9

STRUCTURAL EQUATION MODELING WITH ROBUST COVARIANCES

*Ke-Hai Yuan**
*Peter M. Bentler**

Existing methods for structural equation modeling involve fitting the ordinary sample covariance matrix by a proposed structural model. Since a sample covariance is easily influenced by a few outlying cases, the standard practice of modeling sample covariances can lead to inefficient estimates as well as inflated fit indices. By giving a proper weight to each individual case, a robust covariance will have a bounded influence function as well as a nonzero breakdown point. These robust properties of the covariance estimators will be carried over to the parameter estimators in the structural model if a technically appropriate procedure is used. We study such a procedure in which robust covariances replace ordinary sample covariances in the context of the Wishart likelihood function. This procedure is easy to implement in practice. Statistical properties of this procedure are investigated. A fit index is given based on sampling from an elliptical distribution. An estimating equation approach is used to develop a variety of robust covariances, and consistent covariances of these robust estimators, needed for standard errors and test statistics, follow from this approach. Examples illustrate the inflated statistics and distorted parameter estimates obtained by using sample covariances when

This work was supported by National Institute on Drug Abuse Grants DA01070 and DA00017. We gratefully acknowledge the help of Kenneth Bollen, Jodie Ullman, and Yiu-Fai Yung in supplying the raw data sets used in this paper, as well as the valuable feedback by Adrian Raftery and three referees, which led to an improved version of the paper.
*University of California, Los Angeles

compared with those obtained by using robust covariances. The merits of each method and its relevance to specific types of data are discussed.

1. INTRODUCTION

Measurements in behavioral and social sciences generally contain random errors. Regression models and other models with independent or predictor variables that contain errors of measurements are problematic because they lead to inconsistent estimators such as estimated regression coefficients. Incorrect inference in these situations can be avoided by modeling the unobserved latent variables and the error process. Since data collections in many areas contain measurement errors, modeling with latent variables is emerging as a useful tool in diversified fields (e.g., Catalano and Ryan 1992; Kendler et al. 1992; Muthén 1992; Tosteson et al. 1989). In this paper we restrict ourselves to latent variable models of the covariance structure type since these are among the most commonly used multivariate procedures, as was shown by Gnanadesikan and Kettenring (1984) in their comprehensive survey of multivariate techniques in applications. They reported that factor analysis was the most widely used multivariate method, and that in education, psychology, and sociology, factor analysis was much more extensively used than any other multivariate method. With the advance of factor analytic models to more flexible and confirmatory models, and with the help of popular software like LISREL (Jöreskog and Sörbom 1993) and EQS (Bentler 1995), the literature on structural equation modeling has increased dramatically in the past decade (e.g., see Austin & Calderón 1996). A new journal *Structural Equation Modeling* has even been created for the development and application of this class of models.

Classical structural equation models assume normality of latent variables as well as measurement errors (Bollen 1989). In the context of covariance structure analysis, parameters in a structural model can be estimated by minimizing the Wishart likelihood function

$$F(S,\Sigma(\theta)) = \mathrm{tr}\{\Sigma^{-1}(\theta)S\} - \log|\Sigma^{-1}(\theta)S| - p \tag{1}$$

for $\hat{\theta}_n$, where p is the number of observed variables, S is the sample covariance matrix, and $\Sigma(\theta)$ is the population covariance matrix that is presumed to be generated by a hypothetical latent structure model such as a confirmatory factor model. Under these conditions, $T_{ML} = nF(S,\Sigma(\hat{\theta}_n))$ can be used as a test statistic for evaluating the quality of the structural model. Unfortunately, observed data rarely are normally distributed, implying that the assumption of normal latent variables and errors is also unlikely to be true. Conditions exist for normal theory inference to be valid for nonnormal data with some specific models (e.g., Anderson and Amemiya 1988; Amemiya and Anderson 1990; Satorra and Bentler 1990), but it is not known how to verify these conditions in practice, and T_{ML} is not reliable at all with violated conditions (Hu et al. 1992). As reviewed by Bentler and Dudgeon (1996), one solution to this problem has been the development of the asymptotically distribution free procedure of Browne (1982) and Chamberlain (1982), and, more recently, a finite sample variant thereof (Yuan and Bentler 1997). While this methodology requires no distributional assumptions, it was not developed to deal with gross pathologies in the data such as miscoded scores or outlier cases. For example, Yuan and Bentler (1996) recently showed that a small proportion of outliers leads to inflated fit indices and biases in the estimates of parameters even if the model is correct for the majority of the data. This is because the sample covariance matrix S is unduly influenced by a small proportion of outliers and can be very inefficient when the sample comes from a distribution with heavy tails (Tyler 1983). To deal with such problems, Jöreskog (1977) and Browne (1982) suggested the possibility of using a robust covariance estimator S_n in (1) instead of the sample covariance matrix S. Huba and Harlow (1987) followed up these suggestions, and reported having experimented extensively with the use of resistant covariance matrices in place of ordinary sample covariance matrices in standard structural modeling programs. In spite of these early observations, made one to two decades ago, no technical development has emerged to clarify the precise effects of substituting a robust estimator for S in (1) in the structural equation literature. For example, the recent comprehensive review of mean and covariance structure methodology by Browne and Arminger (1995) and the newly published book by Gnanadesikan (1997) on multivariate methods did not discuss this topic, which is the main focus of this paper.

There has been a limited development of robust methodology for latent variable modeling. Yamaguchi and Watanabe (1993) used a covariance S_n of MLE from a multivariate t-distribution along with (1). Their empirical results indicate that $\hat{\theta}_n$ based on this covariance matrix is more efficient than the corresponding estimator based on the sample covariance matrix S. Yuan and Bentler (forthcoming) studied methods of minimum chi-square and maximum likelihood based on an elliptical density. However, using a robust covariance S_n in (1) is another approach to robustness and can be expected to give more stable solutions than the minimum chi-square method. Another advantage of using S_n in (1) is that such an approach could be easily adapted into existing software such as EQS and LISREL, and hence can be made readily available to practitioners. This provides another reason for studying this method.

Two important concepts in describing the robustness of a statistic are the influence function and breakdown point. For an estimator T based on data from a population F, the associated influence function $IF(x)$ is a measure of the relative change in T with respect to the proportion of observations that are not from F but from a "contamination" point at the x. The $IF(x)$ gives an idea of how the estimator T responds when an extra observation at x is added to a sample. For example, when an observation x is added to a sample with size n, the proportion of x is $1/(n+1)$, the relative change in the sample mean is $\{(n\bar{X}+x)/(n+1)-\bar{X}\}/\{1/(n+1)\}$, which is unbounded in x for any n. This is because the influence function associated with $\bar{X}$ is $IF(x) = x - \mu$, the limit of the relative change as n goes to infinity, where μ is the population mean of F. The influence function associated with the sample covariance S is a quadratic function. For a sample with size n and an estimator T, the finite sample breakdown point ϵ_n^* is the smallest proportion of the n observations which can render the estimator T meaningless (e.g., becoming infinity). The (asymptotic) breakdown point of T is the limiting value ϵ^* of ϵ_n^* as n goes to infinity. For example, for both $\bar{X}$ and S, $\epsilon_n^* = 1/n$ and $\epsilon^* = 0$. Generally, a good robust estimator will have a bounded influence function as well as a nonzero breakdown point—that is, for some constant c, $|IF(x)| < c$ for any x and $\epsilon^* > 0$. We refer readers to Staudte and Sheather (1990) and Hoaglin et al. (1983) for a discussion of the theory and application of robust methods.

The unbounded influence function and zero breakdown point of the sample covariance indicate that S is very sensitive to outliers. On the other hand, most robust covariances S_n have bounded influence functions as well as nonzero breakdown points, and they are generally more efficient than S when the underlying distribution has heavy tails. Since $\hat{\theta}_n$ is obtained through minimizing (1), these nice properties of S_n will be inherited by $\hat{\theta}_n$ when S_n is used in (1). However, using a robust S_n in (1) means dropping the normality assumption, so that standard errors based on the information matrix and the likelihood ratio test T_{ML} for testing $\Sigma = \Sigma(\theta)$ cannot be used any more. In fact, even if the sample is from $N_p(\mu, \Sigma)$, S_n will be different from the sample covariance matrix S, so that inference on model structure and parameters based on normal theory cannot be used without modification. In this paper, we will study inference problems associated with using a robust S_n in equation (1).

Outside the field of structural equation modeling, robust estimation of population means and covariances has attracted much attention. In particular, the influence functions and breakdown points of a variety of estimators, including M-estimators, S-estimators, and other estimators, have been well studied (e.g., Maronna 1976; Lopuhaä 1989). Both M-estimators and S-estimators have bounded influence functions, the difference between them being that the breakdown point of an M-estimator is at most $1/(p + 1)$, which is restrictive in practice when the dimension p is high. On the other hand, the breakdown point of an S-estimator is not limited by the dimension of the data and can be as high as approximately $1/2$. Generally, the asymptotic efficiency of an S-estimator is related to its breakdown point and it is impossible to obtain a highly efficient estimator with a breakdown point near $1/2$. Among all the classes of robust estimators, M-estimators represent the primary class that has been used in practical data analysis. Within the family of elliptical distributions, Campbell (1980) and Devlin et al. (1981) empirically studied principal component analysis based on robust covariances. Being primarily illustrations of exploratory data analysis, these references are only indirectly related to structural equation modeling. However, Tyler (1983) studied how to test explicit constraints on the elements of Σ. Our work is related to his prior work as we shall show.

In order to make the following material self-contained, we would like to give some details of notations that will be used later. If A is a $p \times p$ matrix, $\text{vec}(A)$ is the p^2-dimensional vector formed by stacking the columns of A while $\text{vech}(A)$ is the $p^* = p(p + 1)/2$-dimensional vector formed by the nonduplicated elements of A when A is symmetric. For example, when $A = (a_{ij})$ is a 3×3 matrix, $\text{vec}(A) = (a_{11}\, a_{21}\, a_{31}\, a_{12}\, a_{22}\, a_{32}\, a_{13}\, a_{23}\, a_{33})'$ and $\text{vech}(A) = (a_{11}\, a_{21}\, a_{31}\, a_{22}\, a_{32}\, a_{33})'$. There exists a unique $p^2 \times p^*$ matrix D_p such that $\text{vec}(A) = D_p\, \text{vech}(A)$ and $\text{vech}(A) = D_p^+\, \text{vec}(A)$, where $D_p^+ = (D_p' D_p)^{-1} D_p'$ is the generalized inverse of D_p. If A is a $m \times n$ matrix and B is a $p \times q$ matrix, then the $mp \times nq$ matrix $(a_{ij} B)$ is called the Kronecker product of A and B and is denoted by $A \otimes B$. A good reference to these notations is Magnus and Neudecker (1988). We will use $\sigma(\theta) = \text{vech}(\Sigma(\theta))$ and $\bar{\sigma}(\theta) = \text{vec}(\Sigma(\theta))$. A function with dot on top means derivative—for example, $\dot{\bar{\sigma}}(\theta) = \partial\bar{\sigma}(\theta)/\partial\theta'$, $\dot{G}(x, \mu, \sigma) = \partial G(x, \mu, \sigma)/\partial(\mu', \sigma')$, $\ddot{F}(S_n, \Sigma(\theta)) = \partial^2 F(S_n, \Sigma(\theta))/\partial\theta\partial\theta'$, and $\dot{\Sigma}_j(\theta) = \partial\Sigma(\theta)/\partial\theta_j$, the latter being the derivative of the matrix function with respect to the jth element of θ.

When a function is evaluated at the true value of the parameter, we often replace the argument with a subscript index such as $\sigma_0 = \sigma(\theta_0)$ or $\dot{\sigma}_0 = \dot{\sigma}(\theta_0)$. There are three probabilistic notations: $t_n = o_p(a_n)$ means $a_n t_n$ approaches zero in probability as n approaches infinity; $t_n = O_p(a_n)$ means $a_n t_n$ is bounded in probability; and $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution. A classically good reference to these notations is Bishop et al. (1975, chap. 14).

In Section 2, we will study inference problems associated with using S_n. Specific types of robust covariances will be given in Section 3. Applications of these results to covariance structures for some real data sets will be presented in Section 4. In the development, we will emphasize results that are relevant to applications. For regularity conditions, we will assume that the population covariance matrix Σ_0 exists and is nonsingular, and that the model $\Sigma(\theta)$ is identified. In order to minimize technical details, we also implicitly assume some other standard regularity conditions that are hard to verify but are generally satisfied in applications. Readers can refer to Maronna (1976) and Lopuhaä (1989) for conditions on robust covariances and Yuan and Bentler (1997) for conditions on structural models.

2. INFERENCE ON STRUCTURAL MODELS

A variety of robust procedures have been proposed for estimating the mean vector and covariance matrix in a multivariate distribution. In the development of most of these procedures, an elliptically symmetric population has been a basic assumption (e.g., Maronna 1976), though this assumption is seldom checked in applications (e.g., Campbell 1980; Devlin et al. 1981). The density of an elliptical distribution has a form

$$f(x) = |\Sigma_0|^{-1/2} h\{(x-\mu_0)'\Sigma_0^{-1}(x-\mu_0)\}. \tag{2}$$

By adjusting the function $h(t)$, we can assume $\mathrm{cov}(X) = \Sigma_0$ in (2). A robust estimator $\hat{\Sigma}_n$ generally does not converge to the population covariance; it converges instead to a constant times the population covariance matrix—$\alpha\Sigma_0$, where α is a positive scalar. So the term "robust covariance" is vague, and sometimes another term, "robust dispersion," is used instead. However, the term—structural equation modeling with robust covariances—is still well defined as long as the structural model $\Sigma(\theta)$ is invariant under a constant scaling factor (ICSF). That is, for any parameter vector θ and positive constant α, there exists a parameter vector θ^* such that $\Sigma(\theta^*) =$

$\alpha\Sigma(\theta)$. So if a structural model $\Sigma(\theta)$ is ICSF, the model that holds for a covariance matrix Σ also holds for a rescaled version of the covariance matrix $\alpha\Sigma$. As noted by Browne (1984:73), "Nearly all models for covariances matrices in current use are ICSF. Amongst the few exceptions are models which require certain elements of Σ to have fixed nonzero values and restricted factor analysis models for Σ with some fixed nonzero factor loadings and fixed factor variances. These are seldom employed in practice because of the difficulty of prespecifying nonzero parameter values."

Even though an ICSF model that holds for Σ also holds for $\alpha\Sigma$, θ^* generally depends on α. So the ambiguity of robust covariance leaves its heritage onto the parameter θ^*. In order to fully appreciate the effect of this heritage, let us take the factor analysis model

$$X = \mu + \Lambda f + e, \tag{3}$$

as an example. With the typical hypothesis that f and e are uncorrelated, we have the following covariance structure

$$\Sigma(\theta) = \Lambda\Phi\Lambda' + \Psi, \tag{4}$$

where Λ is a factor loading matrix, $\Phi = \mathrm{cov}(f)$, $\Psi = \mathrm{cov}(e)$, and $\theta = (\theta_\lambda', \theta_\phi', \theta_\psi')'$ consists of the unknown elements in Λ, Φ, and Ψ, respectively. Two kinds of parameterizations in (4) are generally used in order for the model to be identified. The first is to fix all the factor variances at 1 so that Φ is a correlation matrix; the second is to fix one factor loading at 1 corresponding to each factor. Assuming that (4) is ICSF—that is, there is no prefixed nonzero element in Λ, Φ or Ψ besides those required for identification purposes—we will look at the relationship of θ^* and θ. When Φ is a correlation matrix, then

$$\alpha\Sigma(\theta) = (\sqrt{\alpha}\Lambda)\Phi(\sqrt{\alpha}\Lambda)' + \alpha\Psi,$$

so $\theta_\lambda^* = \sqrt{\alpha}\theta_\lambda$, $\theta_\phi^* = \theta_\phi$, and $\theta_\psi^* = \alpha\theta_\psi$ in this case. When fixing one factor loading at 1 for each factor, then

$$\alpha\Sigma(\theta) = \Lambda(\alpha\Phi)\Lambda' + \alpha\Psi$$

and $\theta_\lambda^* = \theta_\lambda$, $\theta_\phi^* = \alpha\theta_\phi$, and $\theta_\psi^* = \alpha\theta_\psi$ in this case. In either case, the correlations among the variables (from Σ or $\alpha\Sigma$) remain invariant; the correlations among factors (from Φ or $\alpha\Phi$) and residuals (from Ψ or $\alpha\Psi$) remain identical; and especially relevant, the standardized factor loadings, useful for interpretive reasons, remain invariant. The latter can be seen from the rescaling of $R(\theta) = D(\alpha\Sigma(\theta))D$, with $R(\theta)$ being a correlation matrix and

$$R(\theta) = (D\sqrt{\alpha}\Lambda)\Phi(\Lambda'\sqrt{\alpha}D) + \alpha D\Psi D = \tilde{\Lambda}\Phi\tilde{\Lambda}' + \alpha D\Psi D,$$

where $\tilde{\Lambda}$ is the standardized factor pattern matrix. In addition, for either of these parameterizations, there will be no effect of α on the following interesting relationships among parameters: (i) some factor loadings being zero, say, $\lambda_{ij} = 0$; (ii) comparison among factor loadings, say, $\lambda_{i_1 j_1} = \lambda_{i_2 j_2}$; and (iii) magnitude of the coefficients of reliability $\rho_i = (\mathrm{var}(x_i) - \mathrm{var}(e_i))/\mathrm{var}(x_i)$. It is also easy to see that the test statistic T_{ML} as well as a rescaled version of it that will be introduced in the next section will not be influenced by the fact that θ^* depends on α. Some functions of the parameters will change with α—e.g., the ratio of a factor loading over the factor variance—but it can be argued that this kind of function will never have any practical interest.

We have used a factor model as an example to demonstrate the effect of using a robust covariance. The effect on other types of models, such as LISREL types of models and linear covariance structure models, is similar: That is, the ambiguity of a robust covariance may change parameter estimates in a systematic way, but it does not change inference that typically has substantive interest. Actually, the effect of using a robust covariance for structural equation modeling is very similar to those of using robust covariances with other multivariate methods (e.g., Campbell 1980; Devlin et al. 1981; Tyler 1983). For example, even though the variances of principal components change with α, there will be no effect of α on discovering a lower-dimensional space or on rules used for reduction of dimensionality.

Let $s_n = \mathrm{vech}(S_n)$ be a robust covariance estimator and satisfy

$$\sqrt{n}(s_n - \xi_0) \xrightarrow{\mathcal{L}} N(0, \Gamma) \tag{5}$$

with $\xi_0 = \alpha\sigma_0$ for a positive α. We can then use S_n in equation (1) based on our above discussion. We do not need to know the value of α if the model $\Sigma(\theta)$ is ICSF. Since using κS_n and $\kappa^2\Gamma$ for any positive κ will lead to the same evaluation of the structural model and relationships among parameters of interest, we may assume $\alpha = 1$ in the rest of this paper without loss of generality. For simplicity, we will keep the notation $\sigma_0 = \sigma(\theta_0)$ in place of $\xi_0 = \xi(\theta_0^*)$. So parameter θ_0 possesses some kind of generic meaning, but it is specific with a specific robust covariance matrix estimate.

Besides the ICSF condition, condition (5) is also essential for our development. As we shall see in the next section, many robust estimators satisfy (5). It is obvious that sample covariances based on normal samples as well as those based on nonnormal samples also satisfy (5), assuming finite fourth-order moments with the latter. When condition (5) is satisfied and we minimize $F(S_n, \Sigma(\theta))$ for $\hat{\theta}_n$, then $\hat{\theta}_n$ is consistent as long as $\Sigma(\theta)$ is identified (Kano 1986). Such an $\hat{\theta}_n$ will also have the same breakdown point as that of S_n and a bounded influence function as long as the influence function of S_n is bounded. We will mainly deal with the asymptotic distribution of $\hat{\theta}_n$ and the test statistics associated with $F(S_n, \Sigma(\hat{\theta}_n))$ for evaluating the structure $\Sigma_0 = \Sigma(\theta_0)$. Let

$$g_{nj}(\theta) = \text{vec}'(\dot{\Sigma}_j(\theta))\{\Sigma^{-1}(\theta) \otimes \Sigma^{-1}(\theta)\}\text{vec}(S_n - \Sigma(\theta)) \tag{6}$$

and $g_n(\theta) = (g_{n1}(\theta), \cdots, g_{nq}(\theta))'$, where q is the number of unknown parameters in θ. Then in practice $\hat{\theta}_n$ is obtained through solving $g_n(\hat{\theta}_n) = 0$, which is the normal equation associated with minimizing (1). Using a Taylor expansion on $g_n(\hat{\theta}_n) = 0$, we obtain

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = A^{-1}(\theta_0)\sqrt{n}g_n(\theta_0) + o_p(1), \tag{7}$$

where $A(\theta) = (a_{jk}(\theta))$ with

$$a_{jk}(\theta) = \text{vec}'(\dot{\Sigma}_j(\theta))\{\Sigma^{-1}(\theta) \otimes \Sigma^{-1}(\theta)\}\text{vec}(\dot{\Sigma}_k(\theta)).$$

So matrix A corresponds to the information matrix associated with MLE based on the normality assumption. Equations (5), (6), and (7) lead to

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Omega), \tag{8}$$

where $\Omega = A^{-1}\Pi A^{-1}$ with

$$\Pi = \dot{\sigma}_0'(\Sigma_0^{-1} \otimes \Sigma_0^{-1})D_p \Gamma D_p'(\Sigma_0^{-1} \otimes \Sigma_0^{-1})\dot{\sigma}_0.$$

When $S_n = S$ is the sample covariance based on a sample from $N(\mu, \Sigma_0)$, then $\Pi = 2A$, and (8) characterizes the asymptotic distribution of the normal theory MLE. When $S_n = S$ but the data are nonnormal, Ω in (8) is of the form of the well-known sandwich type covariance matrix (e.g., Browne 1984; Bentler and Dijkstra 1985; Arminger and Schoenberg 1989; Arminger and Sobel 1990). Generally, $\Pi \neq 2A$ for a robust S_n even if the sample is from a normal distribution. Inference on parameters can be done with (8) using a consistent estimator $\hat{\Omega}$, which is obtained in practice by replacing A and Π by their consistent estimates.

Because latent variable models usually involve many structural assumptions and hypothetical latent variables that cannot be observed, it is critical, especially in applications, to have a goodness-of-fit index that permits an evaluation of the hypothesized model $\Sigma = \Sigma(\theta)$. This can be obtained from knowledge of the distribution of $nF(S_n, \Sigma(\hat{\theta}_n))$, as will be developed next. Using a Taylor expansion on $F(S_n, \Sigma(\hat{\theta}_n))$ at θ_0 and with $\Sigma_0 = \Sigma(\theta_0)$, we obtain

$$F(S_n, \Sigma(\hat{\theta}_n)) = F(S_n, \Sigma_0) + \dot{F}'(S_n, \Sigma_0)(\hat{\theta}_n - \theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)'\ddot{F}(S_n, \Sigma(\bar{\theta}_n))(\hat{\theta}_n - \theta_0), \tag{9}$$

where $\bar{\theta}_n$ is a vector lying between θ_0 and $\hat{\theta}_n$. Notice that

$$-\log|\Sigma_0^{-1} S_n| = -\log|I_p - (I_p - \Sigma_0^{-1} S_n)|,$$

using (5) and a matrix version of Taylor expansion (e.g., Muirhead 1982: 363, eq. 15),

$$-\log|\Sigma_0^{-1} S_n| = \mathrm{tr}(I_p - \Sigma_0^{-1} S_n) + \frac{1}{2}\mathrm{tr}(I_p - \Sigma_0^{-1} S_n)^2 + O_p\left(\frac{1}{n^{3/2}}\right). \tag{10}$$

It follows from (10) that

$$F(S_n, \Sigma_0) = \frac{1}{2}\mathrm{tr}(I_p - \Sigma_0^{-1} S_n)^2 + O_p\left(\frac{1}{n^{3/2}}\right). \tag{11}$$

Similarly,

$$\dot{F}'(S_n, \Sigma_0)(\hat{\theta}_n - \theta_0) = -\mathrm{vec}'(S_n - \Sigma_0)P\,\mathrm{vec}(S_n - \Sigma_0) + o_p\left(\frac{1}{n}\right) \tag{12}$$

and

$$\frac{1}{2}(\hat{\theta}_n - \theta_0)'\ddot{F}(S_n, \Sigma(\bar{\theta}_n))(\hat{\theta}_n - \theta_0)$$
$$= \frac{1}{2}\operatorname{vec}'(S_n - \Sigma_0)P\operatorname{vec}(S_n - \Sigma_0) + o_p\left(\frac{1}{n}\right), \tag{13}$$

where

$$P = (\Sigma_0^{-1} \otimes \Sigma_0^{-1})\dot{\sigma}_0\{\dot{\sigma}_0'(\Sigma_0^{-1} \otimes \Sigma_0^{-1})\dot{\sigma}_0\}^{-1}\dot{\sigma}_0'(\Sigma_0^{-1} \otimes \Sigma_0^{-1}).$$

By putting (11), (12), and (13) into (9), we obtain

$$nF(S_n, \Sigma(\hat{\theta}_n)) = Z_n' Q Z_n + o_p(1),$$

where $Z_n = \Gamma^{-1/2}\sqrt{n}\operatorname{vech}(S_n - \Sigma_0)$ and

$$Q = \frac{1}{2}\Gamma^{1/2} D_p'\{(\Sigma_0^{-1} \otimes \Sigma_0^{-1}) - P\}D_p\Gamma^{1/2}. \tag{14}$$

It follows from (5) that $Z_n \xrightarrow{\mathcal{L}} N_{p^*}(0, I)$. When $S_n = S$ is the sample covariance based on a sample from $N_p(\mu, \Sigma)$, $\Gamma = 2D_p^+(\Sigma_0 \otimes \Sigma_0)D_p^{+\prime}$, Q is a projection matrix of rank $p^* - q$ and consequently, $nF(S, \Sigma(\hat{\theta}_n)) \xrightarrow{\mathcal{L}} \chi^2_{p^*-q}$. Generally, Q is only a nonnegative definite matrix of rank $p^* - q$. Let $\lambda_1 \geq \cdots \geq \lambda_{p^*-q} > 0$ be the nonzero eigenvalues of Q, then

$$nF(S_n, \Sigma(\hat{\theta}_n)) \xrightarrow{\mathcal{L}} \sum_{j=1}^{p^*-q} \lambda_j \chi^2_{1j}, \tag{15}$$

where χ^2_{1j} are independent chi-square variates with degree of freedom 1. Unless $\lambda_1 = \cdots = \lambda_{p^*-q}$, no commonly used distribution is available to describe the behavior of the right-hand side of (15), which is a mixture of chi-square distributions. Several approximations to a mixture of chi-square distributions were developed by Box (1954) and Satterthwaite (1941). Another was studied by Bentler (1994) for approximating (15) with $S_n = S$ based on a sample from a nonnormal distribution. When S is based on a sample from an elliptical distribution, Browne (1984) proposed a rescaling factor to $nF(S, \Sigma(\hat{\theta}_n))$ using Mardia's coefficient of kurtosis (Mardia et al. 1979:31). Satorra and Bentler (1994) proposed a more general rescaling factor to the likelihood ratio statistic. With a rescaling factor, Tyler (1983) studied the likelihood ratio test for a specific function of elements of the population covariance matrix. As we shall see in the next section, within the family of elliptical distributions, a similar correction applies to $nF(S_n, \Sigma(\hat{\theta}_n))$.

3. ROBUST MEANS AND COVARIANCE MATRICES

We will emphasize two classes of robust estimators of mean and covariance in this section. One is the class of M-estimators, the other is the class of S-estimators. This is because robust properties of these estimators are well studied and they satisfy condition (5), which is necessary for their applications in covariance structure analysis. We will use an estimating equation approach in presenting these estimators. An advantage of this approach is that a consistent estimator for the matrix Γ can be obtained by identifying the corresponding estimating equation for each estimator. In addition, we will state a common property of these estimators within the family of elliptical distributions. This will motivate us to find a way to make use of $nF(S_n, \Sigma(\hat{\theta}_n))$ in testing the quality of the model. Some of these estimators will be used in our examples in the next section.

Let $X_1, \cdots, X_n$ be a given sample; many robust estimators $(\hat{\mu}_n, \hat{\Sigma}_n)$ can then be obtained by simultaneously solving

$$\frac{1}{n}\sum_{i=1}^{n} G_1(X_i, \hat{\mu}_n, \hat{\Sigma}_n) = 0 \tag{16}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} G_2(X_i, \hat{\mu}_n, \hat{\Sigma}_n) = 0, \tag{17}$$

where $G_1(x, \mu, \Sigma)$ is a $p \times 1$ vector function and $G_2(x, \mu, \Sigma)$ is a $p \times p$ matrix function. For example, if we let

$$d(x, \mu, \Sigma) = \{(x - \mu)'\Sigma^{-1}(x - \mu)\}^{1/2},$$

an M-estimator will then correspond to

$$G_1(x, \mu, \Sigma) = u_1\{d(x, \mu, \Sigma)\}(x - \mu) \tag{18}$$

and

$$G_2(x,\mu,\Sigma) = u_2\{d^2(x,\mu,\Sigma)\}(x-\mu)(x-\mu)' - \Sigma, \tag{19}$$

where $u_1(t)$ and $u_2(t)$ are univariate weight functions that will be given when discussing different estimators below. Equations such as (16) and (17) are often called estimating equations because they are used to define parameter estimators. A nice property of using these estimating equations is that the asymptotic distribution of $(\hat{\mu}_n, \hat{\Sigma}_n)$ can be presented in a unified way. If

$$G(x,\mu,\sigma) = (G_1'(x,\mu,\Sigma), \mathrm{vech}'\{G_2(x,\mu,\Sigma)\})',$$

then

$$\sqrt{n}\begin{pmatrix}\hat{\mu}_n - \mu_0\\ \hat{\sigma}_n - \sigma_0\end{pmatrix} \xrightarrow{\mathcal{L}} N(0,V), \tag{20}$$

where $V = H^{-1}BH'^{-1}$ with

$$H = \mathrm{E}\{\dot{G}(X,\mu_0,\sigma_0)\} \qquad \text{and} \qquad B = \mathrm{E}\{G(X,\mu_0,\sigma_0)G'(X,\mu_0,\sigma_0)\}.$$

Note that (20) holds as long as μ_0 and Σ_0 satisfy $\mathrm{E}\{G(X,\mu_0,\sigma_0)\} = 0$ and the expectations with H and B exist. Also, since a proper weight is attached to each individual case, we generally need to assume only that the second moment of the sampling population exists in order for B to exist for most of the robust estimators. On the other hand, we need to assume that the fourth-order moment will be finite for the sampling distribution when using the sample covariance. A consistent estimate of V can be obtained by using consistent estimates for H and B; these are given by

$$\hat{H}_n = \frac{1}{n}\sum_{i=1}^{n}\dot{G}(X_i,\hat{\mu}_n,\hat{\sigma}_n) \qquad \text{and}$$

$$\hat{B}_n = \frac{1}{n}\sum_{i=1}^{n}G(X_i,\hat{\mu}_n,\hat{\sigma}_n)G'(X_i,\hat{\mu}_n,\hat{\sigma}_n).$$

Let V_{22} be the submatrix of V corresponding to the asymptotic covariance of $\sqrt{n}(\hat{\sigma}_n - \sigma_0)$; then $\Gamma = V_{22}$ and $\hat{\Gamma} = \hat{V}_{22}$, which will be used in our applications.

Different G_1 and G_2 correspond to different estimators. Within the class of M-estimators with G_1 and G_2 being given by (18) and (19), a variety of weight functions $u_1(t)$ and $u_2(t)$ lead to different types of M-estimators (e.g., Hoaglin et al. 1983). The well-known Huber-type M-estimator corresponds to

$$u_1(d) = \begin{cases} 1, & \text{if } d \le r \\ r/d, & \text{if } d > r \end{cases}$$

and $u_2(d^2) = \{u_1(d)\}^2/\beta$ (e.g., Tyler 1983), where r^2 is given by $P(\chi_p^2 > r^2) = \alpha$, α is the percentage of outliers one wants to control in the data, and β is a constant such that $E\chi_p^2 u_2(\chi_p^2) = p$, which makes the estimator $\hat{\Sigma}_n$ unbiased if sampling from a p-variate normal distribution. If using $u_1(t) = u_2(t^2) = -2\dot{h}(t^2)/h(t^2)$ in (18) and (19), we obtain the maximum likelihood estimator of (μ, Σ) based on the elliptical density in (2). When a p-variate t-distribution with degrees of freedom m is used, $u_1(d) = u_2(d^2) = (p + m)/(m + d^2)$. The $\hat{\Sigma}_n$ based on such a weight can be rescaled by $S_n = \kappa\hat{\Sigma}_n$ so that S_n is consistent for the population covariance when sampling from a multivariate normal distribution. The constant κ is a solution to $E\{(\kappa\chi_p^2)/(m + \kappa\chi_p^2)\} = p/(m + p)$. However, the constants β and κ are unnecessary if only considering inference on models and parameters as long as $\Sigma(\theta)$ is ICSF. We shall use these scalings in our examples in the next section just for obtaining comparable parameter estimates.

For the functions given in (18) and (19), $(\hat{\mu}_n, \hat{\Sigma}_n)$ defined in (16) and (17) satisfy

$$\mu = \sum_{i=1}^{n} u_1\{d(X_i, \mu, \Sigma)\}X_i \Big/ \sum_{i=1}^{n} u_1\{d(X_i, \mu, \Sigma)\},$$

and

$$\Sigma = \sum_{i=1}^{n} u_2\{d^2(X_i, \mu, \Sigma)\}(X_i - \mu)(X_i - \mu)'/n.$$

The above two equations are the usual way to define an M-estimator and give an iterative algorithm for obtaining $(\hat{\mu}_n, \hat{\Sigma}_n)$. Motivated by the unbiasedness of the sample covariance, Campbell (1980) defined another form of M-estimator:

$$\mu = \sum_{i=1}^{n} w_i X_i \Big/ \sum_{i=1}^{n} w_i \tag{21}$$

and

$$\Sigma = \sum_{i=1}^{n} w_i^2 (X_i - \mu)(X_i - \mu)' \bigg/ \left(\sum_{i=1}^{n} w_i^2 - 1 \right), \tag{22}$$

where $w_i = w\{d(X_i, \mu, \Sigma)\}$ for a weight function $w(t)$. The estimators defined in (21) and (22) correspond to the estimating equations (16) and (17) with

$$G_1(x, \mu, \Sigma) = w(x, \mu, \sigma)(x - \mu)$$

and

$$G_2(x, \mu, \Sigma) = w^2(x, \mu, \sigma)(x - \mu)\{(x - \mu)(x - \mu)' - \Sigma\} + \frac{1}{n}\Sigma.$$

Campbell used $w(d) = \omega(d)/d$ and

$$\omega(d) = \begin{cases} d, & \text{if } d \leq d_0 \\ d_0 \exp\{-\frac{1}{2}(d - d_0)^2/b_2^2\}, & \text{if } d > d_0 \end{cases}, \tag{23}$$

where $d_0 = \sqrt{p} + b_1/\sqrt{2}$, b_1 and b_2 are constants. Based on extensive empirical experience, proposed choices for b_1 and b_2 were given by Campbell (1980): (a) $b_1 = \infty$ corresponding to the usual sample covariance; (b) $b_1 = 2$, $b_2 = \infty$ corresponding to a Huber-type M-estimator; (c) $b_1 = 2$, $b_2 = 1.25$ corresponding to a Hampel-type redescending M-estimator (Hampel 1974).

Since the breakdown point of an M-estimator is limited by the dimension of the data, other types of estimators with high breakdown points have been proposed. One such estimator is the well-studied S-estimator. Let $\rho(t)$ be a continuously differentiable symmetric function that is also strictly increasing on $[0, c_0]$ and constant on $[c_0, \infty]$. For a constant b_0 ($0 < b_0 < a_0 = \rho(c_0)$), an S-estimator of (μ, Σ) is defined by minimizing $|\Sigma|$ subject to the constraint

$$\frac{1}{n}\sum_{i=1}^{n} \rho\{d(X_i, \mu, \Sigma)\} = b_0.$$

The breakdown point of the S-estimator is approximately given by $b_0/\rho(c_0)$ and can be as large as $1/2$. Using Lagrange multipliers, Lopuhaä (1989) showed that an S-estimator $(\hat{\mu}_n, \hat{\Sigma}_n)$ satisfies the estimating equations (16) and (17) with

$$G_1(x,\mu,\Sigma) = u\{d(x,\mu,\Sigma)\}(x-\mu)$$

and

$$G_2(x,\mu,\Sigma) = pu\{d(x,\mu,\Sigma)\}(x-\mu)(x-\mu)' - v(d)\Sigma,$$

where $u(t) = \dot{\rho}(t)/t$ and $v(t) = t\dot{\rho}(t) - \rho(t) + b_0$. A recommended $\rho(t)$ is Tukey's biweight function (e.g., Ruppert 1992)

$$\rho_{c_0}(t) = \begin{cases} t^2/2 - t^4/(2c_0^2) + t^6/(6c_0^4), & \text{if} \quad t \le c_0 \\ c_0^2/6, & \text{if} \quad t > c_0 \end{cases}.$$

Note that equations (16) and (17) corresponding to an S-estimator will have multiple solutions. A practical algorithm is needed for solving these equations in order to obtain the S-estimator. We will use the SURREAL algorithm developed by Ruppert (1992) for our examples in the next section.

The family of elliptical distributions has been well explored in robustness studies. This is because the distributions represented by (2) include multivariate normal as well as many multivariate nonnormal distributions with heavy or light tails. Assuming that the sample is from (2), then for both M- and S-estimators the asymptotic distribution of $(\hat{\mu}_n, \hat{\Sigma}_n)$ has a nice property. In particular, $\hat{\mu}_n$ and $\hat{\Sigma}_n$ are asymptotically independent with

$$\sqrt{n}(\hat{\mu}_n - \mu_0) \xrightarrow{\mathcal{L}} N(0, c\Sigma) \qquad \text{and} \qquad \sqrt{n}(\hat{\sigma}_n - \sigma_0) \xrightarrow{\mathcal{L}} N(0, \Gamma),$$

where

$$\Gamma = 2aD_p^+(\Sigma_0 \otimes \Sigma_0)D_p^{+\prime} + bD_p^+ \operatorname{vec}(\Sigma_0)\operatorname{vec}'(\Sigma_0)D_p^{+\prime},$$

and a, b, and c are constants depending on the specific G_1 and G_2 used in the estimating equations (16) and (17). Specific forms of a, b, and c can be found in Maronna (1976) and Tyler (1982) for M-estimators and in Lopuhaä (1989) for S-estimators; these are not required in our applications. Within the elliptical family, the statistic $nF(S_n, \Sigma(\hat{\theta}_n))$ also converges to a nice form if the structure $\Sigma(\theta)$ satisfies

$$\Sigma(\theta_0) = c_1\dot{\Sigma}_1(\theta_0) + \cdots + c_q\dot{\Sigma}_q(\theta_0) \tag{24}$$

for some constants $c_1, \cdots, c_q$. Condition (24) is implied by the ICSF condition and is consequently satisfied by almost all structural models in current use. Under condition (24), it can be easily shown that all the nonzero eigenvalues of the matrix Q in (14) are equal to a. In such a case,

$$nF(S_n, \Sigma(\hat{\theta}_n))/a \xrightarrow{\mathcal{L}} \chi^2_{p^*-q}.$$

A consistent estimate of a is given by $\hat{a} = \mathrm{tr}(\hat{Q})/(p^* - q)$. So

$$T_1 = nF(S_n, \Sigma(\hat{\theta}_n))/\hat{a} \tag{25}$$

can be used as a statistic for testing the hypothetical structure $\Sigma(\theta)$. When the sample is not from an elliptically symmetric distribution, T_1 generally will not approach $\chi^2_{p^*-q}$ but will approach a distribution with mean $p^* - q$ instead. When $S_n = S$ is the sample covariance and $\hat{\Gamma} = S_Y$ is the sample covariance of $Y_i = \mathrm{vech}\{(X_i - \overline{X})(X_i - \overline{X})'\}$, T_1 is equivalent to the statistic T_{SB} proposed by Satorra and Bentler (1994). Existing simulation studies indicate that T_{SB} is very insensitive to various violations of elliptical symmetry of the underlying distributions (Hu et al. 1992; Curran et al. 1996). We suspect a similar behavior for T_1 using a robust covariance matrix S_n. More research in this direction is needed.

4. EMPIRICAL EXAMPLES

We shall use some real data sets to demonstrate applications of the procedures developed in the last two sections. In particular, we shall compare estimates and test statistics by using different robust estimators of covariances.

The first is a classical data set from Holzinger and Swineford (1939). The data set consists of mental ability tests scores of seventh- and eighth-grade children from two different schools. There are 26 variables and 145 subjects from the Grant-White school. Jöreskog (1969) used 9 of the 26 variables in studying the correlation structure with normal theory maximum-likelihood method. We shall also use these 9 variables in our application. The 9 variables are 1. Visual Perception, 2. Cubes, 3. Lozenges, 4. Paragraph Comprehension, 5. Sentence Completion, 6. Word Meaning, 7. Addition, 8. Counting Dots, and 9. Straight-Curved Capitals. In Holzinger and Swineford's original report, variables 1, 2, and 3 were designed to measure the spatial ability of the subjects; variables 4, 5, and 6 were designed to measure the verbal ability of the subjects; and variables 7, 8, and 9 were designed to measure the speed factor of the subjects in performing the tasks. Let X represent the 9 observed variables; the confirmatory factor model, presented in (3), then represents the hypothesis of the original design with

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{42} & \lambda_{52} & \lambda_{62} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{73} & \lambda_{83} & \lambda_{93} \end{pmatrix}'. \qquad (26)$$

We assume that the measurement errors are uncorrelated, with $\Psi = \mathrm{cov}(e)$ being a diagonal matrix; for identification purposes, we also set Φ in (4) to be a 3×3 correlation matrix. So there are $q = 21$ unknown parameters in θ, nine of which are factor loadings.

Previous analyses of this classical data set have always assumed normality, and hence we may wonder if a robust method is necessary. As is well known, a robust method generally gives smaller weights to observations that deviate from the majority of the data. After working on a number of real and simulated data sets, Campbell (1980) concluded that the ith case is an atypical observation if the associated weight w_i^2 in (22) is less than .30 with $b_1 = 2.0$ and $b_2 = 1.25$. The two smallest weights for this data set are $w_{24}^2 = .072$ and $w_{106}^2 = .121$. So cases 24 and 106 may indicate the long tails of the underlying distribution of this data set, and a robust covariance estimator may provide a better estimate of the population covariance.

Several M-estimates and an S-estimate of covariances are used in (1) for fitting the factor model to this data set. The three M-estimates are based on the multivariate t-distribution with degree of freedom 1 [Mt(1)], the Huber-type weight with $q = .2$ [Huber(.2)], and Campbell's weight with $b_1 = 2.0$ and $b_2 = \infty$ [Campbell$(2,\infty)$]. For the S-estimate, the biweight function with $c_0 = 11.105$ is used; this corresponds to a breakdown point of .2. For comparison purposes, we also include the normal theory method with the sample covariance matrix S. The estimated factor loadings with their standard errors as well as the fit indices based on (25) are given in Table 1. As can be seen in the top part of the table, all the covariances give a similar pattern in estimates of factor loadings. With respect to fit indices, given in the last row of the table, the normal theory method gives the largest test statistic for the structural model with $T_{ML} = 51.19$. Referring to the chi-square distribution with 24 degrees of freedom, this corresponds to a p-value of .001. So we may reject the model structure in (4) and (26) if based on the normal theory method. However, using the S_n based on the multivariate Cauchy density, [Mt(1)] yields a fit statistic T_1 that corresponds to a p-value of .03, indicating that the factor model is a

TABLE 1
Fit Indices and Estimated Factor Loadings

λ_{ij}	Mt(1)	Huber(.2)	Campbell(2,∞)	S(.2)	ML
λ_{11}	4.606 (.693)	4.709 (.649)	4.678 (.795)	4.702 (.691)	4.678 (.624)
λ_{21}	1.966 (.421)	2.160 (.392)	2.258 (.395)	2.244 (.387)	2.296 (.408)
λ_{31}	5.679 (.864)	5.494 (.749)	5.681 (.761)	5.730 (.754)	5.769 (.751)
λ_{42}	2.703 (.301)	2.789 (.256)	2.893 (.266)	2.880 (.258)	2.922 (.237)
λ_{52}	3.876 (.396)	3.857 (.350)	3.876 (.336)	3.888 (.340)	3.856 (.333)
λ_{62}	6.314 (.697)	6.262 (.568)	6.457 (.604)	6.482 (.588)	6.567 (.569)
λ_{73}	15.37 (2.197)	15.67 (1.972)	15.71 (1.904)	15.79 (1.905)	15.68 (2.012)
λ_{83}	17.15 (1.824)	15.77 (1.668)	16.33 (1.761)	16.57 (1.743)	16.71 (1.752)
λ_{93}	22.30 (3.427)	24.15 (2.996)	25.01 (3.082)	25.05 (3.082)	25.97 (3.117)
Test	38.85	43.10	44.40	46.50	51.19

Source: Based on Psychological Data from Holzinger and Swineford (1939).

reasonable model. This supports the design of the original measurements. The statistics corresponding to the other three types of robust covariances correspond to p-values of .01, .007, and .004, respectively, which suggests that the model in (4) and (26) ranges from acceptable to barely acceptable. This reflects differences among the various robust covariances.

For this example, we used the model (4) with Λ in (26), which represents the original design by Holzinger and Swineford (1939). Other models may also provide a good explanation of the measured variables, as described in Jöreskog (1969). Our purpose here is to show that when abnormal observations exist in a data set, using robust covariances with corrected inference procedures may lead to a better evaluation of model structure. On the other hand, the normal theory method is easily influenced by outliers so that a reasonable model that would fit the majority of the data may be discredited by a few influential observations.

The data for our second example comes from Dukes et al. (1995), who studied the effect of the Drug Abuse Resistance Education (D.A.R.E.) program based on data obtained across four years from 440 classrooms and 10,000 students. Outcomes were evaluated by Dukes et al. on the classroom level, using a Solomon four-group design with latent variables. For each cohort, the design had an experimental group and a control group that were pretested and posttested, as well as an experimental group and a control group that were not pretested (posttest only), permitting an isolation of pretest on posttest as well as maturation effects. Here we study only the posttest experimental group (group C) with sample size $n = 122$, using a standard confirmatory factor analysis model based on 12 variables and 4 correlated common factors. The factors in this model represent self-esteem; resistance to peer pressure; family, police, and teacher bonds; and acceptance of risky behavior, and have 2, 2, 3, and 5 univocal indicators. The model was justified by theoretical considerations, and it gives a reasonable explanation of their data. As in the analysis of the first example, we want to see whether the use of a robust method yields different conclusions from that found with the normal theory method. Using Campbell's weight scheme with $b_1 = 2.0$ and $b_2 = 1.25$ (Campbell[2,1.5]), the three smallest weights w_i^2 as in (22) are $w_{99}^2 = 6.34 \times 10^{-7}$, $w_{79}^2 = .041$, and $w_{43}^2 = .14$. So cases 99, 79, and 43 are definitely atypical observations based on the criterion of Campbell (1980). This may indicate that a robust covariance approach would be useful for analyzing this data set.

As in the first example, four robust covariances as well as the sample covariance are used to analyze this data set. The four robust covariances are based on Huber(.2), Campbell$(2,\infty)$, Mt(1), and S(.2) with Tukey's biweight function $\rho(t)$. The estimates of fit indices and factor loadings are given in Table 2. The estimates of factor loadings by different methods are comparable. Basically, all the methods give similar solutions to the estimates. As in our first example, there exist large and meaningful differences among the fit indices. The statistic corresponding to the normal theory method is $T_{ML} = 102.10$ with a p-value $= 8.9 \times 10^{-6}$ when referred to the chi-square distribution with 48 degrees of freedom. This implies that the model hardly fits the data. On the other hand, the robust covariances obtained by Mt(1) and Campbell$(2,\infty)$ lead to test statistics 79.28 and 78.62 respectively, much smaller though still technically significant. However, all the fit statistics based on robust covariances are below two times the degrees of freedom, while the normal theory test statistic is much larger than 2×44.

The third data set is from Bollen (1989:30–31). It consists of three estimates of percent cloud cover for 60 slides. This data set was introduced for outlier identification purposes. Bollen and Arminger (1991) further used a one-factor model to fit this data to study observational residuals in factor analysis. An interesting feature of this data set is that using the

TABLE 2
Fit Indices and Estimates of Factor Loadings

λ_{ij}	Mt(1)	Huber(.2)	Campbell$(2,\infty)$	S(.2)	ML
λ_{11}	.689 (.165)	.673 (.122)	.698 (.125)	.698 (.120)	.690 (.118)
λ_{21}	.239 (.079)	.280 (.064)	.324 (.086)	.316 (.069)	.330 (.060)
λ_{32}	.155 (.028)	.142 (.027)	.153 (.038)	.152 (.030)	.153 (.029)
λ_{42}	.269 (.037)	.279 (.037)	.289 (.038)	.289 (.035)	.287 (.034)
λ_{53}	.199 (.029)	.192 (.026)	.189 (.027)	.192 (.026)	.186 (.026)
λ_{63}	.271 (.034)	.251 (.029)	.250 (.033)	.254 (.030)	.247 (.030)
λ_{73}	.206 (.026)	.182 (.023)	.183 (.023)	.188 (.023)	.183 (.022)
λ_{84}	.090 (.012)	.091 (.011)	.096 (.012)	.096 (.012)	.096 (.010)
λ_{94}	.076 (.009)	.073 (.008)	.075 (.008)	.076 (.008)	.075 (.010)
$\lambda_{10,4}$	.087 (.011)	.082 (.009)	.081 (.009)	.083 (.009)	.080 (.012)
$\lambda_{12,4}$	.118 (.013)	.115 (.013)	.117 (.013)	.119 (.014)	.118 (.013)
$\lambda_{11,4}$	.093 (.013)	.090 (.012)	.093 (.012)	.094 (.012)	.094 (.009)
Test	79.28	81.82	78.62	87.92	102.10

Source: Based on Drug Abuse Resistance Education Data D.A.R.E., from Dukes et al. (1995).

typical sample covariance matrix based on all the 60 cases leads to an improper solution (a negative error variance) with

$$\Lambda = (1.0,.97,1.18)', \qquad \Phi = 1052, \qquad \text{and}$$
$$\Psi = \mathrm{diag}(249,474,-51.4).$$

After removing the three cases (52, 40, and 51) corresponding to the three largest residuals in Bollen and Arminger's (1991) analysis, the solution to the factor model is

$$\Lambda = (1.0,1.14,1.12)', \qquad \Phi = 1023, \qquad \text{and}$$
$$\Psi = \mathrm{diag}(106,157,58.2).$$

Notice that the outliers have a big effect on the estimates of error variances but little effect on the factor loadings. This phenomenon is also reflected in using robust covariances for this data set.

Using Campbell(2,1.25), the three smallest weights are $w_{52}^2 = 4.2 \times 10^{-7}$, $w_{40}^2 = 2.1 \times 10^{-5}$, and $w_{51}^2 = 2.1 \times 10^{-3}$. So cases 52, 40, and 51 are also atypical by Campbell's criterion. All other weight functions also give these three cases the smallest weights. In order to fully understand the mechanism of the different robust procedures, we list the five smallest weights corresponding to several commonly used robust covariances: $u_2(d_i^2)$ in the Huber(.2); $u_2(d_i^2)$ in Mt(1) and Mt(5); $u(d_i)$ in S(.2); and w_i^2 in Campbell(2,1.5) and Campbell(2,∞). Since using the sample covariance corresponds to giving each individual a weight of 1, and each removed outlier corresponds to a weight of 0, we also add these in Table 3 for better comparison. There ML corresponds to the maximum likelihood weights based on the sample covariance, and ORML corresponds to outlier removal followed by maximum likelihood. Even though all the methods give the same order of weights for these five most outlying cases, the weights change in different ways according to different methods. Among the six robust procedures, only S(.2) gives a 0 weight to the most outlying case 52, which is equivalent to removing this case as an outlier. This may reflect the fact that an S-estimator has a higher breakdown point than a general M-estimator. The weights associated with multivariate t-density changes in the smoothest way, reflecting that the t-distribution admits a long tail in the underlying distribution of the sample. Sitting between these two extremes are the Huber-type weights and Campbell's weights. The Campbell(2,1.5) gives the three

TABLE 3
Five Smallest Weights in Various Covariances

Method	Case Number 52	40	51
Huber(.2)	.089	.095	.154
Mt(1)	6.55×10^{-4}	7.22×10^{-4}	1.09×10^{-3}
Mt(5)	.107	.118	.173
Campbell(2,1.5)	4.22×10^{-7}	2.14×10^{-5}	2.10×10^{-3}
Campbell(2,∞)	.337	.371	.575
S(.2)	0	5.70×10^{-3}	.119
ML	1	1	1
ORML	0	0	0

Method	Case Number 31	43
Huber(.2)	.272	.353
Mt(1)	1.92×10^{-3}	2.69×10^{-3}
Mt(5)	.293	.346
Campbell(2,1.5)	.696	.742
Campbell(2,∞)	1	1
S(.2)	.540	.565
ML	1	1
ORML	1	1

Source: Based on cloud cover data from Bollen (1989).

most outlying cases minuscule, though not zero weights, while Huber(.2) and Campbell(2,∞) change the weights in a relatively smooth way.

The estimated parameters based on the robust covariances are given in Table 4. As was done by Bollen and Arminger (1991), we fix the first factor loading at 1.0 for identification purposes. From these estimates, we can see that all methods give very similar estimates of factor loadings but that significant differences exist among the various estimates of error variances and factor variance. These can be compared with the difference observed when using the sample covariances with and without outliers, as is again added in the last two rows. The smallest estimates of factor variance and error variances are given by Mt(1), while the largest of these estimates are given by S(.2) and Campbell(2,∞). These reflect the different weights used by the various robust methods. Rescaling the substantially smaller

TABLE 4
Estimates of Model Parameters

Method	λ_2	λ_3	Φ	ψ_1	ψ_2	ψ_3
Huber(.2)	1.067	1.068	1148	69.67	136.7	38.44
Mt(1)	1.118	1.051	31.31	1.198	2.388	1.063
Mt(5)	1.101	1.094	1137	84.85	161.3	31.90
Campbell(2,1.5)	1.136	1.113	1040	99.90	147.9	57.74
Campbell(2,∞)	1.059	1.142	1045	163.8	305.6	13.88
S(.2)	1.128	1.110	1417	125.6	198.8	65.32
ML	.970	1.176	1052	248.8	473.8	−51.44
ORML	1.143	1.123	1023	105.8	157.3	58.15

Source: Based on cloud cover data from Bollen (1989).

estimates of factor and error variances by Mt(1), as is permissible, would make the values more similar to those by the other methods; however, the model-implied correlations among the variables would not change so there is not much reason to do this. Also notice that proportions of factor variance to error variances do not differ strikingly among methods, so there is not much difference among methods regarding model inference. There is zero degree of freedom associated with the model, so the statistic T_1 is not relevant here.

Our fourth example is the industrialization and political democracy panel data set introduced by Bollen (1989), who studied various models for this data set. Bollen and Arminger (1991) further used this data set to study observational residuals in structural equation models. This data set consists of eight political democracy variables $Y = (y_1, \cdots, y_8)'$ and three industrialization variables $X = (x_1, x_2, x_3)'$ in 75 developing countries up to the 1960s. The variables y_1 to y_4 are indicators of political democracy in 1960, and y_5 to y_8 are the same variables measured in 1965. Assuming that political democracy in 1965 is a function of 1960 political democracy and industrialization, and that the 1960 industrialization level also affects the 1960 political democracy level, the model proposed by Bollen (1989) is

$$Y = \Lambda_Y \eta + \varepsilon, \quad X = \Lambda_X \xi + \delta,$$

and

$$\eta = B\eta + \Gamma\xi + \zeta,$$

where

$$\Lambda_Y = \begin{pmatrix} 1 & \lambda_1 & \lambda_2 & \lambda_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda_1 & \lambda_2 & \lambda_3 \end{pmatrix}', \quad \Lambda_X = (1 \quad \lambda_4 \quad \lambda_5)',$$

$$B = \begin{pmatrix} 0 & 0 \\ \beta_{21} & 0 \end{pmatrix}, \quad \Gamma = (\gamma_{11} \quad \gamma_{21})',$$

ε, δ, and ζ are vectors of errors. It is allowed that some error terms in ε are correlated based on theoretical considerations; see Bollen (1989) for details about the data and the model. Our interest here is to see the estimates and fit indices associated with different methods.

Using Campbell(2,1.25) on this data set finds $w_i^2 = 1$ for all the cases, which suggests no particular atypical cases. This is in accord with the findings by Bollen and Arminger (1991). Similarly, no case is heavily downweighted by any of the other weight functions—for example, the smallest weight with Huber(.2) and S(.2) are .76 and .71, respectively. The test statistics and estimates of factor loadings and regression coefficients are given in Table 5. Among the fit indices, the one given by Mt(1) is the smallest and the one given by Campbell$(2,\infty)$ is the largest. With $p = 11$ and $q = 28$ in this model, the degrees of freedom are 38. So the model fits the data pretty well judged by any of the fit indices. The differences among the fit indices reflect the differences of the weight functions. The fit index

TABLE 5
Fit indices, estimates of factor loadings and regression coefficients

θ	Mt(1)	Huber(.2)	Campbell$(2,\infty)$	S(.2)	ML
λ_1	1.219 (.160)	1.175 (.134)	1.191 (.133)	1.186 (.135)	1.191 (.139)
λ_2	1.066 (.123)	1.140 (.120)	1.175 (.118)	1.156 (.118)	1.175 (.120)
λ_3	1.192 (.113)	1.214 (.122)	1.251 (.121)	1.233 (.120)	1.251 (.117)
λ_4	1.989 (.176)	2.140 (.154)	2.180 (.143)	2.157 (.147)	2.180 (.138)
λ_5	1.748 (.190)	1.794 (.147)	1.818 (.139)	1.807 (.145)	1.818 (.152)
β_{21}	.855 (.059)	.867 (.055)	.865 (.053)	.865 (.054)	.865 (.064)
γ_{11}	2.005 (.521)	1.493 (.379)	1.471 (.351)	1.516 (.375)	1.471 (.392)
γ_{21}	.672 (.228)	.643 (.226)	.600 (.205)	.619 (.213)	.600 (.218)
Test	26.57	34.72	42.50	38.10	39.64

Source: Based on industrialization and political democracy data from Bollen (1989).

with the ML method is not the largest for this data set, which is related to the fact that no particular observation is atypical. Regarding the estimates of parameters, the regression coefficients γ_{11} and γ_{21} by Mt(1) are the largest while β_{21} by Mt(1) is the smallest. Since each case in Campbell(2,∞) gets a unit weight, the associated parameter estimates are identical to those by ML. The fit index by Campbell(2,∞) is a little larger than that by ML. This reflects the fact that the estimated $\hat{a}$ in (25) associated with Campbell(2,∞) is less than one. In general, the importance of this example is that it shows that our robust methods perform acceptably even when they are technically not needed.

From the above four examples, we can see some differences between the classical method and the robust methods on fit indices. There also exist differences on parameter estimates among the various methods, though these are not as impressive as those on the fit indices. In order to illustrate a more dramatic variation in parameter estimates by different methods, we present another example based on a real data set but with some added artificial outliers. Mardia et al. (1979, table 1.2.1) give test scores of $n = 88$ students on five topics (Mechanics, Vectors, Algebra, Analysis, and Statistics). Because the first two topics were tested with closed book exams and the last three topics with open book exams, this data set is sometimes referred to as the open-closed book data. Since these two examination methods may tap different abilities, a two-factor model was proposed and confirmed by Tanaka et al. (1991). The first two variables are indicators of the closed book factor, and the last three variables are indicators of the open book factor. The eighty-first case has previously been identified as the most influential point by various authors (Tanaka et al. 1991; Cadigan 1995; Lee and Wang 1996). Our analysis of the 88 cases shows that the weight associated with this case using Campbell(2,1.25) is $w_{81}^2 = .415$, indicating that even the most influential point may not be really atypical. Actually, the two-factor model fits the data well judged by either the ML method or any of the robust methods.

In order to see the effect of different methods, we added a vector

$$X_{89} = k(0, 0, 0, 11.49, 12.52)$$

to the original data set to make it a sample with size $n = 89$. An illuminating approach is to compare the different methods when k changes. Fixing $\phi_{11} = \phi_{22} = 1$ for identification purposes, we use the ML method as well as different robust methods to evaluate the model. Since the results by different robust methods gives similar information, we report only those by Huber(.2) to save space. The results are presented in Table 6 in which we use $k = 0$ to denote the original data set without the artificial outlier. The ML results are in the left part of the table. The effect of the one outlier on the ML method is obvious. As k increases, ψ_{44} and ψ_{55} dramatically increase at first, then ϕ_{12} decreases and ψ_{11} becomes negative (a Heywood case), finally λ_{32}, ϕ_{12}, ψ_{22} become negative (another Heywood case). In the right part of the table are the estimates by Huber(.2). As k increases, there is a minor effect on parameter estimates at the beginning and almost no difference when k moves from 5 to 25. The effect on fit statistics is similar to that on parameter estimates. With 4 degrees of freedom in this model, one outlier totally discredits an apparently good model when evaluated by the ML method, and the more extreme the outlier, the worse the ML test statistic is at describing the majority of the data. If evaluated by one of the robust methods, the effect of the outlier is basically eliminated. The unbounded influence function and zero breakdown point of the sample covariance S is also well illustrated through this example.

In summary, using robust covariances leads to smaller fit indices for the first two data sets. Even though the statistic corresponding to a robust method is still significant, a decrease in the fit index (from 102.10 to 78.62 in the second example) gives us more statistical support for using a theoretically justified model. For the third data set, using robust procedures automatically leads to a proper solution of estimates. These effects reflect that a robust covariance can effectively downweight individual observations that deviate from a proposed structure. In the fourth data set, the normal theory method works well and no case was heavily downweighted by any of the robust methods, indicating that no particular observation is atypical. The last example demonstrates that there can be a dramatic difference between the ML method and one of the robust methods, depending on how outlying the atypical observation is. It also illustrates that a single outlier can totally ruin the performance of the classical method.

After seeing the effect of different estimators in these examples, it would be helpful to give some guidance on choice of estimators in practice. In Huber-type estimators, the effect of abnormal cases is downweighted but not eliminated. If data are near normal, the estimators based on Huber-type weights are still highly efficient. So Huber-type weight functions—for example, Huber(.2) and Campbell(2,∞)—are better used for data sets whose distributions are not too far away from normal. In redescending weight functions, the effect of outlying cases is much smaller than that in Huber-type weights (see the numerical weights in Table 3).

TABLE 6
Fit Indices and Estimates of Model Parameters

	ML					Huber(.2)				
k	0	5	10	15	25	0	5	10	15	25
λ_{11}	12.25	12.78	12.79	18.35	2.719	11.79	11.87	11.81	11.78	11.76
λ_{21}	10.38	11.57	11.56	8.059	54.40	10.01	10.14	10.03	9.996	9.980
λ_{32}	9.834	10.80	10.73	2.209	−1.022	9.479	9.577	9.429	9.384	9.356
λ_{42}	11.49	9.978	7.126	18.34	26.79	11.18	10.99	11.04	11.10	11.17
λ_{52}	12.52	10.54	7.432	19.56	33.06	12.74	12.45	12.55	12.64	12.72
ϕ_{12}	.818	.866	.888	.160	−.031	.861	.868	.864	.862	.861
ψ_{11}	155.6	156.0	155.7	−17.56	312.0	156.3	154.3	154.4	154.5	154.8
ψ_{22}	65.04	65.68	65.95	134.7	−2760	65.90	65.79	65.72	65.62	65.50
ψ_{33}	16.19	23.77	25.19	135.5	139.3	17.74	19.31	19.59	19.63	19.62
ψ_{44}	88.35	119.6	219.4	58.79	150.7	84.56	86.78	87.63	87.82	87.91
ψ_{55}	141.1	187.9	316.3	149.7	24.95	139.3	144.5	144.6	144.3	143.9
Test	2.073	14.45	46.22	61.07	61.63	1.658	2.750	2.783	2.814	2.823

Source: Based on open-closed book data from Mardia et al. (1979).

So Hampel-type weight functions—for example, Campbell(2,1.5)—can be used for data sets with far outlying cases, but the estimators will lose more efficiency than those based on Huber-type weight functions when data are approximately normal. The weight function based on a multivariate *t*-distribution is best used for data sets whose discrepancy can be approximately described by the *t*-distribution, or for data sets with long-tailed distribution but no obvious outliers (few cases sit far away from the cloud formed by the majority of the cases). S-estimators are designed for high-dimensional data with many outliers—that is, data that are collected with many mishandlings. Of course, the best method to use would be the normal theory method if we know that a data set is fairly normal. The above discussion and recommendation are based on our limited experience with these different weight functions. In practice, if not much is known about the quality or distribution of a data set, we suggest experimenting with several of these methods, including the sample covariance as well as the S-estimators, before finding the one that suits the data set and theoretical models. If no case is particularly downweighted with a robust approach (e.g., Campbell[2,1.25]), using the sample covariance is probably enough. When many cases are heavily downweighted, it is necessary to use highly robust estimators, such as the S-estimators, in getting a reliable inference.

We have used Campbell's weight to define atypical observations, primarily because he explicitly connected the magnitude of w_i^2 to abnormal observations. Any other robust procedure may work equally well for a given data set.

5. DISCUSSION

Data in social and behavioral sciences are seldom normal, yet appropriate procedures for dealing with such nonnormal data in the context of latent variable structural modeling are only rarely used. When the data-generating mechanism is smooth and there are no atypical observations, various methods exist (reviewed by Bentler and Dudgeon 1996) that would give appropriate inference in this situation. However, outlying and grossly distorting observations seem to be typical in the social sciences. As noted by Wilcox (1996:xv) "recent investigations (cited in the text) indicate that outliers are very common in the social sciences, and outliers can substantially reduce power and give a distorted view of data based on conventional measures of location and scale." Thus motivated, we studied the technical and practical aspects of a procedure that uses robust covariances in structural equation modeling. Our examples show important differences between the classical method and the various robust methods.

When facing a data set that comes from a distribution with long tails or with outliers, the procedure of removing outliers followed by a classical method is often used (e.g., Berkane and Bentler 1988). As discussed in Huber (1981:4–5), outlier rejection followed by a classical method may not be a good statistical procedure, since the most influential points may not be real outliers; these may only indicate the long tails of the underlying distribution. Actually, it is hard to make a clear distinction between outliers and typical observations. For example, Campbell (1980) suggested that a weight w_i^2 below .30 is associated with an atypical observation, but he did not give a suggestion about what to do with a weight of .31. On the other hand, using a robust covariance approach automatically generates a proper weight to each of the cases. The outlying cases are automatically downweighted. Also, the statistical theory for our robust procedures is well developed, as given in Sections 2 and 3, while the theory that would hold for a two-step estimator involving outlier removal is not so clear.

A situation in which outlier detection may be useful is when one is familiar with the data collection process behind each individual case. In such a situation, identifying the most influential points may help the researcher to find extra information to explain these abnormal points. Even though using both the sample covariance and the robust covariances identified the same set of most influential points in Table 3, this result will not always be observed. Sometimes there may exist masking effects of multiple outliers, as demonstrated by Rousseeuw and van Zomeren (1990) in several interesting examples. So a robust procedure is still preferable in practice, if only to identify the "real outliers."

Since most applications of structural equation modeling involve covariance structures, we only studied the procedure of using a robust covariance in the Wishart likelihood function. It will be apparent from equation (11) that our development is equally applicable to the situation in which robust covariances are used in the normal theory generalized least squares (GLS) function $\frac{1}{2}\mathrm{tr}\{W(S_n - \Sigma(\theta))\}^2$, where W is a consistent estimator of Σ_0^{-1}. For example $W = S_n^{-1}$ and the iteratively updated $W = \hat{\Sigma}_0^{-1}$ define classical GLS and iteratively reweighted GLS functions (Bentler 1995). The functions then can be corrected by $\hat{a}$ as in (25) to yield an asymptotically equivalent test statistic.

Sometimes, a mean structure is also of substantive interest. Since the influence function associated with the sample mean is unbounded, it would be preferable to develop a procedure that uses robust means and covariances simultaneously in the normal theory likelihood function. However, the fit statistic T_1 cannot be generalized directly to mean and covariance structures. This aspect is still under further study.

As far as we know, no existing structural equation modeling software can compute the robust procedures described above. In order to make these methods readily available, they are being incorporated into the new release 6.0 of EQS. Furthermore, to ensure that researchers can utilize these statistics as they desire, all the key components (e.g., case weights, robust means and covariances, parameter estimates, asymptotic covariance matrices) will be writable to an external file. The methods will also be integrated into EQS's simulation module to permit Monte Carlo and resampling research on the performance of these methods under varying conditions.

REFERENCES

Amemiya, Yasuo, and Theodore W. Anderson. 1990. "Asymptotic Chi-Square Tests for a Large Class of Factor Analysis Models." *Annals of Statistics* 18:1453–63.

Anderson, Theodore W., and Yasuo Amemiya. 1988. "The Asymptotic Normal Distribution of Estimators in Factor Analysis Under General Conditions." *Annals of Statistics* 16:759–71.

Arminger, Gerhard, and Ronald J. Schoenberg. 1989. "Pseudo Maximum Likelihood Estimation and a Test for Misspecification in Mean- and Covariance-Structure Models." *Psychometrika* 54:409–25.

Arminger, Gerhard, and Michael Sobel. 1990. "Pseudo Maximum Likelihood Estimation of Mean and Covariance Structures with Missing Data." *Journal of the American Statistical Association* 85:195–203.

Austin, James T., and Robert F. Calderón. 1996. "Theoretical and Technical Contributions to Structural Equation Modeling: An Updated Annotated Bibliography." *Structural Equation Modeling* 3:105–75.

Bentler, Peter M. 1994. "A Testing Method for Covariance Structure Analysis." Pp. 123–35 in *Multivariate Analysis and its Applications*, vol. 24, edited by T. W. Anderson, K. T. Fang and I. Olkin. Hayward, CA: Institute of Mathematical Statistics.

———. 1995. *EQS Structural Equations Program Manual*. Encino, CA: Multivariate Software.

Bentler, Peter M., and Theo Dijkstra. 1985. "Efficient Estimation via Linearization in Structural Models." Pp. 9–42 in *Multivariate Analysis VI*, edited by P. R. Krishnaiah. Amsterdam: North-Holland.

Bentler, Peter M., and Paul Dudgeon. 1996. "Covariance Structure Analysis: Statistical Practice, Theory, Directions." *Annual Review of Psychology* 47:563–92.

Berkane, Maia, and Peter M. Bentler. 1988. "Estimation of Contamination Parameters and Identification of Outliers in Multivariate Data." *Sociological Methods and Research* 17:55–64.

Bishop, Yvonne M. M., Stephen E. Fienberg, and Paul W. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Bollen, Kenneth A., and Gerhard Arminger. 1991. "Observational Residuals in Factor Analysis and Structural Equation Models." Pp. 235–62 in *Sociological Methodology 1991*, edited by Peter V. Marsden. Cambridge, MA: Blackwell Publishers.

Box, George E. P. 1954. "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problem: I. Effect of Inequality of Variance in the One-way Classification." *Annals of Mathematical Statistics* 25:290–302.

Browne, Michael W. 1982. "Covariance Structures." Pp. 72–141 in *Topics in Multivariate Analysis*, edited by D. M. Hawkins. Cambridge, England: Cambridge University Press.

———. 1984. "Asymptotic Distribution-Free Methods for the Analysis of Covariance Structures." *British Journal of Mathematical and Statistical Psychology* 37:62–83.

Browne, Michael W., and Gerhard Arminger. 1995. "Specification and Estimation of Mean- and Covariance-Structure Models." Pp. 185–249 in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, edited by G. Arminger, C. C. Clogg and M. E. Sobel. New York: Plenum.

Cadigan, Noel G. 1995. "Local Influence in Structural Equation Models." *Structural Equation Modeling* 2:13–30.

Campbell, Norm A. 1980. "Robust Procedures in Multivariate Analysis I: Robust Covariance Estimation." *Applied Statistics* 29:231–37.

Catalano, Paul J., and Louise M. Ryan. 1992. "Bivariate Latent Variable Models for Clustered Discrete and Continuous Outcomes." *Journal of the American Statistical Association* 87:651–58.

Chamberlain, Gary. 1982. "Multivariate Regression Models for Panel Data." *Journal of Econometrics* 18:5–46.

Curran, Patrick S., Stephen G. West, and John F. Finch. 1996. "The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis." *Psychological Methods* 1:16–29.

Devlin, Susan J., Ramanathan Gnanadesikan, and Jon R. Kettenring. 1981. "Robust Estimation of Dispersion Matrices and Principal Components." *Journal of the American Statistical Association* 76:354–62.

Dukes, Richard L., Jodie Ullman, and Judy A. Stein. 1995. "An Evaluation of D.A.R.E. (Drug Abuse Resistance Education), Using a Solomon Four-Group Design with Latent Variables." *Evaluation Review* 19:409–35.

Gnanadesikan, Ramanathan. 1997. *Methods for Statistical Data Analysis of Multivariate Observations*. New York: Wiley.

Gnanadesikan, Ramanathan, and Jon R. Kettenring. 1984. "A Pragmatic Review of Multivariate Methods in Applications." Pp. 309–41 in *Statistics: An Appraisal*, edited by H. A. David and H. T. David. Ames, Iowa: Iowa State University Press.

Hampel, Frank R. 1974. “The Influence Curve and Its Role in Robust Estimation.” *Journal of the American Statistical Association* 69:383–93.

Hoaglin, David C., Frederick Mosteller, and John W. Tukey. 1983. *Understanding Robust and Exploratory Data Analysis*. New York: Wiley.

Holzinger, Karl J., and Frances Swineford. 1939. *A Study in Factor Analysis: The Stability of a Bi-factor Solution*. University of Chicago, Supplementary Educational Monographs, No. 48.

Hu, Li-tze, Peter M. Bentler, and Yutaka Kano. 1992. “Can Test Statistics in Covariance Structure Analysis Be Trusted?” *Psychological Bulletin* 112:351–62.

Huba, George J., and Lisa L. Harlow. 1987. “Robust Structural Equation Models: Implications for Developmental Psychology.” *Child Development* 58:147–66.

Huber, Peter J. 1981. *Robust Statistics.* New York: Wiley.

Jöreskog, Karl G. 1969. “A General Approach to Confirmatory Maximum Likelihood Factor Analysis.” *Psychometrika* 34:183–202.

———. 1977. “Structural Equation Models in the Social Sciences: Specification, Estimation and Testing.” Pp. 265–87 in *Applications of Statistics,* edited by P. R. Krishnaiah. Amsterdam: North Holland.

Jöreskog, Karl G., and Dag Sörbom. 1993. *LISREL 8 User's Reference Guide*. Chicago: Scientific Software International.

Kano, Yutaka. 1986. “Conditions on Consistency of Estimators in Covariance Structure Model.” *Journal of the Japan Statistical Society* 16:75–80.

Kendler, Kenneth S., Andrew C. Heath, Michael C. Neale, Ronald C. Kessler, and Lindon J. Eaves. 1992. “A Population-Based Twin Study of Alcoholism in Women.” *Journal of the American Medical Association* 268:1877–82.

Lee, Sik-Yum., and Shu-Jia Wang. 1996. “Sensitivity Analysis of Structural Equation Models.” *Psychometrika* 61:93–108.

Lopuhaä, Hendrik P. 1989. “On the Relation Between S-estimators and M-estimators of Multivariate Location and Covariances.” *Annals of Statistics* 17:1662–83.

Magnus, Jan R., and Heinz Neudecker. 1988. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley.

Mardia, Kanti V., John T. Kent, and John M. Bibby. 1979. *Multivariate Analysis*. New York: Academic Press.

Maronna, Ricardo A. 1976. “Robust M-estimators of Multivariate Location and Scatter.” *Annals of Statistics* 4:51–67.

Muirhead, Robb J. 1982. *Aspects of Multivariate Statistical Theory*. New York: Wiley.

Muthén, Bengt O. 1992. “Latent Variable Modeling in Epidemiology.” *Alcohol Health and Research World* 16:286–92.

Rousseeuw, Peter J., and Bert C. van Zomeren. 1990. “Unmasking Multivariate Outliers and Leverage Points (with discussion).” *Journal of the American Statistical Association* 85:633–51.

Ruppert, David. 1992. “Computing S Estimators for Regression and Multivariate Location/Dispersion.” *Journal of Computational and Graphical Statistics* 1:253–70.

Satorra, Albert, and Peter M. Bentler. 1990. “Model Conditions for Asymptotic Robustness in the Analysis of Linear Relations.” *Computational Statistics and Data Analysis* 10:235–49.

———. 1994. “Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis.” Pp. 399–419 in *Latent Variables Analysis: Applications for Developmental Research,* edited by A. von Eye and C. C. Clogg. Thousand Oaks, CA: Sage.

Satterthwaite, Franklin E. 1941. “Synthesis of Variance.” *Psychometrika,* 6:309–16.

Staudte, Robert G., and Simon J. Sheather. 1990. *Robust Estimation and Testing*. New York: Wiley.

Tanaka, Yutaka, Shingo Watadani, and Sung H. Moon. 1991. “Influence in Covariance Structure Analysis: With an Application to Confirmatory Factor Analysis.” *Communication in Statistics-Theory and Method* 20:3805–21.

Tosteson, Tor, Leonard A. Stefanski, and Daniel W. Schafer. 1989. “A Measurement Error Model for Binary and Ordinal Regression.” *Statistics in Medicine* 8:1139–47.

Tyler, David E. 1982. “Radial Estimates and the Test for Sphericity.” *Biometrika* 69:429–36.

———. 1983. “Robustness and Efficiency Properties of Scatter Matrices.” *Biometrika* 70:411–20.

Wilcox, Rand R. 1996. *Statistics for the Social Sciences*. San Diego: Academic Press.

Yamaguchi, Kazunori, and Michko Watanabe. 1993. “The Efficiency of Estimation Methods for Covariance Structure Models Under Multivariate t Distribution.” Pp. 281–84 in *Proceedings of the Asian Conference on Statistical Computing*, International Association for Statistical Computing.

Yuan, Ke-Hai, and Peter M. Bentler. Forthcoming. “Robust Methods for Mean and Covariance Structure Analysis.” *British Journal of Mathematical and Statistical Psychology.*

———. 1996. “Effect of Outliers on Estimators and Test Statistics in Covariance Structure Analysis.” Under review.

———. 1997. “Mean and Covariance Structure Analysis: Theoretical and Practical Improvements.” *Journal of the American Statistical Association* 92:767–74.