Title
Robust Methods for Mean and Covariance Structure Analysis

Authors
K. H. Yuan
P. M. Bentler

Publication Date
2011-10-25

UCLA Statistical Series
Report No. 195

Robust Methods for Mean and Covariance Structure Analysis*

Ke-Hai Yuan and Peter M. Bentler

October 15, 1995

*Ke-Hai Yuan is Statistician, and Peter M. Bentler is Professor, Department of Psychology and Center for Statistics, University of California, Los Angeles, Box 951563, Los Angeles, CA 90095-1563. This work was supported by National Institute on Drug Abuse Grants DA01070 and DA00017.

Abstract

Covariance structure analysis plays an important role in social and behavioral sciences to evaluate hypothesized influences among unmeasured latent and observed variables. Existing methods for analyzing these data rely on unstructured sample means and covariances estimated under normality, and evaluate a proposed structural model using statistical theory based on normal theory MLE and generalized least squares (GLS) with a weight matrix obtained from inverting a matrix based on sample fourth moments and covariances. Since the influence functions associated with these methods are quadratic, a few outliers can make these classical procedures a total failure. Considering that data collected in social and behavioral sciences are not so accurate, some robust methods are necessary in estimation and testing. Even though the theory for robustly estimating multivariate location and scatter has been developed extensively, very little has been accomplished in robust mean and covariance structure analysis. While robust principal components and canonical variates have been described many years ago, this methodology is essentially exploratory in nature and does not provide tests of model fit nor the covariance matrix of the estimator that are essential to covariance structure analysis. In this paper, several robust methods in model fitting and testing are proposed. These include direct estimation of M-estimators of structured parameters and a two-stage procedure based on robust M- and S-estimators of population covariances. The large sample properties of these estimators are obtained. The equivalence between a direct M-estimator and a two-stage estimator based on an M-estimator of population covariance is established when sampling from an elliptical distribution. Two test statistics are presented in judging the adequency of a hypothesized model; both are asymptotically distribution free if using distribution free weight matrices. So these test statistics possess both small sample and large sample robustness. The two-stage procedures can be easily adapted into standard software packages by modifying existing GLS procedures. To demonstrate the easy application of the two-stage procedure, M-estimators under six different weight functions are calculated for a real data set. All the weight functions give the smallest weight to the case which has been formerly identified as the most influential point.

KEY WORDS: Outliers, M-estimators, S-estimators, Two-stage estimators; Robust Tests, Elliptical theory, Distribution-free.

1 Introduction

Mean and covariance structure analysis plays a very important role in understanding the underlying structure of social and multivariate data and in dimension reduction (Bentler and Dudgeon 1996). A special case of it is covariance structure analysis in which the unknown mean is a nuisance parameter. In a typical covariance structure analysis setting, we have a random sample X_1, ..., X_n with $\mathrm{E}X_i = \mu_0$ and $var(X_i) = \Sigma_0 = \Sigma(\theta_0)$. A classical application is the confirmatory factor analysis model $\Sigma = \Lambda\Phi\Lambda^T + \Psi$, where the elements of Λ, Φ, and Ψ are elements of the parameter vector θ_0. The primary interest is in getting a good estimator of θ_0 and in judging the adequancy of the model $\Sigma = \Sigma(\theta)$. Because of the complexity of the data, a variety of methods have been proposed to estimate θ_0 and evaluate the structural model. The most well known and often used method is classical normal theory maximum likelihood (ML). In this approach, the normal theory likelihood based on the unstructured ML sample covariance S is maximized with respect to θ to obtain the structured estimator $\hat{\Sigma} = \Sigma(\hat{\theta}_n)$. Since the underlying data may not be normal, a generalized least squares method called asymptotically distribution free (ADF) was proposed by Browne (1982, 1984) and Chamberlain (1982). The idea in ADF is to fit $\Sigma(\theta)$ by S using an asymptotically correct weight matrix based on the sample fourth moments. Even though ADF gives correct inference when sample size is very large, say 5000, for small to medium sample sizes, ADF usually rejects the true model too often (Hu, Bentler and Kano, 1992). Yuan and Bentler (1995) proposed using a residual based regression approach. One variant of this approach effectively corrects Browne's (1984) ADF test statistic in such a way as to perform much better in empirical simulations.

All the above mentioned methods assume that the data are good, and model the sample covariance S by $\Sigma(\theta)$. In practice, real data may not be well-behaved. Hence, the standard methods may be susceptible to substantially worse performance than shown in simulations which have relied upon smooth and well-behaved data generation mechanisms. In a regular case, it will be appropriate to base the analysis on the sample covariance S based on the standard formula in which every observation is treated with equal weight. However, when the data generating mechanism is not well-behaved, S may not be an adequate unstructured estimator of Σ. Since the influence function associated with S is a quadratic function, one single case with gross error or an outlier can totally distort the sample covariance. Devlin, Gnanadesikan and Kettenring (1981) gave such an example in principal component analysis. Campbell (1980) also considered the effect of outliers on estimated covariances through examples.

In the setting of covariance structure analysis, Tanaka, Watadani and Moon (1991) and Lee and Wang (1995) considered identifying the most influential points. Using outlier detection to eliminate cases before computing the sample covariance S hopefully would improve the performance of normal theory ML or ADF. As discussed in Huber (1981, pp. 4-5), however, outlier rejection followed by a classical method may not be a good statistical procedure. For example, if the data is from a long tailed distributions, the most influential points may not be outliers. After such points have been identified, a decision about them thus remains hard to make. Furthermore, theoretical properties of such a procedure are not clear. A natural approach would thus be to use robust statistics in structural models.

In the literature on estimating the population mean and the population covariance matrix, a variety of robust procedures have been proposed. Examples are: various

types of M-estimators investigated by Maronna (1976), Huber (1977), Tyler (1983, 1987), the Stahel-Donoho estimator independently proposed by Stahel (1981) and Donoho (1982), the minimum distance estimator investigated by Tamura and Boos (1986) and Donoho and Liu (1988a, 1988b), the minimum covariance determinant (MCD) estimator and the minimum volume ellipsoid (MVE) estimator introduced by Rousseeuw (1983), the type of S-estimators investigated by Davies (1987) and Lopuhaä (1989), the τ-estimators introduced by Lopuhaä (1991a), and many others. Applications of robust procedures in principal component analysis, canonical variates, and related correlational problems have been considered by Devlin et al (1981), Campbell (1980, 1982), and others (e.g., Wilcox 1994). In this paper, we will consider using a robust procedure to estimate θ_0 and to evaluate the general structure. In some sense, our work is an extension and application of the theory developed in estimating unstructured means and covariances and parallels the work of Devlin et al and Campbell. However, the procedures considered by Devlin et al and Campbell are basically exploratory in nature. In covariance structure analysis, on the other hand, the emphasis is on evaluating a specific hypothesized structure $\Sigma(\theta)$ and evaluating the significance of the parameter estimates. These problems require the use of a valid test of model structure $\Sigma(\theta)$ and a coordinated method for determining the covariance matrix of the estimated $\hat{\theta}_n$. Hence, previously developed robust methods for exploratory data analysis have to be extended to confirmatory mean and covariance structure analysis.

We will use the following notations: If A is a $p \times p$ matrix, $vec(A)$ is the p^2-dimensional vector formed by stacking the columns of A while $vech(A)$ is a $p^* = p(p+1)/2$-dimensional vector formed by the nonduplicated elements of A when A is symmetric. $\sigma(\theta) = vech(\Sigma(\theta))$ and $\bar{\sigma}(\theta) = vec(\Sigma(\theta))$. When we refer to unstructured means and covariances, $\beta = (\mu^T, vech^T(\Sigma))^T$ will be used. A function with a dot on top means derivatives, e.g. $\dot{\mu}(\theta) = \partial\mu/\partial\theta^T$, $\dot{h}_\theta(x,\theta) = \partial h(x,\theta)/\partial\theta^T$. When a function

is evaluated at the population value, we often omit its argument, e.g. $\dot{\mu} = \dot{\mu}(\theta_0)$.

We will consider the M-estimator of a structured parameter in Section 2. In Section 3, we give a summary of the properties of the existing robust estimators. The S-estimators of multivariate mean and covariance will be detailed, especially the different estimators of the asymptotic covariance under different assumptions on the populations. Section 4 will present a two stage estimating process based on estimators of the population covariance. In Section 5, we consider two test statistics in judging the adequacy of a structured model. All the proofs of the theorems are in the appendix. A real data example will be considered in Section 6.

2 M-estimator

Let $X_1, \ldots, X_n$ be a sample of a p-dimensional random vector X. In this section, we will develop an M-estimator of the structured parameter of a mean and covariance based on such a sample. We consider both structured means and structured covariances in the first part, and structured covariances with unstructured means in the second part. Although it is not our main emphasis, when assuming the sample is from an elliptical distribution, the asymptotic covariance involved can be substantially simplified. A distribution with density $f(x) = |\Sigma_0|^{-\frac{1}{2}} h\{(x-\mu_0)^T \Sigma_0^{-1}(x-\mu_0)\}$, where $h(t)$ is a function independent of p, is refered to as an elliptical distribution or elliptical symmetric distribution (Fang, Kotz and Ng, 1990).

2.1 Structured mean and covariance

Maronna (1976) defined an M-estimator of μ_0 and Σ_0. Now we have a structure

$\mu_0 = \mu(\theta_0)$ and $\Sigma_0 = \Sigma(\theta_0)$. Following the maximum likelihood estimator of θ_0 in an elliptical distribution and the M-estimator of μ_0 and Σ_0 defined by Maronna (1976), we define an M-estimator $\hat{\theta}_n$ of θ_0 by

$$\frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\theta}_n) = 0, \tag{2.1}$$

where

$$\begin{aligned} g_i(\theta) = g(X_i, \theta) \quad &= u_1(d_i)\dot{\mu}^T(\theta)\Sigma^{-1}(\theta)(X_i - \mu(\theta)) \\ &+\tfrac{1}{2}\dot{\sigma}^T(\theta)\{u_2(d_i^2)vec(T(X_i,\theta)) - vec(\Sigma^{-1}(\theta))\} \end{aligned}$$

with

$$d_i = d(X_i, \theta) = \{(X_i - \mu(\theta))^T\Sigma^{-1}(\theta)(X_i - \mu(\theta))\}^{\frac{1}{2}}$$

$$T(x, \theta) = \Sigma^{-1}(\theta)(x - \mu(\theta))(x - \mu(\theta))^T\Sigma^{-1}(\theta).$$

The functions $u_1(t)$ and $u_2(t)$ are weight functions. In the context of maximum likelihood in an elliptical family, $u_1(t) = u_2(t^2) = -2\dot{h}(t^2)/h(t^2)$. In defining $\hat{\theta}_n$ in (2.1), we assume that the population θ_0 satisfies $\mathrm{E}g(X, \theta_0) = 0$. In an elliptically symmetric family, this is justified if $\Sigma(\theta)$ is invariant under a constant scaling factor as defined and discussed in Browne (1982, p.77).

When there is no structure on μ and Σ, exact assumptions for the existence and uniqueness of $\hat{\mu}_n$ and $\hat{\Sigma}_n$ were given by Maronna (1976). As noted by both Lopuhaä (1991b) and Tyler (1991), when $\psi_2(t) = tu_2(t)$ is a monotone function, which is assumption (C) of Maronna (1976, p. 53), there exists a unique M-estimator $\hat{\Sigma}_n$. However, as commented by Huber (1981, p. 223), the assumptions for uniqueness of the solutions to Maronna's equation (1.1) and (1.2) may be unrealistic for a specific set of data. For one dimensional data generated from a Cauchy distribution, Fu (1989) demonstrated that the maximum likelihood equation can have multiple solutions at local maxima.

Since (2.1) is an estimating equation, we will use the results of Yuan and Jennrich (1995) and show that there is always a consistent solution of (2.1) for large n. We need the following assumptions.

Assumptions 2:

2.1. $\mu(\theta)$ and $\Sigma(\theta)$ are twice continuously differentiable and $\Sigma(\theta) > \epsilon I$ in a neighborhood of θ_0 for some $\epsilon > 0$.

2.2. With probability one $u_1(d(X,\theta))$ and $u_2(d^2(X,\theta))$ are continuously differentiable with respect to θ and $u_1(d(x,\theta))$, $u_2(d^2(x,\theta))$, $\dot{u}_1(d(x,\theta))$ and $\dot{u}_2(d^2(x,\theta))$ are bounded.

2.3. X has finite fourth moment.

2.4. $\mathrm{E}g(X,\theta_0) = 0$, $V = var(g(X,\theta_0)) > 0$, $A = \mathrm{E}\dot{g}_\theta(X,\theta_0)$ is nonsingular.

Assumptions 2.1 and 2.3 are similar as those made for maximum likelihood estimators based on normal theory. It is obvious that $u_1(t)$ and $u_2(t)$ corresponding to a multivariate t-distribution satisfy assumption 2.2. Since most of the weight functions used in linear models have a finite number of discontinuity points and are bounded together with their derivatives, they also satisfy assumption 2.2. Assumption 2.4 are moment assumptions about the model.

Let $G_n(\theta) = \frac{1}{n}\sum_{i=1}^n g_i(\theta)$. With this definition, we have the following theorem.

Theorem 2.1. *Under assumptions 2.1 to 2.4, for any neighborhood N of θ_0, with probability one there are $\hat{\theta}_n \in N$ such that $G_n(\hat{\theta}_n) = 0$ for all n sufficiently large.*

Theorem 1 claims the existence of roots of (2.1). The following theorem is on consistency of the roots $\hat{\theta}_n$.

Theorem 2.2. *Under assumptions 2.1 to 2.4, with probability one there are zeros* $\hat{\theta}_n$ *of* $G_n(\theta)$ *such that* $\hat{\theta}_n \to \theta_0$.

Theorems 2.1 and 2.2 give the existence and consistency of a solution. If the solution for (2.1) is unique, it must be consistent. When equation (2.1) has multiple solutions, one needs to have a good algorithm to find the consistent solution, as demonstrated by Fu (1989) for the maximum likelihood estimator of a Cauchy distribution. Since it is not our purpose here to explore algorithms for obtaining $\hat{\theta}_n$, we will assume that a consistent $\hat{\theta}_n$ has been obtained in the rest of this paper.

In order to evaluate hypothesies about θ_0, we need to obtain the asymptotic distribution of $\hat{\theta}_n$. This is given in the following theorem.

Theorem 2.3. *Let* $G_n(\hat{\theta}_n) = 0$ *and* $\hat{\theta}_n \xrightarrow{P} \theta_0$, *then under assumptions 2.1 to 2.4,*

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

where

$$\Omega = A^{-1} V A^{-T}.$$

To apply this result, we need to obtain a consistent estimator of Ω. This can be obtained through consistent estimators of A and V which are given respectively by

$$A_n = \frac{1}{n} \sum_{i=1}^{n} \dot{g}_i(\hat{\theta}_n) \tag{2.2}$$

$$V_n = \frac{1}{n} \sum_{i=1}^{n} g_i(\hat{\theta}_n) g_i^T(\hat{\theta}_n). \tag{2.3}$$

The consistency of A_n and V_n does not depend on the underlying distribution of the sample as long as assumption 2.4 holds. Estimators with this property will be refered

to as "distribution free" estimators later.

If the data are from an elliptical distribution, the expression for A and V can be much more detailed.

Corollary 2.1. *Assuming that the underlying distribution of X is elliptical symmetric, then*

$$\begin{aligned} A(\theta_0) &= -\{(\tfrac{1}{p}E[\dot{u}_1(d)d] + Eu_1(d))\dot{\mu}^T\Sigma^{-1}\dot{\mu} \\ &\quad +(\tfrac{1}{2} + \tfrac{1}{p(p+2)}E[\dot{u}_2(d^2)d^4])\dot{\bar{\sigma}}^T(\Sigma^{-1}\otimes\Sigma^{-1})\dot{\bar{\sigma}} \\ &\quad +\tfrac{1}{2p(p+2)}E[\dot{u}_2(d^2)d^4]\dot{\bar{\sigma}}vec(\Sigma^{-1})vec^T(\Sigma^{-1})\dot{\bar{\sigma}}\} \end{aligned} \tag{2.4}$$

and

$$\begin{aligned} V(\theta_0) &= \tfrac{1}{p}E[u_1^2(d)d^2]\dot{\mu}^T\Sigma^{-1}\dot{\mu} + \tfrac{1}{2p(p+2)}E[u_2^2(d^2)d^4]\dot{\bar{\sigma}}^T(\Sigma^{-1}\otimes\Sigma^{-1})\dot{\bar{\sigma}} \\ &\quad +\tfrac{1}{4}\{\tfrac{1}{p(p+2)}E[u_2^2(d^2)d^4] - 1\}\dot{\bar{\sigma}}vec(\Sigma^{-1})vec^T(\Sigma^{-1})\dot{\bar{\sigma}}, \end{aligned} \tag{2.5}$$

where $d = d(X, \theta_0)$.

So in an elliptical family, there are several choices for a consistent $\hat{\Omega}_n$. For example, both $V(\hat{\theta}_n)$ and $\frac{1}{n}\sum_{i=1}^n g_i(\hat{\theta}_n)g_i^T(\hat{\theta}_n)$ are consistent estimators for V. Similarly, we can estimate A by either $A(\hat{\theta}_n)$ or $\frac{1}{n}\sum_{i=1}^n \dot{g}_i(\hat{\theta}_n)$. Inside these estimators, Σ can be estimated by either $\Sigma(\hat{\theta}_n)$ or $\hat{\Sigma}_n$ from an unstructured model. Even though asymptotically each combination gives a consistent estimator for Ω, these various alternatives may have nonignorable differences with small to medium sample sizes.

The influence function of $\hat{\theta}_n$ is

$$IF_\theta(x) = -A^{-1}g(x, \theta_0),$$

which can be obtained by either equation (4.2.9) and (4.2.10) of Hampel et al (1986, p. 230) or using the technique demonstrated in Serfling (1980, Section 6.5). It can be seen that in order to get an M-estimator with a bounded influence function, both

$u_1(t)$ and $u_2(t)$ should go to zero at least as fast as $1/t$. We can choose $u_1(t)$ and $u_2(t)$ either from the weight functions used in linear models, e.g., the Huber-type ψ-function, Hampel's redescending ψ-function, etc., or by the use of some elliptical distribution, e.g., the multivariate t-distribution (Kano 1994; Kent, Tyler, and Vardi 1994). We will give more details in Section 6.

2.2 Structured covariance only

As we have mentioned, an important special case is a structured covariance with an unstructured mean. Let $\gamma^T = (\mu^T, \theta^T)$. The equation (2.1) becomes the following two equations

$$\frac{1}{n}\sum_{i=1}^{n} u_1(d_i)(X_i - \mu) = 0, \tag{2.6}$$

$$\frac{1}{n}\sum_{i=1}^{n} \dot{\sigma}^T(\theta)\{u_2(d_i^2)vec(T(X_i,\gamma)) - vec(\Sigma^{-1}(\theta))\} = 0. \tag{2.7}$$

Let

$$g_i(\gamma) = \begin{pmatrix} u_1(d_i)(X_i - \mu) \\ \dot{\sigma}^T(\theta)\{u_2(d_i^2)vec(T(X_i,\gamma)) - vec(\Sigma^{-1}(\theta))\} \end{pmatrix}.$$

The asymptotic distribution of $\hat{\gamma}_n$, which satisfies (2.6) and (2.7), is given by

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{\mathcal{L}} N(0, \Omega),$$

where $\Omega = A^{-1}VA^{-T}$ with A=E$\dot{g}_i(\gamma_0)$ and $V = var(g_i(\gamma_0))$.

An interesting specialization of the asymptotic distribution of $\hat{\gamma}_n$ occurs when the sample is from an elliptical distribution. Let

$$a = \frac{1}{p(p+2)}E[\dot{u}_2(d^2)d^4],$$

$$b = \frac{1}{p(p+2)}E[u_2^2(d^2)d^4],$$

$$c = \frac{pE[u_1^2(d)d^2]}{\{pEu_1(d) + E[\dot{u}_1(d)d]\}^2},$$

and

$$p_1 = vec^T(\Sigma^{-1})\dot{\bar{\sigma}}\{\dot{\bar{\sigma}}^T(\Sigma^{-1} \otimes \Sigma^{-1})\dot{\bar{\sigma}}\}^{-1}\dot{\bar{\sigma}}^T vec(\Sigma^{-1}).$$

Corollary 2.2. *Assume the underlying distribution of X is elliptical symmetric. Then $\hat{\mu}_n$ and $\hat{\theta}_n$ are asymptotically independent with*

$$\sqrt{n}(\hat{\mu}_n - \mu_0) \xrightarrow{\mathcal{L}} N(0, c\Sigma),$$

and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Omega_{22}),$$

where

$$\Omega_{22}^{-1} = \frac{(1+2a)^2}{2b}\dot{\bar{\sigma}}^T(\Sigma^{-1} \otimes \Sigma^{-1})\dot{\bar{\sigma}} + \kappa(p_1)\dot{\bar{\sigma}}^T vec(\Sigma^{-1})vec^T(\Sigma^{-1})\dot{\bar{\sigma}}, \tag{2.8}$$

and

$$\kappa(p_1) = \frac{(1+2a)^2 + 2a^2b(2+p_1) - b}{2b[(2+p_1)b - p_1]}.$$

In an elliptical family, the influence functions of $\hat{\mu}_n$ and $\hat{\theta}_n$ can be separated. The influence function of $\hat{\mu}_n$ was given by Maronna (1976), which is

$$IF_{\hat{\mu}}(x) = \frac{u_1(d(x,\theta_0))(x-\mu_0)}{Eu_1(d(X,\theta_0)) + \frac{1}{p}E[\dot{u}_1(d(X,\theta_0))d(X,\theta_0)]}.$$

The influence function of $\hat{\theta}_n$ is

$$IF_{\hat{\theta}_n}(x) = A_{22}^{-1}\dot{\bar{\sigma}}\{u_2(d(x)vec(T(x,\theta_0)) - vec[\Sigma^{-1}(\theta_0)]\}, \tag{2.9}$$

where

$$A_{22} = (1+2a)\dot{\bar{\sigma}}^T(\Sigma^{-1} \otimes \Sigma^{-1})\dot{\bar{\sigma}} + a\dot{\bar{\sigma}}^T vec(\Sigma^{-1})vec^T(\Sigma^{-1})\dot{\bar{\sigma}}.$$

It can be shown that after some standardization and setting $u_2(t) = 1$, (2.9) is equivalent to the influence function obtained by Tanaka et al (1991) in their equation

(6). They investigated the empirical influence on $\hat{\theta}_n$ of individual observations. Since $u_2(t) = 1$, $IF_{\hat{\theta}_n}(x)$ is quadratic in x.

The mean parameter μ is a nuisance parameter in covariance structure analysis, and $\hat{\mu}_n$ and $\hat{\theta}_n$ are asymptotically independent, so we can use any $\sqrt{n}$-consistent estimator $\hat{\mu}_n$ in equation (2.7) and solve for $\hat{\theta}_n$ only. This will not influence the asymptotic distribution of $\hat{\theta}_n$. For example, we may use the marginal median or the marginal trimmed mean as $\hat{\mu}_n$. If one is worried about losing the equivariance of $\hat{\theta}_n$, one can solve first for $\hat{\mu}_n$ and $\hat{\Sigma}_n$ as defined in equations (1.1) and (1.2) of Maronna (1976) to obtain an equivariant estimator $\hat{\mu}_n$ and then solving for $\hat{\theta}_n$ through (2.7). But this may not be a lot simpler than solving (2.6) and (2.7) simultaneously.

3 S-estimators of multivariate location and covariance

Even though M-estimators have bounded influence functions by choosing proper weight functions $u_1(t)$ and $u_2(t)$, an unsatisfactory aspect of M-estimators is their low breakdown point in high dimensions. It has been demonstrated (e.g., Maronna 1976, Huber 1981) that the breakdown point of an M-estimator is at most $1/(p+1)$. Because of the relationship between S-estimators and M-estimators (Lopuhaä 1989), Lopuhaä (1991b) and Tyler (1991) argued that when an estimating equation defining M-estimators enjoys multiple solutions, some solutions may have higher breakdown points, though it is not clear how to identify such solutions.

Because of the limited breakdown point of M-estimators in high dimensional data, other robust procedures have been proposed to estimate μ_0 and Σ_0. Most of the meth-

ods mentioned in Section 1 possess a high breakdown point. In order to obtain better robustness, we discuss some alternatives. Letting $\hat{\Sigma}_n$ be any consistent estimator of σ_0, we need to assume that

$$a_n(\hat{\sigma}_n - \sigma(\theta_0)) \xrightarrow{\mathcal{L}} N(0, C(\theta)) \tag{3.1}$$

for some $a_n \to \infty$. Among all the high breakdown point estimators mentioned above, only the MCD estimator, S-estimator and τ-estimator satisfy such an assumption. The other estimators, either because of unknown asymptotic properties or a non-normal limiting distribution, do not satisfy (3.1), according to Lopuhaä (1991b) and Tyler (1991). For example, the minimum distance estimator is biased according to Tamura and Boos (1986).

In order to obtain a high breakdown point estimator and for motivating the next section, a brief review of S-estimators, based mainly on Lopuhaä (1989, 1991b), will be given. Since the form of the asymptotic covariance of an S-estimator is almost the same as that of an M-estimator, a consistent covariance matrix of the estimators is easily obtained. As in the last section, we will contrast the different estimators and their underlying assumptions.

Let $\rho(t)$ be a function on real line which satisfies:

Assumptions 3:

3.1. $\rho(t)$ is symmetric, has continuous derivative $\psi(t)$ and $\rho(0) = 0$.

3.2. There exists a finite $c_0 > 0$ such that $\rho(t)$ is strictly increasing on $[0, c_0]$ and constant on $[c_0, \infty)$.

Under these assumptions, Lopuhaä (1989) defined S-estimators $(\hat{\mu}_n, \hat{\Sigma}_n)$ as a solution

to minimizing $\det(\Sigma)$ subject to

$$\frac{1}{n}\sum_{i=1}^{n}\rho[\{(X_i-\mu)^T\Sigma^{-1}(X_i-\mu)\}^{\frac{1}{2}}]=b_0, \tag{3.2}$$

where $0<b_0<\sup\rho(t)$. Further conditions on existence, consistency, and asymptotic normality of $(\hat{\mu}_n,\hat{\Sigma}_n)$ were given in Lopuhaä (1989). Especially, $(\hat{\mu}_n,\hat{\Sigma}_n)$ satisfy the following two equations:

$$\frac{1}{n}\sum_{i=1}^{n}u(d_i)(X_i-\mu)=0 \tag{3.3}$$

$$\frac{1}{n}\sum_{i=1}^{n}pu(d_i)(X_i-\mu)(X_i-\mu)^T-v(d_i)\Sigma=0, \tag{3.4}$$

where $u(t)=\psi(t)/t$ and $v(t)=t\psi(t)-\rho(t)+b_0$. It will be apparent that (3.3) and (3.4) are equations of the type that define M-estimators. Hence, the asymptotic distribution of $(\hat{\mu}_n,\hat{\Sigma}_n)$ will be similar to that of M-estimators. Let

$$\Psi(x,\beta)=\begin{pmatrix} u(d)(x-\mu) \\ pu(d)vech[(x-\mu)(x-\mu)^T]-v(d)vech(\Sigma)\end{pmatrix},$$

then the asymptotic distribution of $\hat{\beta}_n$ is given by

$$\sqrt{n}(\hat{\beta}_n-\beta_0)\xrightarrow{\mathcal{L}}N(0,\Omega),$$

where $\Omega=A^{-1}\Pi A^{-T}$ with $A=\mathrm{E}\dot{\Psi}(\beta_0)$ and $\Pi=var(\Psi(X_i,\beta_0))$. Consistent estimators of A and Π are

$$A_n=\frac{1}{n}\sum_{i=1}^{n}\dot{\Psi}_\beta(X_i,\hat{\beta}_n), \tag{3.5}$$

$$\Pi_n=\frac{1}{n}\sum_{i=1}^{n}\Psi(X_i,\hat{\beta}_n)\Psi^T(X_i,\hat{\beta}_n). \tag{3.6}$$

The consistency of A_n and Π_n does not depend on the underlying distribution of X_i as long as the expectations $A=\mathrm{E}\dot{\Psi}_\beta(X,\beta_0)$ and $\Pi=var(\Psi(X_i,\beta_0))$ exist. So

$$\sqrt{n}(\hat{\sigma}_n-\sigma_0)\xrightarrow{\mathcal{L}}N(0,\Omega_{22}), \tag{3.7}$$

where Ω_{22} is the lower right $p^* \times p^*$ submatrix of Ω. It can be consistently estimated by the corresponding submatrix of any consistent $\hat{\Omega}_n$.

The asymptotic variance Ω can be further simplified if the sample is from an elliptical distribution. In this case $\hat{\mu}_n$ and $\hat{\Sigma}_n$ are asymptotically independent with

$$\sqrt{n}(\hat{\mu}_n - \mu_0) \xrightarrow{\mathcal{L}} N(0, \delta_1 \Sigma), \tag{3.8}$$

where $\delta_1 = pE\psi^2(d)/\{(p-1)Eu(d) + E\dot{\psi}(d)\}^2$, and

$$\sqrt{n}(\hat{\sigma}_n - \sigma_0) \xrightarrow{\mathcal{L}} N(0, \Omega_{22}), \tag{3.9}$$

where

$$\Omega_{22} = \delta_2 D_p^+ (\Sigma \otimes \Sigma) D_p^{+T} + \delta_3 D_p^+ vec(\Sigma) vec^T(\Sigma) D_p^{+T} \tag{3.10}$$

with

$$\delta_2 = \frac{2p(p+2)E\psi^2(d)d^2}{[E\dot{\psi}(d)d^2 + (p+1)E\psi(d)d]^2},$$

$$\delta_3 = -\frac{2}{p}\delta_2 + \frac{4E(\rho(d) - b_0)^2}{[E\psi(d)d]^2},$$

and D_p is a $p^2 \times p^*$ constant duplication matrix as defined in Magnus and Neudecker (1988, p. 49).

As with M-estimators, the influence functions of S-estimators are bounded by choosing a proper function $\rho(t)$. The breakdown point of S-estimators 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 high efficiency with a breakdown point around 1/2. However, there exist functions $\rho(t)$ which can obtain the same efficiency as an M-estimator but a higher breakdown point. This is demonstrated in Lopuhaä (1989), by using Tukey's biweight function as the $\rho(t)$.

In principle, we could define an S-estimator of θ_0 as was done by Lopuhaä (1989) or Davis (1987). Since minimizing $\det(\Sigma(\theta))$ under a structured constraint as in (3.2) in the space of θ does not have the obvious statistical meaning as with standard S-estimators of (μ, Σ), we prefer using an S-estimator $\hat{\Sigma}_n$ to do covariance structure analysis as in the next section.

4 A two-stage estimation procedure

Suppose a robust estimator $\hat{\sigma}_n$ satisfying (3.1) is available. For each n, let M_n be a weight matrix. We define

$$F_n(\theta) = (\hat{\sigma}_n - \sigma(\theta))^T M_n (\hat{\sigma}_n - \sigma(\theta)) \tag{4.1}$$

and the estimator $\tilde{\theta}_n$ that satisfies $F_n(\tilde{\theta}_n) = min_{\theta \in \Theta} F_n(\theta)$ will be called a minimum chi-square estimator. This term follows from Ferguson (1958) who defined a minimum chi-square estimator when, in (3.1), $a_n = \sqrt{n}$. So we can use Ferguson's (1958) result for consistency and asymptotic normality of $\tilde{\theta}_n$ if $a_n = \sqrt{n}$. Since some robust covariance with unknown theoretical properties may find to possess a_n-consistency, we will present our application in a little more general setting. For example, we do not need the asymptotic normality of $\hat{\Sigma}_n$ in order to obtain consistency of $\tilde{\theta}_n$. The following assumptions are required.

Assumptions 4:

4.1. Θ is compact.

4.2. $\Sigma(\theta)$ is identified, that is $\Sigma(\theta_1) = \Sigma(\theta_2)$, $\theta_1, \theta_2 \in \Theta$ implies $\theta_1 = \theta_2$.

4.3. $\Sigma(\theta)$ is continuously differentiable on Θ and $\dot{\sigma}(\theta_0)$ is of full column rank.

4.4. $M_n \xrightarrow{P} M > 0$.

The following theorem is on the consistency of the minimum chi-square estimator $\tilde{\theta}_n$.

Theorem 4.1. *Under assumptions 4.1, 4.2, 4.4 and assuming the continuity of* $\sigma(\theta)$, *if* $a_n(\hat{\sigma}_n - \sigma(\theta_0)) = O_p(1)$ *for some* $a_n \to \infty$, *then* $\tilde{\theta}_n \xrightarrow{P} \theta_0$.

For obtaining $\tilde{\theta}_n$, we usually need to solve

$$\dot{\sigma}^T(\theta) M_n(\hat{\sigma}_n - \sigma(\theta)) = 0 \tag{4.2}$$

and the solution is indeed $\tilde{\theta}_n$ if it is unique. For the asymptotic normality of $\tilde{\theta}_n$, we need to assume the asymptotic normality of $\hat{\Sigma}_n$.

Theorem 4.2. *Under assumptions 4.3 and 4.4, and assuming* $a_n(\hat{\sigma}_n - \sigma(\theta_0)) \xrightarrow{\mathcal{L}} N(0, C)$ *for some* $a_n \to \infty$, *then*

$$a_n(\tilde{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Omega),$$

where

$$\Omega = (\dot{\sigma}^T M \dot{\sigma})^{-1} \dot{\sigma}^T M C M \dot{\sigma} (\dot{\sigma}^T M \dot{\sigma})^{-1}. \tag{4.3}$$

When C is nonsingular and a consistent estimator $\hat{C}_n$ is available, we can choose $M_n = \hat{C}_n^{-1}$ and (4.3) is simplified to

$$\Omega = (\dot{\sigma}^T C^{-1} \dot{\sigma})^{-1}. \tag{4.4}$$

This is the covariance matrix of the minimum variance estimator.

The robust properties of $\tilde{\theta}_n$ will follow from those of $\hat{\sigma}_n$. Let $H(\theta, \sigma) = \dot{\sigma}^T(\theta) M(\sigma - \sigma(\theta))$, then $\dot{H}_\theta(\theta, \sigma)$ is nonsingular in a neighborhood of $\tilde{\theta}_n$. So $\tilde{\theta}_n$ is an implicit function of $\hat{\sigma}_n$ decided by $H(\tilde{\theta}_n, \hat{\sigma}_n) = 0$. Since

$$\dot{\theta}(\sigma) = -(\dot{H}_\theta(\theta, \sigma))^{-1} \dot{H}_\sigma(\theta, \sigma),$$

which approaches $(\dot{\sigma}^T M \dot{\sigma})^{-1}\dot{\sigma}^T M$ at the estimated value, the influence function of the minimum chi-square estimator $\tilde{\theta}_n$ is

$$IF_\theta(x) = \{(\dot{\sigma}^T M \dot{\sigma})^{-1}\dot{\sigma}^T M\} IF_\sigma(x).$$

Thus $\tilde{\theta}_n$ inherits the local robust properties of $\hat{\Sigma}_n$. Actually $\tilde{\theta}_n$ depends on both $\hat{\sigma}_n$ and M_n. Since $\tilde{\theta}_n$ can not be expressed as an implicit function of M, we can not give the influence function of $\tilde{\theta}_n$ with respect to that of M_n. Since $\tilde{\theta}_n$ is more sensitive to $\hat{\sigma}_n$ than to M_n, which can be seen from the speed of convergence required for $\hat{\sigma}_n$ and M_n, $\tilde{\theta}_n$ will have a bounded influence function as long as both $\hat{\sigma}_n$ and M_n do. And, since θ is a continuous function of σ, and $\dot{\theta}(\sigma)$ is bounded in a neighborhood of $\hat{\sigma}_n$, it is obvious that $\tilde{\theta}_n$ also possesses the global sensitivity of $\hat{\sigma}_n$ as long as M_n is of high breakdown point.

If $\hat{\sigma}_n$ is from an M-estimator instead of from an S-estimator, we can proceed in the same way to get an estimator $\tilde{\theta}_n$ of θ_0 through the above two stage process. An immediate question is which estimator is more efficient: a direct M-estimator $\hat{\theta}_n$, or a two stage estimator $\tilde{\theta}_n$ based on an M-estimator $\hat{\sigma}_n$. We answer this question in the context of an elliptical distribution. For this we need the asymptotic distribution of an M-estimator of $\hat{\Sigma}_n$, which can be obtained through Corollary 2 by letting $\theta = \sigma$, or using the result in Tyler (1982),

$$\sqrt{n}(\hat{\sigma}_n - \sigma_0) \xrightarrow{\mathcal{L}} N(0, C),$$

where

$$C^{-1} = \frac{(1+2a)^2}{2b} D_p^T(\Sigma^{-1} \otimes \Sigma^{-1}) D_p + \kappa(p) D_p^T vec(\Sigma^{-1}) vec^T(\Sigma^{-1}) D_p. \tag{4.5}$$

Using (4.4), the asymptotic covariance Ω of the minimum chi-square estimator $\tilde{\theta}_n$ is given by

$$\Omega^{-1} = \frac{(1+2a)^2}{2b} \dot{\dot{\sigma}}^T(\Sigma^{-1} \otimes \Sigma^{-1})\dot{\dot{\sigma}} + \kappa(p)\dot{\dot{\sigma}}^T vec(\Sigma^{-1}) vec^T(\Sigma^{-1})\dot{\dot{\sigma}}. \tag{4.6}$$

Comparing (4.6) with (2.8), we can see that the asymptotic efficiency of a minimum chi-square estimator based on an M-estimator $\hat{\Sigma}_n$ and the direct M-estimator defined in Section 2 will be the same if $p_1 = p$. As we show next, this condition will be satisfied in a variety of models.

Theorem 4.3. *In an elliptical distribution, the asymptotic efficiency of a minimum chi-square estimator based on an M-estimator $\hat{\Sigma}_n$ and the direct M-estimator defined in Section 2 will be the same if*

$$\Sigma(\theta_0) = c_1\dot{\Sigma}_1 + \ldots + c_q\dot{\Sigma}_q, \tag{4.7}$$

where $\dot{\Sigma}_j = \dot{\Sigma}_{\theta_j}(\theta_0)$.

It is obvious that linear covariance structure models satisfies (4.7). For the AR(1) structure $\Sigma(\theta) = \theta_1(\theta_2^{|i-j|})$ since $\theta_{10}\dot{\Sigma}_1 = \Sigma(\theta_0)$, (4.7) is satisfied. For most factor analysis models

$$\Sigma = \Lambda\Phi\Lambda^T + \Psi,$$

(4.7) is satisfied unless parameters beyond the minimal set to achieve identification are set to prespecified nonzero values.

5 Test

As mentioned earlier, we need some quantitative method to judge the adequacy of a proposed model. Two such procedures will be proposed in this section. The first one is in the context of the M-estimator $\hat{\theta}_n$ defined in Section 2. The second one is the minimum chi-square test which can be applied to any a_n-consistent and

asymptotically normal estimators of covariance.

5.1 Test for M-estimators

We will present our test statistics in the context of structured means and structured covariances. Models with structured covariances but unstructured means are special cases. Let $\hat{\theta}_n$ be the M-estimator of θ_0 defined in Section 2. Let $\hat{\beta}_n$ be the M-estimator of β_0 defined by

$$\frac{1}{n}\sum_{i=1}^{n} s_i(\hat{\beta}_n) = 0,$$

where

$$s_i(\beta) = \begin{pmatrix} u_1(d_i)\Sigma^{-1}(X_i-\mu) \\ \frac{1}{2}D_p^T\{u_2(d_i^2)vec[\Sigma^{-1}(X_i-\mu)(X_i-\mu)^T\Sigma^{-1}] - vec(\Sigma^{-1})\} \end{pmatrix}.$$

The asymptotic variance of $\hat{\beta}_n$ is $\Gamma = B^{-1}\Xi B^{-T}$ with $\Xi = var(s_i(\beta_0))$ and $B = E\dot{s}_i(\beta_0)$. Let $\hat{\Gamma}_n$ be any consistent estimator of Γ. Let $\dot{\beta}_c(\theta)$ be a $(p+p^*)\times(p+p^*-q)$ matrix whose columns are orthogonal to those of $\dot{\beta}(\theta)$.

Theorem 5.1. *Under the assumptions 2.1 to 2.4 for both the structured and unstructured models, if $\hat{\theta}_n$ and $\hat{\beta}_n$ are consistent solutions, then*

$$T_n = n(\hat{\beta}_n - \beta(\hat{\theta}_n))^T\dot{\beta}_c(\hat{\theta}_n)\{\dot{\beta}_c^T(\hat{\theta}_n)\hat{\Gamma}_n\dot{\beta}_c(\hat{\theta}_n)\}^{-1}\dot{\beta}_c^T(\hat{\theta}_n)(\hat{\beta}_n - \beta(\hat{\theta}_n)) \qquad (5.1)$$

is distributed as $\chi^2_{p+p^*-q}$ under the null hypothesis of a correct structure.

Under a sequence of local alternatives, it can be shown that T_n approaches a noncentral chi-square variate as shown in Chapter 3 of Yuan (1995). We will not pursue the details here. Using Lemma 1 of Khatri (1966), (5.1) can be rewritten as

$$\begin{aligned} T_n &= n(\hat{\beta}_n - \beta(\hat{\theta}_n))^T\hat{\Gamma}_n^{-1}(\hat{\beta}_n - \beta(\hat{\theta}_n)) \\ &\quad -n(\hat{\beta}_n - \beta(\hat{\theta}_n))^T\hat{\Gamma}_n^{-1}\dot{\beta}(\hat{\theta}_n)\{\dot{\beta}^T(\hat{\theta}_n)\hat{\Gamma}_n^{-1}\dot{\beta}(\hat{\theta}_n)\}^{-1}\dot{\beta}^T(\hat{\theta}_n)\hat{\Gamma}_n^{-1}(\hat{\beta}_n - \beta(\hat{\theta}_n)) \end{aligned} \qquad (5.2)$$

Comparing with (5.1), (5.2) is more easy to compute.

5.2 Minimum chi-square test

In Section 4, we presented a two stage procedure through a minimum chi-square estimator. The minimized index $F_n(\tilde{\theta}_n)$ can also be used as a test statistic. Let C be nonsingular with a consistent estimator $\hat{C}_n$ and $M_n = \hat{C}_n^{-1}$.

Theorem 5.2. *Under assumptions 4.1 to 4.4, if $M = C^{-1}$ and $\tilde{\theta}_n$ is a consistent solution, then $a_n^2 F_n(\tilde{\theta}_n) \xrightarrow{\mathcal{L}} \chi^2_{p^*-q}$ under the null hypothesis of a correct structure.*

By a straight forward argument, it can be shown that $a_n^2 F_n(\tilde{\theta}_n)$ is distributed as a noncentral chi-square under some local alternatives $\theta_1 = \theta_0 + O_p(\delta/a_n)$. Since noncentral distributions are not so interesting from a practical point of view, we will not pursue this topic further here.

Note that both statistics in this section are asymptotically distribution free if $\hat{\Gamma}_n$ and $\hat{C}_n$ are distribution free estimators. Furthermore, as discussed in Sections 2 and 3, several consistent estimators for Γ and C exist if the sample is from an elliptical distribution. These will correspond to different test statistics. Even though they are all asymptotically chi-square distributed, with small to medium sample sizes, these statistics may have significantly different type I errors as demonstrated in Yuan and Bentler (1995). This problem is now under further study through simulation.

6 An Example

In this section, we will apply some of the robust procedures developed previously to

the open and closed book data set from Mardia, Kent, and Bibby (1979, Table 1.2.1). The data set consists of n=88 cases and p=5 variables. The covariance structure of this data set has been confirmed by Tanaka et al (1991) as a two factors model with the first two variables depend on the first fator and the last three variables depend on the second factor. Specifically, let $X = (x_1, \ldots, x_5)^T$, then $X = \Lambda F + E$, where

$$\Lambda^T = \Lambda^T(\theta) = \begin{pmatrix} \theta_1 & \theta_2 & 0 & 0 & 0 \\ 0 & 0 & \theta_3 & \theta_4 & \theta_5 \end{pmatrix},$$

$$var(F) = \Phi(\theta) = \begin{pmatrix} 1 & \theta_6 \\ \theta_6 & 1 \end{pmatrix},$$

and $var(E) = \Psi(\theta) = diag(\theta_7, \ldots, \theta_{11})$. So the structured covariance is

$$\Sigma(\theta) = \Lambda(\theta)\Phi(\theta)\Lambda^T(\theta) + \Psi(\theta) \tag{6.1}$$

The normal theory maximum likelihood estimator has been obtained by Tanaka et al and sensitivity analysies by Tanaka et al (1991), Cadigan (1995), and Lee and Wang (1995) indicate that the 81st case is the most influential point.

We will fit (6.1) by the two stage procedure through minimum chi-square as discussed in Sections 4 and 5, with M-estimators $\hat{\Sigma}_n$. We use this illustration because no efficient algorithms are available for most high breakdown point estimators as noted by Tyler (1991). Also, M-estimators of Σ are straightforward and have been shown to work well in a variety of settings, e.g. principal component analysis (Devlin et al 1981) and canonical variates (Campbell 1980; 1981). Another reason for us to choose the two-stage procedure is because the algorithm for estimating θ_0 can be easily implemented in standard statistical packages by simple modifications to the covariance matrix and the "distribution free" weight matrix used classically.

Six different weights were used for M-estimators $\hat{\Sigma}$. Two of these are from $t_p(df, \mu, \Sigma)$, a p-variate t-distribution with degrees of freedom df. The corresponding

weight functions are

$$w_1(d_i) = w_2(d_i^2) = (p + df)/(df + d_i^2). \tag{6.2}$$

We choose $df = 1$ and $df = 5$ respectively, with $df = 1$ corresponding to the Cauchy distribution and $df = 5$ corresponding to the minimum integer degrees of freedom for which the 4th moment of a multivariate t-distribution exists. Note that since we assume that the 4th moment of X exists, we can not assume that the data is from a Cauchy distribution. Here we use the Cauchy density only to get weight functions that guard against observations with large residuals.

Two of the Huber(q) type of weights were used with

$$w_1(d_i) = \begin{cases} 1, & if\ d_i \le k_1 \\ k/d_i, & if\ d_i > k_1 \end{cases} \tag{6.3}$$

and $w_2(d_i^2) = \{w_1(d_i)\}^2/k_2$, where k_1 is the qth percentile of χ_p^2 and k_2 is a constant satisfying $\mathrm{E}\chi_p^2 w_2(\chi_p^2) = p$. We choose $q_1 = 5$ and $q_2 = 10$ respectively for the Huber weights.

Campbell (1980) defined the M-estimator $(\hat{\mu}_n, \hat{\Sigma}_n)$ through

$$\hat{\mu}_n = \sum_{i=1}^{n} w_i X_i / \sum_{i=1}^{n} w_i, \tag{6.4}$$

$$\hat{\Sigma}_n = \sum_{i=1}^{n} w_i^2 (X_i - \hat{\mu}_n)(X_i - \hat{\mu}_n)^T / (\sum_{i=1}^{n} w_i^2 - 1), \tag{6.5}$$

where

$$w_i = w(d_i) = \omega(d_i)/d_i,$$

and

$$\omega(d_i) = \begin{cases} d_i, & if\ d_i \le d_0 \\ d_0 exp\{-.5(d_i - d_0)^2/b_2^2\}, & if\ d_i > d_0 \end{cases} \tag{6.6}$$

with $d_0 = \sqrt{p} + b_1/\sqrt{2}$. Campbell's recommendation for the constants are: (1) $b_1 = 2$, $b_2 = \infty$ which corresponds to Huber-type weights; (2) $b_1 = 2$, $b_2 = 1.25$ which corresponds to Hampel-type weights.

The "distribution free" weight matrices were used for all these six weights. The 81st case was identified as the one with the smallest weight with all these six weight functions. The estimated parameters together with their standard errors and the robust test statistics are in Table 1, where C-Ha (Campbell-Hampel) and C-Hu (Campbell-Huber) correspond to the weight functions used by Campbell with $b_1 = 2$ and $b_2 = 1.25$, ∞ respectively.

From Table 1, we can see that different weight functions generally give different estimators and different test statistics. Huber(10) gives the largest chi-square statistic with a corresponding p-value of .042, which suggests that the model fits the data reasonably well. The fit indices corresponding to other weight function are far from significant, so the null hypothesis (6.1) is not rejected. The multi-t(1) gives the smallest fit index. In most of the estimators by robust procedures, the standard errors are usually smaller than those of the corresponding normal theory MLE. The factor loading and unique variance estimators corresponding to multi-t(1) are smaller than those by other procedures, while the estimator of the factor correlation by multi-t(1) is the largest. Huber weights and Campbell weights give larger factor loadings than those by a multivariate t-distribution. No overall preference in solution is obvious when comparing the two Huber weights, the Campbell-Huber and Campbell-Hampel weights, and the different Huber and Campbell-Huber and Campbell-Hampel weights. A possible choice among these weights could be made by choosing the estimator with minimum overall variance, e.g., based on the smallest trace or determinant of the estimated asymptotic covariance matrix of the estimator. This possibility is not explored further here.

7 Discussion

In social and behavioral sciences, the data collection process is usually rough. Bad data or outliers commonly exist. Researchers definitely are aware of this problem, but a coherent strategy for dealing with the problem has not been advanced. Outlier detection or sensitivity analysis has been recommended as a precursor to a formal estimation procedure, though such a method has its own drawback (e.g., Bollen 1989). Assuming the quality of the data is good enough, the GLS procedure gives correct inference when sample size is large enough. However, if most of the data in a sample are good while a few outliers exist, the GLS procedure can totally fail because of its quadratic influence functions. Furthermore, when a sample is from a distribution with medium to large kurtosis, the GLS estimators can be highly inefficient, while proper downweighting to cases with large residuals usually should lead to more efficient estimators.

Robust methods have been developed and used in many different areas of statistics. Theory for robustly estimating multivariate location and scatter also has been studied extensively. While robust methods have been used in principal components and canonical variates, the application of robust methods to covariance structure analysis has been rather slow. This may be because robust methods in principal components and canonical variates are just eigenvalue-eigenvector decompositions of robustified matrices, which are rather straight forward, while a complete theory for robust covariance structure analysis involving estimation and testing is relatively more complicated. As we have mentioned, our two-stage procedure can be easily adapted into EQS (Bentler 1995) or LISREL (Jöreskog and Sörbom 1993), the major software packages in this field. Greater availability of such new methods hopefully could

promote the application of robust procedures in covariance structure analysis in the social and behavioral sciences.

Covariance structure analysis is usually used to model high-dimensional data sets. In this situation, a high breakdown point estimator like the S-estimator is necessary when a sample has problematic observations. Since computational problems exist for all known estimators of multivariate location and scatter with high breakdown point, this may limit their immediate use in covariance structure analysis. We hope some efficient algorithms for this situation will be developed in the near future. Since the data set we used is of dimension 5, and only the 81st case has been identified as an influential point, there may not be much difference between the S-estimator and an M-estimator for this data set. But in general applications of covariance structure analysis, high breakdown point estimators like the S-estimator will be necessary to get a robust analysis.

The results we obtained in our empirical study yield no clear winner among various weighting functions used to define our robust M-estimators. These results may or may not apply to other data sets. Further study, which may based on extensive empirical simulation, on different models, and on different underlying distributions is necessary to obtain a better understanding of the general merits of these different weights.

8 Appendix

All the theorems will be proved in this appendix. We need some lemmas for the proof of Theorems 2.1 and 2.2. The following lemma is from Theorem 2 of Jennrich (1969). Essentially the same result has been stated in Theorem 4.2.1 of Amemiya (1985).

Lemma A.1. *Let X_1, X_2, ... be an i.i.d. sample, and let $h(x,\theta)$ be measurable in x and continuous in $\theta \in \Theta$, a compact set. If $E\sup_{\theta\in\Theta} h(X,\theta) < \infty$, then with probability one $\frac{1}{n}\sum_{i=1}^{n} h(X_i,\theta)$ converges to $Eh(X,\theta)$ uniformly for all $\theta \in \Theta$.*

Let $h_i(\theta)$, $i = 1, 2, \ldots$ be independent p-variate stochastic functions of $\theta \in \Theta$ and

$$H_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} h_i(\theta).$$

Some results of Yuan and Jennrich (1995) will be used. We need the following assumptions for the next two lemmas.

Assumptions A:

A.1. $H_n(\theta_0) \to 0$ with probability one.

A.2. There is a neighborhood N of θ_0 on which with probability one $H_n(\theta)$ are continuously differentiable and the Jacobians $\dot{H}_n(\theta)$ converge uniformly to a nonstochastic function which is nonsingular at θ_0.

The following two lemmas are from Theorems 1 and 2 of Yuan and Jennrich (1995).

Lemma A.2. *Under assumptions A.1 and A.2, with probability one for any $r > 0$ there are $\hat{\theta}_n \in B_r(\theta_0)$ such that $G_n(\hat{\theta}_n) = 0$ for all n sufficiently large.*

Lemma A.3. *Under assumptions A.1 and A.2, with probability one there are zeros $\hat{\theta}_n$ of $H_n(\theta)$ such that $\hat{\theta}_n \to \theta_0$.*

Proof of Theorems 2.1 and 2.2: Since X_1, ..., X_n is an i.i.d. sample, assumption 2.4 implies assumption A.1. So we only need to check that assumptions 2.1 to 2.4 imply assumption A.2. Assumption 2.4 implies that $\mathrm{E}\dot{g}_\theta(X,\theta)$ exist in a neighborhood of θ_0 and $A{=}\mathrm{E}\dot{g}_\theta(X,\theta_0)$ is nonsingular, by Lemma A.1 we only need to

show that $\dot{g}_\theta(x,\theta)$ is bounded by an integrable function. Let

$$A_1(x,\theta) = -\frac{\dot{u}_1(d(x,\theta))}{2d(x,\theta)} \quad \begin{array}{l}\{2\dot{\mu}^T(\theta)T(x,\theta)\dot{\mu}(\theta) \\ +\dot{\mu}^T(\theta)\Sigma^{-1}(\theta)(x-\mu)vec^T(T(x,\theta))\dot{\bar{\sigma}}(\theta)\}\end{array},$$

$$A_2(x,\theta) = u_1(d(x,\theta))\{[\frac{\partial}{\partial\theta^T}(\dot{\mu}^T(\theta)\Sigma^{-1}(\theta))](x-\mu(\theta)) - \dot{\mu}^T(\theta)\Sigma^{-1}(\theta)\dot{\mu}(\theta)\},$$

$$A_3(x,\theta) = \frac{1}{2}[\frac{\partial}{\partial\theta^T}\dot{\bar{\sigma}}^T(\theta)]\{u_2(d^2(x,\theta))vec(T(x,\theta)) - vec(\Sigma^{-1}(\theta))\},$$

$$A_4(x,\theta) = -\tfrac{1}{2}\dot{u}_2(d^2(x,\theta))\dot{\bar{\sigma}}^T(\theta)vec(T(x,\theta)) \quad \begin{array}{l}\{2(x-\mu)^T\Sigma^{-1}(\theta)\dot{\mu}(\theta) \\ +vec^T(T(x,\theta))\dot{\bar{\sigma}}(\theta)\}\end{array},$$

$$A_5(x,\theta) = \frac{1}{2}\dot{\bar{\sigma}}^T(\theta)\{u_2(d^2(x,\theta))vec(\dot{T}_\theta(x,\theta)) - (\Sigma^{-1}(\theta)\otimes\Sigma^{-1}(\theta))\dot{\bar{\sigma}}(\theta)\},$$

then

$$\dot{g}_\theta(x,\theta) = A_1(x,\theta) + A_2(x,\theta) + A_3(x,\theta) + A_4(x,\theta) + A_5(x,\theta),$$

where

$$\begin{array}{rl} vec(\dot{T}_\theta(x,\theta)) & = -(\{T(x,\theta)\otimes\Sigma^{-1}(\theta)\}\dot{\bar{\sigma}}(\theta) + \{\Sigma^{-1}(\theta)\otimes T(x,\theta)\}\dot{\bar{\sigma}}(\theta) \\ & +\{[\Sigma^{-1}(\theta)(x-\mu)]\otimes\Sigma^{-1}(\theta)\}\dot{\mu}(\theta) \\ & +\{\Sigma^{-1}(\theta)\otimes[\Sigma^{-1}(\theta)(x-\mu)]\}\dot{\mu}(\theta))\end{array}.$$

Since X has a finite fourth moment, $\Sigma(\theta) > \epsilon I$ in a neighborhood of θ_0, both $d(x,\theta)$, $d^2(x,\theta)$ are bounded by integrable functions in a neighborhood of θ_0. Since

$$\begin{array}{rl} ||A_1(x,\theta)|| & \le |\dot{u}_1(d(x,\theta))|\{d(x,\theta)||\dot{\mu}^T(\theta)\Sigma^{-1}(\theta)\dot{\mu}(\theta)|| \\ & +\frac{1}{2}d^2(x,\theta)||\dot{\mu}^T(\theta)\Sigma^{-1}(\theta)\dot{\mu}(\theta)||^{\frac{1}{2}}||\Sigma^{-\frac{1}{2}}(\theta)||^2||\dot{\bar{\sigma}}||\}\end{array},$$

there is a neighborhood of θ_0 on which $A_1(X,\theta)$ is bounded by an integrable function. Similarly as $A_1(x,\theta)$, it can be shown that $A_2(x,\theta)$ to $A_5(x,\theta)$ are bounded by integrable functions in a neighborhood of θ_0, so is $\dot{g}_\theta(X,\theta)$.

Proof of Theorem 2.3: Using a Taylor expansion on $G_n(\theta)$, we have

$$G_n(\hat{\theta}_n) = G_n(\theta_0) + \dot{G}_n(\bar{\theta}_n)(\hat{\theta}_n - \theta_0), \qquad (A.1)$$

where $\dot{G}_n(\bar{\theta}_n)$ denotes that each row of $\dot{G}_n(\theta)$ is evaluated at a consistent $\bar{\theta}_n$ of θ_0. Since $V = var(g_i(\theta_0)) > 0$, the multivariate version of classical central limit theorem implies

$$\sqrt{n}G_n(\theta_0) \xrightarrow{\mathcal{L}} N(0, \Pi).$$

Since $\dot{G}_n(\bar{\theta}_n) \xrightarrow{P} A = E\dot{g}_i(\theta_0)$ and A is nonsingular, (A.1) can be written as

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -A^{-1}\sqrt{n}G_n(\theta_0) + o_p(1). \tag{A.2}$$

So the theorem follows from (A.2) and the Slutsky theorem.

Proof of Corollaries 2.1 and 2.2: Since the proof of these two corollaries are similar, and both involve tedious but direct calculations, we only list the key properties of an elliptical distribution used in the proofs. Let $X = (x_1, \ldots, x_p)$ obey an elliptical distribution, then

$$Ex_i^{r_i}x_j^{r_j}x_k^{r_k}x_l^{r_l} = 0, \tag{A.3}$$

if $r = r_i + r_j + r_k + r_l$ is an odd number. Let $d(X) = \{(X-\mu_0)^T\Sigma_0^{-1}(X-\mu_0)\}^{\frac{1}{2}}$ and $U = \Sigma_0^{-\frac{1}{2}}(X-\mu_0)/d(X)$, then $d(X)$ and U are independent and

$$EUU^T = \frac{1}{p}I \tag{A.4}$$

and

$$Evec(UU^T)vec^T(UU^T) = \frac{1}{p(p+2)}\{I_{p^2} + K_p + vec(I_p)vec^T(I_p)\}, \tag{A.5}$$

where K_p is the commutation matrix defined in Magnus and Neudecker (1988, p. 46). Equations (A.3) and (A.4) can be found in Fang et al (1989, p. 34) and equation (A.5) was proved in Magnus and Neudecker (1979). Another property used is $Ed^2(X)u_2(d^2(X)) = p$, this follows since the estimating equation is unbiased. With these properties, the proofs of these two corollaries are straight forward.

Proof of Theorem 4.1:

Since $a_n \to \infty$, $\sigma(\theta)$ is continuous, and M_n is consistent, we have

$$\begin{aligned} F_n(\theta) - F_n(\theta_0) &= 2(\hat{\sigma}_n - \sigma(\theta_0))^T M_n(\sigma(\theta_0) - \sigma(\theta)) \\ &\quad +(\sigma(\theta) - \sigma(\theta_0))^T M_n(\sigma(\theta) - \sigma(\theta_0)) \\ &\quad \xrightarrow{P} (\sigma(\theta) - \sigma(\theta_0))^T M(\sigma(\theta) - \sigma(\theta_0)) \end{aligned} \tag{A.6}$$

uniformly in $\theta \in \Theta$. Let $\tilde{\theta}_{n_i}$ be any convergent subsequence of $\tilde{\theta}_n$ with limit θ', then

$$F_{n_i}(\tilde{\theta}_{n_i}) - F_{n_i}(\theta_0) \leq 0. \tag{A.7}$$

Let $n_i \to \infty$ in (A.7), then

$$(\sigma(\theta') - \sigma(\theta_0))^T M(\sigma(\theta') - \sigma(\theta_0)).$$

So $\theta' = \theta_0$. Since $\tilde{\theta}_{n_i}$ can be any convergent subsequence, $\tilde{\theta}_n \xrightarrow{P} \theta_0$.

Proof of Theorem 4.2:

From

$$\dot{\sigma}^T(\tilde{\theta}_n) M_n(\hat{\sigma}_n - \sigma(\tilde{\theta}_n)) = 0,$$

it follows that

$$\begin{aligned} \dot{\sigma}^T(\tilde{\theta}_n) M_n(\hat{\sigma}_n - \sigma(\theta_0)) &= \dot{\sigma}^T(\tilde{\theta}_n) M_n(\sigma(\tilde{\theta}_n) - \sigma(\theta_0)) \\ &= \dot{\sigma}^T(\tilde{\theta}_n) M_n \dot{\sigma}(\bar{\theta}_n)(\tilde{\theta}_n - \theta_0), \end{aligned} \tag{A.8}$$

where $\dot{\sigma}(\bar{\theta}_n)$ denotes that each row of $\dot{\sigma}(\theta)$ is evaluated at a consistent estimator $\bar{\theta}_n$ of θ_0. Since $\dot{\sigma}^T(\tilde{\theta}_n) M_n \dot{\sigma}(\bar{\theta}_n) \xrightarrow{P} \dot{\sigma}^T M \dot{\sigma}$, which is positive definite, (A.8) can be rewritten as

$$a_n(\tilde{\theta}_n - \theta_0) = [\dot{\sigma}^T(\tilde{\theta}_n) M_n \dot{\sigma}(\bar{\theta}_n)]^{-1} \dot{\sigma}^T(\tilde{\theta}_n) M_n a_n(\hat{\sigma}_n - \sigma(\theta_0)). \tag{A.9}$$

The theorem follows from (A.9) and the Slutsky theorem.

Proof of Theorem 4.3:

Since

$$vec(\Sigma^{-1}) = (\Sigma^{-\frac{1}{2}} \otimes \Sigma^{-\frac{1}{2}})vec(I),$$

$p_1 = vec^T(I)P_\Delta vec(I)$, where $P_\Delta = \Delta(\Delta^T\Delta)^{-1}\Delta^T$ is the usual projection matrix formed by $\Delta = (\Sigma^{-\frac{1}{2}} \otimes \Sigma^{-\frac{1}{2}})\dot{\sigma}$. If $vec(I)$ is in the space of Δ, then $p_1 = p$. From

$$\Delta = (vec(\Sigma^{-\frac{1}{2}}\dot{\Sigma}_1\Sigma^{-\frac{1}{2}}), \ldots, vec(\Sigma^{-\frac{1}{2}}\dot{\Sigma}_q\Sigma^{-\frac{1}{2}})),$$

$vec(I)$ being in the space of Δ is equivalent to

$$c_1\Sigma^{-\frac{1}{2}}\dot{\Sigma}_1\Sigma^{-\frac{1}{2}} + \ldots + c_q\Sigma^{-\frac{1}{2}}\dot{\Sigma}_q\Sigma^{-\frac{1}{2}} = I$$

for some c_1, ..., c_q. That is Σ can be expressed as

$$\Sigma = c_1\dot{\Sigma}_1 + \ldots + c_q\dot{\Sigma}_q.$$

Proof of Theorem 5.1:

From (A.2),

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -A^{-1}\sqrt{n}G_n(\theta_0) + o_p(1). \tag{A.10}$$

Since an unstructured model is also a structure, we have

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = -B^{-1}\sqrt{n}S_n(\beta_0) + o_p(1), \tag{A.11}$$

where $S_n(\beta) = \frac{1}{n}\sum_{i=1}^n s_i(\beta)$. Direct calculation shows

$$G_n(\theta) = \dot{\beta}^T(\theta)S_n(\beta(\theta)). \tag{A.12}$$

From (A.10), (A.11) and (A.12), we have

$$\begin{aligned}\sqrt{n}(\beta(\hat{\theta}_n) - \beta(\theta_0)) &= \dot{\beta}\sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1) \\ &= \dot{\beta}A^{-1}\dot{\beta}^T B\sqrt{n}(\hat{\beta}_n - \beta_0) + o_p(1),\end{aligned} \tag{A.13}$$

and consequently

$$\sqrt{n}(\hat{\beta}_n - \beta(\hat{\theta}_n)) = (I - \dot{\beta}A^{-1}\dot{\beta}^T B)\sqrt{n}(\hat{\beta}_n - \beta_0) + o_p(1). \qquad (A.14)$$

Since

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{\mathcal{L}} N(0, \Gamma), \qquad (A.15)$$

(A.14) and (A.15) imply

$$\sqrt{n}(\hat{\beta}_n - \beta(\hat{\theta}_n)) \xrightarrow{\mathcal{L}} N(0, V), \qquad (A.16)$$

where

$$V = (I - \dot{\beta}A^{-1}\dot{\beta}^T B)\Gamma(I - \dot{\beta}A^{-1}\dot{\beta}^T B)^T.$$

Since the columns of $\dot{\beta}_c(\theta)$ are orthogonal to those of $\dot{\beta}(\theta)$, then

$$\dot{\beta}_c^T(\hat{\theta}_n)\sqrt{n}(\hat{\beta}_n - \beta(\hat{\theta}_n)) \xrightarrow{\mathcal{L}} N(0, \dot{\beta}_c^T(\theta_0)V\dot{\beta}_c(\theta_0)),$$

and

$$\dot{\beta}_c^T(\theta_0)V\dot{\beta}_c(\theta_0) = \dot{\beta}_c^T(\theta_0)\Gamma\dot{\beta}_c(\theta_0).$$

So the statistic

$$T_n = n(\hat{\beta}_n - \beta(\hat{\theta}_n))^T\dot{\beta}_c(\hat{\theta}_n)\{\dot{\beta}_c^T(\hat{\theta}_n)\hat{\Gamma}_n\dot{\beta}_c(\hat{\theta}_n)\}^{-1}\dot{\beta}_c^T(\hat{\theta}_n)(\hat{\beta}_n - \beta(\hat{\theta}_n))$$

is distributed as $\chi^2_{p+p^*-q}$ under the null hypothesis of a correct structure.

Proof of Theorem 5.2:

From (A.9) it follows

$$a_n(\sigma(\tilde{\theta}_n) - \sigma(\theta_0)) = \dot{\sigma}(\dot{\sigma}^T C^{-1}\dot{\sigma})^{-1}\dot{\sigma}^T C^{-1} a_n(\sigma_n - \sigma(\theta_0)) + o_p(1). \qquad (A.17)$$

So

$$\begin{aligned} a_n(\hat{\sigma}_n - \sigma(\tilde{\theta}_n)) &= a_n(\hat{\sigma}_n - \sigma(\theta_0)) + a_n(\sigma(\theta_0) - \sigma(\tilde{\theta}_n)) \\ &= \{I - \dot{\sigma}(\dot{\sigma}^T C^{-1}\dot{\sigma})^{-1}\dot{\sigma}^T C^{-1}\} a_n(\sigma_n - \sigma(\theta_0)) + o_p(1), \end{aligned}$$

and $a_n^2 F_n(\tilde{\theta}_n)$ can be written as

$$a_n^2 F_n(\tilde{\theta}_n) = a_n(\sigma_n - \sigma(\theta_0))^T C^{-\frac{1}{2}} P C^{-\frac{1}{2}} a_n(\sigma_n - \sigma(\theta_0)) + o_p(1), \qquad (A.18)$$

where

$$P = \{I - C^{-\frac{1}{2}}\dot{\sigma}(\dot{\sigma}^T C^{-1}\dot{\sigma})^{-1}\dot{\sigma}^T C^{-\frac{1}{2}}\}$$

is a projection matrix of rank $p^* - q$. The theorem follows from (A.18).

References

[1] Amemiya, T. (1985), *Advanced Econometrics,* Massachusetts: Harvard University Press.

[2] Bentler, P. M. (1995), *EQS Structural Equations Program Manual,* Encino, CA: Multivariate Software.

[3] Bentler, P. M., and Dudgeon, P. (1996), "Covariance structure analysis: Statistical practice, theory, and directions," *Annual Review of Psychology,* **47**, 541–570.

[4] Bollen, K. A. (1989), *Structural Equations with Latent Variables,* New York: Wiley.

[5] Browne, M. W. (1982), "Covariance structures," in *Topics in Applied Multivariate analysis*, ed. D. M. Hawkins, Cambridge: Cambridge University Press, pp. 72–141.

[6] Browne, M. W. (1984), "Asymptotic distribution-free methods for the analysis of covariance structures," *British Journal of Mathematical and Statistical Psychology,* **37**, 62–83.

[7] Cadigan, N. G. (1995), "Local influence in structural equation models," *Structural Equation Modeling,* **2**, 13–30.

[8] Campbell, N. A. (1980), "Robust procedures in multivariate analysis I: Robust covariance estimation," *Applied Statistics,* **29**, 231–237.

[9] Campbell, N. A. (1982), "Robust procedures in multivariate analysis II: Robust canonical variate analysis," *Applied Statistics,* **31**, 1–8.

[10] Chamberlain, G. (1982), "Multivariate regression models for panel data," *Journal of Econometrics,* **18**, 5–46.

[11] Davies, P. L. (1987), "Asymptotic behavior of S-estimators of multivariate location parameters and dispersion matrices," *The Annals of Statistics,* **15**, 1269–1292.

[12] Devlin, S. J., Gnanadesikan, R., and Kettenring, J. R. (1981), "Robust estimation of dispersion matrices and principal components," *Journal of the American Statistical Association,* **76**, 354–362.

[13] Donoho, D. L. (1982), "Breakdown properties of multivariate location estimators," Ph.D. qualifying paper, Department of Statistics, Harvard University.

[14] Donoho, D. L., and Liu, R. C. (1988a), "The "automatic" robustness of minimum distance functions," *The Annals of Statistics,* **16**, 552–586.

[15] Donoho, D. L., and Liu, R. C. (1988b), "Pathologies of some minimum distance estimators," *The Annals of Statistics,* **16**, 587–608.

[16] Fang, K. T., Kotz, S., and Ng, K. (1989), *Symmetric Multivariate and Related Distributions,* London: Chapman & Hall.

[17] Ferguson, T. (1958), "A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities," *Annals of Mathematical Statistics,* **29**, 1046–1062.

[18] Fu, J. C. (1989), "Method of KIM-ZAM: an algorithm for computing the maximum likelihood estimator," *Statistics & Probability Letters,* **8**, 289–296.

[19] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986), *Robust Statistics: The Approach based on Influence Functions,* New York: Wiley.

[20] Hu, L, Bentler, P. M., and Kano, Y. (1992), "Can test statistics in covariance structure analysis be trustedΓ" *Psychological Bulletin,* **112**, 351–362.

[21] Huber, P. J. (1977), "Robust covariances," in *Statistical Decision Theory and Related Topics,* Vol. 2, eds. S. S. Gupta and D. S. Moore, New York: Academic Press, pp. 165–191.

[22] Huber, P. J. (1981), *Robust Statistics,* New York: Wiley.

[23] Jennrich, R. I. (1969), "Asymptotic properties of non-linear least squares estimators," *Annals of Mathematical Statististics,* **40**, 633–643.

[24] Jöreskog, K. G., and Sörbom, D. (1993), *LISREL 8 User's Reference Guide,* Chicago: Scientific Software International.

[25] Kano, Y. (1994), "Consistency property of elliptical probability density functions," *Journal of Multivariate Analysis,* **51**, 139–147.

[26] Kent, J. T., Tyler, D. E., and Vardi, Y. (1994), "A curious likelihood identity for the multivariate t-distribution," *Communications in Statistics-Simulation and Computation,* **23**, 441–453.

[27] Khatri (1966), "A note on a MANOVA model applied to problems in growth curves," *Annals of the Institute of Statistical Mathematics,* **18**, 75–86.

[28] Lee, S. Y., and Wang, S. J. (1995), "Sensitivity analysis of structural equation models," *Psychometrika,* in press.

[29] Lopuhaä, H. P. (1989), "On the relation between S-estimators and M-estimators of multivariate location and covariances," *The Annals of Statistics,* **17**, 1662–1683.

[30] Lopuhaä, H. P. (1991a), "τ-estimators for location and scatter," *The Canadian Journal of Statististics,* **19**, 307–321.

[31] Lopuhaä, H. P. (1991b), "Breakdown point and asymptotic properties of multivariate S-estimators and τ-estimators: a summary," in *Directions in Robust Statistics and Diagnostics: Part I*, eds. W. Stahel and S. Weisberg, New York: Springer-Verlag, pp. 167–182.

[32] Magnus, J. R., and Neudecker, H. (1979), "The commutation matrix: some properties and applications," *The Annals of Statistics,* **7**, 381–394.

[33] Magnus, J. R., and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics,* New York: Wiley.

[34] Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis,* New York: Academic Press.

[35] Maronna, R. A. (1976), "Robust M-estimators of multivariate location and scatter," *The Annals of Statistics,* **4**, 51–67.

[36] Rousseeuw, P. J. (1983), "Multivariate estimation with high breakdown point," Fourth Pannonian Symposium on Mathematical Statistics, Bad Tatzmanndorf, Austria.

[37] Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics,* John Wiley & Sons.

[38] Stahel, W. A. (1981), "Breakdown of covariance estimators," Research report 31, Fachgruppe für Statistik, E.T.H., Zurich.

[39] Tamura, R. N., and Boos, D. D. (1986), "Minimum Hellinger distance estimation for multivariate location and covariance," *Journal of the American Statistical Association,* **81**, 223–229.

[40] Tanaka, Y., Watadani, S., and Moon, S. H. (1991), "Influence in covariance structure analysis: with an application to confirmatory factor analysis," *Communication in Statistics-Theory and Method,* **20**, 3805–3821.

[41] Tyler, D. E. (1982), "Radial estimates and the test for sphericity," *Biometrika,* **69**, 429–436.

[42] Tyler, D. E. (1983), "Robustness and efficiency properties of scatter matrices," *Biometrika,* **70**, 411–420.

[43] Tyler, D. E. (1987), "A distribution-free M-estimator of multivariate scatter," *The Annals of Statistics,* **15**, 234–251.

[44] Tyler, D. E. (1991), "Some issues in the robust estimation of multivariate location and scatter," in *Directions in Robust Statistics and Diagnostics: Part II*, eds. W. Stahel and S. Weisberg, New York: Springer-Verlag, pp. 145–157.

[45] Wilcox, R. R. (1994), "The percentage bend correlation coefficient," *Psychometrika,* **59**, 601–616.

[46] Yuan, K.-H. (1995), *Asymptotics for Nonlinear Regression Models with Applications,* unpublished Ph.D. dissertation, UCLA, Dept. of Mathematics.

[47] Yuan, K.-H., and Bentler, P. M. (1995), "Mean and covariance structure analysis: theoretical and practical improvements," under editorial review.

[48] Yuan, K.-H., and Jennrich, R. I. (1995), "Estimating equation asymptotics," under editorial review.