British Journal of Mathematical and Statistical Psychology (2000), **53**, 31–50 *Printed in Great Britain*

Robust transformation with applications to structural equation modelling

Ke-Hai Yuan*

University of North Texas, USA

Wai Chan

The Chinese University of Hong Kong, Hong Kong, PRC

Peter M. Bentler

University of California, Los Angeles, USA

Data sets in social and behavioural sciences are seldom normal. Influential cases or outliers can lead to inappropriate solutions and problematic conclusions in structural equation modelling. By giving a proper weight to each case, the influence of outliers on a robust procedure can be minimized. We propose using a robust procedure as a transformation technique, generating a new data matrix that can be analysed by a variety of multivariate methods. Mardia's multivariate skewness and kurtosis statistics are used to measure the effect of the transformation in achieving approximate normality. Since the transformation makes the data approximately normal, applying a classical normal theory based procedure to the transformed data gives more efficient parameter estimates. Three procedures for parameter evaluation and model testing are discussed. Six examples illustrate the various aspects with the robust transformation.

1. Introduction

As one of the major tools for studying the relationships among latent constructs, structural equation modelling (SEM) has been used extensively in social and behavioural sciences. This is reflected in the dramatic increase in the literature on SEM in the past decade (see Austin & Calderón, 1996; Austin & Wolfle, 1991; Bentler & Dudgeon, 1996; Tremblay & Gardner, 1996). The classical statistics associated with SEM are based on the assumption of multivariate normality (Bollen, 1989; Jöreskog, 1969). Since data sets in social and behavioural sciences are seldom normal (Micceri, 1989), Browne (1984) developed a generalized least-squares (GLS) approach which is asymptotically distribution-free (ADF). The normal theory maximum likelihood (ML) and the ADF methods are the two main approaches to SEM and hence are implemented in major SEM software programs such as LISREL (Jöreskog & Sörbom, 1993) and EQS (Bentler, 1995). These approaches can give consistent parameter

* Requests for reprints should be addressed to Dr Ke-Hai Yuan, Department of Psychology, University of North Texas, P.O. Box 311280, Denton, TX 76203-1280, USA (e-mail: kyuan@unt.edu).

estimates for a variety of distributions. However, for typical data sets with excess kurtosis in the social and behavioural sciences, estimates by these methods are generally not efficient. It is known that the sample covariance matrix **S** is efficient only when data are normal. With influential cases or outliers in a sample, **S** will be inefficient or biased. Actually, not just the normal theory and ADF methods, but any methods based on **S** will inherit the problem of inaccurate model evaluation, such as biased parameter estimates and incorrect test statistics. There are at least two types of approaches to dealing with this problem. The first involves the use of transformations to achieve multivariate normality, and the second involves the use of robust statistics. In this paper, we combine these two approaches.

In pioneering work, Mooijaart (1993) proposed the use of univariate Box–Cox transformations of badly distributed variables to achieve better performance in structural modelling, and showed that the method could be implemented effectively through ML. He concluded that optimally transforming the variables could lead to substantially different, and presumably better, models. It is likely that Mooijaart's approach could be improved if it were possible to replace the separate univariate transformations by multivariate transformations, and if outliers could also be handled within the methodology. In the context of multivariate analysis, Velilla (1995) explored a generalization of the Box–Cox transformation family to the multivariate case. In theory, SEM can proceed based on the transformed data in a way parallel to that pioneered by Mooijaart. However, transformation to normality based on Box–Cox transformations requires the estimation of many extra parameters besides the mean and covariance matrix. More importantly, a conceptual issue is that with the Box–Cox transformation it is hard to know what the relationship is between the covariance structure of the transformed data and that of the original data which are of primary interest.

For a data set with outliers, robust estimation of the population covariance matrix has been studied and recommended by a variety of authors in the statistical literature (Hampel, Ronchetti, Rousseeuw, & Stahel, 1986; Huber, 1981; Wilcox, 1997). Besides diminishing the influence of outliers, a robust covariance is generally more efficient than the sample covariance for a data set from a distribution with heavy tails. Yuan and Bentler (1998a, 1998b) recently proposed several robust methods for SEM. Compared with the classical approaches, these robust approaches have the following advantages. First, robust approaches can still give reasonable solutions with problematic data, while classical approaches lead to inappropriate solutions such as Heywood cases. Second, by downweighting the influence of outliers, robust approaches lead to smaller chi-square statistics which give more support to a theoretically interesting model. Third, for data sets that are approximately normal, both the classical and robust approaches lead to basically the same conclusions. Based on the above comparisons, and considering that data collection procedures in the social and behavioural science are susceptible to outliers, robust methods should be at least as relevant to SEM as they are to regression, where a variety of robust procedures have been developed.

In this paper, by comparing the formula for a robust covariance with that of the ordinary sample covariance, we propose using the robust estimation process as a data transformation procedure. Comparing this transformation to the Box–Cox transformation, there is no need to estimate any extra transformation parameters besides the mean vector and the covariance matrix. Since the sample covariance of the transformed data is just the robust covariance matrix, as showed in Yuan and Bentler (1998b), the covariance structure of the transformed sample will be the same as the covariance structure of the original sample when the data are approximately elliptically distributed and the model structure is invariant under a constant

scaling factor (ICSF) (Browne, 1982). In order to operationalize our approach, we will use Mardia's multivariate skewness and kurtosis statistics to evaluate the degree of normality of the transformed data. As we shall see in some of the examples, outliers create skewness and kurtosis in the sample, and a robust transformation makes the data approximately normal. After the transformation, any procedures that could be applied to the original sample can also be applied to the transformed sample. Since the transformed sample approximately follows a multivariate normal distribution, standard normal theory based procedures should provide quite efficient parameter estimates and reliable model evaluation. For comparison purposes and in cases when the transformed sample may still not be normally distributed, we also apply two rescaled test statistics and sandwich-type standard errors (Satorra & Bentler, 1988, 1994; Yuan & Bentler, 1998b) to the transformed sample.

2. Robust transformation to normality

The multivariate normal distribution has been the key assumption for almost all of the classical multivariate techniques in data analysis (Anderson, 1984). This assumption also justifies the use of the sample covariance matrix $\mathbf{S}$ because it is the most efficient estimate of the population covariance matrix Σ when data are normally distributed. Since data sets in practice may not follow multivariate normal distributions, various attempts to generate multivariate non-normal distributions have been made (Fang, Kotz & Ng, 1990; Olkin, 1994). Among the generalizations to multivariate non-normal distributions, the class of elliptical distributions has been well studied and found applicable in many different circumstances (Kano, Berkane, & Bentler, 1993; Lange, Little, & Taylor, 1989; Little, 1988).

The density of an elliptical distribution is given by

$$f(\mathbf{x}) = |\Sigma|^{-1/2} h\{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\}, \tag{2.1}$$

where $h(\cdot)$ is a scalar function that does not depend on $\boldsymbol{\mu}$ and Σ. The multivariate normal distribution corresponds to $h(r) = (2\pi)^{-p/2}\exp(-r/2)$. By choosing different $h(\cdot)$, a variety of distributions with heavier or lighter tails than those of a normal distribution can be obtained (Fang *et al.*, 1990). Notice that the sample covariance matrix is not the most efficient estimate of Σ unless the distribution is normal.

Within the family of elliptical distributions, several methods have been proposed to estimate $\boldsymbol{\mu}$ and Σ. For a p-variate sample $\mathbf{X}_1, \ldots, \mathbf{X}_N$, let

$$d^2(\mathbf{X}_i, \boldsymbol{\mu}, \Sigma) = (\mathbf{X}_i - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X}_i - \boldsymbol{\mu})$$

be the Mahalanobis distance and $u_1(t)$ and $u_2(t)$ be non-negative scalar functions. Maronna (1976) defined the robust M-estimator $(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$ by solving the equations

$$\hat{\boldsymbol{\mu}} = \sum_{i=1}^{N} u_1\left\{d(\mathbf{X}_i, \hat{\boldsymbol{\mu}}, \hat{\Sigma})\right\}\mathbf{X}_i \Big/ \sum_{i=1}^{N} u_1\left\{d(\mathbf{X}_i, \hat{\boldsymbol{\mu}}, \hat{\Sigma})\right\} \tag{2.2a}$$

and

$$\hat{\Sigma} = \sum_{i=1}^{N} u_2\left\{d^2(\mathbf{X}_i, \hat{\boldsymbol{\mu}}, \hat{\Sigma})\right\}(\mathbf{X}_i - \hat{\boldsymbol{\mu}})(\mathbf{X}_i - \hat{\boldsymbol{\mu}})'/N. \tag{2.2b}$$

Different weight functions lead to different estimators. When $u_1(d) = u_2(d^2) = 1$, then (2.2) defines the classical sample mean and sample covariance. When $u_1(d) = u_2(d^2) = (p+m)/$

$(m + d^2)$, then the solution to (2.2) is the maximum likelihood estimate (MLE) based on a p-variate t distribution with degrees of freedom m. More generally, when $u_1(t) = u_2(t^2) = -2\dot{h}(t^2)/h(t^2)$, where $\dot{h}(\cdot)$ is the derivative of $h(\cdot)$, then (2.2) defines the MLE corresponding to distribution (2.1). The well-known Huber-type M-estimator will be obtained with

$$u_1(d) = \begin{cases} 1, & \text{if } d \le r, \\ r/d, & \text{if } d > r, \end{cases} \tag{2.3}$$

and $u_2(d^2) = \{u_1(d)\}^2/\tau$ (Tyler, 1983), where r^2 is given by $P(\chi_p^2 > r^2) = \alpha$, α is the proportion of outliers one wants to control assuming the massive data cloud follows a multivariate normal distribution, and τ is a constant such that $E\{\chi_p^2 u_2(\chi_p^2)\} = p$ which makes the estimator $\hat{\Sigma}$ unbiased if sampling from a p-variate normal distribution. Within the family of elliptical distributions, theoretical properties of $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ have been studied extensively by various authors (Huber, 1977; Maronna, 1976; Tyler, 1983, 1987). Even though data sets in practice may not exactly follow an elliptical distribution, robust methods have shown their effectiveness in applications such as principal components (Ammann, 1989; Devlin, Gnanadesikan, & Kettenring, 1981), canonical correlation (Campbell, 1980, 1982), discriminant analysis (Kharin, 1996), SEM (Huba & Harlow, 1987; Yuan & Bentler, 1998a, 1998b), and repeated measures (Lange *et al.*, 1989; Little, 1988). As we shall see in our applications, the robust transformation can reduce kurtosis as well as skewness in the original sample.

The solution to (2.2) needs an iterative procedure. For example, letting $\boldsymbol{\mu}^{(j)}$ and $\Sigma^{(j)}$ be the jth-stage solutions, the solutions at the $(j+1)$th stage are given by

$$\hat{\boldsymbol{\mu}}^{(j+1)} = \sum_{i=1}^{N} u_1\{d(\mathbf{X}_i, \boldsymbol{\mu}^{(j)}, \Sigma^{(j)})\}\mathbf{X}_i / \sum_{i=1}^{N} u_1\{d(\mathbf{X}_i, \boldsymbol{\mu}^{(j)}, \Sigma^{(j)})\} \tag{2.4a}$$

and

$$\hat{\Sigma}^{(j+1)} = \sum_{i=1}^{N} u_2\{d^2(\mathbf{X}_i, \boldsymbol{\mu}^{(j)}, \Sigma^{(j)})\}(\mathbf{X}_i - \boldsymbol{\mu}^{(j)})(\mathbf{X}_i - \boldsymbol{\mu}^{(j)})'/N. \tag{2.4b}$$

The estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ are obtained at the convergence of this process. The iterative process (2.4) is the well-known iteratively reweighted least squares (IRLS) algorithm whose convergence properties have been studied by Holland and Welsch (1977), Rubin (1983) and Green (1984). The sample mean and sample covariance matrix provide convenient initial estimates. Our experience with a variety of real data sets is that the process converges in only a few steps.

Now we are ready for our key proposal. Let $u_{2i} = u_2\{d^2(\mathbf{X}_i, \hat{\boldsymbol{\mu}}, \hat{\Sigma})\}$ and

$$\mathbf{X}_{bi} = \sqrt{u_{2i}}(\mathbf{X}_i - \hat{\boldsymbol{\mu}}). \tag{2.5a}$$

Then we can rewrite (2.2b) as

$$\hat{\Sigma} = \sum_{i=1}^{N} \mathbf{X}_{bi}\mathbf{X}_{bi}'/N, \tag{2.5b}$$

which is just the sample covariance matrix of the transformed sample $\mathbf{X}_{bi}$. When $u_1 = u_2 = 1.0$, (2.5*a*) corresponds to the centred sample $\mathbf{X}_{ai} = \mathbf{X}_i - \bar{\mathbf{X}}$. As mentioned earlier, if the sample comes from (2.1) and we have used $u_1(t) = u_2(t^2) = -2\dot{h}(t^2)/h(t^2)$ in (2.2), $\hat{\Sigma}$ would be the most efficient estimate for the population Σ. It is known that the sample covariance matrix is the most efficient estimate of the population covariance matrix only when the sample is from a normal distribution. This implies that we will get an almost normal sample $\mathbf{X}_{bi}$ if weights $u_1(t)$ and $u_2(t)$ in (2.2) are correctly chosen. In practice, we do not know the distributional form of $h(t)$. However, as will be demonstrated in Section 4, an approximately normal sample $\mathbf{X}_{bi}$ will be obtained if $u_1(t)$ and $u_2(t)$ are properly chosen. So we may regard (2.2) as a transformation to achieve near-normality.

As with other transformation procedures, we may treat the $\mathbf{X}_{bi}$ as raw data for any given statistical application. This approach can be compared with treating the typical deviation data $\mathbf{X}_{ai} = \mathbf{X}_i - \bar{\mathbf{X}}$ as raw data. Even though the $\mathbf{X}_{bi}$ or the $\mathbf{X}_{ai}$ are not independent samples, the correlation among the individual cases is only in the magnitude of $1/N$. We may also think of u_{2i} and $\boldsymbol{\mu}$ as transformation parameters which are somewhat like the power parameters in the Box–Cox transformation (Velilla, 1995). Besides treating $\mathbf{X}_{bi}$ as transformed data, we may also regard the u_{2i} and $\boldsymbol{\mu}$ as nuisance parameters. The nuisance parameters in weight u_{2i} are for downweighting the outliers, while the parameter $\boldsymbol{\mu}$ is the location vector which does not influence the covariance parameter estimates. Treating the weight parameters as fixed at the value of convergence in the IRLS estimation in (2.4) has been suggested by Huber (1973) and Gross (1977) in the context of robust estimation and by Lee and Jennrich (1979) in the context of covariance structure analysis.

A formal index is needed to evaluate the success of the transformation. Mardia (1970, 1974) developed two statistics for measuring the skewness and kurtosis of a multivariate distribution, which are respectively

$$b_{1,p} = \frac{1}{N^2}\sum_{i,j=1}^{N}\{(\mathbf{X}_i - \bar{\mathbf{X}})'\mathbf{S}^{-1}(\mathbf{X}_j - \bar{\mathbf{X}})\}^3$$

and

$$b_{2,p} = \frac{1}{N}\sum_{i=1}^{N}\{(\mathbf{X}_i - \bar{\mathbf{X}})'\mathbf{S}^{-1}(\mathbf{X}_i - \bar{\mathbf{X}})\}^2.$$

When the sample is from a multivariate normal distribution, the asymptotic distributions of the standard versions of these two statistics were given by Mardia (1970) as

$$M_1 = Nb_{1,p}/6 \sim \chi_f^2, \qquad f = p(p+1)(p+2)/6 \tag{2.6a}$$

and

$$M_2 = \frac{\{b_{2,p} - p(p+2)\}}{\{8p(p+2)/N\}^{1/2}} \sim N(0,1). \tag{2.6b}$$

We will rely on these two statistics to evaluate the multivariate normality of a sample. If our proposed transformation is successful, the data $\mathbf{X}_{bi}$ should be approximately normal.

To implement the transformation procedure, a specific weight function needs to be chosen. A variety of weight functions exist, as listed in Table 11-1 of Hoaglin, Mosteller, and Tukey (1983), and only minor practical differences exist among these different functions (Yuan &

Bentler, 1998b). In Section 4, we use the Huber-type weight function as in (2.3) because of its flexibility in controlling the tails. For example, by choosing the parameter α in the Huber-type weight function, we can control the protection against the proportion of outliers in the sample. As discussed by Huber (1981, p. 3), 1–10% gross errors may often exist in practical data sets. For data sets with only slightly heavy tails, we may choose $\alpha = 0.01$; for a sample with relatively heavy tails, we may choose $\alpha = 0.1$. We may also increase α to an even larger number such as 0.20, if necessary. The α plays the role of a tuning constant similar to the power parameter in the Box–Cox transformation. An advantage here is that we do not need to estimate α and that the results are often quite insensitive to its precise value. The effect of different weighting schemes can be evaluated through Mardia's statistics.

After the transformation, we can proceed with a multivariate procedure based on the transformed sample. We will mainly deal with SEM here. When treating $\mathbf{X}_{bi}$ as a sample, any methods for SEM that could be applied to the $\mathbf{X}_i$ can also be used on the $\mathbf{X}_{bi}$. Because the $\mathbf{X}_{bi}$ approximately follows a multivariate normal distribution, we will only recommend the normal theory MLE. Even when just based on the MLE, there are still many inference procedures (Yuan & Bentler, 1997, 1998c). In the next section, we only outline three of them, which will be adequate for parameter evaluation and model testing in most practical situations.

3. Inference with transformed data

Inferences with robust procedures are generally more complicated than those with classical procedures. As discussed by Huber (1973), Gross (1977) and Carroll (1979), there is no unique best solution to the problems of standard errors and test statistics even for the simplest one-dimensional robust location estimator. Some of the proposed procedures are based on asymptotics, some on IRLS or GLS treating the weights at convergence as fixed, and some others on mixture formulae with components from both the asymptotic expansion and the GLS. After comparing four different standard error estimation methods through extensive empirical studies, Gross (1977) found only slight differences among them. In the literature of mean and covariance structure analysis, standard errors and test statistics are not unique either. For example, for the normal theory MLE, both standard errors based on the information matrix and a sandwich-type covariance matrix give consistent estimates (Yuan & Bentler, 1997). In addition, there are a variety of statistics for evaluating a model structure that may or may not perform equivalently in all situations (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992; Yuan & Bentler, 1998c) with the same estimation method.

In the following, we outline three procedures for inferences. These are respectively the standard normal theory procedure, a GLS approach treating the transformation parameters as being fixed, and a procedure based on asymptotic expansion. We need to introduce some standard notation for this section. For a $p \times p$ symmetric matrix $\mathbf{A}$, let vech ($\mathbf{A}$) be the $p^* = p(p+1)/2$-dimensional vector by stacking the columns of $\mathbf{A}$, omitting the elements above its diagonal, and $\mathbf{D}_p$ be the duplication matrix as defined in Magnus and Neudecker (1988). Denote $\sigma = \text{vech}(\Sigma)$ and

$$\mathbf{W} = 2^{-1}\mathbf{D}_p'(\Sigma^{-1} \otimes \Sigma^{-1})\mathbf{D}_p.$$

We use a dot on top of a function to denote the derivative (e.g., $\dot{\boldsymbol{\sigma}}(\boldsymbol{\theta}) = \partial\boldsymbol{\sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$). We may omit the argument of a function when evaluated at the population value (e.g., $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\theta}_0)$).

The standard normal theory procedure is motivated by the fact that the transformed sample approximately follows a multivariate normal distribution. In this method, parameter estimates $\hat{\boldsymbol{\theta}}$ are obtained by minimizing

$$F(\mathbf{S}_n, \Sigma(\boldsymbol{\theta})) = \text{tr}\{\Sigma^{-1}(\boldsymbol{\theta})\mathbf{S}_n\} - \log|\Sigma^{-1}(\boldsymbol{\theta})\mathbf{S}_n| - p, \tag{3.1}$$

where $\mathbf{S}_n = \hat{\Sigma}$ is a robust covariance matrix. This method also implies using

$$T_{\text{ML}} = nF(\mathbf{S}_n, \Sigma(\hat{\boldsymbol{\theta}})) \sim \chi^2_{p^*-q}$$

for model testing, where q is the number of unknown parameters in $\boldsymbol{\theta}$ and $n = N - 1$. The asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ can be obtained by inverting the corresponding information matrix

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Pi), \tag{3.2}$$

where $\Pi = (\dot{\boldsymbol{\sigma}}'\mathbf{W}\dot{\sigma})^{-1}$. Standard error estimates of $\hat{\boldsymbol{\theta}}$ follow from (3.2) by replacing the unknown parameters in Π by $\hat{\boldsymbol{\theta}}$.

Even though the transformed sample generally follows a multivariate normal distribution more closely than does the original sample, multivariate normality as judged by the Mardia's statistics may be only an approximation. Motivated by the GLS approach studied by Gross (1977), we may treat the weights u_{2i} and $\hat{\boldsymbol{\mu}}$ in (2.5b) as being fixed. Then $\mathbf{X}_{bi}$ is just an ordinary sample with sample covariance matrix $\hat{\Sigma}$. Let $\mathbf{s}_n = \text{vech}(\hat{\Sigma})$. The central limit theorem implies

$$\sqrt{n}(\mathbf{s}_n - \sigma) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Gamma). \tag{3.3}$$

Based on (3.3), we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Omega), \tag{3.4}$$

where $\Omega = \Pi(\dot{\sigma}'\mathbf{W}\Gamma\mathbf{W}\dot{\sigma})\Pi$. Let $\mathbf{Y}_i = \text{vech}[(\mathbf{X}_{bi} - \bar{\mathbf{X}}_b)(\mathbf{X}_{bi} - \bar{\mathbf{X}}_b)']$ and $\mathbf{S}_{\mathbf{Y}}$ be the sample covariance matrix of $\mathbf{Y}_i$, where $\bar{\mathbf{X}}_b$ is the sample mean of $\mathbf{X}_{bi}$. We may use the sample fourth-order moment matrix $\mathbf{S}_{\mathbf{Y}}$ to estimate its population counterpart Γ. Let

$$\mathbf{U} = \mathbf{W} - \mathbf{W}\dot{\sigma}(\dot{\sigma}'\mathbf{W}\dot{\sigma})^{-1}\dot{\boldsymbol{\sigma}}'W$$

and $\hat{c}_1 = \text{tr}(\mathbf{U}\hat{\mathbf{S}}_{\mathbf{Y}})/(p^* - q)$. Satorra and Bentler (1988, 1994) proposed using

$$T_{\text{SB}} = T_{\text{ML}}/\hat{c}_1 \sim \chi^2_{p^*-q} \tag{3.5}$$

for model inference and using

$$\hat{\Omega}^{(1)} = \Pi(\hat{\dot{\sigma}}'\hat{\mathbf{W}}\mathbf{S}_{\mathbf{Y}}\hat{\mathbf{W}}\hat{\dot{\sigma}})\hat{\Pi} \tag{3.6}$$

to estimate standard errors of $\hat{\boldsymbol{\theta}}$. Empirical studies support this procedure for a variety of distributions (Curran *et al.*, 1996; Hu *et al.*, 1992; Yuan & Bentler, 1997, 1998c). Recent results in Yuan and Bentler (1999) imply that (3.5) may still asymptotically follow $\chi^2_{p^*-q}$ even for a highly skewed sample.

Another approach to inference is large sample theory, which was used in Yuan and Bentler (1998a, 1998b). Here we will give a brief outline of this procedure. Notice that the solution to

equation (2.2) satisfies the generalized estimating equation

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{G}(\mathbf{X}_i, \hat{\boldsymbol{\mu}}, \hat{\Sigma}) = \mathbf{0},$$

where

$$\mathbf{G}(\mathbf{x}, \boldsymbol{\mu}, \Sigma) = \begin{pmatrix} u_1\{d(\mathbf{x}, \boldsymbol{\mu}, \Sigma)\}(\mathbf{x} - \boldsymbol{\mu}) \\ u_2\{d^2(\mathbf{x}, \boldsymbol{\mu}, \Sigma)\}\mathrm{vech}\{(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'\} - \sigma \end{pmatrix}.$$

According to the theory of generalized estimating equations (Liang & Zeger, 1986; Yuan & Jennrich, 1998),

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0 \\ \hat{\sigma} - \sigma_0 \end{pmatrix} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{V}),$$

where $\mathbf{V} = \mathbf{H}^{-1}\mathbf{B}\mathbf{H}'^{-1}$ with

$$\mathbf{H} = E\{\dot{\mathbf{G}}(\mathbf{X}, \boldsymbol{\mu}_0, \Sigma_0)\}, \qquad \mathbf{B} = E\{\mathbf{G}(\mathbf{X}, \boldsymbol{\mu}_0, \Sigma_0)\mathbf{G}'(\mathbf{X}, \boldsymbol{\mu}_0, \Sigma_0)\}.$$

A consistent estimate $\hat{\mathbf{V}}$ can be obtained by using consistent estimates for $\mathbf{H}$ and $\mathbf{B}$; these are given by

$$\hat{\mathbf{H}} = \frac{1}{N}\sum_{i=1}^{N}\dot{\mathbf{G}}(\mathbf{X}_i, \hat{\boldsymbol{\mu}}, \hat{\Sigma}), \qquad \hat{\mathbf{B}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{G}(\mathbf{X}_i, \hat{\boldsymbol{\mu}}, \hat{\Sigma})\mathbf{G}'(\mathbf{X}_i, \hat{\boldsymbol{\mu}}, \hat{\Sigma}).$$

Let $\hat{\mathbf{V}}_{22}$ be the submatrix of $\hat{\mathbf{V}}$ corresponding to σ. Then $\hat{\mathbf{V}}_{22}$ is another estimate of Γ. Let $\hat{c}_2 = \mathrm{tr}(\hat{\mathbf{U}}\hat{\mathbf{V}}_{22})/(p^* - q)$. Yuan and Bentler (1998b) suggested the rescaled statistic

$$T_{\mathrm{R}} = T_{\mathrm{ML}}/\hat{c}_2 \tag{3.7}$$

for model testing and

$$\hat{\Omega}^{(2)} = \hat{\Pi}(\hat{\dot{\sigma}}'\hat{\mathbf{W}}\hat{\mathbf{V}}_{22}\hat{\mathbf{W}}\hat{\dot{\sigma}})\hat{\Pi} \tag{3.8}$$

to estimate Ω in (3.4).

Inference based on (3.7) and (3.8) is essentially the procedure developed in Yuan and Bentler (1998b). Notice that the above three procedures depend on the same parameter estimate $\hat{\boldsymbol{\theta}}$—differences are only in standard errors and test statistics. If a transformed sample closely follows a mutlivariate normal distribution, these procedures will be asymptotically equivalent, and numerical differences among them are only due to finite sample behaviour. If a transformed sample is still not normally distributed, then the normal theory based method may not perform satisfactorily. In such a case, the methods based on the two different estimates of the Γ matrix may perform better. As remarked earlier, these three methods are not the only methods that can be applied to a transformed sample. For a transformed sample that is still far away from normally distributed, the various versions of ADF methods as in Browne (1984) and Yuan and Bentler (1998c), or the minimum chi-square method as in Yuan and Bentler (1998a), may be more appropriate.

4. Applications

We use the Huber-type weight function with $\alpha = 0.05$ for each of the samples and evaluate

the normality of the transformed sample by Mardia's statistics. A larger α is used if the transformed sample is still significantly different from normality. Notice that the sample size N is used in the denominator on the right-hand side of (2.2b) and (2.5). While this is appropriate, in order to adhere to the common practice of using unbiased estimators, in this section the denominator $n = N - 1$ will be used when calculating the sample covariance matrices of the transformed samples. Two types of standard error estimate are reported when the normal theory method is applied to a raw sample. One is based on the information matrix ($\mathrm{Std}_{\mathrm{I}}$) as in (3.2), the other is based on the sandwich-type covariance matrix ($\mathrm{Std}^{(1)}_{\mathrm{SW}}$) as in (3.6). When the normal theory method is applied to a transformed sample, standard errors ($\mathrm{Std}^{(2)}_{\mathrm{SW}}$) based on (3.8) will also be reported in addition to $\mathrm{Std}_{\mathrm{I}}$ and $\mathrm{Std}^{(1)}_{\mathrm{SW}}$. Corresponding to the three types of standard errors, we have test statistics T_{ML}, T_{SB} and T_{R}, which are also reported below the standard errors.

Our first example is based on the data set in Table 2.5 of Bollen (1989, pp. 30–31). This data set consists of three estimates of percentage cloud cover for 60 slides. It was introduced for outlier identification purposes and was further used by Bollen and Arminger (1991) to study observational residuals in factor analysis. An unusual feature of this data set is that using the sample covariance matrix to fit a one-factor model leads to an improper solution with a negative error variance. Mardia's statistics for the original sample are respectively $M_1 = 0.06$ and $M_2 = 5.09$. This indicates that the data may come from a distribution with tails heavier than those of a normal distribution when referring the statistics to distributions χ^2_{10} and $N(0, 1)$, respectively. Mardia's statistics for the transformed sample are $M_1 = 0.05$ and $M_2 = 0.23$, so we may regard the transformed data as from an approximately normal distribution. Estimates based on the transformed data will generally be more efficient. Applying the ML method to the transformed data also leads to a set of reasonable solutions. Bollen and Arminger (1991) suggested removing the three most influential cases, $\mathbf{X}_{52}$, $\mathbf{X}_{40}$ and $\mathbf{X}_{51}$, which also leads to a set of reasonable solutions.

Parameter estimates and their standard errors are given in Table 1, where we refer to the data set with the three outliers removed as the OR data. Factor loading estimates based on the three samples are comparable. However, there exists a large difference for error variance estimates. Carefully examining the parameter estimates, we may notice that all the parameter estimates based on H (0.05) data lie between corresponding parameters based on the raw data and the OR data. This reflects the degree of influence of these three cases. In the raw data the three cases are given weights 1.0; their weights are 0 in the OR data, and somewhere between 0 and 1 in the H (0.05) data. Standard error estimates for factor loadings are also more comparable than those for error variances. There is a large discrepancy between $\mathrm{Std}_{\mathrm{I}}$ and $\mathrm{Std}^{(1)}_{\mathrm{SW}}$ within the raw sample, especially for ψ_i. These two types of standard error are much more comparable within the OR data. The $\mathrm{Std}_{\mathrm{I}}$ and $\mathrm{Std}^{(1)}_{\mathrm{SW}}$ within the H(0.05) data are also quite comparable, but $\mathrm{Std}^{(2)}_{\mathrm{SW}}$ gives relatively large standard errors for ψ_i, especially ψ_2. Even though $\hat{\psi}_3$ is negative based on the raw sample, when evaluated by either of the standard errors it is not significant. However, $\hat{\psi}_3$ is significant at level 0.05 with the OR data. Thus, although there are differences among the three types of standard errors with H(0.05) data, evaluation for significance of parameters is about the same for any of them. For example, $\hat{\psi}_3$ is not significant, when evaluated by any of the standard errors in H(0.05). Because there are zero degrees of freedom, no test statistic is available.

Our second example is based on the data of Holzinger and Swineford (1939). This classic data set consists of mental ability tests scores of seventh- and eighth-grade children from two

Table 1. Parameter estimates and standard errors for the cloud cover data from Bollen (1989)

	Raw data			OR data			H(0.05) data			
$\boldsymbol{\theta}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\text{Std}_{\text{SW}}^{(2)}$
λ_1	32.43	3.68	2.88	31.99	3.34	2.96	32.27	3.33	2.81	2.90
λ_2	31.46	4.10	3.59	36.57	3.86	2.00	34.98	3.78	2.42	2.80
λ_3	38.15	3.46	2.35	35.91	3.56	2.42	36.34	3.41	2.23	2.32
ψ_1	248.77	61.21	113.32	105.80	28.52	42.28	135.67	34.61	47.30	63.58
ψ_2	473.78	95.23	163.70	157.29	39.88	42.74	237.60	51.84	61.35	117.70
ψ_3	−51.44	56.97[a]	80.37[a]	58.15	27.89	27.32	28.52	31.08[a]	31.67[a]	41.05[a]

[a] z-scores not significant at level $\alpha = 0.05$.

different schools. There are 24 variables and 145 subjects from the Grant-White school. Jöreskog (1969) used 9 of the 24 variables in studying the correlation structure with normal theory ML: (1) visual perception, (2) cubes, (3) lozenges, (4) paragraph comprehension, (5) sentence completion, (6) word meaning, (7) addition, (8) counting dots, (9) straight-curved capitals. We will also use these same variables in our application. Mardia's statistics for these nine variables are $M_1 = 0.77$ and $M_2 = 3.04$, which indicates that the data may come from a distribution with slightly heavier tails than those of a normal distribution when referring these statistics to χ^2_{165} and $N(0,1)$, respectively. On the other hand, Mardia's statistics for the transformed data with $\alpha = 0.05$ are $M_1 = 0.91$ and $M_2 = -0.53$, both far from significant. So it is more appropriate to apply the normal theory method to the transformed sample.

In the original report of Holzinger and Swineford (1939), variables 1, 2 and 3 were designed to measure spatial ability, variables 4, 5 and 6 were designed to measure verbal ability, and variables 7, 8 and 9 were designed to measure a speed factor in performing the tasks. Let $\mathbf{X}$ represent the nine observed variables. Then the confirmatory factor model

$$\mathbf{X} = \Lambda\mathbf{f} + \mathbf{e} \qquad \text{cov}(\mathbf{X}) = \Lambda\Phi\Lambda' + \Phi, \tag{4.1}$$

with

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{42} & \lambda_{52} & \lambda_{62} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{73} & \lambda_{83} & \lambda_{93} \end{pmatrix}', \quad \Phi = \begin{pmatrix} 1.0 & \phi_{12} & \phi_{13} \\ \phi_{21} & 1.0 & \phi_{23} \\ \phi_{31} & \phi_{32} & 1.0 \end{pmatrix}, \tag{4.2}$$

represents the hypothesis of the original design. We assume the measurement errors are uncorrelated, with $\Phi = \text{cov}(\mathbf{e})$ being a diagonal matrix. There are $q = 21$ unknown parameters, 9 of which are factor loadings. The model has 24 degrees of freedom.

Parameter estimates, their standard errors as well as the associated test statistics for both samples are given in Table 2. Estimates of factor correlations and error variances based on the two samples are quite similar, so, to save space, we do not report these. The parameter estimates based on the transformed sample are about the same as those obtained from the original sample; so are the various standard error estimates. There exist minor differences among the test statistics for this example. Even the smallest $T_{\text{R}} = 46.47$ is still highly significant for test statistics based on H (0.05) data. All statistics yield the same conclusion regarding model fit.

Table 2. Parameter estimates, standard errors and test statistics (df = 24) for the psychological data from Holzinger and Swineford (1939)

	Raw data			H(0.05) data			
$\boldsymbol{\theta}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\text{Std}_{\text{SW}}^{(2)}$
λ_{11}	4.68	0.62	0.70	4.72	0.62	0.65	0.72
λ_{21}	2.30	0.41	0.38	2.22	0.40	0.37	0.38
λ_{31}	5.77	0.75	0.73	5.55	0.74	0.70	0.73
λ_{42}	2.92	0.24	0.25	2.84	0.23	0.24	0.26
λ_{52}	3.86	0.33	0.33	3.87	0.33	0.33	0.34
λ_{62}	6.57	0.57	0.57	6.36	0.55	0.53	0.57
λ_{73}	15.68	2.00	1.84	15.65	2.00	1.85	1.88
λ_{83}	16.71	1.75	1.78	16.09	1.66	1.58	1.65
λ_{93}	25.96	3.11	3.09	24.83	3.03	2.78	2.91
T		51.19[a]	49.37[a]		49.40[a]	51.02[a]	46.47[a]

[a] Significant at level $\alpha = 0.01$.

Since model (4.2) does not fit the sample, Jöreskog (1969) proposed a variety of alternative models. One of these models is

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} & 0 & 0 & 0 & 0 & \lambda_{81} & \lambda_{91} \\ 0 & 0 & 0 & \lambda_{42} & \lambda_{52} & \lambda_{62} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{73} & \lambda_{83} & \lambda_{93} \end{pmatrix}', \qquad \Phi = \begin{pmatrix} 1.0 & \phi_{12} & 0 \\ \phi_{21} & 1.0 & \phi_{23} \\ 0 & \phi_{32} & 1.0 \end{pmatrix}. \tag{4.3}$$

Both the original sample and the transformed one are used for this model. As can be seen from Table 3, there is little difference for parameter and standard error estimates among the various methods. All the model test statistics are far from being significant. For example, the p-value for $T_{\text{ML}} = 25.75$ with the original sample is about 0.31. Because the transformation reduces the kurtosis, the statistics based on the transformed data give more support for model (4.3). However, the fit statistics are only minimally smaller.

This data set has only slightly heavier tails as compared to normal data. With $\alpha = 0.05$ in the transformation, the kurtosis changed from 3.04 for the original sample to -0.53 for the transformed sample. This implies that a smaller α will be good enough to make M_2 non-significant. Because of the near-normality of the original sample, there exist only minor differences for parameter estimates and test statistics before and after the transformation.

Our third data set is from the EQS manual (Bentler, 1995, p. 117). This 6-variable and 50-case artificial data set was introduced to demonstrate the effect of outliers on the classical model fitting procedure. The first 49 cases were generated from a multivariate normal distribution and the 50th case is an outlier. Mardia's statistics for this data set are $M_1 = 206.28$ and $M_2 = 6.82$, and when referred to χ^2_{56} and $N(0,1)$ both are highly significant. Without prior knowledge about this data set, we might have regarded it as representing a skewed distribution with long tails. Mardia's statistics for the transformed sample are $M_1 = 7.25$ and $M_2 = -1.04$. Neither is significant when referred to χ^2_{56} and $N(0,1)$, respectively.

As in the EQS manual, we fit both samples using a two-factor model with three indicators

Table 3. Parameter estimates, standard errors and test statistics (df = 23) for the psychological data from Holzinger and Swineford (1939)

	Raw data			H(0.05) data			
$\boldsymbol{\theta}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\text{Std}_{\text{SW}}^{(2)}$
λ_{11}	5.03	0.59	0.63	5.08	0.58	0.59	0.64
λ_{21}	2.17	0.40	0.37	2.09	0.39	0.37	0.38
λ_{31}	5.53	0.72	0.68	5.33	0.71	0.64	0.68
λ_{81}	6.17	1.63	1.61	5.85	1.54	1.43	1.50
λ_{91}	20.82	2.93	3.16	19.99	2.87	2.84	3.03
λ_{42}	2.90	0.24	0.25	2.82	0.23	0.24	0.26
λ_{52}	3.85	0.33	0.33	3.86	0.33	0.32	0.33
λ_{62}	6.51	0.57	0.57	6.31	0.55	0.53	0.57
λ_{73}	19.95	2.45	2.53	19.73	2.39	2.32	2.41
λ_{83}	13.98	1.92	1.84	13.49	1.80	1.70	1.74
λ_{93}	16.62	2.89	2.88	16.50	2.82	2.74	2.83
T		25.75	24.73		22.86	23.80	21.54

for each factor. To save space, we only report the factor loading estimates and their standard errors in Table 4 together with the test statistics. There is a relatively large difference between parameter estimates based on the original sample and those based on the transformed sample. Actually, because of the influence of the outliers, parameter estimates for λ_{31} and λ_{62} based on the original sample are not even significant as evaluated by either Std_{I} or $\text{Std}_{\text{SW}}^{(1)}$. On the other side of the table are the results based on the transformed sample. All the λ_{ij} are significant when evaluated by any of the three types of standard errors. Parameter estimates and their standard errors based on the transformed sample are very comparable to those based on the OR data in which the 50th case is removed. There is also a large difference between the test statistics based on the raw sample and the other two samples. However, there is little difference among the test statistics within either the OR data or the H (0.05) data. With $T_{\text{ML}} = 16.69$ referring to χ_8^2, the corresponding p-value is about 0.03. For this small sample

Table 4. Parameter estimates, standard errors and test statistics (df = 8) for the artificial data from the EQS manual

	Raw data			OR data			H(0.05) data			
$\boldsymbol{\theta}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\text{Std}_{\text{SW}}^{(2)}$
λ_{11}	1.56	0.35	0.64	0.69	0.17	0.20	0.72	0.18	0.20	0.20
λ_{21}	0.49	0.17	0.14	0.51	0.15	0.13	0.54	0.15	0.14	0.14
λ_{31}	0.19	0.18[a]	0.37[a]	1.02	0.19	0.17	0.94	0.19	0.18	0.23
λ_{42}	0.67	0.16	0.21	0.43	0.15	0.13	0.44	0.15	0.13	0.14
λ_{52}	0.88	0.16	0.16	0.81	0.18	0.18	0.82	0.18	0.18	0.20
λ_{62}	0.18	0.14[a]	0.24[a]	0.46	0.15	0.13	0.44	0.14	0.13	0.14
T		16.69[b]	13.93		2.70	3.16		2.77	3.20	2.24

[a] z-scores not significant at level $\alpha = 0.05$.
[b] Significant at level $\alpha = 0.05$.

size we should get a larger p-value, as discussed in the EQS manual. On the other hand, all the statistics based on the transformed sample definitely imply that the two-factor model is a correct model.

Even though both M_1 and M_2 are highly significant in the original sample, they are caused by only one outlier. The transformation with $\alpha = 0.05$ greatly reduces the effect of this outlier. The negative M_2 for the transformed data may imply that a smaller α value may work equally well in transforming the original sample to approximate normality. This example also demonstrates that the robust transformation makes the data more symmetric as well.

Our fourth example is based on the alcohol and psychological symptom data set from Neumann (1994). This data set consists of 10 variables and 335 cases. The two variables in $\mathbf{X} = (x_1, x_2)$ are respectively family history of psychopathology and family history of alcoholism, which are indicators for a latent construct of family history. The eight variables in $\mathbf{Y} = (y_1, \ldots, y_8)$ are respectively the age of first problem with alcohol, age of first detoxification from alcohol, alcohol severity score, alcohol use inventory, SCL-90 psychological inventory, the sum of the Minnesota Multiphasic Personality Inventory scores, the lowest level of psychosocial functioning during the past year, and the highest level of psychosocial functioning during the past year. With two indicators for each latent construct, these eight variables are respectively measuring: age of onset, alcohol symptoms, psychopathology symptoms, and global functioning. Neumann's (1994) theoretical model for this data set is

$$\mathbf{X} = \Lambda_{\mathbf{X}}\boldsymbol{\xi} + \boldsymbol{\delta}, \qquad \mathbf{Y} = \Lambda_{\mathbf{Y}}\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{4.4a}$$

and

$$\eta = \mathbf{B}\eta + \Gamma\xi + \zeta, \tag{4.4b}$$

where

$$\Lambda_{\mathbf{X}} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}, \qquad \Lambda_{\mathbf{Y}} = \begin{pmatrix} 1 & \lambda_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \lambda_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_6 \end{pmatrix}', \tag{4.4c}$$

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta_{21} & 0 & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 & 0 \\ 0 & \beta_{42} & \beta_{43} & 0 \end{pmatrix}, \qquad \Gamma = \begin{pmatrix} \gamma_{11} \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{4.4d}$$

and $\boldsymbol{\epsilon}$, $\boldsymbol{\delta}$ and ζ are vectors of errors whose elements are all uncorrelated.

The parameter estimates and test statistics based on the raw sample are given in Table 5. With model degrees of freedom being 29, the p-values for $T_{\mathrm{ML}} = 48.96$ and $T_{\mathrm{SB}} = 47.34$ are respectively about 0.01 and 0.02, indicating that model (4.4) is not statistically acceptable. Mardia's statistics for the raw sample are $M_1 = 76.06$ and $M_2 = 14.76$; when referred to χ^2_{220} and $N(0, 1)$ respectively, M_2 is highly significant, indicating that the data set comes from a distribution with heavy tails. In contrast, Mardia's statistics for the H(0.05) transformed sample are $M_1 = 3.26$ and $M_2 = 2.25$. Parameter estimates for this transformed sample, shown in Table 5, are about the same as those based on the original sample. However, the two

Table 5. Parameter estimates, standard errors and test statistics (df = 29) for the alcohol data from Neumann (1994)

	Raw data			H(0.05) data				H(0.10) data			
$\boldsymbol{\theta}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\text{Std}_{\text{SW}}^{(2)}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\text{Std}_{\text{SW}}^{(2)}$
λ_1	0.77	0.13	0.15	0.70	0.12	0.12	0.13	0.68	0.12	0.11	0.12
λ_2	1.69	0.26	0.26	1.43	0.21	0.19	0.21	1.40	0.21	0.18	0.20
λ_3	0.99	0.10	0.10	1.02	0.10	0.10	0.10	1.03	0.10	0.10	0.10
λ_4	1.27	0.11	0.10	1.26	0.10	0.10	0.10	1.25	0.10	0.09	0.10
λ_5	0.15	0.02	0.02	0.15	0.02	0.02	0.02	0.15	0.02	0.02	0.02
λ_6	0.93	0.18	0.19	0.94	0.19	0.19	0.20	0.95	0.19	0.19	0.20
β_{21}	−0.65	0.09	0.09	−0.67	0.09	0.10	0.10	−0.68	0.09	0.09	0.10
β_{31}	0.51	0.40[a]	0.42[a]	0.49	0.40[a]	0.40[a]	0.41[a]	0.48	0.41[a]	0.40[a]	0.43[a]
β_{32}	2.42	0.34	0.36	2.46	0.34	0.35	0.36	2.46	0.34	0.35	0.37
β_{42}	−0.08	0.04	0.04	−0.08	0.03	0.03	0.04	−0.08	0.03	0.03	0.03
β_{43}	−0.04	0.01	0.01	−0.04	0.01	0.01	0.01	−0.04	0.01	0.01	0.01
γ_{11}	−3.39	0.72	0.56	−3.46	0.71	0.57	0.58	−3.42	0.71	0.56	0.59
T		48.96[b]	47.34[b]		42.42	45.66[b]	41.17		40.98	45.31[b]	40.15

[a] z-scores not significant at level $\alpha = 0.05$.
[b] Significant at level $\alpha = 0.05$.

test statistics decrease to $T_{\text{ML}} = 42.42$ and $T_{\text{SB}} = 45.66$ with p-values 0.05 and 0.03. In particular, $T_{\text{R}} = 41.17$ corresponds to a p-value of 0.07, indicating that model (4.4) is marginally acceptable.

Since $M_2 = 2.25$ is still significant for the H(0.05) sample, we further use the transformation H(0.10). This leads to $M_1 = 2.30$, $M_2 = -0.11$, and $T_{\text{ML}} = 40.98$, $T_{\text{SB}} = 45.31$, $T_{\text{R}} = 40.15$ corresponding to p-values 0.07, 0.03 and 0.08. Thus, fitting the model (4.4) to the sample using H(0.10) leads more support to the theoretically justified model (4.4). Even though there exist noticeable differences among the three test statistics within a transformed sample, the differences are not substantial, and all of the statistics imply that (4.4) is a reasonable model.

Standard errors within either of the samples are very comparable. There are very few differences in parameter estimates and standard errors among the different samples. It will be seen that parameter β_{31} is not statistically significant, when evaluated by any of the standard errors within any of the samples.

Our fifth data set is from a longitudinal study of adolescent development conducted by Newcomb and Bentler (1988) and Stein, Newcomb, and Bentler (1996). This data set consists of 12 variables and 350 cases. The four variables in $\mathbf{X} = (x_1, x_2, x_3, x_4)'$ are respectively relation with parents, self-acceptance, depression (negatively scored), and relation with family measured during years 4–5 of the study. They are the indicators of a relationship construct. There are eight subsequent dependent indicators in $\mathbf{Y} = (y_1, \ldots, y_8)$. Variables y_1 to y_4 are measures of licit drug use at years 6–9 of the study, and variables y_5 to y_8 are measures of deviancy also at years 6–9 of the study. The substantive theory suggests the model

$$\mathbf{X} = \Lambda_{\mathbf{X}}\xi + \delta, \qquad \mathbf{Y} = \Lambda_{\mathbf{Y}}\eta + \varepsilon \tag{4.5a}$$

and

$$\eta = \Gamma\xi + \zeta, \tag{4.5b}$$

where

$$\Lambda_{\mathbf{X}} = \begin{pmatrix} 1 & \lambda_1 & \lambda_2 & \lambda_3 \end{pmatrix}', \qquad \Lambda_{\mathbf{Y}} = \begin{pmatrix} 1 & \lambda_4 & \lambda_5 & \lambda_6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda_7 & \lambda_8 & \lambda_9 \end{pmatrix}', \tag{4.5c}$$

$$\Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{21} \end{pmatrix}', \qquad \mathrm{cov}(\zeta) = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix}, \tag{4.5d}$$

and where ε and δ are vectors of errors whose elements are all uncorrelated.

Mardia's statistics for the original sample are $M_1 = 8.63$ and $M_2 = 56.55$, which indicates that the sample comes from a distribution with quite heavy tails when referred to χ^2_{364} and $N(0, 1)$, respectively. The transformation with $\alpha = 0.05$ yields $M_1 = 4.70$ and $M_2 = 11.88$, a substantial improvement towards normality. The parameter estimates, their standard errors and test statistics are given in Table 6. Parameter estimates and the various standard errors within and between the two samples are quite similar. However, there exist some large differences between the test statistics. In particular, the difference between T_{R} for the transformed sample and T_{ML} for the original sample is about 61.7. Unfortunately, T_{R}, is still highly significant when referred to χ^2_{51}.

Since the sample with H(0.05) still has a quite significant kurtosis ($M_2 = 11.88$), it is possible that the significant statistics may arise from the heavy tails of the sample. We further use $\alpha = 0.20$ in the Huber-type weight function. This results in $M_1 = 2.36$ and $M_2 = 1.48$; neither is significant when referred to χ^2_{364} and $N(0, 1)$, respectively. However, as can be seen from the right-hand side of Table 6, all the test statistics are still highly significant. This may

Table 6. Parameter estimates, standard errors and test statistics (df = 51) for the longitudinal data with 12 variables

	Raw data			H(0.05) data				H(0.20) data			
$\boldsymbol{\theta}$	$\hat{\boldsymbol{\theta}}$	$\mathrm{Std}_{\mathrm{I}}$	$\mathrm{Std}^{(1)}_{\mathrm{SW}}$	$\hat{\boldsymbol{\theta}}$	$\mathrm{Std}_{\mathrm{I}}$	$\mathrm{Std}^{(1)}_{\mathrm{SW}}$	$\mathrm{Std}^{(2)}_{\mathrm{SW}}$	$\hat{\boldsymbol{\theta}}$	$\mathrm{Std}_{\mathrm{I}}$	$\mathrm{Std}^{(1)}_{\mathrm{SW}}$	$\mathrm{Std}^{(2)}_{\mathrm{SW}}$
λ_1	1.15	0.15	0.25	1.09	0.16	0.20	0.26	1.05	0.15	0.17	0.23
λ_2	0.89	0.12	0.23	0.74	0.12	0.16	0.22	0.65	0.11	0.13	0.19
λ_3	0.70	0.09	0.17	0.64	0.10	0.11	0.14	0.58	0.09	0.09	0.12
λ_4	1.09	0.11	0.14	1.18	0.11	0.13	0.15	1.20	0.11	0.12	0.14
λ_5	−1.29	0.12	0.14	−1.35	0.13	0.15	0.16	−1.37	0.12	0.14	0.16
λ_6	1.20	0.13	0.09	1.20	0.14	0.10	0.10	1.19	0.14	0.09	0.10
λ_7	1.06	0.06	0.06	1.04	0.05	0.05	0.05	1.04	0.05	0.04	0.05
λ_8	1.05	0.06	0.06	1.05	0.05	0.05	0.05	1.03	0.05	0.04	0.05
λ_9	1.10	0.06	0.07	1.10	0.06	0.05	0.06	1.08	0.05	0.05	0.05
γ_{11}	−0.29	0.07	0.07	−0.27	0.07	0.07	0.08	−0.24	0.07	0.08	0.08
γ_{21}	−0.03	0.04[a]	0.03[a]	−0.02	0.03[a]	0.02[a]	0.02[a]	−0.02	0.03[a]	0.02[a]	0.02[a]
T		247.89[b]	184.01[b]		212.40[b]	233.85[b]	186.20[b]		203.90[b]	253.52[b]	196.62[b]

[a] z-scores not significant at level $\alpha = 0.05$.
[b] Significant at level $\alpha = 0.001$.

indicate that the problem with the bad fit is not due to poorly distributed data, but rather because model (4.5) does not accurately reflect the structural relationship in the population. As with the other two samples, standard errors by the various methods are quite comparable. In particular, γ_{21} is not statistically significant, judged by any of the standard errors within any of the samples.

Our final example is based on the industrialization and political democracy panel data introduced by Bollen (1989), who proposed various models for this data set. Bollen and Arminger (1991) also used this data set to study observational residuals in structural equation models. This data set consists of eight political democracy variables $\mathbf{Y} = (y_1, \ldots, y_8)'$ and three industrialization variables $\mathbf{X} = (x_1, x_2, x_3)'$ in 75 developing countries during 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 indicators measured in 1965. Assuming that political democracy in 1965 is predicted by 1960 political democracy, and both are further predicted by 1960 industrialization, Bollen (1989) proposed the model

$$\mathbf{X} = \Lambda_{\mathbf{X}}\xi + \delta, \qquad \mathbf{Y} = \Lambda_{\mathbf{Y}}\eta + \varepsilon \tag{4.6a}$$

and

$$\eta = \mathbf{B}\eta + \Gamma\xi + \zeta, \tag{4.6b}$$

where

$$\Lambda_{\mathbf{X}} = \begin{pmatrix} 1 & \lambda_1 & \lambda_2 \end{pmatrix}', \qquad \Lambda_{\mathbf{Y}} = \begin{pmatrix} 1 & \lambda_3 & \lambda_4 & \lambda_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda_3 & \lambda_4 & \lambda_5 \end{pmatrix}', \tag{4.6c}$$

$$\mathbf{B} = \begin{pmatrix} 0 & 0 \\ \beta_{21} & 0 \end{pmatrix}, \qquad \Gamma = \begin{pmatrix} \gamma_{11} \\ \gamma_{21} \end{pmatrix} \tag{4.6d}$$

and again ε, δ and ζ are vectors of errors. Based on theoretical considerations, some error terms in ε are allowed to be correlated; see Bollen (1989) for details about the data and the model.

Table 7. Parameter estimates, standard errors and test statistics (df = 38) Industrialization and Political Democracy Data from Bollen (1989)

	Raw data			H(0.05) data			
$\boldsymbol{\theta}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\hat{\boldsymbol{\theta}}$	Std_{I}	$\text{Std}_{\text{SW}}^{(1)}$	$\text{Std}_{\text{SW}}^{(2)}$
λ_1	2.18	0.14	0.14	2.17	0.14	0.14	0.15
λ_2	1.82	0.15	0.14	1.81	0.15	0.14	0.14
λ_3	1.19	0.14	0.13	1.19	0.14	0.13	0.13
λ_4	1.17	0.12	0.12	1.17	0.12	0.11	0.13
λ_5	1.25	0.12	0.12	1.24	0.12	0.12	0.13
β_{21}	0.87	0.06	0.05	0.86	0.06	0.05	0.05
γ_{11}	1.47	0.39	0.35	1.48	0.39	0.35	0.35
γ_{21}	0.60	0.22	0.20	0.61	0.22	0.21	0.21
T		39.64	43.07		39.82	43.61	40.46

Mardia's statistics for this data set are $M_1 = 8.23$ and $M_2 = -1.22$. Neither is significant when referred to χ^2_{286} and $N(0,1)$, respectively. Bollen and Arminger did not find any statistically significant residual with model (4.6). Thus, we would expect the normal theory method with the raw data to give efficient estimates and reliable model evaluation. Actually, both T_{ML} and T_{SB} with the original sample support this model. It would be interesting to see the effect of our transformation even if such a transformation is unnecessary. Mardia's statistics for the transformed data are $M_1 = 8.67$ and $M_2 = -1.42$, which are almost the same as for the original sample. As can be seen in Table 7, parameter and standard error estimates based on the transformed data are also almost identical to those based on the original sample. The differences among the three test statistics within the transformed data are also comparable to those between T_{ML} and T_{SB} within the original sample. This example demonstrates that transformation H(0.05) has almost no effect on a data set that is approximately normally distributed.

5. Discussion

We propose using robust estimation as a data transformation procedure. The sample covariance matrix of the transformed sample is just the robust covariance matrix. This covariance matrix is generally more efficient than the sample covariance matrix when the sampling distribution has heavier tails than those of a normal distribution. Since the transformed data more closely follow a multivariate normal distribution, the classical normal theory based procedure is more appropriate for the transformed data than for the original data. Parameter estimates are more efficient and inference based on T_{ML} is also more accurate in evaluating a model structure. Since the transformation is straightforward to implement (an SAS IML program is available from the authors), there is no difficulty in implementing the standard normal theory based procedure with any SEM software. The Satorra–Bentler procedure is implemented in EQS, and there is also no difficulty in implementing this procedure with the EQS software. Implementation of the procedure-based (3.7) and (3.8) may require the user to master a programing language before this procedure is implemented into commercial software.

Our examples show that for skewed samples with heavy tails, the transformed samples will have less skewness and kurtosis. If the skewness and kurtosis are created by a few outliers, the effect of these outliers basically disappears after the transformation. We have mainly used $\alpha = 0.05$, the proportion of outliers to control, but a larger α may be necessary for data showing heavy-tailed distributions, as in the fourth and fifth data sets in our applications. Even though a smaller α may be enough for data sets with slightly heavier tails than the normal distribution, our examples show that there is not really much difference between the transformed sample with $\alpha = 0.05$ and the original sample if the original data set is approximately normal. Considering the fact that data sets in social and behavioural sciences are often characterized by skewness and extra kurtosis, it makes sense to recommend the use of $\alpha = 0.05$ in the Huber-type weight function as a routine transformation for data sets to be used in SEM applications. Of course, if the transformed sample is still significantly non-normal, it is necessary to increase the value of α. Importantly, at any stage we can use methods such as Mardia's statistics to evaluate the effect of the transformation in achieving near-normality.

In many instances, SEM methodology is but one approach to data analysis with a given

data set. An advantage of the current transformation approach over the previously proposed robust methodology of Yuan and Bentler (1998a, 1998b) is that a new data matrix is created that can be used for a variety of other analysis methods. For example, clustering of cases can be accomplished with the transformed data, and one would expect that the resulting clusters are less likely to be influenced by outlying data points than would be obtained by clustering the original data. Similarly, our transformed data can be made available to many other multivariate procedures such as principal components, factor analysis, and discriminant analysis.

In order to achieve a more thorough analysis, we recommend the user to try all three inference procedures, as outlined in Section 3, before drawing conclusions about a model. If the transformation is successful in making the resulting sample normally distributed, the standard normal theory procedure will be enough for model inference. In such a case, actually, all the three inference procedures should give similar conclusions about the model. Of course it is entirely possible with some data sets that even the transformed samples remain non-normally distributed. In that case, it would be more appropriate to use the two alternative procedures in Section 3. Of course, any other procedures applied to the transformed data should also give better model evaluations than those based on the raw sample. Our preference for these three procedures is based on our limited experience with a variety of data sets. Further research and more applications will be needed to provide a thorough evaluation of the proposed methodology. Among the various SEM procedures, it would be especially valuable to identify one that can give most reliable model inference based on robust samples. Since there are numerous types of violations to multivariate normality, definitive conclusions may require extensive Monte Carlo studies. Finally, resampling procedures for model testing and standard errors may also be used on transformed samples. This topic also deserves further research.

Acknowledgements

This project was supported by a Direct Grant for Research from the Chinese University of Hong Kong, and in part by grants DA01070 and DA00017 from the National Institute on Drug Abuse at the US National Institutes of Health. This research was facilitated while the first author was visiting the Psychology Department at the Chinese University of Hong Kong during summer 1998. We appreciate the constructive comments by the editor and the two referees, which led to an improved version of the paper.

References

Ammann, L. P. (1989). Robust principal components. *Communications in Statistics: Simulation and Computation, 18,* 857–874.

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (2nd ed.). New York: Wiley.

Austin, J. T., & Calderón, R. F. (1996). Theoretical and technical contributions to structural equation modeling: An updated annotated bibliography. *Structural Equation Modeling, 3,* 105–175.

Austin, J. T., & Wolfle, D. (1991). Annotated bibliography of structural equation modelling: Technical work. *British Journal of Mathematical and Statistical Psychology, 44,* 93–152.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, directions. *Annual Review of Psychology, 47,* 563–592.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollen, K. A., & Arminger, G. (1991). Observational residuals in factor analysis and structural equation models. *Sociological Methodology, 21,* 235–262.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge: Cambridge University Press.

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

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

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

Carroll, R. J. (1979). On estimating variances of robust estimators when the errors are asymmetric. *Journal of the American Statistical Association, 74,* 674–679.

Curran, P. S., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. *Psychological Methods, 1,* 16–29.

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

Fang, K.-T., Kotz, S., & Ng, K. W. (1990). *Symmetric multivariate and related distributions.* London: Chapman & Hall.

Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives (with discussion). *Journal of the Royal Statistical Society, Series B, 46,* 149–192.

Gross, A. M. (1977). Confidence intervals for bisquare regression estimates. *Journal of the American Statistical Association, 72,* 341–354.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions.* New York: Wiley.

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). *Understanding robust and exploratory data analysis.* New York: Wiley.

Holland, P. W., & Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Communications in Statistics: Theory and Methods, A6,* 813–827.

Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution.* University of Chicago: Supplementary Educational Monographs, No. 48.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112,* 351–362.

Huba, G. J., & Harlow, L. L. (1987). Robust structural equation models: Implications for developmental psychology. *Child Development, 58,* 147–166.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Annals of Statistics, 1,* 799–821.

Huber, P. J. (1977). Robust covariances. In S. S. Gupta & D. S. Moore (Eds.), *Statistical decision theory and related topics* (Vol. 2, pp. 165–191). New York: Academic Press.

Huber, P. J. (1981). *Robust statistics.* New York: Wiley.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8 user's reference guide.* Chicago: Scientific Software International.

Kano, Y., Berkane, M., & Bentler, P. M. 1993). Statistical inference based on pseudo-maximum likelihood estimators in elliptical populations. *Journal of the American Statistical Association, 88,* 135–143.

Kharin, Y. S. (1996). Robustness in discriminant analysis. In H. Rieder (Ed.), *Robust statistics, data analysis, and computer intensive methods* (pp. 225–234). New York: Springer-Verlag.

Lange, K. L., Little, R. J. A., & Taylor, J. M. G. (1989). Robust statistical modeling using the *t* distribution. *Journal of the American Statistical Association, 84,* 881–896.

Lee, S.-Y., & Jennrich, R. I. (1979). A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. *Psychometrika, 44,* 99–114.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73,* 13–22.

Little, R. J. A. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics, 37,* 23–38.

Magnus, J. R., & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. New York: Wiley.

Mardia, K. V. (1970). Measure of multivariate skewness and kurtosis with applications. *Biometrika, 57,* 519–530.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā Series B, 36,* 115–128.

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

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105,* 156–166.

Mooijaart, A. (1993). Structural equation models with transformed variables. In K. Haagen, D. J. Bartholomew, & M. Deistler (Eds.), *Statistical modeling and latent variables* (pp. 249–258). Amsterdam: North-Holland.

Neumann, C. S. (1994). *Structural equation modeling of symptoms of alcoholism and psychopathology*. Dissertation, University of Kansas, Lawrence.

Newcomb, M. D., & Bentler, P. M. (1988). *Consequences of adolescent drug use: Impact on the lives of young adults*. Beverly Hills, CA: Sage.

Olkin, I. (1994). Multivariate non-normal distributions and models of dependency. In T. W. Anderson, K. T. Fang, & I. Olkin (Eds.), *Multivariate analysis and its applications* (pp. 37–53). Hayward, CA: Institute of Mathematical Statistics.

Rubin, D. B. (1983). Iteratively reweighted least squares. In N. L. Johnson & S. Kotz (Eds.), *Encyclopedia of statistical sciences* (Vol. 4, pp. 272–275). New York: Wiley.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. In *American Statistical Association 1988 Proceedings of Business and Economics Sections* (pp. 308–313). Alexandria, VA: American Statistical Association.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Stein, J. A., Newcomb, M. D., & Bentler, P. M. (1996). Initiation and maintenance of tobacco smoking: Changing personality correlates in adolescence and young adulthood. *Journal of Applied Social Psychology, 26,* 160–187.

Tremblay, P. F., & Gardner, R. C. (1996). On the growth of structural equation modeling in psychological journals. *Structural Equation Modeling, 3,* 93–104.

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

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

Velilla, S. (1995). Diagnostics and robust estimation in multivariate data transformations. *Journal of the American Statistical Association, 90,* 945–951.

Wilcox, R. R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.

Yuan, K.-H., & Bentler, P. M. (1997). Improving parameter tests in covariance structure analysis. *Computational Statistics and Data Analysis, 26,* 177–198.

Yuan, K.-H., & Bentler, P. M. (1998a). Robust mean and covariance structure analysis. *British Journal of Mathematical and Statistical Psychology, 51,* 63–88.

Yuan, K.-H., & Bentler, P. M. (1998b). Structural equation modeling with robust covariances. *Sociological Methodology, 28,* 363–396.

Yuan, K.-H., & Bentler, P. M. (1998c). Normal theory based test statistics in structural equation modelling. *British Journal of Mathematical and Statistical Psychology, 51,* 289–309.

Yuan, K.-H., & Bentler, P.M. (1999). On normal theory and associated test statistics in covariance structure analysis under two classes of nonnormal distributions. *Statistica Sinica, 9,* 831–853.

Yuan, K.-H., & Jennrich, R. I. (1998). Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis, 65,* 245–260.

Received 14 December 1998; revised version received 28 September 1999