Annu. Rev. Psychol. 1996. 47:563–92

COVARIANCE STRUCTURE ANALYSIS: Statistical Practice, Theory, and Directions

Peter M. Bentler

Department of Psychology, University of California, Los Angeles, Box 951563, Los Angeles, California 90095-1563

Paul Dudgeon

Department of Psychology, University of Melbourne, and National Health & Medical Research Council Schizophrenia Research Unit, Mental Health Research Institute, Royal Park Hospital, Parkville, Victoria 3052 Australia

KEY WORDS: multivariate analysis, structural modeling, latent variables, LISREL, EQS

ABSTRACT

Although covariance structure analysis is used increasingly to analyze nonexperimental data, important statistical requirements for its proper use are frequently ignored. Valid conclusions about the adequacy of a model as an acceptable representation of data, which are based on goodness-of-fit test statistics and standard errors of parameter estimates, rely on the model estimation procedure being appropriate for the data. Using analogies to linear regression and anova, this review examines conditions under which conclusions drawn from various estimation methods will be correct and the consequences of ignoring these conditions. A distinction is made between estimation methods that are either correctly or incorrectly specified for the distribution of data being analyzed, and it is shown that valid conclusions are possible even under misspecification. A brief example illustrates the ideas. Internet access is given to a computer code for several methods that are not available in programs such as EQS or LISREL.

0066-4308/96/0201-0563$08.00

CONTENTS

INTRODUCTION

Most psychological data are multivariate in nature. An important approach to understanding such data is to develop and evaluate a model of how the data might have been generated. In the case of experiments, the explanatory variables are design variables whose values are controlled. By their nature, these variables are presumed to be understood. The dependent variables, on the other hand, typically are best understood with help of a model. If interest lies primarily in the means of these variables, the standard linear model and its statistical implementation via analysis of variance (anova) or its multivariate version provide good insight. Of course, assumptions need to be made, such as independence of observations, linearity and additivity of effects, homogeneity of variances, normally distributed errors, etc. There is substantial agreement on the performance characteristics of these methods when the assumptions are met, as well as, to a lesser extent, on the consequences of violation of assumptions. The same cannot be said for methods in the analysis of nonexperimental data. This chapter addresses some of the consequences of violation of assumptions in covariance structure analysis, and relates a few results to the comparable situation from anova or regression. Although most of the results reviewed here are very old, they have not yet permeated the practice of covariance structure analysis.

Nonexperimental data are inherently more difficult to analyze and understand because various variables may have different effects and directions of influence, their effects may not be independent, observed variables may be influenced by unmeasured latent variables, omitted variables may bias the observed effects, and so on. To understand such influences, typically one considers a general linear structural model for a p-variate vector of variables $\underset{\sim}{x}$

as $\underset{\sim}{x} = \mathbf{A}\xi$, where the matrix $\mathbf{A} = \mathbf{A}(\gamma)$ is a function of a basic vector γ of parameters, and the underlying k ($k \geq p$) generating variables ξ may represent measured, latent, or residual variables (e.g. Anderson 1994, Bentler 1983a, Satorra 1992). In typical applications of anova, one can assume the model is correct, and the statistical problem is to isolate true nonzero (i.e. significant) effects from zero effects. In contrast, in nonexperimental contexts the basic model setup itself may be a source of contention. That is, the matrix **A** or its parameters γ may be misspecified; an inappropriate or incomplete set of variables ξ may be hypothesized; or the moments, i.e. means or covariances, of these ξ variables may be incorrectly specified. Hence an important part of the methodology involves evaluating the quality of the model as a representation of the data. The standard questions of parameter significance are important only if the model itself is plausible. In this chapter we review issues in model and parameter evaluation, though we accept the basic model setup, which, of course, in some contexts may itself be questioned. For example, one could question the linearity assumption or the absence of nonlinear or interaction terms (e.g. Jöreskog & Yang 1995, Kenny & Judd 1984, Mooijaart & Bentler 1986), though the most popular model variants assume simple linear relations among variables. Nonlinear relations arise naturally with categorical data models (e.g. Jöreskog 1994, Lee et al 1992, 1995, Muthén 1984), but for simplicity we deal only with continuous variable models. Other interesting questions, such as causality (see e.g. Bullock et al 1994, Sobel 1995, Steyer 1993), equivalent models (e.g. Bekker et al 1994, Lee & Hershberger 1990, MacCallum et al 1993), or model modification (e.g. MacCallum et al 1992) also are not addressed.

As just noted, under a linear structural model, understanding the observed variables $\underset{\sim}{x}$ hinges on understanding the parameters γ and the generating variables ξ. In practice, one is satisfied with knowing how the means and the covariances (i.e. variances and correlations) among the $\underset{\sim}{x}$ variables are generated. Under the model, this requires estimating and testing for significance the parameters γ and the means and covariances of the generating ξ variables. That is, if the means and covariances among the $\underset{\sim}{x}$ variables are given by μ and Σ, an appropriate model would be based on a more basic set of parameters θ, such that $\mu = \mu(\theta)$ and $\Sigma = \Sigma(\theta)$. The q parameters in θ represent elements of γ as well as the intercepts, regression coefficients, or variances and covariances of the ξ variables. Specific versions of such models are given by the equations of the confirmatory factor analysis (Jöreskog 1969, Lockhart 1967), factor analytic simultaneous equation (Jöreskog & Sörbom 1993, Wiley 1973), Bentler-Weeks (1980), or RAM (McArdle & McDonald 1984) models. These models have a wide range of application, from individual growth modeling (Willett & Sayer 1994) to decomposition of trait and method variance (e.g. Dudgeon 1994). They can be taken as mean and covariance structure models (e.g.

Browne & Arminger 1995, Satorra 1992), though more general models, in which higher-order moments are also of concern, have been developed (Bentler 1983a) but are only rarely studied (Mooijaart 1985, Mooijaart & Bentler 1986).

In this review we concentrate on the most typical applications, those of covariance structure models. In such models, μ is unstructured and hence can be estimated (in practice, at the sample mean), which allows the parameters of the covariance structure, $\Sigma = \Sigma(\theta)$, to be treated separately. Covariance structure models have become extremely popular in psychology and other social sciences since the previous *Annual Review* chapter on this topic (Bentler 1980). Widely known computer programs, such as LISREL (Jöreskog & Sörbom 1993), EQS (Bentler 1995, Bentler & Wu 1995a,b), and others (see Browne & Arminger 1995, pp. 241–42; Ullman 1995), have made the models easily accessible to applied researchers, and good general introductory texts on the topic now exist (e.g. Bollen 1989, Byrne 1994, Dunn et al 1993, Hoyle 1995, Loehlin 1992). Steiger (1994) and Faulbaum & Bentler (1994) provide perspective overviews. A new journal, *Structural Equation Modeling*, covers recent developments.

As noted above, covariance structure models are typically motivated by linear models in hypothesized variables ξ. The distribution of the ξ variables affects the distributions of the measured variables $\underset{\sim}{x}$. Typically, one assumes that ξ and hence $\underset{\sim}{x}$ are multivariate normally distributed. This assumption simplifies the statistical theory. A test of the model structure and of hypotheses on particular parameters thus are easy to obtain. However, in practice, the normality assumption will often be incorrect. For example, Micceri (1989) reported that among 440 large-sample achievement and psychometric measures taken from journal articles, research projects, and tests, all were significantly nonnormally distributed. Yet, as noted by Breckler (1990) and Gierl & Mulvenon (1995), practitioners generally do not bother to evaluate this very strong assumption and simply accept normal theory statistics as if the data were normal. As a result, conclusions that are drawn about model adequacy (from the goodness-of-fit test statistic) and parameters (from *z*-statistics based on standard errors) are often liable to be incorrect as well. This is an alarming state of affairs in view of the increasing reliance on covariance structure models for understanding relationships among nonexperimental data.

It is not that there are no alternatives to normal theory statistics. Several have been developed and have been available for some time in certain computer programs, especially in EQS. Others have been developed but are not yet available to applied researchers. In general, software publishers have fallen substantially behind the theoretical developments. For example, a distribution-free test [see Equation (10) below] developed by Browne about 15 years ago (Browne 1982, 1984) is not available in any extant computer program, includ-

ing Browne's own program RAMONA (Browne et al 1994). Similarly, a test based on heterogeneous kurtosis theory was published a half decade ago (Kano et al 1990), yet it has not been incorporated into any programs, including Bentler's EQS program. Several other valuable statistics are similarly unavailable. Therefore, in addition to reviewing and discussing the alternatives, we provide a high-level code that can be accessed via the Appendix to implement some of the newer statistics.

THE MODELING PROCESS

Two aspects of the modeling process are important in understanding judgments about the adequacy of models and the significance of parameters. One, as touched on already, is the distributional assumptions made about the variables forming the model—this is the main focus of the review. The other is more fundamental to the process itself and has to do with models being approximate rather than exact representations. We examine this aspect briefly now because it places the first in a broader context. Consider the most popular covariance structure model, the confirmatory factor model. In this model, $\underset{\sim}{x} = \Lambda\xi + \epsilon$ explains the measured variables $\underset{\sim}{x}$ as a linear combination with weights Λ of common factors ξ and unique variates ("errors") ϵ. Factors are allowed to correlate with covariance matrix $\mathcal{E}(\xi\xi') = \Phi$, errors are uncorrelated with factors, i.e. $\mathcal{E}(\xi\epsilon') = 0$, and various error variates are uncorrelated and have a diagonal covariance matrix $\mathcal{E}(\epsilon\epsilon') = \Psi$. As a result, $\Sigma = \Sigma(\theta) = \Lambda\Phi\Lambda' + \Psi$, and the elements of θ are the unknown free parameters in the Λ, Φ, and Ψ matrices. The distribution of the variables ξ and ϵ is an important part of the specification of the model, but these are typically unknown and only the distribution of the measured variables $\underset{\sim}{x}$ is available for evaluation. As shown below, misspecification of the distribution of $\underset{\sim}{x}$ affects inferences on the null hypothesis $\Sigma = \Sigma(\theta)$ as well as on θ. Although assuring that the parameters in θ are identified is not a minor matter, to avoid getting sidetracked on this important problem we assume that uniqueness of parameter specification is not an issue so that for two possibly different parameter vectors θ_1 and θ_2, $\Sigma(\theta_1) = \Sigma(\theta_2) \Rightarrow \theta_1 = \theta_2$. Thus equality of the two covariance matrices implies equality of the parameters that generate them.

Let **S** represent the $p \times p$ sample covariance matrix obtained from $\underset{\sim}{x}'$ $(=x_1,...,x_p)$ variables, each independently observed $N = n + 1$ times. There are situations in which the assumption of independent observations is implausible, in which case special methods are needed (see e.g. Lee 1990, Muthén & Satorra 1989, Weng & Bentler 1995). However, independence generally can be reasonably assumed, and we can estimate the values of the parameters in θ from **S** and in some circumstances test the fit of the model $\Sigma(\theta)$, by minimizing some scalar function $F = F[\mathbf{S}, \Sigma(\theta)]$ which indicates the discrepancy between **S**

and the covariance matrix $\Sigma(\theta)$ reproduced from the fitted model. Discrepancy functions have the following properties: (*a*) the value of *F* will be greater than or equal to zero; (*b*) *F* will only equal zero if $\Sigma(\theta) = \mathbf{S}$; and (*c*) F must be twice differentiable with respect to both **S** and $\Sigma(\theta)$. The parameter estimates, signified by $\hat{\theta}$, are obtained at the minimum of *F*, signified by $\hat{F} = F[\mathbf{S}, \Sigma(\hat{\theta})]$, where the matrix $\Sigma(\hat{\theta})$ (or more conveniently signified by $\hat{\Sigma}$) indicates the covariance matrix reconstructed from the estimated parameters of the specified model. Normal theory maximum likelihood (ML) and generalized least squares (GLS) provide typical examples of discrepancy functions. For now, it is not important to give the full expression for these functions; they are given later in the paper.

A primary aim of covariance structure analysis is to specify enough restrictions in $\Sigma(\theta)$ so that, substantively, it becomes a sufficiently simple and acceptable representation for the theoretical or interpretative issue being investigated (McDonald 1989). Technically, also, the model should improve precision, i.e. reduce variance in the parameter estimator, at the expense of little or no bias in the estimator (de Leeuw 1988, Kano et al 1993). Even in the implausible situation in which we know the population covariance matrix Σ, say Σ_o, the onus would still be on us to specify some simplifying model $\Sigma(\theta_o)$ for representing the relationships in that matrix, otherwise "there is no point in using a model" (McDonald 1989, p. 101). Although in reality we employ a sample covariance matrix **S** as a consistent estimator of Σ_o, it is useful to consider briefly this implausible situation, for it helps make the problem of choosing a discrepancy function clear. Ideally, we would like to define a unique set of parameter values θ_o for our structural model such that

$$\Sigma(\theta_o) = \Sigma_o. \tag{1}$$

This implies that for our known population covariance matrix we also (implausibly) know the true model that generated the covariance matrix. In this instance all overidentifying restrictions in the model are correctly specified, and as a consequence these restrictions (and therefore the model) hold exactly. If we fit this model to Σ_o, then at the minimum of the ML, GLS, or any other function meeting the three requirements just defined, we would find that

$$F\left[\Sigma_o, \Sigma\left(\hat{\theta}_o\right)\right] = 0 \tag{2}$$

and therefore $\hat{\theta}_o = \theta_o$.

In reality we will never know $\Sigma(\theta_o)$. But let us still continue to assume that we know Σ_o. Then in these circumstances the results in Equation (2) will no longer necessarily hold. Although we could fit some hypothesized model $\Sigma(\theta_\phi)$ with different parameters θ_ϕ by minimizing some function $F[\Sigma_o, \Sigma(\theta_\phi)]$, the conjectured true model $\Sigma(\theta_\phi)$ would not necessarily equal the actual true

$\Sigma(\theta_o)$. While we still regard the set of true parameter values θ_o as the value of θ_ϕ acquired at the minimum, $\Sigma(\theta_\phi)$ would in all probability be an approximation to $\Sigma(\theta_o)$. Because the ideal defined by (1) no longer necessarily occurs, we would now find that

$$F\left[\Sigma_o, \Sigma\left(\hat{\theta}_\phi\right)\right] \geq 0 \tag{3}$$

although this relationship would almost invariably be a strict inequality. More importantly, we would also find that the values of both the (approximated) true parameters in $\hat{\theta}_\phi$ and the minimum discrepancy function value $F[\Sigma_o, \Sigma(\hat{\theta}_\phi)]$ would vary according to the particular discrepancy function used.[1]

In practice we estimate our models from the sample covariance matrix S rather than fit them to Σ_o, and this introduces additional discrepancies besides that arising from (3). Cudeck & Henly (1991) provide a very good discussion of the various forms of discrepancy involved in fitting models (see also Browne & Cudeck 1993 for ways to evaluate those discrepancies). While the details of Cudeck and Henly's paper are outside our present scope, the important point to be made here about the modeling process is that the choice of a discrepancy function will influence the assessment of models and parameter estimates not only because we work with sample data that are often nonnormal, but also because the null hypothesis (1) never holds. As such, our models are only approximate rather than exact representations of the reality being envisaged. A statistical basis for making an appropriate choice is therefore needed.

Test Statistic on Model Hypothesis

If we make the right choice of discrepancy function, and if the modeling assumptions are correct and the sample size is large enough, then at the minimum, $T = n\hat{F}$ is distributed under the null hypothesis (1) as a goodness-of-fit χ^2 variate with $(p^* - q)$ degrees of freedom,[2] where $p^* = p(p + 1)/2$. T can be used as a test statistic to evaluate the null hypothesis. The null hypothesis is rejected if T exceeds a critical value in the χ^2 distribution at an α-level of

[1] This can be shown easily by fitting any model in LISREL or EQS and saving the reproduced covariance matrix (it does not have to be a particularly good fit). If we then refit the same model, but now to the reproduced covariance matrix rather than to the sample covariance matrix, then the fitting function value will always be zero and the parameter estimates and standard errors will be the same for, say, ML or GLS estimation. If we change the model slightly to some different specification of parameters, but still use the reproduced covariance matrix, then the fitting function value, the parameter estimates, and the standard errors will vary according to ML or GLS estimation.

[2] More generally, the degrees of freedom are further increased by one for each independent equality restriction that might be imposed (Lee & Bentler 1980), but for simplicity in this paper we assume that there are no equality restrictions.

significance. Otherwise, the model cannot be rejected, and the null hypothesis is accepted.

Because a numerical value for T is computed and printed out by all computer programs, there is a strong tendency to treat it as a χ^2 variate whether or not that is its actual distribution. In fact, except for unusual circumstances associated with the specialized "asymptotic robustness theory" (see below), when T is based on the assumption of multivariate normality of variables but the data are not normal (the typical case in practice as noted above), T will not be χ^2 distributed. As a result, incorrect conclusions about model adequacy often are obtained. And, as shown by Hu et al (1992), asymptotic robustness theory is not robust to violation of its assumptions, so it cannot be used to justify an inappropriate choice of fit function and test statistic.

Statistics on Parameter Hypotheses

Once a model null hypothesis is accepted, typical practice involves interpreting the relative size and significance levels of particular parameter estimates $\hat{\theta}_i$ to see if they differ significantly from zero. Typically, this involves evaluating the hypothesis $\theta_i = 0$ using the statistic $Z = (\hat{\theta}_i - \theta_i) / \mathrm{SE}(\hat{\theta}_i)$, where the denominator is an estimate of the standard error. In practice, computer programs calculate this standard error estimate from the square root of the appropriate element from the inverse of the "information matrix." Unfortunately, aside from tests on regression coefficients that may be correct due to asymptotic robustness theory—which cannot be relied upon to apply to a given data analysis situation—this is the correct expression only when the distributional assumption used in defining the discrepancy function is correct. Thus, tests of parameters based on z will be incorrect in the typical case where a normal theory method is used, but the data are not normal. As a result, incorrect substantive conclusions about the meaning of a model may well be drawn. The situation is the same when sets of parameters are evaluated simultaneously using the Wald test (e.g. Bentler & Dijkstra 1985; Dijkstra 1981, Theorem 8; Lee 1985). Similar problems occur when missing parameters are evaluated using the Lagrange Multiplier test (e.g. Bentler 1995). Satorra (1989) provides an excellent general discussion.

To gain some insight into this situation, consider the regression model $\underset{\sim}{y} = \mathbf{X}\beta + \epsilon$ with dependent variable $\underset{\sim}{y}$ and fixed design matrix $\mathbf{X}$. When the errors ϵ are independent, normal, and homoscedastic, as in typical applications of anova, the information matrix of $\hat{\beta}$ is proportional to $(\mathbf{X}'\mathbf{X})$. Thus the standard errors of $\hat{\beta}$ are given, up to a constant that involves sample size, by the square roots of the diagonal elements of the inverse $(\mathbf{X}'\mathbf{X})^{-1}$ of this information matrix. These are the standard errors given by available regression and anova programs. Unfortunately, when the assumptions are not true, this formula does not give the correct standard errors (see, e.g. Arminger 1995). It is likely that

many applications of regression thus give incorrect tests on parameters, as do many covariance structure applications. We shall use the regression analogy several times.

Theoretical or Empirical Robustness to Violation of Assumptions

We see, then, that the typical requirements for a covariance structure statistic to be trustworthy under the null hypothesis $\Sigma = \Sigma(\theta)$ are that the parameters of the model are identified; the observations or scores for different subjects are independent; the sample size is very large; and either a discrepancy function consistent with the actual distribution of variables is optimized, or a method of model and parameter testing is chosen that is robust to violation of certain of these assumptions. Unfortunately, these several conditions are often difficult to meet in practice. As noted above, we do not discuss identification or independence; we simply assume these conditions since they can often be arranged by design. The other points bear some discussion prior to developing the technical details.

Sample size turns out to be critical because all of the statistics known in covariance structure analysis are "asymptotic," that is, are based on the assumption that N becomes arbitrarily large. Since this situation can rarely be obtained, except perhaps by large national testing services and censuses, it becomes important to evaluate whether N may be large enough in practice for the theory to work reasonably well. Different data and discrepancy functions have different robustness properties with respect to sample size. Basically, sample size requirements increase as data become more nonnormal, models become larger, and more assumption-free discrepancy functions are used (e.g. Chan et al 1995, Chou et al 1991, Curran et al 1994, Hu et al 1992, Muthén & Kaplan 1992, West et al 1995, Yung & Bentler 1994).

In principle, if one matches a discrepancy function to the distribution of variables, the resulting T and z test statistics should be well behaved. However, some of these functions cannot be applied with large models because the computational demands are simply too heavy. Also, some provide test statistics that do not work well except at impractically large sample sizes. Thus alternatives have been developed that may potentially work in more realistic sized samples, i.e. be robust to violation of the asymptotic sample size requirement of all known methods. Unfortunately, not much is known about their actual performance in practice. These topics are discussed below.

Ideally, one could specify conditions under which even the technically wrong method could lead to correct statistical inferences. This is the hope of researchers who use normal theory methods when their data are nonnormal. The only known theoretical justification for such a practice is that of asymptotic robustness (e.g. Browne 1987), which we illustrate but do not review in

detail. Anderson & Amemiya (1988) and Amemiya & Anderson (1990) found, for example, that the asymptotic χ^2 goodness-of-fit test in factor analysis can be insensitive to violations of the assumption of multivariate normality of both common and unique factors, if all factors are independently distributed and the elements of the covariance matrices of common factors are all free parameters. With an additional condition of the existence of the fourth-order moments of both unique and common factors, Browne & Shapiro (1988) and Mooijaart & Bentler (1991) also demonstrated the robustness of normal theory methods in the analysis of a general class of linear latent variate models. Satorra & Bentler (1990, 1991) obtained similar results for a wider range of discrepancy functions, estimators, and test statistics. Browne (1990) and Satorra (1992) extended this theory to mean and covariance structure models, and Satorra (1993) to multiple samples. Unfortunately, asymptotic robustness theory cannot be relied upon in practice, because it is practically impossible to evaluate whether its conditions are met. Thus we cannot use this theory to avoid taking a more detailed look at the statistics of covariance structure analysis, to which we now turn.

STATISTICS BASED ON CORRECTLY SPECIFIED DISTRIBUTIONS

Ideally, there would be many classes of multivariate distributions that are realistic and practical models for data analysis, but this is not the case (Olkin 1994). In covariance structure analysis, only three types of specific distributions have been considered: multivariate normal, elliptical, and heterogeneous kurtotic. In this section we provide basic definitions for these cases and define the test statistics and parameter estimator covariance matrices that result from the correct specification of distributional forms.

To simplify matters, we note that the distribution of the data induces a distribution of the sample statistics under consideration in covariance structure analysis, namely the distribution of the elements of the sample covariance matrix **S**. Hence, we can focus on the distribution of **S** instead of, or in addition to, the distribution of the raw data. Since **S** contains redundant elements, we need only be concerned with the nonduplicated elements. Let $\underset{\sim}{s}$ and $\sigma(\theta)$ be p^* $\times$ 1 column vectors formed from the nonduplicated elements of **S** and $\Sigma(\theta)$, respectively. We are interested in the asymptotic distribution of $\sqrt{n}[\underset{\sim}{s} - \sigma(\theta)]$. We shall assume that typical regularity conditions hold and that the model is correct, so that asymptotically $\sqrt{n}[\underset{\sim}{s} - \sigma(\theta)]$ is multivariate normally distributed with a mean of zero and a covariance matrix given by

$$\operatorname{acov}\left\{\sqrt{n}\left[\underset{\sim}{s} - \sigma(\theta)\right]\right\} = \Gamma \tag{4}$$

The notation "acov" means asymptotic covariance, i.e. as *n* becomes arbitrarily large. It implies that the covariance matrix of the data $\underset{\sim}{s}$, based on a sample of size *N*, is given by Γ/n, where the divisor reflects the typical reduction of variances with increasing sample size. Now the specific form of Γ, that is, the detailed mathematical expressions for the elements of this matrix, depends upon the distribution of the variables $\underset{\sim}{x}$ that are being modeled. Let us abstractly consider these matrices to be given by $\Gamma_{N,}$ Γ_E, and Γ_{HK} for normal, elliptical, and heterogeneous kurtotic distributions. Explicit expressions are given below.

Note that any specific discrepancy function $F[\mathbf{S}, \Sigma(\theta)]$ applied to a covariance structure model is associated with two matrices:

1. $\hat{\mathbf{W}}$, which is a consistent and unbiased estimator of some population weight matrix $\mathbf{W}$ having the property that, except possibly for a constant, the matrix of expected values of the second derivatives is given by

$$\mathcal{E}\left(\frac{\partial^2 F}{\partial\theta\partial\theta'_{|\theta=\sigma}}\right) = \mathbf{W}. \tag{5}$$

This could be called the information matrix (adopting a standard usage from maximum likelihood theory) for a saturated model. $\mathbf{W}$ is fixed by the estimation method chosen, that is, the specific discrepancy function $F[\mathbf{S}, \Sigma(\theta)]$ to be optimized. Although data may be used to estimate $\mathbf{W}$, via $\hat{\mathbf{W}}$, this matrix does not necessarily depend on the actual distribution of the data variables $\underset{\sim}{x}$, which may be different from that assumed.

2. Γ, the true covariance matrix of the sample covariances, which depends on the actual distribution of the data. Under specific distributions, this can be denoted as Γ_N,Γ_E, or Γ_{HK}, depending on the actual distribution of the variables that generates the sample $\underset{\sim}{s}$. Although there is a single true (typically unknown) Γ, different choices of estimators (e.g. $\hat{\Gamma}_N$) imply different discrepancy functions.

As a result, to be more precise, we shall now define the particular discrepancy function chosen for analysis as $F[(\mathbf{S}, \Sigma(\theta)) \mid \mathbf{W},\Gamma]$. With this notation, following Browne (1984) we can define a discrepancy function for a correctly specified distribution as one in which $\mathbf{W} = \Gamma^{-1}$, i.e. the class of functions $F[(\mathbf{S}, \Sigma(\theta)) \mid \mathbf{W} = \Gamma^{-1},\Gamma]$. These functions are called asymptotically optimal by Satorra (1989). For all such functions and data, test statistics have a simple form.

If we estimate θ so as to minimize $F[(\mathbf{S}, \Sigma(\theta)) \mid \mathbf{W} = \Gamma^{-1},\Gamma]$, at the minimum we have $F[(\mathbf{S}, \Sigma(\hat{\theta})) \mid \mathbf{W} = \Gamma^{-1},\Gamma]$. For such correctly specified discrepancy functions, if the sample size is large enough, under the model hypothesis we have

$$T = n\hat{F} = nF\left[\left(\mathbf{S}, \Sigma\left(\hat{\theta}\right)\right) \mid \mathbf{W} = \Gamma^{-1}, \Gamma\right] \sim \chi^2_{(p^*-q)}. \tag{6}$$

Thus there exists a simple test of the model. In addition, the estimators $\hat{\theta}$ are asymptotically efficient, i.e. have the smallest possible sampling variances among estimators using the same information from the data. The covariance matrix of $\hat{\theta}$ is given by the inverse of the optimal information matrix (adopting this name from ML theory), namely

$$\text{acov}\left(\hat{\theta}\right) = \left[\mathcal{E}\left(\frac{\partial^2 F}{\partial\theta\partial\theta'_{|\theta=\theta_o}}\right)\right]^{-1} = n^{-1}\left(\Delta'\Gamma^{-1}\Delta\right)^{-1}, \tag{7}$$

where $\Delta = (\partial\sigma(\theta)/\partial\theta')_{|\theta=\theta_o}$ is the matrix of partial derivatives of the model with respect to the parameters. Standard errors are then the square roots of the diagonal elements of (7). In practice, consistent estimators of the matrices in (7) are used.

To make (7) a bit more intuitive, consider again the linear regression model. For that model, the covariance matrix of the residual ϵ, up to a constant, is $\Gamma = \mathbf{I}$, and with $\Delta = \mathbf{X}$, the matrix in (7) is proportional to $(\mathbf{X}'\mathbf{X})^{-1}$. This is the usual result, but it also holds more generally. Suppose in regression that the covariance matrix of the ϵ is Γ, not $\mathbf{I}$, and that GLS with weight matrix Γ^{-1} is used rather than least-squares estimation. Then the covariance matrix of the estimator is (7), i.e. proportional to $(\mathbf{X}'\Gamma^{-1}\mathbf{X})^{-1}$. However, as we shall see, when the distribution of variables is misspecified, (7) does not give the covariance matrix of $\hat{\theta}$. Unfortunately, in practice, researchers seem to use Equation (7) whether or not it is the correct formula to use.

STATISTICS BASED ON MISSPECIFIED DISTRIBUTIONS

Now we consider the more general case, in which the discrepancy function used in an analysis is misspecified, yet we desire to compute correct statistics. We define a misspecified function as one in which $W \neq \Gamma^{-1}$, i.e. the class of functions $F[(\mathbf{S},\Sigma(\theta)) \mid \mathbf{W} \neq \Gamma^{-1},\Gamma]$. Perhaps the most typical example is one in which $\mathbf{W} = \Gamma_N^{-1}$, but $\Gamma \neq \Gamma_N$. That is, a normal theory method is used, but the data are not normally distributed. In such a case, (6) and (7) do not hold. Specifically, T is not χ^2 distributed, and the matrix (7) is not relevant nor computed. It also means that the estimator generally is not asymptotically efficient, i.e. it will not have the smallest possible sampling variability. This might be an argument for using discrepancy functions that are correctly specified, but this may be impractical. As noted by Bentler & Dijkstra (1985), Dijkstra (1981), Shapiro (1983), and especially by Satorra & Bentler (1986,

1988, 1994), the general distribution of T is in fact not χ^2, but rather a mixture

$$T \xrightarrow{\mathcal{L}} \Sigma_1^{df} \alpha_i \tau_i, \tag{8}$$

where α_i is one of the *df* (degrees of freedom) nonnull eigenvalues of the matrix $\mathbf{U}\Gamma$, τ_i is one of the *df* independent χ_1^2 variates, and, when there are no constraints on free parameters,

$$\mathrm{U} = \mathbf{W} - \mathbf{W}\Delta(\Delta'\mathbf{W}\Delta)^{-1}\Delta'\mathbf{W} \tag{9}$$

is the residual weight matrix under the model and the weight matrix $\mathbf{W}$ used in the estimation.[3] Even though the distribution (8) has been known for over a decade, its usage has been considered impractical, and to our knowledge, the test statistic (8) was first used in covariance structure analysis by Bentler (1994). It is not available in any extant program.

Another test statistic that should hold generally, yet has not become available in any program, is the general quadratic form test statistic of Browne (1984, Proposition 4, Equation 2.20). Unlike tests based on T (see Equation 6), the test statistic

$$T_{\mathrm{QF}} = n\hat{F}_{\mathrm{QF}} = n\left[\mathbf{s} - \sigma\left(\hat{\theta}\right)\right]' \hat{\mathbf{U}}_\Gamma \left[\mathbf{s} - \sigma\left(\hat{\theta}\right)\right], \tag{10}$$

is $\chi^2_{p^*-q}$ distributed. Here, $\hat{\mathbf{U}}_\Gamma$ is given by the matrix defined in (9), and $\mathbf{W} = \Gamma^{-1}$ is based on the asymptotically distribution free (ADF—see Equation 14 below) estimated weight matrix. This test statistic can be used without any assumption that the matrix $\hat{\mathbf{W}}$ used in a minimum discrepancy function (see Equation 5) has been correctly specified. Browne (1984, p. 82) noted that although Equation (10) is theoretically correct, it lacked empirical investigation. Remarkably, this is still true today. Its chief appeal lies in it enabling the more tractable ML or GLS estimation methods to be employed for obtaining the parameter estimates. Whether (10) suffers from the problems of poor performance in small samples, like the ADF test statistic (see below), is unknown but is certainly a possibility. As noted by Browne (1984, p. 70), Bentler's (1983b) linearized ADF estimator yields a χ^2 test that is an alternative to (10). It is the default ADF method and statistic in EQS.

Although not strictly relevant to this section, we should note that a version of the quadratic form test statistic based on that of Browne (1984) was developed by Bentler for use with normal theory least-squares estimation. Since the

[3] The arrow in Equation 8 indicates that as sample size increases indefinitely, the distribution of T becomes equivalent to the distribution of the right-hand side term..

typical test statistic (6) is not available, he used a variant of (10) based on normal theory, namely, where

$$T_{\mathrm{NQF}} = n\hat{F}_{\mathrm{NQF}} = n\left[\mathrm{s} - \sigma\left(\hat{\theta}\right)\right]' \hat{\mathbf{U}}_{\Gamma_{\mathrm{N}}}\left[\mathrm{s} - \sigma\left(\hat{\theta}\right)\right], \tag{11}$$

and $\hat{\mathbf{U}}_{\Gamma_{\mathrm{N}}}$ is given by the matrix defined in (9) with $\mathbf{W} = \hat{\Gamma}_{\mathrm{N}}^{-1}$ based on the normal theory estimated weight matrix. Applied to least-squares estimation, this test has been in EQS and its documentation since 1989 (Bentler 1995). Similar tests hold, of course, for elliptical and heterogeneous kurtotic distributions by suitable use of $\mathbf{W} = \hat{\Gamma}_{\mathrm{E}}^{-1}$ or $\mathbf{W} = \hat{\Gamma}_{\mathrm{HK}}^{-1}$. Tests of this form also were discussed by Satorra & Bentler (1990). Satorra & Bentler (1994) showed in a small study that the test could work quite well. Of course, tests such as (11) require correct distributional specification, which Browne's test (10) was designed to avoid.

Under the correct model, but with distributional misspecification, the matrix (7) also does not describe the variability of the estimator $\hat{\theta}$. The correct large-sample covariance matrix is given by

$$\mathrm{acov}\left(\hat{\theta}\right) = n^{-1}\left(\Delta'\mathbf{W}\Delta\right)^{-1}\left(\Delta'\mathbf{W}\Gamma\mathbf{W}\Delta\right)\left(\Delta'\mathbf{W}\Delta\right)^{-1}. \tag{12}$$

This covariance matrix has been known to be the correct covariance matrix of the estimator for almost 15 years (e.g. Arminger & Schoenberg 1989; Bentler 1983a; Bentler & Dijkstra 1985; Browne 1982, 1984; Chamberlain 1982; Dijkstra 1981; Shapiro 1983; see also Kano 1993), but even today it seems to be computed only in the EQS and LINCS (Schoenberg & Arminger 1990) programs. In EQS, where (12) has been available since 1989, it is known as the "robust" covariance matrix. In contrast, by default extant programs calculate

$$\mathrm{acov}\left(\hat{\theta}\right) = n^{-1}\left(\Delta'\mathbf{W}\Delta\right)^{-1} \tag{13}$$

which is the inverse of the information matrix. Even though (13) does not give correct standard errors, it is the formula used in typical practice, e.g. in ML estimation without normal data.

Emphasizing again the parallel to linear regression with $y = \mathbf{X}\beta + \epsilon$, the information matrix (13) does not give the covariance matrix of the least-squares estimator $\hat{\beta}$ if cov(ϵ) is not proportional to I. The correct covariance matrix is given by (12), with $\mathbf{W}$ = I (due to least-squares estimation), $\Delta = \mathbf{X}$, and Γ as the true covariance matrix of the ϵ. More generally, if $\hat{\beta}$ is the generalized least-squares estimator based on the incorrect assumption that cov (ϵ) = $\mathbf{W}^{-1}$, the information matrix formula (13) does not give the standard errors, whereas (12) does.

Although the tests (8) and (10) and the covariance matrix (12) define statistics that are always correct, irregardless of the distribution of variables,

they are not the only options, nor necessarily the best options in any given situation. Chou et al (1991), Chou & Bentler (1995), and Finch et al (1995) did find the robust covariance matrix to give good estimates of sampling variability. Bentler (1994) found the mixture test (8) performed very well in most instances, but was destroyed by a certain type of nonnormality. The source of good or poor performance is not understood, but it is clear from the formulas that poor estimates of Γ may make these statistics behave badly in practice. Since the various results summarized in this paper rely on large-sample theory, it is also possible that these statistics can be outperformed in small samples by other methods. Additional research is clearly needed.

If the distributional assumption is correct so that $\mathbf{W} = \Gamma^{-1}$, the robust covariance matrix given in (12) reduces to the usual inverse of the information matrix as given in both (7) and (13). More generally, the standard error estimates obtained from (13) cannot be smaller than those of (12), because the difference between (12) and (13) is nonnegative definite. This means that using the usual and incorrect information matrix expression (13) under distributional misspecification will understate the variability of the estimator. This bias can be substantial, as was clearly shown by Finch et al (1995). In practice, this would make the parameter estimates appear to be more significant in *z*-statistics than they really are. We now review the test statistics used in practice and point to some problems and potentials.

SOME SPECIFIC TEST STATISTICS

If a distribution-free method can be used, the results will be optimal because the discrepancy function would then always be correctly specified. This is the ideal situation introduced into covariance structure analysis by the ADF method of Browne (1982) and the minimum distance method of Chamberlain (1982), which are identical. They proposed minimizing the quadratic form discrepancy function $F_{\mathrm{QD}} = [s - \sigma(\theta)]'\hat{\mathbf{W}}[s - \sigma(\theta)]$ in which $\mathbf{W} = \Gamma^{-1}$ without any assumption on the distribution of variables. To implement this, an old result was used, namely that

$$\Gamma_{ij,k\ell} = \sigma_{ijk\ell} - \sigma_{ij}\sigma_{k\ell}, \tag{14}$$

where

$$\sigma_{ijk\ell} = \mathbf{E}(x_{ti} - \mu_i)(x_{tj} - \mu_j)(x_{tk} - \mu_k)(x_{t\ell} - \mu_\ell)$$

is the fourth-order multivariate moment of variables xi about their means μi, and σ_{ij} is an element of Σ. In practice, sample moment estimators

$$s_{ijk\ell} = N^{-1}\Sigma_1^{\mathrm{N}}(x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j)(x_{tk} - \bar{x}_k)(x_{t\ell} - \bar{x}_\ell)$$

and

$$s_{ij} = n^{-1}\Sigma_1^N \left(x_{ti} - \bar{x}_i\right)\left(x_{tj} - \bar{x}_j\right)$$

are used to consistently estimate σ_{ijkl} and σ_{ij} to provide the elements of $\hat{\mathbf{W}}$. Alternative ADF estimators, based on linearization, also are available (Bentler 1983a,b, Bentler & Dijkstra 1985). The ADF methods for the first time provided a way of attaining the $\chi 2$ test of model fit (6) without an assumption on the distribution of variables; they also provided for optimal and correct standard errors via (7). The standard ADF method is now available in most structural modeling programs under various names: arbitrary distribution generalized least squares (AGLS) in EQS and weighted least squares (WLS) in LISREL. EQS gives the linearized ADF method by default.

Unfortunately, this great theoretical advance has not proven to be practically useful. Although the χ^2 test of model fit (6) is in principle always available via $T_{\text{ADF}} = n\hat{F}_{\text{QD}}$, the sample size may need to be impractically large for the theory to work well in practice. For example, in the simulation study of Hu et al (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. Discouraging results are typical (Chan et al 1995, Chou & Bentler 1995, Chou et al 1991, Curran et al 1994, Muthén & Kaplan 1992). Yung & Bentler (1994) proposed some computationally intensive modifications to the ADF test statistic, which improve but do not fully cure its performance deficiency.

At the most restrictive end of the distributional continuum, the ML discrepancy function based on the assumed normality of variables is

$$F_{\text{ML}} = \log|\Sigma| - \log|\mathbf{S}| + tr(\mathbf{S}\Sigma^{-1}) - p.$$

As shown by Browne (1974), for this discrepancy function

$$\Gamma = \Gamma_{\text{N}} = 2K_p'(\Sigma \otimes \Sigma)K_p,$$

where K_p is a transition matrix of known 0, 1/2, or 1 values that reduces the $p^2 \times p^2$ matrix $(\Sigma \otimes \Sigma)$ to order p^*, and $\otimes$ is the Kronecker product. Although it is not obvious, implicitly $\mathbf{W} = \Gamma_{\text{N}}^{-1}$, so, if the data are truly multivariate normal $T_{\text{ML}} = \hat{F}_{\text{ML}}$ meets (6). But if the data are not normal, T_{ML} is generally not a χ^2 variate (though it may be so, e.g. via asymptotic robustness theory). Assuming T_{ML} to be a χ^2 variate is the typical mistake in applications of covariance structure analysis. The vices and virtues of the ML statistics are shared by normal theory GLS statistics based on minimizing $F_{\text{GLS}} = .5tr\{[\mathbf{S} - \Sigma(\theta)]\mathbf{V}^{-1}\}^2$. If the data are non-normal, this function is always incorrectly specified, and (6) cannot be guaranteed to hold. However, if the data are normal, this function may or may not be correctly specified, depending on the choice of the

weight matrix **V**. The choice of **V** = **I**, as in least-squares (LS) analysis, is misspecified, so (6) will not hold. The typical choice of GLS is **V** = **S**; then, $\mathbf{W} = \Gamma_N^{-1}$, and $T_{GLS} = n\hat{F}_{GLS}$ will meet (6). If $\mathbf{V} = \hat{\Sigma}$ is iteratively updated, F_{GLS} is the reweighted least-squares function F_{RLS}, yielding $T_{RLS} = n\hat{F}_{RLS}$. This also meets (6). In fact, Browne (1974) has shown that if **V** converges in probability to Σ (e.g. **V** = **S** or **V** = $\hat{\Sigma}$) then GLS and ML estimators are asymptotically equivalent.

Estimators and tests whose requirements fall between the normal and distribution-free theory are given by elliptical and heterogeneous kurtosis theory. In elliptical theory (Browne 1982, 1984), all marginal distributions of a multivariate distribution are symmetric and have the same relative kurtosis. This is more general than normal theory, yet estimators and test statistics can be obtained by simple adjustments to the statistics derived from normal theory methods. Let $\kappa = \sigma_{iiii} / 3\sigma_{ii}^2 - 1$ be the common kurtosis parameter of a distribution from the elliptical class. Multivariate normal distributions are members of this class with $\kappa = 0$. The fourth-order multivariate moments σ_{ijkl} are related to κ by

$$\sigma_{ijk\ell} = (\kappa + 1)\left(\sigma_{ij}\sigma_{k\ell} + \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk}\right).$$

As a result of this simplification, the F_{QD} discrepancy function for an elliptical distribution simplifies to

$$F_E = \frac{1}{2}(\kappa + 1)^{-1}\text{tr}\left[\mathbf{S} - \Sigma(\theta)\right]\mathbf{V}^{-1}{}^2 - \delta\ tr\left[\mathbf{S} - \Sigma(\theta)\right]\mathbf{V}^{-1}{}^2,$$

where as before V is any consistent estimator of Σ and

$$\delta = \kappa/[4(\kappa + 1)^2 + 2p\kappa(\kappa + 1)]$$

(Bentler 1983a). The selection of **V** as a consistent estimator of Σ and a kurtosis estimator such as

$$\hat{\kappa} + 1 = \Sigma_1^N\left[\left(\underset{\sim}{x} - \bar{\underset{\sim}{x}}\right)'\mathbf{S}^{-1}\left(\underset{\sim}{x} - \bar{\underset{\sim}{x}}\right)\right]^2 / Np(p+2)$$

leads, under the model and assumptions, to an asymptotically efficient estimator of θ with $T_E = n\hat{F}_E$ meeting (6). If $\mathbf{V} = \hat{\Sigma}$ is iteratively updated, and the model is invariant with respect to a constant scaling factor, at the minimum of F_E the second term drops out yielding $T_E = T_{ERLS}$ (see Browne 1984, Shapiro & Browne 1987).

Heterogeneous kurtosis (HK) theory (Kano et al 1990) defines a still more general class of multivariate distributions that allows marginal distributions to have heterogeneous kurtosis parameters. The elliptical distribution is a special case of this class of distributions. Let $\kappa_i^2 = \sigma_{iiii} / 3\sigma_{ii}^2$ represent a measure of

excess kurtosis of the *i*-th variable, and the fourth-order moments have the structure

$$\sigma_{ijk\ell} = \left(a_{ij}a_{k\ell}\right)\sigma_{ij}\sigma_{k\ell} + \left(a_{ik}a_{j\ell}\right)\sigma_{ik}\sigma_{j\ell} + \left(a_{i\ell}a_{jk}\right)\sigma_{i\ell}\sigma_{jk},$$

where $a_{ij} = (\kappa_i + \kappa_j)/2$. If the covariance structure $\Sigma(\theta)$ is fully scale invariant and the modeling and distributional assumptions are met, the F_{QD} discrepancy function can be expressed as $F_{HK} = .5tr\{[\mathbf{S} - \Sigma(\theta)]\hat{\mathbf{C}}^{-1}\}^2$, where $\hat{\mathbf{C}} = \hat{\mathbf{A}} * \hat{\Sigma}$, and * denotes the elementwise (Hadamard) product of the two matrices of the same order. In practice, $\hat{\mathbf{A}} = (a_{ij}) = (\hat{\kappa}_i + \hat{\kappa}_j)/2$ using the usual moment estimators for each variable $\hat{\kappa}_i^2 = s_{iiii}/3s_{ii}^2$, with $\hat{\mathbf{C}} = \hat{\mathbf{A}} * \mathbf{S}$. [For another estimator, see Bentler et al (1991).] The HK estimator is asymptotically efficient, and the associated test statistic $T_{HK} = n\hat{F}_{HK}$, at the minimum meets (6). An attractive feature of the Kano et al theory is that fourth-order moments of the measured variables do not need to be computed as they do in ADF theory, because these moments are just a function of the variances and covariances and the univariate kurtoses. As a result, the HK method can be used on models with many measured variables. While ADF cannot be implemented with more than about 30–40 variables due to the large size of its weight matrix, this is not a limitation of the Kano et al HK method. Koning et al (1993) studied a generalization of F_{HK} in which the matrix C is unrestricted but, unfortunately, it must be related to an estimated ADF weight matrix.

Based on the general distribution (8), Satorra & Bentler (1986, 1988, 1994) developed two modifications of any standard goodness-of-fit statistic test T (T_{ML}, T_{HK}, etc.) so that its distributional behavior should more closely approximate χ^2. The mean of the asymptotic distribution of T is given by $tr(\mathbf{U}\Gamma)$, where $\mathbf{U}$ is defined in (9). Letting $c = (df)^{-1}\, tr(\hat{\mathbf{U}}\hat{\Gamma})$, where $\hat{\mathbf{U}}$ is a consistent estimator of $\mathbf{U}$ based on $\hat{\theta}$ and $\hat{\Gamma}$ is an estimator based on the ADF matrix (14), the Satorra-Bentler scaled test statistic is

$$\bar{T} = c^{-1}\, T. \tag{15}$$

The scaling constant c effectively corrects the statistic T so that the mean of the sampling distribution of $\bar{T}$ will be closer to the expected mean under the model (see also Kano 1992). The scaled statistic that has been implemented in EQS since 1989 is based on the use of T_{ML} in (15). Chou & Bentler (1995), Chou et al (1991), and Curran et al (1994) found the scaled statistic to work well in simulation studies. In the study by Hu et al (1992), the Satorra-Bentler scaled statistic performed best overall under a wide variety of conditions of varied distributions and sample sizes, outperforming the ADF method at all but the largest sample sizes, where it performed equally well. Even though current evidence shows that the scaled statistic (15) performs better than others currently available in modeling programs, in principle some of the alternative

statistics, such as those based on (8) and (10) might perform better because their sampling distributions are well-specified. However, Bentler (1994) found (8) to break down under conditions where the scaled statistic remained well-behaved, and essentially nothing is known about how (10) compares to (15).

Satorra & Bentler also reported the development of a test statistic that adjusts not only the mean, but also the variance, of the statistic to more closely approximate a χ^2 distribution. The adjusted test statistic $\bar{\bar{T}}$ is obtained by computing the integer d nearest to d', defined by $d' = [tr(U\Gamma)]^2 \div tr[(U\Gamma)^2]$, and computing the statistic

$$\bar{\bar{T}} = \frac{d}{tr(U\Gamma)} T. \tag{16}$$

The effect is to scale with a degrees of freedom adjustment, since (16) is χ^2 distributed with d *df*. If facilities permit χ^2 to be computed for noninteger *df*, one can calculate $\bar{\bar{T}} = \{d'/[tr(U\Gamma)]\}\, T$ and evaluate it with fractional *df*. Satorra & Bentler (1994) showed with an illustrative example that their statistic (16) can work well, but we are not aware of any systematic study of this statistic.

Because many of the potentially valuable statistics are not available in standard computer programs, in the Appendix we show how a standard matrix language available in SAS or SPSS can be used, along with extant programs such as LISREL and EQS, to yield some of the potentially useful tests. We now illustrate the similarities and differences with a short example.

AN EXAMPLE: TEACHER STRESS

In a study examining stress among school teachers, Bell et al (1990) used an 11 item measure of somatic complaints (e.g. dizziness, shortness of breath, headaches, etc.). The questions were answered on a 5 point response scale for frequency of occurrence (1 = *rarely or never* to 5 = *very often*). The Pearson correlation matrix for these 11 items, as well as their standard deviations and measures of relative skewness ($g_{1(i)}$) and relative kurtosis ($g_{2(i)}$) are given in Table 1 for 362 primary school teachers out of the total teacher sample of 956 from primary, technical, and secondary schools. The data are quite obviously non-normal, with excessive kurtosis and skewness being evident for all items. Although structural models can be fitted to data of this kind using polychoric correlations, in applying such models one assumes that the underlying latent distribution for responses to each item is normal. For the sorts of data being considered in this example, the validity of that assumption may well be questioned.

The 11 items were fitted to a single factor model by LISREL using HK estimation. Details on how this is done are in the Appendix: In essence we minimize the F_{QD} discrepancy function (i.e. WLS in LISREL nomenclature)

but supply our own computed weight matrix **W**; **W** makes use of Browne's (1977) normal theory relation $\mathbf{W} = \Gamma_N^{-1}$, where $\Gamma_N = 2K_p'(V \otimes V)K_p$, but where the general matrix **V** is in this instance given by $\hat{\mathbf{C}} = \hat{\mathbf{A}} * \boldsymbol{\Sigma}$ from HK theory.

The HK parameter estimates and standard errors are displayed in Table 2. For comparative purposes, the results of using both ML and ADF (i.e. WLS) estimation are also provided in Table 2, along with the Satorra-Bentler scaled test of fit (Equation 15) and robust standard errors (Equation 12) for maximum likelihood. (Details of how these robust procedures can be computed for LISREL output are also given in the Appendix.)

The model test statistics for both ML (T_{ML} =79.89, p. < .001) and ADF (T_{ADF} =66.82, p.=0.015) are considerably higher than either the HK (T_{HK} =36.14, p.=0.794) or the scaled ML statistics ($\bar{T}_{ML}$ =45.89, p.=0.394). If we assume that the null hypothesis of a single factor model holds exactly, then the test statistics for both the HK estimator and the robust scaled ML estimator are quite acceptable. The corresponding probabilities for ML and ADF estimation

Table 1 Correlation matrix of 11 somatic complaints items and measures of item distribution (N = 362).

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1.	1.000										
2.	0.503	1.000									
3.	0.375	0.388	1.000								
4.	0.542	0.498	0.459	1.000							
5.	0.444	0.471	0.451	0.472	1.000						
6.	0.419	0.411	0.382	0.364	0.369	1.000					
7.	0.291	0.302	0.272	0.262	0.274	0.380	1.000				
8.	0.468	0.509	0.295	0.455	0.402	0.450	0.245	1.000			
9.	0.274	0.224	0.230	0.270	0.285	0.232	0.173	0.256	1.000		
10.	0.351	0.294	0.200	0.308	0.284	0.340	0.256	0.402	0.280	1.000	
11.	0.396	0.399	0.399	0.445	0.382	0.409	0.342	0.394	0.290	.0.342	1.000
s.d.	0.648	0.809	1.210	0.809	0.837	1.005	0.917	0.877	1.045	0.671	1.224
$g_{1(i)}$	2.949	2.148	0.757	2.466	2.260	1.584	1.115	2.170	2.127	2.717	0.588
$g_{2(i)}$	9.229	4.547	-0.389	6.012	4.677	1.793	0.645	4.245	3.588	7.135	-0.641

Note: $g_{1(i)} = N^{1/2}\Sigma_1^N (x_{it} - \bar{x}_i)^3 / \left[\Sigma_1^N (x_{it} - \bar{x}_i)^2\right]^{3/2}$ and

$g_{2i} = N\Sigma_1^N (x_{it} - \bar{x}_i)^4 / \left[\Sigma_1^N (x_{it} - \bar{x}_i)^2\right]^2 - 3$

are both less than 0.02, thereby indicating that the null hypothesis would have been rejected if the latter two discrepancy functions had been employed in practice. Interestingly, the adjusted ML statistic ($\bar{\bar{T}}_{ML}$ =51.07, p.=0.391), with fractional *df* = 48.98, has almost exactly the same probability of acceptance as the scaled $\bar{T}_{ML}$ statistic. The results here are consistent with those of Hu et al (1992) who found that the HK estimator tended to slightly underestimate the expected χ^2 statistic relative to the robust scaled ML statistic, but that the HK test statistic was more correct than either the ML or ADF statistics, which were inflated well above their expected values. If we inspect the parameter estimates and standard errors in Table 2, we see that the HK and ML values are more similar compared to the ADF estimates, although the difference is more pronounced for factor loadings than for unique variances. The ML standard

Table 2 Unstandardized parameter estimates and standard error values for the single factor model of somatic complaints under different estimation methods (*df* = 44).

	ML		ADF		HK		Robust ML
	$\hat{\theta}$	(S.E.)	$\hat{\theta}$	(S.E.)	$\hat{\theta}$	(S.E.)	(S.E.)
λ_1	0.453	(0.031)	0.289	(0.058)	0.403	(0.054)	(0.069)
λ_2	0.561	(0.039)	0.416	(0.056)	0.512	(0.058)	(0.064)
λ_3	0.702	(0.062)	0.632	(0.053)	0.751	(0.066)	(0.056)
λ_4	0.571	(0.039)	0.356	(0.060)	0.541	(0.060)	(0.067)
λ_5	0.542	(0.042)	0.425	(0.058)	0.531	(0.060)	(0.068)
λ_6	0.620	(0.051)	0.597	(0.064)	0.653	(0.063)	(0.068)
λ_7	0.411	(0.049)	0.373	(0.054)	0.459	(0.056)	(0.062)
λ_8	0.578	(0.043)	0.466	(0.061)	0.509	(0.063)	(0.071)
λ_9	0.418	(0.056)	0.388	(0.074)	0.414	(0.076)	(0.078)
λ_{10}	0.330	(0.035)	0.228	(0.048)	0.312	(0.054)	(0.054)
λ_{11}	0.762	(0.061)	0.661	(0.049)	0.795	(0.064)	(0.050)
ψ_1	0.215	(0.019)	0.167	(0.026)	0.209	(0.036)	(0.031)
ψ_2	0.340	(0.029)	0.281	(0.038)	0.325	(0.044)	(0.047)
ψ_3	0.971	(0.078)	0.839	(0.067)	0.855	(0.073)	(0.074)
ψ_4	0.329	(0.029)	0.194	(0.034)	0.295	(0.046)	(0.050)
ψ_5	0.407	(0.033)	0.302	(0.039)	0.382	(0.052)	(0.056)
ψ_6	0.626	(0.051)	0.439	(0.062)	0.551	(0.062)	(0.075)
ψ_7	0.672	(0.052)	0.571	(0.047)	0.611	(0.056)	(0.059)
ψ_8	0.434	(0.036)	0.325	(0.040)	0.384	(0.052)	(0.064)
ψ_9	0.918	(0.070)	0.734	(0.101)	0.879	(0.102)	(0.117)
ψ_{10}	0.341	(0.027)	0.198	(0.035)	0.302	(0.046)	(0.052)
ψ_{11}	0.916	(0.075)	0.829	(0.073)	0.861	(0.071)	(0.075)
χ^2 value	79.89		66.82		36.14		45.89

errors are lower than corresponding ADF and HK values, whereas the latter two are slightly lower on average than the robust ML standard errors.

It is also instructive to consider the effect of applying Browne's (1984) general quadratic test of fit (Equation 10) to the incorrectly specified ML discrepancy function. Under maximum likelihood, this statistic is 70.11 and is therefore an improvement on the inflated likelihood ratio statistic of 79.89. However, it is still not comparable to the robust scaled statistic. The HK estimator statistic under Browne's general quadratic test of fit is much higher at 68.23, compared with its corresponding T_{HK} value. Using the robust scaled ML statistic as a benchmark, these results suggest that Browne's general quadratic test of fit does not adequately correct for employing an inefficient estimation method in this example. When the complete sample of 956 teachers is analyzed (these results are not shown here), the ADF test statistic (T_{ADF}=101.27) is now much closer to the scaled ML statistic ($\bar{T}_{ML}$=96.90) and the HK statistic (T_{HK}=76.02), compared to the ML statistic (T_{ML}=166.76). Assuming from simulation studies such as those of Hu et al (1992) that the robust scaled statistic is the least biased test of fit of those available, the ADF estimator therefore appears to require a large number of cases to obtain relatively accurate tests of fit even in the present instance of a small-to-moderate number of variables.

DISCUSSION

It is remarkable that in spite of substantial technical innovation in the statistics of covariance structure analysis during the past 15 years, only a few of these developments have found their way into texts, commercial computer programs, or general knowledge among users. For example, one of the authors inquired about the robust covariance matrix through the SEMNET special interest internet user's group and found only scattered awareness about its existence. It seems that only the most dedicated methodologists will know about the technical inadequacies in the methods routinely available for application. This review aims at broadening the knowledge of proven and potentially useful statistics for this field, and, with the code accessible via the Appendix, permitting the applied researcher to incorporate some of the more promising into their own favorite computer program. Although we have concentrated on methods for a single group, the same principles hold for multiple population covariance structure models (e.g. Bentler et al 1987) and to mean and covariance structure models (e.g. Browne & Arminger 1995, Satorra 1992). Some parallel results have been reported in the econometric literature (Newey & McFadden 1994) and in the statistical literature on nonlinear regression (Yuan 1995). Clearly, these newer methods should prove useful

toward the more accurate evaluation of psychological theories with nonexperimental data.

With a small number of variables to model, clearly there are several alternative test statistics that hold under potential misspecification. Currently, only the Satorra-Bentler scaled statistic (15) is known to behave well empirically under a wide variety of distributional misspecifications. Yet, there are other potentially useful statistics that have hardly been studied, e.g. the mixture test (8) and the quadratic form test based on (10). It is not clear why this should be so. Of course, simulation work is time consuming and few theoretical statisticians will undertake it, yet without such work we will never know the actual performance of statistics under less than textbook conditions. Certainly, future research should determine whether these ignored tests have any role to play in real data analysis. In such situations, there seems to be little excuse for not using the robust covariance matrix (12), at least until feasible and improved alternatives become available. The current practice of typically using the wrong formula (Equation 13) to evaluate parameter significance seems especially unfortunate because it tends to give misleadingly optimistic results.

The situation is more difficult for models based on say, 40 or more variables, where computer limitations make some estimators infeasible. Although not enough research has gone into establishing the types of nonnormal distributions that might be fruitfully modeled by the Kano et al (1990) HK approach, this method should be studied further in the future since it is one of the few that holds any promise of handling models based on a huge number of nonnormal variables. In such circumstances, methods that require computation of a distribution-free estimate of the true covariance matrix Γ of the sample covariances will become unavailable, and normal, elliptical, and HK theory (or transformations to such — see Mooijaart 1993) seem to be the only alternatives. Aside from asymptotic robustness theory, which if it could be applied wisely would suggest when use of normal theory methods would work (see, e.g. Hu et al 1992), HK theory would seem to be a promising alternative. In the Appendix we describe a way to implement this method. It could be implemented easily in other programs if software distributors followed Schoenberg & Arminger's (1990) LINCS program of permitting input of the matrix V in a normal theory method, since this method can be adapted to yield the HK method. The rationale for the HK method requires scale-invariant models (e.g. Krane & McDonald 1978), so equality restrictions could create a problem in practice (see also O'Brien & Reilly 1995). Extension of the theory to correlation structure models would permit a wider range of models to be used. The standard error estimates for the HK theory are easy to compute, though of course they are only strictly correct under an HK distributional assumption. Research will have to establish the robustness of the HK statistics to violation of its assumptions. Certainly, because of the need to estimate a very large Γ,

the robust covariance matrix (12) will be difficult to compute. Perhaps computer-intensive resampling methods will have to be used instead (e.g. Bollen & Stine 1993, Ichikawa & Konishi 1995, Yung & Bentler 1995).

In addition to distributional misspecification, and the number of variables and parameters in a model, sample size is a major factor influencing the quality of statistics in covariance structure models. Although projection of asymptotic theory onto small sample data analysis is an old problem (Boomsma 1983, Tanaka 1987), it remains a continuing one: More and more statistics rely on fourth-order moments of the data and these are unstable at the relatively low samples sizes that characterize most real data. Clearly, further research should be directed toward improving estimators of weight matrices that require such moments.

A FINAL NOTE

Since this review went to press, Yuan and Bentler (1995) developed some new test statistics for mean and covariance structure analysis. One of these can be computed as a simple Bartlett-type correction to the ADF test statistic. Specifically, considering a model fitted under arbitrary distributions using optimal ADF estimation, their corrected statistic can be computed as

$$T_{\mathrm{CADF}} = T_{\mathrm{ADF}} / (1 + T_{\mathrm{ADF}} / n).$$

In a small simulation study with a covariance structure model, they replicated earlier results showing that the standard ADF test statistic is essentially unusable in small to intermediate sized samples. On the other hand, their corrected statistic yielded a dramatic improvement in performance. It behaved close to nominally at all sample sizes, though there was a tendency for the rejection rate under the null hypothesis to be somewhat too small. As sample size gets very large, as is obvious from the formula, $T_{\mathrm{CADF}} \rightarrow T_{\mathrm{ADF}}$, i.e., the Yuan-Bentler statistic becomes equivalent to the Browne-Chamberlain statistic in covariance structure analysis. This optimistic development clearly bears further study.

APPENDIX

This Appendix contains an overview of several procedures that have been written in the MATRIX command language of SPSS Release 4 (SPSS 1990) for carrying out a number of the above test statistics which are not available currently in structural modeling programs like LISREL or EQS. These procedures can be readily adapted to other matrix programming languages such as PROC IML (SAS Institute 1990) or GAUSS (Aptech Systems 1992) or incorporated into general matrix procedures for covariance structure analysis (see,

e.g. Cudeck et al 1993). Files containing the MATRIX commands, as well as detailed instructions and test data, are available on the Internet via anonymous ftp at ftp.stat.ucla.edu in the directory /pub/code/statlib/csm or from either author. For those unfamiliar with using ftp, brief instructions are provided at the end of this Appendix. Further details concerning the use of the procedures are provided in files at the ftp site.

Heterogenous Kurtosis Estimation

Heterogenous kurtosis estimation can be obtained from standard covariance structure modeling programs if the program includes ADF estimation as an option. To implement the HK estimator, we make use of the equivalence between the normal theory GLS discrepancy function

$$F_{\text{GLS}} = \frac{1}{2} tr\ \left[\mathbf{S} - \Sigma(\theta)\right]\mathbf{V}^{-1}\ ^{2} \tag{A1}$$

and the quadratic discrepancy function

$$F_{\text{QD}} = \left[\mathbf{s} - \sigma(\theta)\right]' \hat{\mathbf{W}} \left[\mathbf{s} - \sigma(\theta)\right] \tag{A2}$$

where in the latter

$$\begin{aligned} \hat{\mathbf{W}} &= \left[2\mathrm{K}_{\mathrm{p}}'(\mathbf{V} \otimes \mathbf{V})\mathrm{K}_{\mathrm{p}}\right]^{-1} \\ &= \tfrac{1}{2}\mathrm{K}_{\mathrm{p}}^{-}\left(\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}\right)\mathrm{K}_{\mathrm{p}}^{-\prime} \end{aligned} \tag{A3}$$

and $\mathrm{K}_{\mathrm{p}}^{-}$ is the Moore-Penrose inverse of K_{p} (see, e.g. Browne 1974, 1984). It is useful to remember that in this instance, $\hat{\mathbf{W}} = \Gamma_N^{-1}$ because the discrepancy function is correctly specified. In any desired application of the HK discrepancy function to a particular covariance structure model employing SPSS MATRIX commands, we need to (*a*) calculate the estimates of $\hat{\kappa}_i = (\mathrm{s}_{iiii} / 3\mathrm{s}_{ii}^2)^{1/2}$ for the observed variables i = 1,...,p in the model, (*b*) calculate the product $\hat{\mathbf{C}} = \hat{\mathbf{A}} * \hat{\Sigma}$, where $\hat{\mathbf{A}} = (\hat{\mathrm{a}}_{ij}) = (\hat{\kappa}_i + \hat{\kappa}_j)/2$, (*c*) compute the required transition matrix K_{p} for the number of observed variables, and (*d*) compute (A3), where we substitute $\hat{\mathbf{C}}$ for $\mathbf{V}$ and signify the resultant weight matrix as $\hat{\mathbf{W}}_{\text{HK}}$. The LISREL or EQS user then minimizes the quadratic form discrepancy function

$$F\left[\left(\mathbf{S}, \Sigma\left(\hat{\theta}\right)\right) \middle| \mathbf{W} = \Gamma_{\text{HK}}{}^{-1}, \Gamma_{\text{HK}}\right]$$

for heterogenous kurtosis estimation by taking the lower symmetric form of the resultant estimate of $\hat{\mathbf{W}}_{\text{HK}}$, and inputting it as the external weight matrix file associated with WLS or AGLS estimation respectively. It should be noted

that the weight matrix in both LISREL and EQS is inverted after it has been read in from an external file, so the SPSS MATRIX commands produce the noninverted form of (A3). Similar methods can be used to obtain estimates under elliptical distribution in LISREL or to compare the equivalence of estimation under the ML discrepancy function F_{ML} to that of the reweighted least-squares function F_{RLS}.

Scaled Test Statistics and Robust Standard Errors

The Satorra-Bentler scaled test statistic and robust standard errors are available already in EQS. To obtain these statistics for any covariance structure model that has been fitted by programs such as LISREL, we require (*a*) the ADF weight matrix, signified here as $\hat{\Gamma}$ and based on sample estimators for (14), (*b*) the appropriate consistent estimator of

$$\mathbf{W} = \hat{\Gamma}_{\mathrm{N}}^{-1}, \hat{\Gamma}_{\mathrm{E}}^{-1}, \text{ or } \hat{\Gamma}_{\mathrm{HK}}^{-1}$$

using Equation (A3) for the particular estimation method being employed, (*c*) the matrix $\hat{\Delta}$ of partial first derivatives, and (*d*) the test statistic *T* and degrees of freedom (*df*) for the fitted model. The ADF weight matrix can be obtained from PRELIS, where it is called the asymptotic covariance matrix, or it can be computed directly using SPSS MATRIX commands. The matrix $\hat{\mathbf{W}}$ can be computed by a SPSS MATRIX procedure using (A3), where we substitute for **V** the appropriate sample matrix for the estimation method chosen in LISREL. Finally, a numerical approximation can be obtained in SPSS MATRIX for the matrix $\hat{\Delta}$ of partial derivatives using a forward finite-difference method (Kennedy & Gentle 1980, Section 10.2.6). Let *h* be a small constant, for instance 10–5, and let c_{i} be a column vector of the same dimension as θ with a value of unity in its *i*-th element and zero values in all remaining elements. Then the $p^* \times 1$ vector of partial derivatives for the *i*-th value of θ is given by

$$\frac{\partial \sigma\left(\hat{\theta}\right)}{\partial \theta_i'} = \frac{\sigma\left(\hat{\theta} + \mathrm{c}_i \mathrm{h}\right) - \sigma\left(\hat{\theta}\right)}{\mathrm{h}} = \sigma\left(\theta_i'\right).$$

The $p^* \times q$ matrix $\hat{\Delta}$ can be derived from the concatenation of successive columns of $\sigma\left(\theta_i'\right)$. The test statistic *T* can be found in the printout of the model.

Let us assume, for example, that we have estimated a model using maximum likelihood, and that we have calculated the three required matrices noted above (i.e. $\hat{\Gamma}, \mathbf{W} = \hat{\Gamma}_{\mathrm{N}}^{-1}$, and $\hat{\Delta}$) and also obtained the values for T_{ML} and the *df* from the LISREL output. Then we can obtain (*a*) the Satorra-Bentler scaled test statistic $\overline{T}_{\mathrm{ML}}$ by simple substitution to obtain **U** in Equation (9) and then solve Equation (15); and (*b*) robust standard errors by simple substitution into Equation (12), in which $\Gamma = \hat{\Gamma}$ and $\mathbf{W} = \hat{\Gamma}_{\mathrm{N}}^{-1}$. The Satorra-Bentler ad-

justed test statistic (16) can be readily obtained in a similar manner. Note that SPSS MATRIX has a χ^2 cumulative distribution function routine that takes noninteger *df* values, so that it can compute the more precise form of (16).

Accessing the SPSS MATRIX files by FTP

For persons who have not used anonymous ftp (file transfer protocol) procedures before, the following gives a few basic instructions. You will need access to the Internet and appropriate software for performing ftp. The hostname of the computer where the SPSS MATRIX files can be found is ftp.stat.ucla.edu. The account name that you must give to gain access is "anonymous" and the password is (usually) your email address. Once you have gained access, go to the subdirectory /pub/code/statlib/csm where the files are located (typically, this is done by typing "cd pub/code/statlib/csm"). All files are in ASCII format and can be transferred to your own computer by the appropriate commands for the particular ftp software implementation being used (typically, this is by typing "mget *.*"). Some ftp software implementations utilize graphical user interfaces, and so there is no way of covering all the possibilities of how to ftp the files, beyond the basic approach given here. It would be best to consult your local computer advisory service if you are uncertain about what to do.

Literature Cited

Amemiya Y, Anderson TW. 1990. Asymptotic chi-square tests for a large class of factor analysis models. *Ann. Stat.* 18:1453–63

Anderson TW. 1994. See Anderson et al 1994, pp. 1–20

Anderson TW, Amemiya Y. 1988. The asymptotic normal distribution of estimators in factor analysis under general conditions. *Ann. Stat.* 16:759–71

Anderson TW, Fang KT, Olkin I, eds. 1994. *Multivariate Analysis and Its Applications.* Hayward, CA: Inst. Math. Stat.

Aptech Systems. 1992. *GAUSS System, Version 3.* Kent, WA: Aptech Syst.

Arminger G. 1995. See Arminger et al 1995, pp. 77–183

Arminger G, Clogg CC, Sobel ME, eds. 1995. *Handbook of Statistical Modeling for the Social and Behavioral Sciences.* New York: Plenum

Arminger G, Schoenberg RJ. 1989. Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika* 53: 409–25

Bekker PA, Merckens A, Wansbeek TJ. 1994. *Identification, Equivalent Models, and Computer Algebra.* Boston: Academic

Bell RC, Wearing AJ, Conn M, Dudgeon P, McMurray N, Stanley GV. 1990. *Teacher Stress in Victoria.* Melbourne, Australia: Victorian Minist. Educ.

Bentler PM. 1980. Multivariate analysis with latent variables: causal modeling. *Annu. Rev. Psychol.* 31:419–56

Bentler PM. 1983a. Some contributions to efficient statistics for structural models: specification and estimation of moment structures. *Psychometrika* 48:493–517

Bentler PM. 1983b. Simultaneous equations as moment structure models: with an introduction to latent variable models. *J. Econometr.* 22:13–42

Bentler PM. 1994. See Anderson et al 1994, pp. 123–36

Bentler PM. 1995. *EQS Structural Equations*

Program Manual. Encino, CA: Multivariate Softw.

Bentler PM, Berkane M, Kano Y. 1991. Covariance structure analysis under a simple kurtosis model. In *Computing Science and Statistics*, ed. EM Keramidas, pp. 463–65. Fairfax Station, VA: Interface Found

Bentler PM, Dijkstra T. 1985. Efficient estimation via linearization in structural models. In *Multivariate Analysis VI*, ed. PR Krishnaiah, pp. 9–42. Amsterdam: North Holland

Bentler PM, Lee SY, Weng J. 1987. Multiple population covariance structure analysis under arbitrary distribution theory. *Comm. Stat. Theory* 16:1951–64

Bentler PM, Weeks DG. 1980. Linear structural equations with latent variables. *Psychometrika* 45:289–308

Bentler PM, Wu EJC. 1995a. *EQS for Macintosh User's Guide*. Encino, CA: Multivariate Softw.

Bentler PM, Wu EJC. 1995b. *EQS for Windows User's Guide*. Encino, CA: Multivariate Softw.

Bollen KA. 1989. *Structural Equations with Latent Variables*. New York: Wiley

Bollen KA, Long JS, eds. 1993. *Testing Structural Equation Models*. Newbury Park, CA: Sage

Bollen KA, Stine RA. 1993. See Bollen & Long 1993, pp. 111–35

Boomsma A. 1983. *On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality*. PhD thesis. Univ. Groningen

Borg I, Mohler PPh, eds. 1994. *Trends and Perspectives in Empirical Social Research*. Berlin: Walter de Gruyter

Breckler SJ. 1990. Applications of covariance structure modeling in psychology: Cause for concern? *Psychol. Bull.* 107: 260–73

Browne MW. 1974. Generalized least squares estimators in the analysis of covariance structures. *S. Afr. Stat. J.* 8:1–24

Browne MW. 1982. Covariance structures. In *Topics in Applied Multivariate Analysis*, ed. DM Hawkins, pp. 72–141. Cambridge: Cambridge Univ. Press

Browne MW. 1984. Asymptotically distribution-free methods for the analysis of covariance structures. *Br. J. Math. Stat. Psychol.* 37:62–83

Browne MW. 1987. Robustness of statistical inference in factor analysis and related models. *Biometrika* 74:375–84

Browne MW. 1990. Asymptotic robustness of normal theory methods for the analysis of latent curves. In *Statistical Analysis of Measurement Error Models and Applications*, ed. PJ Brown, WA Fuller, 112:211–25. Providence, RI: Am. Math. Soc.

Browne MW, Arminger G. 1995. See Arminger et al 1995, pp. 185–249

Browne MW, Cudeck R. 1993. See Bollen & Long 1993, pp. 136–62

Browne MW, Mels G, Coward M. 1994. Path Analysis: RAMONA. In *SYSTAT for DOS: Advanced Applications, Version 6 Ed.*, pp. 163–224. Evanston, IL: Systat

Browne MW, Shapiro A. 1988. Robustness of normal theory methods in the analysis of linear latent variate models. *Br. J. Math. Stat. Psychol.* 41:193–208

Bullock HE, Harlow LL, Mulaik SA. 1994. Causation issues in structural equation modeling research. *Struct. Eq. Mod.* 1: 253–67

Byrne BM. 1994. *Structural Equation Modeling with EQS and EQS/Windows*. Thousand Oaks, CA: Sage

Chamberlain G. 1982. Multivariate regression models for panel data. *J. Econometr.* 18: 5–46

Chan W, Yung YF, Bentler PM. 1995. A note on using an unbiased weight matrix in the ADF test statistic. *Mult. Behav. Res.* In press

Chou CP, Bentler PM. 1995. See Hoyle 1995, pp. 36–55

Chou CP, Bentler PM, Satorra A. 1991. Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis: a Monte Carlo study. *Br. J. Math. Stat. Psychol.* 44:347–57

Cudeck RL, Henly SJ. 1991. Model size in covariance structures analysis and the 'problem' of sample size. *Psychol. Bull.* 109:512–19

Cudeck RL, Klebe KJ, Henly SJ. 1993. A simple Gauss-Newton procedure for covariance structure analysis with high-level computer languages. *Psychometrika* 58: 211–32

Curran PJ, West SG, Finch JF. 1994. The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. In *Psychological Methods*. In press

de Leeuw J. 1988. Multivariate analysis with linearizable regressions. *Psychometrika* 53: 437–54

Dijkstra T. 1981. *Latent variables in linear stochastic models*. PhD thesis. Univ. Groningen

Dudgeon P. 1994. A reparameterization of the restricted factor analysis model for multitrait-multimethod matrices. *Br. J. Math. Stat. Psychol.* 47:283–308

Dunn G, Everitt B, Pickles A. 1993. *Modelling Covariances and Latent Variables Using EQS*. London: Chapman & Hall

Faulbaum F, Bentler PM. 1994. See Borg & Mohler 1994, pp. 224–49

Finch JF, West SG, MacKinnon DP. 1995. The effects of sample size and non-normality

on the estimation of direct and indirect effects in latent variable models. *Struct. Eq. Model.* In press

Gierl MJ, Mulvenon S. 1995. *Evaluating the application of fit indices to structural equation models in educational research: a review of the literature from 1990 through 1994.* Presented at 1995 Annu. Meet. Am. Educ. Res. Assoc., San Francisco

Haagen K, Bartholomew DJ, Deistler M, eds. 1993. *Statistical Modelling and Latent Variables.* Amsterdam: North Holland

Hoyle R, ed. 1995. *Structural Equation Modeling: Concepts, Issues, and Applications.* Thousand Oaks, CA: Sage

Hu LT, Bentler PM, Kano Y. 1992. Can test statistics in covariance structure analysis be trusted? *Psychol. Bull.* 112:351–62

Ichikawa M, Konishi S. 1995. Application of the bootstrap methods in factor analysis. *Psychometrika* 60:77–93

Jöreskog KG. 1969. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* 34:183–202

Jöreskog KG. 1994. See Anderson et al 1994, pp. 297–310

Jöreskog KG, Sörbom D. 1993. *LISREL 8 User's Reference Guide.* Chicago: Sci. Softw. Int.

Jöreskog KG, Yang F. 1995. See Marcoulides & Schumaker 1995

Kano Y. 1992. Robust statistics for test-of-independence and related structural models. *Stat. Probab. Lett.* 15:21–26

Kano Y. 1993. See Haagen et al 1993, pp. 173–90

Kano Y, Bentler PM, Mooijaart A. 1993. Additional information and precision of estimators in multivariate structural models. In *Statistical Sciences and Data Analysis,* ed. K Matusita, ML Puri, T Hayakawa, pp. 187–96. Zeist: VSP (Int. Sci. Publ.)

Kano Y, Berkane M, Bentler PM. 1990. Covariance structure analysis with heterogeneous kurtosis parameters. *Biometrika* 77: 575–85

Kennedy WJ Jr, Gentle JE. 1980. *Statistical Computing.* New York: Dekker

Kenny DA, Judd CM. 1984. Estimating the nonlinear and interactive effects of latent variables. *Psychol. Bull.* 96:201–10

Koning R, Neudecker H, Wansbeek T. 1993. See Haagen et al 1993, pp. 191–202

Krane WR, McDonald RP. 1978. Scale invariance and the factor analysis of correlation matrices. *Br. J. Math. Stat. Psychol.* 31: 218–28

Lee S, Hershberger S. 1990. A simple rule for generating equivalent models in covariance structure modeling. *Mult. Behav. Res.* 25: 313–34

Lee SY. 1985. On testing functional constraints in structural equation models. *Biometrika* 72:125–31

Lee SY. 1990. Multilevel analysis of structural equation models. *Biometrika* 77:763–72

Lee SY, Bentler PM. 1980. Some asymptotic properties of constrained generalized least squares estimation in covariance structure models. *S. Afr. Stat. J.* 14:121–36

Lee SY, Poon WY, Bentler PM. 1992. Structural equation models with continuous and polytomous variables. *Psychometrika* 57: 89–105

Lee SY, Poon WY, Bentler PM. 1995. A two-stage estimation of structural equation models with continuous and polytomous variables. *Br. J. Math. Stat. Psychol.* In press

Lockhart RS. 1967. Asymptotic sampling variances for factor analytic models identified by specified zero parameters. *Psychometrika* 32:265–77

Loehlin JC. 1992. *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis.* Hillsdale, NJ: Erlbaum

MacCallum RC, Rosnowski M, Necowitz LB. 1992. Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychol. Bull.* 111:490–504

MacCallum RC, Wegener DT, Uchino BN, Fabrigar LR. 1993. The problem of equivalent models in applications of covariance structure analysis. *Psychol. Bull.* 114:185–99

Marcoulides GA, Schumacker RE, eds. 1995. *Advanced Structural Equation Modeling Techniques.* Hillsdale, NJ: Erlbaum. In press

McArdle JJ, McDonald RP. 1984. Some algebraic properties of the Reticular Action Model for moment structures. *Br. J. Math. Stat. Psychol.* 37:234–51

McDonald RP. 1989. An index of goodness-of-fit based on noncentrality. *J. Classif.* 6: 97–103

Micceri T. 1989. The unicorn, the normal curve, and other improbable creatures. *Psychol. Bull.* 105:156–66

Mooijaart A. 1985. Factor analysis for non-normal variables. *Psychometrika* 50: 323–42

Mooijaart A. 1993. See Haagen et al 1993, pp. 249–58

Mooijaart A, Bentler PM. 1986. Random polynomial factor analysis. In *Data Analysis and Informatics IV,* ed. E Diday, Y Escoufier, L Lebart, J Pages, Y Schektman, et al, pp. 241–50. Amsterdam: Elsevier

Mooijaart A, Bentler PM. 1991. Robustness of normal theory statistics in structural equation models. *Stat. Neerland.* 45:159–71

Muthén B. 1984. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* 49:115–32

Muthén B, Kaplan D. 1992. A comparison of some methodologies for the factor analysis of nonnormal Likert variables: a note on the size of the model. *Br. J. Math. Stat. Psychol.* 45:19–30

Muthén B, Satorra A. 1989. Multilevel aspects of varying parameters in structural models. In *Multilevel Analysis of Educational Data*, ed. RD Bock, pp. 87–99. San Diego: Academic

Newey WK, McFadden D. 1994. Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, ed. RF Engle, D McFadden, 4:2112–245. Amsterdam: Elsevier Sci.

O'Brien RM, Reilly T. 1995. Equality in constraints and metric-setting measurement models. *Struct. Eq. Mod.* 2:1–12

Olkin I. 1994. See Anderson et al 1994, pp. 37–53

SAS Institute. 1990. *SAS/IML Software: Usage and References, Version 6.* Cary, NC: SAS Inst.

Satorra A. 1989. Alternative test criteria in covariance structure analysis: a unified approach. *Psychometrika* 54:131–51

Satorra A. 1992. Asymptotic robust inferences in the analysis of mean and covariance structures. In *Sociological Methodology 1992*, ed. PV Marsden, pp. 249–78. Oxford: Blackwell

Satorra A. 1993. See Haagen et al 1993, pp. 283–98

Satorra A, Bentler PM. 1986. Some robustness properties of goodness of fit statistics in covariance structure analysis. *Proc. Bus. Econ. Stat. Am. Stat. Assoc.*, pp. 549–54

Satorra A, Bentler PM. 1988. Scaling corrections for chi-square statistics in covariance structure analysis. *Proc. Am. Stat. Assoc.*, pp. 308–13

Satorra A, Bentler PM. 1990. Model conditions for asymptotic robustness in the analysis of linear relations. *Comput. Stat. Data Anal.* 10:235–49

Satorra A, Bentler PM. 1991. Goodness-of-fit test under IV estimation: asymptotic robustness of a NT test statistic. In *Applied Stochastic Models and Data Analysis*, ed. R Gutiérrez, MJ Valderrama, pp. 555–67. Singapore: World Sci.

Satorra A, Bentler PM. 1994. Corrections to test statistics and standard errors in covariance structure analysis. In *Latent Variables Analysis: Applications for Developmental Research*, ed. A von Eye, CC Clogg, pp. 399–419. Thousand Oaks, CA: Sage

Schoenberg R, Arminger G. 1990. *LINCS: A User's Guide.* Kent, WA: Aptech Syst.

Shapiro A. 1983. Asymptotic distribution theory in the analysis of covariance structures. *S. Afr. Stat. J.* 17:33–81

Shapiro A, Browne MW. 1987. Analysis of covariance structures under elliptical distributions. *J. Am. Stat. Assoc.* 82:1092–97

Sobel ME. 1995. See Arminger et al 1995, pp. 1–38

SPSS. 1990. *SPSS Reference Guide (Release 4).* Chicago: SPSS

Steiger J. 1994. See Borg & Mohler 1994, pp. 201–23

Steyer R. 1993. Principles of causal modeling: a summary of its mathematical foundations and practical steps. In *SoftStat '93: Advances in Statistical Software*, ed. F Faulbaum, pp. 107–14. Stuttgart: Fischer

Tanaka JS. 1987. "How big is big enough?": sample size and goodness of fit in structural equation models with latent variables. *Child Dev.* 58:134–46

Ullman JB. 1995. Structural equation modeling. In *Using Multivariate Statistics*, ed. BG Tabachnick, LS Fidell. New York: Harper & Row. In press

Weng LJ, Bentler PM. 1995. *Covariance Structure Analysis with Intraclass Dependent Observations*. Taipei: Natl. Taiwan Univ.

West SG, Finch JF, Curran PJ. 1995. See Hoyle 1995, pp. 56–75

Wiley DE. 1973. The identification problem for structural equation models with unmeasured variables. In *Structural Equation Models in the Social Sciences*, ed. AS Goldberger, OD Duncan, pp. 69–83. New York: Academic

Willett JB, Sayer AG. 1994. Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychol. Bull.* 116:363–81

Yuan KH. 1995. *Asymptotics for nonlinear regression models with applications.* PhD thesis. Univ. Calif., Los Angeles

Yuan KH, Bentler PM. 1995. Mean and covariance structure analysis: theoretical and practical improvements. *UCLA Stat. Ser. Rep. 180.* UCLA Cent. Stat. Los Angeles, Calif.

Yung YF, Bentler PM. 1994. Bootstrap-corrected ADF test statistics in covariance structure analysis. *Br. J. Math. Stat. Psychol.* 47:63–84

Yung YF, Bentler PM. 1995. See Marcoulides & Schumacker 1995

Annual Review of Psychology
Volume 47, 1996

CONTENTS