolescent cigarette smoking: Health-related behavior or normative transgression? *Prevention Science, 1*, 115-124.
Turner, N., Ialomiteanu, A., & Room, R. (1999). Checkered expectations: Predictors of approval of opening a casino in the Niagara community. *Journal of Gambling Studies, 15*, 45-70.
Tyler, D. E. (1982). Radial estimates and the test for sphericity. *Biometrika, 69*, 429-436.
Tyler, D. E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika, 70*, 411-420.
Ullman, J. B. (2000). Structural equation modeling. In B.G. Tabachnick, & L. S. Fidell (Eds.). *Using multivariate statistics* (4th ed., pp. 653-771). New York: Allyn & Bacon.
Ullman, J. B., & Bentler, P. M. (2003). Structural equation modeling. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (pp. 607-634). Hoboken, NJ: Wiley.
Ullman, J. B., & Newcomb, M. D. (1999). The transition from adolescent to adult: A time of change in general and specific deviance. *Criminal Behaviour and Mental Health, 9*, 74-90.
Ullman, J. B., Stein, J. A., & Dukes, R. L. (2000). Evaluation of D.A.R.E (Drug Abuse Resistance Education) with latent variables in context of a Solomon Four Group Design. In J. S. Rose, L Chassin, (Eds.). *Multivariate applications in substance use research: New methods for new questions* (pp. 203-231). Mahwah, NJ: Erlbaum.
Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika, 48*, 465-471.
van Buuren, S. (1997). Fitting ARMA time series models by structural equation models. *Psychometrika, 62*, 215-236.
Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-69.
van den Oord, E. J. (2000). Framework for identifying quantitative trait loci in association studies using structural equation modeling. *Genetic Epidemiology, 18*, 341-359.
van den Oord E. J. (2001). Estimating effects of latent and measured genotypes in multilevel models. *Statistical Methods in Medical Research, 10*, 393-407.
van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology, 31*, 33-51.

van Duijn, M. A. J., van Busschbach, J. T., & Snijders, T. A. B. (1999). Multilevel analysis of personal networks as dependent variables. *Social Networks, 21*, 187-209.
van Praag, B. M. S., Dijkstra, T. K., & van Velzen, J. (1985). Least-squares theory based on general distributional assumptions with an application to the incomplete observations problem. *Psychometrika, 50*, 25-36.
van Yperen, N. W., & Snijders, T. A. B. (2000). A multilevel analysis of the demands-control model: Is stress at work determined by factors at the group level or the individual level? *Journal of Occupational Health Psychology, 5*, 182-190.
Vautier, S. (2001). *Assessing internal consistency reliability with Jöreskog's rho: New developments*. Under review.
Vautier, S., & Jmel, S. (2003). Transient error or specificity? An alternative to the staggered equivalent split-half. *Psychological Methods*.
Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods, 3*, 231-251.
Velicer, W. F., & Fava, J. L. (2003). Time series analysis. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research methods in psychology* (pp. 581-606). Hoboken, NJ: Wiley.
Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin, 122*, 231-249.
Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology, 48*, 865-885.
Vitaro, F., Brendgen, M., & Tremblay, R. E. (2001). Preventive intervention: Assessing its effects on the trajectories of delinquency and testing for mediational processes. *Applied Developmental Science, 5*, 201-213.
Volk, R. J., & Flori, D. E. (1996). Structural equation modeling. In D. H. Sprenkle, S. M. Moon, (Eds.). *Research methods in family therapy* (pp. 336-362). New York: The Guilford Press.
von Hippel, P. T. (2004). Biases in SPSS 12.0 missing value analysis. *The American Statistician, 58*, 160-164.
Vuksic-Mihaljevic, Z., Mandic, N., Bensic, M., & Mihaljevic, S. (2000). Posttraumatic stress disorder among Croatian veterans: A causal model. *Psychiatry and Clinical Neurosciences, 54*, 625-636.
Wainer, H., & Velleman, P. F. (2001). Statistical graphics: Mapping the pathways of science. *Annual Review of Psychology, 52*, 305-335.
Wainer, H., & Robinson, D. H. (2003). Shaping up the practice of null hypothesis significance testing. *Educational Researcher, 32*(7), 22-30.
Wakaki, H. (1997). Asymptotic expansion of the joint distribution of sample mean vector and sample covariance matrix from an elliptical population. *Hiroshima Mathematics Journal, 27*, 295-305.
Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society, 54*, 426-482.
Wall, M. M., & Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of the American Statistical Association, 95*, 929-940.
Wall, M. M., & Amemiya, Y. (2001). Generalized appended product indicator procedure for nonlinear structural equation analysis. *Journal of Educational and Behavioral Statistics, 26*, 1-29.
Ward, L. C., Axelrod, B. N., & Ryan, J. J. (2000). Observations on the factor structure of the WAIS-R. *Assessment, 7*, 79-86.
Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. *American Statistician, 39*, 95-101.
Waternaux, C. M. (1984). Principal components in the nonnormal case: The test of equality of q roots. *Journal of Multivariate Analysis, 14*, 323-335.
Watkins, K. W., Connell, C. M., Fitzgerald, J. T., Klem, L., Hickey, T., & Ingersoll-Dayton, B. (2000). Effect of adults' self-regulation of diabetes on quality-of-life outcomes. *Diabetes Care, 23*, 1511-1515.
Watson, D., Suls, J., & Haig, J. (2002). Global self-esteem in relation to structural models of personality and affectivity. *Journal of Personality and Social Psychology, 83*, 185-197.
Weakliem, D. L. (2004). Introduction to the special issue on model selection. *Sociological Methods & Research, 33*, 167-187.
Weary, G., & Reich, D. A. (2001). Attributional effects of conflicting chronic and temporary outcome expectancies: A case of automatic comparison and contrast. *Personality and Social Psychology Bulletin, 27*, 562-574.
Webb, N. M., Nemer, K. M., & Zuniga, S. (2003). Short circuits or superconductors? Effects of group composition on high-achieving students' science assessment performance. *American Educational Research Journal, 39*, 943-989.
Wegener, D. T., & Fabrigar, L. R. (2000). Analysis and design for nonexperimental data: Addressing causal and noncausal hypothesis. In H. T. Reis, C. M. Judd, (Eds.). *Handbook of research methods in social and personality psychology* (pp. 412-450). New York: Cambridge University Press.
Wen, Z., Marsh, H. W., & Hau, K. –T. (2002). Interaction effects in growth modeling: A full model. *Structural Equation Modeling, 9*, 20-39.
Wendorf, C. A. (2002). Comparisons of structural equation modeling and hierarchical linear modeling approaches to couples' data. *Structural Equation Modeling, 9*, 126-140.
Weng, L.-J., & Bentler, P. M. (1987). Linear structural equation modeling with dependent observations. *Proceedings of the Social Statistics Section, American Statistical Association*, 498-500.
Weng, L.-J., & Bentler, P. M. (1997). Covariance structure analysis with intraclass dependent observations. In W.-S. Yang (Ed.), *Recent development and application of quantitative methods in social sciences* (pp. 139-183). Taipei, Taiwan, R.O.C.: Sun Yat-Sen Institute for Social Sciences and Philosophy, Academia Sinica.

Weng, L. -J., & Cheng, C. P. (1997). Why might relative fit indices differ between estimators? *Structural Equation Modeling, 4*, 121-128.
Weng, L. -J., & Cheng, C. P. (2000). Effects of response order on Likert-type scales. *Educational & Psychological Measurement, 60*, 908-924.
Werts, C. E., Rock, D. A., & Grandy, J. (1979). Confirmatory factor analysis applications: Missing data problems and comparison of path models between populations. *Multivariate Behavioral Research, 14*, 199-213.
Werts, C. E., Rock, D. A., Linn, R. L., & Jöreskog, K. G. (1976). Comparison of correlations, variances, covariances, and regression weights with or without measurement error. *Psychological Bulletin, 83*, 1007-10013.
Werts, C. E., Rock, D. A., Linn, R. L., & Jöreskog, K. G. (1977). Validating psychometric assumptions within and between several populations. *Educational & Psychological Measurement, 37*, 863-872.
Werts, C. E., Rock, D. A., Linn, R. L., & Jöreskog, K. G. (1978). A general method of estimating the reliability of a composite. *Educational & Psychological Measurement, 38*, 933-938.
West, S. G. (2001). New approaches to missing data in psychological research: Introduction to the special section. *Psychological Methods, 6*, 315-316.
West, S. G., Biesanz, J. C., & Pitts, S. C. (2000). Causal inference and generalization in field settings: Experimental and quasi-experimental designs. In H. T. Reis, C. M. Judd, (Eds.). *Handbook of research methods in social and personality psychology* (pp.40-84). New York: Cambridge University Press.
Wharton, A. S., Rotolo, T., & Bird, S. R. (2000). Social context at work: A multilevel analysis of job satisfaction. *Sociological Forum, 15*, 65-90.
Wheaton, B. (1987). Assessment of fit in overidentified models with latent variables. *Sociological Methods & Research, 16*, 118-154.
Wheaton, B., Muthén, B., Alwin, D. F., & Summers, G. F. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.), *Sociological methodology 1977* (pp. 84-136). San Francisco: Jossey Bass.
White, K. S., & Farrell, A. D. (2001). Structure of anxiety symptoms in urban children: Competing factor models of the revised Children's Manifest Anxiety Scale. *Journal of Consulting and Clinical Psychology, 69*, 333-337.
Whiteman, M. C., Deary, I. J., & Fowkes, F. G. R. (2000). Personality and social predictors of atherosclerotic progression: Edinburgh Artery Study. *Psychosomatic Medicine, 62*, 703-714.
Whiteside-Mansell, L., & Bradley, R. H. (1996). Early home environment and mental test performance: A structural analysis. *Early Education & Development*, 7, 277-295.
Whiteside-Mansell, L., Bradley, R. H., Little, T. D., Corwyn, R. F., & Spiker, D. (2001). An examination of cross-racial comparability of mother-child interaction among African American and Anglo American families. *Journal of Marriage & Family, 63*, 767-778.
Wicherts, J. M., & Dolan, C. V. (2004). A cautionary note on the use of information fit indexes in covariance structure modeling with means. *Structural Equation Modeling, 11*, 45-50.
Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods, 8*, 16-37.
Wiggins, R. D., & Sacker, A. (2002). Strategies for handling missing data in SEM: A user's perspective. In G. A. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 105-120). Mahwah, NJ: Erlbaum.
Wilcox, R. R. (1995). ANOVA: A paradigm for low power and misleading measures of effect size? *Review of Educational Research, 65*, 51-77.
Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist, 53*, 300-314.
Wilcox, R. R. (2001). *Fundamentals of modern statistical methods.* New York: Springer.
Wilde, N. J., Strauss, E., Chelune, G. J., Hermann, B. P., Hunter, M., Loring, D. W., Martin, R. C., & Sherman E. M. S. (2003). Confirmatory factor analysis of the WMS-III inn patients with temporal lobe epilepsy. *Psychological Assessment, 15*, 56-63.
Wiley, D. E. (1973). The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 69-83). New York: Seminar.
Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116*, 363-381.
Williams, G. C., Gagné, M., Ryan, R. M., & Deci, E. L. (2002). Facilitating autonomous motivation for smoking cessation. *Health Psychology, 21*, 40-50.
Williams, L. J., Bozdogan, H., & Aiman-Smith, L. (1996). Inference problems with equivalent models. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 279-314). Mahwah, NJ: Erlbaum.
Williams, L. J., & Holahan, P. J. (1994). Parsimony-based fit indices for multiple indicator models: Do they work? *Structural Equation Modeling, 1*,161-187.
Wills, T. A., & Cleary, S. D. (2000). Testing theoretical models and frameworks in child health research. In D. Drotar, (Ed.). *Handbook of research in pediatric and clinical child psychology: Practical strategies and methods*, (pp. 21-49). New York: Kluwer AcademicPlenum Publishers.
Wills, T. A., Cleary, S., Filer, M., Shinar, O., Mariani, J., & Spera, K. (2001). Temperament related to early-onset substance use: Test of a developmental model. *Prevention Science, 2*, 145-163.
Wills, T. A., & Stoolmiller, M. (2002). The role of self-control in early escalation of substance use: A time-varying analysis. *Journal of Consulting and Clinical Psychology, 70*, 986-997.

Windle, M., & Dumenci, L. (1999). The factorial structure and construct validity of the Psychopathy Checklist--Revised (PCL--R) among alcoholic inpatients. *Structural Equation Modeling*, *6*, 372-393.
Windle, M., & Windle, R. C. (2001). Depressive symptoms and cigarette smoking among middle adolescents: Prospective associations and intrapersonal and interpersonal influences. *Journal of Consulting and Clinical Psychology*, *69*, 215-226.
Winship, C., & Radbill, L. (1994). Sampling weights and regression analysis. *Sociological Methods & Research*, *23*, 230-257.
Woeller, W., Kruse, J., Schmitz, N., & Richter, B. (1998). Prediction of high-risk illness behavior in patients with bronchial asthma. *Psychotherapie Psychosomatik Medizinische Psychologie*, *48*, 101-107.
Wolfe, D. A., Scott, K., Reitzel-Jaffe, D., Wekerle, C., Grasley, C., & Straatman, A. –L. (2001). Development and validation of the conflict in adolescent dating relationships inventory. *Psychological Assessment*, *13*, 277-293.
Wolfle, L. M. (2003). The introduction of path analysis to the social sciences, and some emergent themes: An annotated bibliography. *Structural Equation Modeling*, *10*, 1-34.
Wood, M. D., Read, J. P., Palfai, T. P., & Stevenson, J. F. (2001). Social influence processes and college student drinking: The mediational role of alcohol outcome expectancies. *Journal of Studies on Alcohol*, *62*, 32-43.
Wood, P. K. (2000). Estimation and equivalent models for quadratic and interactive latent variable models. In J. S. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance abuse research: New methods for new questions* (pp. 161-202). Mahwah, NJ: Erlbaum.
Wood, P. K., & Brown, D. (1994). The study of intraindividual differences by means of dynamic factor models: Rationale, implementation, and interpretation. *Psychological Bulletin*, *116*, 166-186.
Woodhouse, B., & Jackson, P. M. (1977). Lower bounds for the reliability of the total score on a test composed of nonhomogeneous items: II. A search procedure to locate the greatest lower bound. *Psychometrika, 42*, 579-591.
Woodward, J. A., & Bentler, P. M. (1978). A statistical lower-bound to population reliability. *Psychological Bulletin*, *85*, 1323-1326.
Wothke, W. (1996). Models for mulitrait-multimethod matrix analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 7-56). Mahwah, NJ: Erlbaum.
Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 219-240). Mahwah, NJ: Erlbaum.
Wothke, W., & Browne, M. W. (1990). The direct product model for the MTMM matrix parameterized as a second order factor analysis model. *Psychometrika, 55*, 255-262.
Wu, H., & Zhang, J. –T. (2002). Local polynomial mixed-effects models for longitudinal data. *Journal of the American Statistical Association, 97*, 883-897.
Xie, J., & Bentler, P. M. (2003). Covariance structure models for gene expression microarray data. *Structural Equation Modeling, 10*, 566-582.
Yalcin, I., & Amemiya, Y. (2001). Nonlinear factor analysis as a statistical method. *Statistical Science*, *16*, 275-294.
Yamaguchi, K., & Watanabe, M. (1993). The efficiency of estimation methods for covariance structure models under multivariate t distribution. *Proceedings of the Asian Conference on Statistical Computing*, 281-284. International Association for Statistical Computing.
Yang, M., Goldstein, H., Browne, W., and Woodhouse, G. (2002). Multivariate multilevel analyses of examination results. *Journal of the Royal Statistical Society, Series A*, *165*, 137-153.
Yang-Wallentin, F. (2001). Comparisons of the ML and TSLS estimators for the Kenny-Judd model. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 425-442). Lincolnwood, IL: Scientific Software International.
Yang-Wallentin, F., & Jöreskog, K. G. (2001). Robust standard errors and chi-squares for interaction models. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 159-172). Mahwah, NJ: Erlbaum.
Ybema, J. F., Kuijer, R. G., Buunk, B. P., DeJong, G. M., & Sanderman, R. (2001). Depression and perceptions of inequity among couples facing cancer. *Personality and Social Psychology Bulletin, 27*, 3-13.
Yeung, A. S., Marsh, H. W., & Suliman, R. (2000). Can two languages live in harmony: Analysis of the National Education Longitudinal Study of 1988 (NELS88) longitudinal data on the maintenance of home language. *American Educational Research Journal, 37*, 1001-1026.
Yuan, K. –H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research*, *40*, 115-148.
Yuan, K. –H., & Bentler, P. M. (1994). Test of linear trend in eigenvalues of k covariance matrices with applications in common principal components analysis. *Communications in Statistics- Theory and Methods*, *23*, 3141-3155.
Yuan, K. –H., & Bentler, P. M. (1996). Mean and covariance structure analysis with missing data. In A. Gupta & V. Girko (Eds.) *Multidimensional statistical analysis and theory of random matrices: Proceedings of Sixth Eugene Lukacs Symposium* (pp. 307-326). Utrecht: VSP.
Yuan, K. –H., & Bentler, P. M. (1997a). Finite sample distribution-free test statistics for nested structural models. *Behaviormetrika*, *24*, 19-26.
Yuan K. –H., & Bentler, P. M. (1997b). Generating multivariate distributions with specified marginal skewness and kurtosis. In W. Bandilla & F. Faulbaum (Eds.), *Softstat '97: Advances in statistical software 6* (pp. 385-391). Stuttgart: Lucius & Lucius.

Yuan, K. –H., & Bentler, P. M. (1997c). Improving parameter tests in covariance structure analysis. *Computational Statistics & Data Analysis*, 26, 177-198.
Yuan, K. –H., & Bentler, P. M. (1997d). Mean and covariance structure analysis: Theoretical and practical improvements. *Journal of the American Statistical Association*, *92*, 767-774.
Yuan, K. –H., & Bentler, P. M. (1998a). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology, 51*, 289-309.
Yuan, K. –H., & Bentler, P. M. (1998b). Robust mean and covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, *51*, 63-88.
Yuan, K. –H., & Bentler, P. M. (1998c). Structural equation modeling with robust covariances. In A. Raftery (Ed.), *Sociological methodology* (pp. 363-396). Malden, MA: Blackwell.
Yuan, K. –H., & Bentler, P. M. (1999a). F tests for mean and covariance structure analysis. *Journal of Educational and Behavioral Statistics*, *24*, 225-243.
Yuan, K. –H., & Bentler, P. M. (1999b). On asymptotic distributions of normal theory MLE in covariance structure analysis under some nonnormal distributions. *Statistics & Probability Letters, 42*, 107-113.
Yuan, K. –H., & Bentler, P. M. (1999c). On normal theory and associated test statistics in covariance structure analysis under two classes of nonnormal distributions. *Statistica Sinica*, *9*, 831-853.
Yuan, K. –H. & Bentler, P. M. (2000a). Inferences on correlation coefficients in some classes of nonnormal distributions. *Journal of Multivariate Analysis, 72*, 230-248.
Yuan, K. –H. & Bentler, P. M. (2000b). On equivariance and invariance of standard errors in three exploratory factor models. *Psychometrika, 65*, 121-133.
Yuan, K. –H. & Bentler, P. M. (2000c). Robust mean and covariance structure analysis through iteratively reweighted least squares. *Psychometrika, 65*, 43-58.
Yuan, K. –H., & Bentler, P. M. (2000d). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology 2000* (pp. 165-200). Washington, DC: American Sociological Association.
Yuan, K. –H., & Bentler, P. M. (2001a). Effect of outliers on estimators and tests in covariance structure analysis. *British Journal of Mathematical & Statistical Psychology, 54*, 161-175.
Yuan, K. –H., & Bentler, P. M. (2001b). A unified approach to multigroup structural equation modeling with nonstandard samples. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 35-56). Mahwah, NJ: Erlbaum.
Yuan, K. –H., & Bentler, P. M. (2002a). On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika, 67*, 251-259.
Yuan, K. –H., & Bentler, P. M. (2002b). On normal theory based inference for multilevel models with distributional violations. *Psychometrika, 67*, 539-561.
Yuan, K. –H., & Bentler, P. M. (2002c). On the asymptotic distributions of two statistics for two-level covariance structure models within the class of elliptical distributions. *Psychometrika*, in press.
Yuan, K. –H., & Bentler, P. M. (2002d; published 2004). On chi-square difference and *z*-tests in mean and covariance structure analysis when the base model is misspecified. *Educational & Psychological Measurement, 64*, 737-757.
Yuan, K. –H., & Bentler, P. M. (2003). Eight test statistics for multilevel structural equation models. *Computational Statistics & Data Analysis, 44*, 89-107.
Yuan, K. –H., & Bentler, P. M. (2004). *Asymptotic robustness of the normal theory likelihood ratio statistic for two-level covariance structure models.* Under review.
Yuan, K.-H., Bentler, P. M., & Chan, W. (2000). *Structural equation modeling with heavy tailed distributions*. Under review.
Yuan, K. –H., Bentler, P. M., & Kano, Y. (1997). On averaging variables in a confirmatory factor analysis model. *Behaviormetrika, 24*, 71-83.
Yuan, K. –H., & Bushman, B. J. (2002). Combining standardized mean differences using the method of maximum likelihood. *Psychometrika, 67*, 589-607.
Yuan, K. –H., & Chan, W. (2001). *On calculation of Bentler's comparative fit index.* Under review.
Yuan, K. –H., & Chan, W. (2002). Fitting structural equation models using estimating equations: A model segregation approach. .*British Journal of Mathematical and Statistical Psychology*, *55*, 41-62.
Yuan, K. –H., Chan, W., & Bentler, P. M. (2000). Robust transformation with applications to structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, *53*, 31-50.
Yuan, K. –H., Fung, W. K., & Reise, S. (in press; published 2004). Three Mahalanobis distances and their role in assessing unidimensionality. *British Journal of Mathematical and Statistical Psychology, 57*, 151-165.
Yuan, K. –H., Guarnaccia, C. A., & Hayslip, B. J. (2003). A study of the distribution of sample coefficient alpha with the Hopkins Symptom Checklist: Bootstrap versus asymptotics. *Educational & Psychological Measurement, 63*, 5-23.
Yuan, K. –H., & Hayashi, K. (2000). *On using an empirical Bayes covariance matrix in bootstrap approach to covariance structure analysis.* Under review.
Yuan, K. –H., & Hayashi, K. (2003). Bootstrap approach to inference and power analysis based on three statistics for covariance structure models. *British Journal of Mathematical and Statistical Psychology, 56*, 93-110.
Yuan, K. –H., Hayashi, K., & Bentler, P. M. (2005). *Normal theory likelihood ratio statistic for mean and covariance structure analysis under alternative hypothesis.* Under review.

Yuan, K. –H., & Jennrich, R. I. (1998). Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis, 65*, 245-260.
Yuan, K. –H., & Jennrich, R. I. (2000). Estimating equations with nuisance parameters: Theory and applications. *Annals of the Institute of Statistical Mathematics, 52*, 343-350.
Yuan, K. H., Lambert, P. L., & Fouladi, R. T. (2004). Mardia's multivariate kurtosis with missing data. *Multivariate Behavioral Research, 39*, 413-437.
Yuan, K. –H., & Marshall, L. L. (2004). A new measure of misfit for covariance structure models. *Behaviormetrika, 31*, 67-90.
Yuan, K. –H., Marshall, L. L., & Bentler, P. M. (2002). A unified approach to exploratory factor analysis with missing data, nonnormal data, and in the presence of outliers. *Psychometrika, 67*, 95-122.
Yuan, K. –H., Marshall, L. L., & Bentler, P. M. (2003). Assessing the effect of model misspecifications on parameter estimates in structural equation models. *Sociological Methodology, 33*, 241-265.
Yuan, K. –H., Marshall, L. L, & Weston, R. (2002). Cross-validation by downweighting influential cases in structural equation modelling. *British Journal of Mathematical and Statistical Psychology, 55*, 125-143.
Yung, Y. –F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika, 62*, 297-330.
Yung, Y. –F., & Bentler, P. M. (1994). Bootstrap-corrected ADF test statistics in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology, 47*, 63-84.
Yung, Y. –F., & Bentler, P. M. (1996a). Bootstrapping techniques in analysis of mean and covariance structures. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling techniques* (pp. 195-226). Hillsdale, NJ: LEA.
Yung, Y. –F., & Bentler, P. M. (1996b). On the hessian and information for ML factor analysis with mean and covariance structures. In F. Faulbaum & W. Bandilla (Eds.), *Softstat '95: Advances in statistical software 5* (pp. 211-218). Stuttgart: Lucius & Lucius.
Yung, Y. –F., & Bentler, P. M. (1999). On added information for ML factor analysis with mean and covariance structures. *Journal of Educational and Behavioral Statistics, 24*, 1-20.
Yung, Y. –F., Chan, W., & Bentler, P. M. (1998). Designing bootstrap tests using assumed information: Application to cross-classification. In W. B. G. Liebrand, A. Nowak, & R. Hegselmann (Eds.), *Computer modeling of social processes* (pp. 144-161). Thousand Oaks, CA: Sage.
Yuo, K.-H., & Wang, H.-C. (1999). Joint estimation of feature transformation parameters and Gaussian mixture model for speaker identification. *Speech Communication, 28*, 227-241.
Zaslow, M. J., Hair, E. C., Dion, A. R., Ahluwalia, S. K., & Sargent, J. (2001). Maternal depressive symptoms and low literacy as potential barriers to employment in a sample of families receiving welfare: Are there two-generational implications? *Women & Health, 32*, 211-251.
Zhu, H.-T., & Lee, S.-Y. (1999). Statistical analysis of nonlinear factor analysis models. *British Journal of Mathematical and Statistical Psychology, 52*, 225-242.
Zhu, L. –X, Fang, K. –T., Bhatti, M. I., & Bentler, P. M. (1995). Testing sphericity of a high-dimensional distribution based on bootstrap approximation. *Pakistan Journal of Statistics, 11*, 49-65.
Zhou, Q. P., O'Brien, B., & Soeken, K. (2001). Rhodes Index of Nausea and Vomiting - Form 2 in pregnant women: A confirmatory factor analysis. *Nursing Research, 50*, 251-257.
Zimmerman, D. W., & Williams, R. H. (2000). Restriction of range and correlation in outlier-prone distributions. *Applied Psychological Measurement, 24*, 267-280.
Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_H: Their relations with each other and two alternate conceptualizations of reliability. *Psychometrika, 70*, 1-11.
Zuroff, D. C., & Duncan, N. (1999). Self-criticism and conflict resolution in romantic couples. *Canadian Journal of Behavioural Science, 31*, 137-149.