Statistica Sinica **9**(1999), 831-853

ON NORMAL THEORY AND ASSOCIATED TEST STATISTICS IN COVARIANCE STRUCTURE ANALYSIS UNDER TWO CLASSES OF NONNORMAL DISTRIBUTIONS

Ke-Hai Yuan and Peter M. Bentler

University of North Texas and University of California, Los Angeles

Abstract: Several test statistics for covariance structure models derived from the normal theory likelihood ratio are studied. These statistics are robust to certain violations of the multivariate normality assumption underlying the classical method. In order to explicitly model the behavior of these statistics, two new classes of nonnormal distributions are defined and their fourth-order moment matrices are obtained. These nonnormal distributions can be used as alternatives to elliptical symmetric distributions in the study of the robustness of a multivariate statistical method. Conditions for the validity of the statistics under the two classes of nonnormal distributions are given. Some commonly used models are considered as examples to verify our conditions under each class of nonnormal distributions. It is shown that these statistics are valid under much wider classes of distributions than previously assumed. The theory also provides an explanation for previously reported Monte-Carlo results on some of the statistics.

Key words and phrases: Covariance structure, kurtosis, likelihood ratio test, nonnormal distribution, scaled statistics, skewness.

1. Introduction

Classical methods for covariance structure analysis are developed under the assumption of multivariate normality. In practice, however, data sets in behavioral and biomedical sciences, in which covariance structure analysis is regularly used, are seldom normal. Browne (1984) developed a generalized least squares (GLS) method that does not require a specific distributional form. Since for most practical models the GLS method requires an extremely large sample size to give a reliable inference, this method is not commonly used in covariance structure practice. Based on the GLS method, Yuan and Bentler (1997a) proposed a new statistic which gives more reliable inference for small to moderate sample sizes. However, for relatively large models and small sample sizes, there may exist problems of nonconvergence in getting the parameter estimator with the GLS method (e.g., Hu, Bentler and Kano (1992), Yuan and Bentler (1997a)). For these reasons, as reviewed by Bentler and Dudgeon (1996), people still prefer the classical normal theory method in modeling covariance structures even if their data are not normal. In some exciting research on the robustness of the normal theory method, Amemiya and Anderson (1990), Anderson (1989), Browne and Shapiro (1988) and Satorra and Bentler (1990) found that some normal theory based statistics for covariance structure models are still valid under some assumptions. Unfortunately there is no effective way to verify these assumptions in practice.

Another direction of research has been to use a correction to the normal theory likelihood ratio (LR) statistic to improve its performance under distributional violation. When a sample is from an elliptical distribution (e.g., Fang, Kotz and Ng (1990)), Muirhead and Waternaux (1980) first proposed a simple correction to the LR statistic in canonical correlation and some specific covariance structures. Tyler (1983) generalized this correction to test statistics based on affine-invariant robust estimates of a scatter matrix. Based on Mardia's multivariate kurtosis parameter (Mardia (1974)), Browne (1984) extended this correction and its application to general covariance structure analysis. The correction for an elliptical distribution was further explored by Shapiro and Browne (1987). Kano (1992) also gave a correction to the normal theory LR statistic for some covariance structure models. Kano's correction incorporated both the independent variables case and the elliptical distribution case, thus is more general in a sense. Satorra and Bentler (1988, 1994) gave another correction. As we shall see, the rescaled statistic of Satorra and Bentler is very general. It is known that these corrected statistics asymptotically follow the nominal chi-square distribution when the data are from an elliptical distribution. However, simulation by Browne (1984) indicates that the statistic based on Mardia's kurtosis parameter performs robustly for skewed data; extensive simulation results of Curran, West and Finch (1996) and Hu et al. (1992) indicate a surprising robustness of the rescaled statistic of Satorra and Bentler. A comprehensive statistical explanation is needed to understand these robust performances.

As an alternative to multivariate normal distributions, the class of elliptical distributions plays an important role in the study of robustness of classical statistical procedures (e.g., Anderson and Fang (1987), Muirhead (1982)). However, data sets in the real world are often characterized by skewness and highly different marginal kurtoses. Kano, Berkane and Bentler (1990) proposed a method of modeling the fourth-order moment matrix by incorporating the marginal kurtoses. However, it is not clear which nonnormal distributions fit this model. Azzalini and Valle (1996), Olkin (1994) and Steyn (1993) proposed various nonnormal distributions, but these do not seem especially relevant to covariance structure analysis. As we shall see, the fourth-order moment matrix plays a key role in the study of covariance structure models. A convenient class of distributions for the study of test statistics should be one whose fourth-order moment matrix is easy to calculate. Further, in order to model skewness and kurtosis in real data analysis, a useful class of nonnormal distributions should also possess the advantage of easy-to-specify marginal skewness and kurtosis. And from the perspective of Monte-Carlo studies, these nonnormal distributions should be easy to generate in simulations. Based on these considerations, we define two new classes of nonnormal distributions which possess the above mentioned qualities. These two classes of nonnormal distributions will be the basis for our study of several different test statistics.

We aim to give explanations for the results of previous Monte-Carlo studies and to give guidance to applications of these statistics in covariance structure analysis. Formal definitions of the statistics will be introduced in Section 2. By some generalizations of the class of elliptical distributions, two classes of nonnormal distributions will be given in Section 3. The validity of the statistics under these two classes of distributions will be discussed in Section 4. Some specific models will be used as examples to verify the asymptotic robustness conditions. This will be given in Section 5. A discussion and conclusion will be given at the end of this paper.

2. Rescaled Normal Theory Test Statistics

We introduce the analytical form of the various statistics at issue in this section. These statistics will be further studied in later sections.

Let X be a p-variate random vector with $E(X) = \mu$ and $\mathrm{Cov}(X) = \Sigma$. For a covariance structure $\Sigma = \Sigma(\theta)$, the normal theory log-likelihood leads to the following function to be minimized

$$F_{ML}(\theta) = \mathrm{tr}\{S\Sigma^{-1}(\theta)\} - \log|S\Sigma^{-1}(\theta)| - p, \tag{2.1}$$

where S is the sample covariance matrix based on a sample from X with size $N = n + 1$. We will assume that X has finite fourth-order moments, and that the standard regularity conditions stated in Yuan and Bentler (1997a) hold for $\Sigma = \Sigma(\theta)$. Let $\hat{\theta}$ be the estimator obtained by minimizing $F_{ML}(\theta)$. Under the assumption of multivariate normality, the statistic

$$T_{ML} = nF_{ML}(\hat{\theta})$$

is asymptotically $\chi^2_{p^*-q}$, where $p^* = p(p+1)/2$ and q is the number of unknown free parameters in θ. Generally, T_{ML} will approach a nonnegative random variable T. Let $\mathrm{vech}(\cdot)$ be an operator which transforms a symmetric matrix into a vector by picking the nonduplicated elements of the matrix, $\sigma = \mathrm{vech}(\Sigma)$, and D_p be the $p^2 \times p^*$ duplication matrix such that $\mathrm{vec}(\Sigma) = D_p\sigma$ (see Magnus and Neudecker (1988, p.49), Schott (1997, p.283)). Also let $\dot{\sigma} = \partial\sigma/\partial\theta'$ and $V = 2^{-1}D_p'(\Sigma^{-1} \otimes \Sigma^{-1})D_p$; then

$$W = V - V\dot{\sigma}(\dot{\sigma}'V\dot{\sigma})^{-1}\dot{\sigma}'V \tag{2.2}$$

is the residual weight matrix, as it was called by Bentler and Dudgeon (1996). Satorra and Bentler (1988) decompose T as $\sum_{i=1}^{p^*-q}\tau_i\chi_{1i}^2$, where the τ_i are the nonzero eigenvalues of the matrix $W\Gamma$ with $\Gamma = \mathrm{Cov}[\mathrm{vech}\{(X-\mu)(X-\mu)'\}]$ and the χ_{1i}^2 are independent chi-square variates, each having one degree of freedom. When X follows an elliptical distribution with kurtosis β, then

$$\Gamma = 2\beta D_p^+(\Sigma\otimes\Sigma)D_p^{+'} + (\beta-1)\sigma\sigma', \tag{2.3}$$

where D_p^+ is the Moore-Penrose inverse of D_p. In such a case all the τ_i's are equal and $\tau = \mathrm{tr}(W\Gamma)/(p^*-q) = \beta$. Let $Y_i = \mathrm{vech}\{(X_i-\bar{X})(X_i-\bar{X})'\}$ and S_Y be the corresponding sample covariance of Y_i. Then S_Y is a consistent estimator of the matrix Γ, and $\hat{\tau} = \mathrm{tr}(\hat{W}S_Y)/(p^*-q)$ is a consistent estimator of τ, where $\hat{W}$ is obtained by substituting $\hat{\theta}$ for θ in $\Sigma = \Sigma(\theta)$ and $\dot{\sigma} = \dot{\sigma}(\theta)$. Based on these, Satorra and Bentler (1988) proposed a rescaled statistic

$$T_{SB} = \hat{\tau}^{-1}T_{ML}.$$

For a p-variate random vector X, the relative Mardia's multivariate kurtosis parameter of X is

$$\eta = E\{(X-\mu)'\Sigma^{-1}(X-\mu)\}^2/\{p(p+2)\}. \tag{2.4}$$

Since $\eta = \beta$ within the class of elliptical distributions, Browne (1984) proposed a rescaled statistic

$$T_B = \hat{\eta}^{-1}T_{ML},$$

where

$$\hat{\eta} = \frac{N+1}{N(N-1)}\sum_{i=1}^{N}\{(X_i-\bar{X})'S^{-1}(X_i-\bar{X})\}^2/\{p(p+2)\}$$

is a consistent estimator of η, and is unbiased when data are normal.

Let the covariance structure $\Sigma = \Sigma(\theta)$ be specified by a linear latent variable model

$$X = \mu + \sum_{j=1}^{J}\Lambda_j(\theta)z_j, \tag{2.5}$$

where the $\Lambda_j(\theta)$ are $p\times k_j$ matrix valued functions of θ, the z_j are k_j-dimensional latent vectors which are uncorrelated with each other, and $E(z_j) = 0$, $\mathrm{Cov}(z_j) = \Phi_j(\theta)$. This generates a covariance structure

$$\Sigma(\theta) = \sum_{j=1}^{J}\Lambda_j(\theta)\Phi_j(\theta)\Lambda_j'(\theta). \tag{2.6}$$

Suppose there exist two matrices P and Q with sizes $p \times m_1$ and $p \times m_2$ respectively such that

$$P'\Lambda_j = 0 \text{ or } Q'\Lambda_j = 0 \tag{2.7}$$

for every j. This will result in

$$P'\Sigma Q = 0. \tag{2.8}$$

Kano (1992) proposed a correction factor

$$\nu = E\{(X-\mu)'P(P'\Sigma P)^{-1}P'(X-\mu)(X-\mu)'Q(Q'\Sigma Q)^{-1}Q'(X-\mu)\}/(m_1 m_2). \tag{2.9}$$

For consistent estimators $\hat{P}$ and $\hat{Q}$ of P and Q, respectively,

$$\hat{\nu} = \frac{1}{N}\sum_{i=1}^{N}\{(X_i-\bar{X})'\hat{P}(\hat{P}'S\hat{P})^{-1}\hat{P}'(X_i-\bar{X})(X_i-\bar{X})'\hat{Q}(\hat{Q}'S\hat{Q})^{-1} \hat{Q}'(X_i-\bar{X})\}/(m_1 m_2)$$

is a consistent estimator of ν. Kano's (1992) test statistic is given by

$$T_K = \hat{\nu}^{-1}T_{ML}.$$

Kano established the validity of T_K for model (2.6) when either the z_j in (2.5) are independent or the X in (2.5) follows an elliptical distribution. As noted by Kano (1992), for a general linear latent variable model, the validity of correction (2.9) depends on whether P and Q satisfying (2.7) exist. One such choice is to choose P and Q in (2.7) to be orthogonal to the matrices $(\Lambda_1, \ldots, \Lambda_{J_1})$ and $(\Lambda_{J_1+1}, \ldots, \Lambda_J)$, respectively for an appropriate J_1; or to choose P to be orthogonal to some columns of Σ and Q to be orthogonal to the other columns in (2.8). Note that (2.7) implies (2.8), but not the reverse. Since we do not relate every covariance structure to a specific latent variable model, unless explicitly stated, we will only assume (2.8) in the rest of the paper.

We will explore the validity of these statistics in larger classes of distributions in Sections 4 and 5. Since the population mean μ is not restricted, it is estimated at its sample mean $\bar{X}$. Thus, without loss of generality we will assume $\mu = 0$ in the rest of this paper.

3. Two Classes of Nonnormal Distributions

Two classes of nonnormal distributions will be introduced in this section. We first give their definitions and discuss some of their interesting properties, then we will obtain their fourth-order moment matrices. These matrices will be used to study various statistics in later sections.

Let $Z \sim N(0, I_p)$ and r be a nonnegative random variable which is independent of Z; then

$$X = r\Sigma^{\frac{1}{2}}Z \tag{3.1}$$

follows an elliptical distribution (e.g., Fang, Kotz and Ng (1990)). Since real data sets often exhibit skewness and different marginal kurtoses, we have no real interest in the nonnormal distribution represented by (3.1), which has no skewness and the same marginal kurtosis for all variables. However, it should be noticed that there are two factors which make the distribution of X in (3.1) elliptical. One is that the distribution of Z is symmetric, the other is that each component of $\Sigma^{\frac{1}{2}}Z$ is multiplied by the same random variable r. Changing any one of these two factors will result in a nonelliptical distribution.

Data Model I. Let $\xi_1, \ldots, \xi_m$ be independent random variables with $E(\xi_i) = 0$, $E(\xi_i^2) = 1$, $E(\xi_i^3) = \zeta_i$, $E(\xi_i^4) = \kappa_i$, and $\xi = (\xi_1, \ldots, \xi_m)'$. Let r be a random variable which is independent of ξ, $E(r^2) = 1$, $E(r^3) = \gamma$, and $E(r^4) = \beta$. Also, let $m \geq p$ and $A = (a_{ij})$ be a $p \times m$ matrix of rank p such that $AA' = \Sigma$. Then the random vector

$$X = rA\xi \tag{3.2}$$

will generally follow a nonelliptical distribution. Since the distribution of ξ is not necessarily symmetric, we do not restrict the r in (3.2) to be a nonnegative random variable. It is easily seen that the population covariance matrix of X is given by Σ. The skewness and kurtosis of x_i are given respectively by

$$\text{skew}(x_i) = \gamma \sum_{j=1}^{m} a_{ij}^3 \zeta_j / \sigma_{ii}^{\frac{3}{2}} \quad \text{and} \quad \text{kurt}(x_i) = \beta\{\sum_{j=1}^{m} a_{ij}^4 (\kappa_j - 3)/\sigma_{ii}^2 + 3\}. \tag{3.3}$$

The constants γ and β are scaling factors which change the marginal skewness and kurtosis proportionally. Discrepancies among the marginal skewnesses and kurtoses can be achieved by changing m, a_{ij}, ζ_j and κ_j. Generating X through (3.2) has two advantages. First, it is easy to implement, and second, the population fourth-order moment matrix is easy to calculate.

When $\kappa_1 = \cdots = \kappa_m = 3$, all the marginal kurtoses of X are equal to 3β. In this case, comparing (3.2) with (3.1), we may call the corresponding distribution of X in (3.2) a *pseudo elliptical distribution*, since its distribution is not elliptical anymore even though the marginal kurtoses are the same. Similarly, we may call the distribution of X in (3.2) a *pseudo normal distribution* if $\beta = 1$ in addition to $\kappa_i = 3$ for $i = 1, \ldots, m$.

Data Model II. Let $\xi_1, \ldots, \xi_m$ be independent random variables with $E(\xi_i) = 0$, $E(\xi_i^2) = 1$, $E(\xi_i^3) = \zeta_i$, $E(\xi_i^4) = 3$, and $\xi = (\xi_1, \ldots, \xi_m)'$. Let $r_1, \ldots, r_p$ be independent random variables with $E(r_i) = \alpha$, $E(r_i^2) = 1$, $E(r_i^3) = \gamma_i$, $E(r_i^4) = \beta_i$, $R = \mathrm{diag}(r_1, \ldots, r_p)$, and R and ξ be independent. Let $A = (a_{ij})$ be a $p \times m$ nonstochastic matrix of rank p and $AA' = G = (g_{ij})$. Another simple generating scheme is

$$X = RA\xi. \tag{3.4}$$

Since the distribution of ξ in (3.4) is not necessarily symmetric, we will not restrict the r_i to be nonnegative random variables. The covariance matrix of X in (3.4) is given by $\Sigma = \alpha^2 G + (1 - \alpha^2) D_G$, where $D_G = \mathrm{diag}(g_{11}, \ldots, g_{pp})$. Obviously, $\sigma_{ii} = g_{ii}$ and $\sigma_{ij} = \alpha^2 g_{ij}$ for $i \neq j$. For a given positive definite matrix Σ, we need to choose a constant α. Since $g_{ii} = \sigma_{ii}$ and $g_{ij} = \sigma_{ij}/\alpha^2$ for $i \neq j$, α can not be too small in order for G to be positive definite. As Σ is positive definite, standard analysis shows that there exists an α_0 such that $0 < \alpha_0 < 1$ and for all $\alpha_0 \leq \alpha \leq 1$ the corresponding $G = \alpha^{-2}\Sigma + (1 - \alpha^{-2}) D_\Sigma$ is positive definite. The marginal skewnesses and kurtoses of X in (3.4) are given by

$$\mathrm{skew}(x_i) = \gamma_i \sum_{j=1}^{m} a_{ij}^3 \zeta_j / \sigma_{ii}^{\frac{3}{2}} \quad \text{and} \quad \mathrm{kurt}(x_i) = 3\beta_i, \tag{3.5}$$

respectively. For a given matrix A, differing marginal skewness in X can be achieved by adjusting the γ_i in (3.5), so we can let $m = p$ and $A = A' = G^{\frac{1}{2}}$ for convenience. The kurtosis of x_i is controlled by β_i. The advantage of generating X through (3.4) is that marginal skewness and kurtosis are easy to control.

It follows from (3.3) and (3.5) that when $\kappa_1 = \cdots = \kappa_m = 3$, $\gamma_1 = \cdots = \gamma_p = \gamma$ and $\beta_1 = \cdots = \beta_p = \beta$ the marginal skewness and kurtosis of x_i in (3.2) are the same as those of the x_i in (3.4). However, the distributions of the two data models are totally different. This is because the r_i in Data Model II are independent. Comparing (3.2) or (3.4) with (2.5), we may relate the columns of A to the Λ_j's in (2.5), and r, R and ξ_j are then latent variables that appear in product, instead of linear, forms. However, our emphasis is more on the observed distributions and their effects on test statistics, rather than on the physical interpretation of the two data models.

The above two generation techniques have been described in more detail in Yuan and Bentler (1997b). They also gave guidelines on how to achieve desired marginal skewness and kurtosis by using existing techniques for generating univariate nonnormal random numbers. In the rest of this section, we give the form of the fourth-order moment matrices Γ that correspond to the two classes of distributions given by Data Models I and II. Proofs are given in an appendix.

Theorem 3.1. *Let $X = rA\xi$ be generated through* (3.2)*, and r, A and ξ satisfy the conditions for Data Model I. Then the fourth-order moment matrix*

$\mathrm{Cov}\{\mathrm{vech}(xx')\}$ *is*

$$\Gamma = 2\beta D_p^+(\Sigma \otimes \Sigma)D_p^{+'} + (\beta - 1)\sigma\sigma' + \beta \sum_{i=1}^{m} (\kappa_i - 3)\mathrm{vech}(a_i a_i')\mathrm{vech}'(a_i a_i'). \quad (3.6)$$

Notice that the first two terms on the RHS of (3.6) correspond to the fourth-order moment matrix from an elliptical distribution with kurtosis β. The last term is contributed by the excess kurtosis that exists in the distribution of ξ_i, which is further inflated by the factor β. Comparing (3.6) with (2.3), we also notice that the Γ matrices corresponding to a pseudo elliptical distribution and an elliptical distribution are the same if they share the same population covariance matrix and marginal kurtosis. Similarly, the Γ matrices corresponding to a pseudo normal distribution and a normal distribution are the same if they have the same population covariance matrix. These observations will facilitate our ability to generalize many of the results in covariance structure analysis within the class of elliptical distributions to much wider classes of distributions.

Let e_i be the ith unit vector of dimension p with unity for its ith element and zeros elsewhere, and $E_{ij} = e_i e_j'$.

Theorem 3.2. *Let* $X = RA\xi$ *be generated through* (3.4)*, and* R*,* A *and* ξ *satisfy the conditions of Data Model II. Then*

$$\begin{aligned}
\Gamma = {} & 2D_p^+(\Sigma \otimes \Sigma)D_p^{+'} + (\alpha^{-2} - 1)\sum_{i=1}^{p}\{2\mathrm{vech}(E_{ii})\mathrm{vech}'(\sigma_i\sigma_i') \\
& +2\mathrm{vech}(\sigma_i\sigma_i')\mathrm{vech}'(E_{ii}) + 3\mathrm{vech}(\sigma_i e_i')\mathrm{vech}'(e_i\sigma_i') + 3\mathrm{vech}(e_i\sigma_i)\mathrm{vech}'(\sigma_i e_i') \\
& +\mathrm{vech}(e_i\sigma_i')\mathrm{vech}'(e_i\sigma_i') + \mathrm{vech}(\sigma_i e_i')\mathrm{vech}'(\sigma_i e_i')\} \\
& +3\sum_{i=1}^{p}(\alpha^{-1}\gamma_i - 2\alpha^{-2} + 1)\sigma_{ii}\{\mathrm{vech}(E_{ii})\mathrm{vech}'(e_i\sigma_i') + \mathrm{vech}(E_{ii})\mathrm{vech}'(\sigma_i e_i') \\
& +\mathrm{vech}(\sigma_i e_i')\mathrm{vech}'(E_{ii}) + \mathrm{vech}(e_i\sigma_i')\mathrm{vech}'(E_{ii})\} \\
& +3\sum_{i=1}^{p}(\beta_i - 4\alpha^{-1}\gamma_i + 8\alpha^{-2} - 2\alpha^{-4} - 3)\sigma_{ii}^2\mathrm{vech}(E_{ii})\mathrm{vech}'(E_{ii}) \\
& +(\alpha^{-4} - 2\alpha^{-2} + 1)\sum_{ij}\sigma_{ij}^2\{2\mathrm{vech}(E_{ii})\mathrm{vech}'(E_{jj}) + 3\mathrm{vech}(E_{ij})\mathrm{vech}'(E_{ji}) \\
& +\mathrm{vech}(E_{ij})\mathrm{vech}'(E_{ij})\}. \qquad (3.7)
\end{aligned}$$

When $\alpha = \gamma_i = \beta_i = 1$, only the first term on the RHS of (3.7) is left. This corresponds to the fourth-order moment matrix of a pseudo normal distribution. For general α, γ_i, and β_i, the terms besides the first term on the RHS of (3.7) come from different marginal kurtoses. Note that even if we let $\kappa_1 = \cdots = \kappa_m = 3$, $\gamma_1 = \cdots = \gamma_p = \gamma$ and $\beta_1 = \cdots = \beta_p = \beta$, (3.7) does not equal (3.6). This reflects the different distributions of (3.2) and (3.4) as we observed earlier. The RHS of (3.7) can be further simplified for specific models. For example, when $\Sigma = \text{diag}(\sigma_{11}, \ldots, \sigma_{pp})$, then

$$\Gamma = 2D_p^+(\Sigma \otimes \Sigma)D_p^{+'} + 3\sum_{i=1}^{p}(\beta_i - 1)\sigma_{ii}^2 \text{vech}(E_{ii})\text{vech}'(E_{ii}). \tag{3.8}$$

When $\beta_i = 1$ and X follows a pseudo normal distribution, the x_i may not be independent even though Σ is diagonal. This result will be used in a study of the uncorrelated variables model in Sections 4 and 5.

Notice that the ζ_j and γ in (3.3) do not appear in (3.6), and the ζ_j in (3.5) do not enter into (3.7) either. This implies that in both Data Model I and Data Model II, the marginal skewnesses of X have no asymptotic effect on the rescaled test statistics in covariance structure analysis. However, we may observe an effect of skewnesses in any given finite sample. Although we do not deal with mean structures in this paper, we would like to note that skewness will be involved in the corresponding test statistics when a mean structure is also of interest.

4. Asymptotic Robustness of Four Statistics

In this section, we will study the asymptotic robustness of the test statistics when samples come from the two classes of nonnormal distributions given in Section 3. Several conditions are relevant to the development of the rescaled statistics. For a covariance structure $\Sigma(\theta)$, if for any parameter vector θ and positive constant α there exists a parameter vector θ^* such that $\Sigma(\theta^*) = \alpha\Sigma(\theta)$, then $\Sigma(\theta)$ is invariant with respect to a constant scaling factor (ICSF). As noted by Browne (1984), nearly all structural models in current use are ICSF. Let $\mathcal{R}(\dot{\sigma})$ represent the range space spanned by the column vectors of $\dot{\sigma}$. If $\sigma \in \mathcal{R}(\dot{\sigma})$, then the model is said to be quasi-linear (QL). It is easily seen that ICSF implies QL (Satorra and Bentler (1986)). So condition QL is also satisfied by almost all of the models in current use. Since the robustness conditions for Data Model I and Data Model II are different, we need to study them separately. The following lemma will be used in several places in this section.

Lemma 4.1. *Let C be a $p \times p$ matrix. If* $\text{vech}(C) \in \mathcal{R}(\dot{\sigma})$*, then* $W\text{vech}(C) = 0$*, where W is the residual weight matrix given in* (2.2).

Proof. There exists a vector a such that $\text{vech}(C) = \dot{\sigma}a$, so $W\text{vech}(C) = W\dot{\sigma}a = 0$.

Let $A = (a_1, \ldots, a_m)$ be the matrix in (3.2). The following condition is used for the study of the test statistics under Data Model I.

Condition 4.1. For each of the $i = 1, \ldots, m$, $\text{vech}(a_i a_i') \in \mathcal{R}(\dot{\sigma})$.

Notice that condition ICSF is implied by Condition 4.1. Unlike the conditions ICSF or QL, Condition 4.1 generally depends on the data generation method and the model structure, as will be further discussed for specific models in the next section. The following lemma characterizes the eigenvalues of $W\Gamma$ with the Γ in (3.6).

Lemma 4.2. *Let* Γ *be given as in* (3.6) *and* W *be given as in* (2.2). *Then under Condition* 4.1 *the* $p^* - q$ *nonzero eigenvalues of* $W\Gamma$ *are equal and are given by* β.

Proof. Let $\Gamma_1 = 2D_p^+(\Sigma \otimes \Sigma)D_p^{+'}$, $\Gamma_2 = \sigma\sigma'$, and $\Gamma_{3i} = \text{vech}(a_i a_i')\text{vech}'(a_i a_i')$. Since Γ_1 corresponds to the fourth-order moment matrix from a normal distribution, it is obvious that the eigenvalues of $W\Gamma_1$ are either 0 or 1 for the matrix W in (2.2). Using (3.6), we need to show that $W\Gamma_2 = 0$ and $W\Gamma_{3i} = 0$ for $i = 1, \ldots, m$, and this follows from Lemma 4.1.

The following lemma gives the scaling factors τ, η, and ν for Data Model I.

Lemma 4.3. *Let* X *be given as in Data Model I. Under Condition* 4.1 *the scaling factors for* T_{SB}, T_B *and* T_K *are given, respectively, by* $\tau = \beta$,

$$\eta = \beta + \frac{\beta}{p(p+2)} \sum_{i=1}^{m} (\kappa_i - 3)(a_i'\Sigma^{-1}a_i)^2, \tag{4.1}$$

and

$$\nu = \beta + \frac{\beta}{m_1 m_2} \sum_{i=1}^{m} (\kappa_i - 3)\{a_i'P(P'\Sigma P)^{-1}P'a_i\}\{a_i'Q(Q'\Sigma Q)^{-1}Q'a_i\}. \tag{4.2}$$

Proof. That $\tau = \beta$ follows from Lemma 4.2. Let $\Gamma_1 = 2D_p^+(\Sigma \otimes \Sigma)D_p^{+'}$, $\Gamma_2 = \sigma\sigma'$, and $\Gamma_{3i} = \text{vech}(a_i a_i')\text{vech}'(a_i a_i')$. Then

$$\text{tr}[\{D_p'(\Sigma^{-1} \otimes \Sigma^{-1})D_p\}\Gamma_1] = 2p^*, \quad \text{tr}[\{D_p'(\Sigma^{-1} \otimes \Sigma^{-1})D_p\}\Gamma_2] = p,$$
$$\text{tr}[\{D_p'(\Sigma^{-1} \otimes \Sigma^{-1})D_p\}\Gamma_{3i}] = (a_i'\Sigma^{-1}a_i)^2,$$

and (4.1) follows from (2.4) and (3.6). Similarly, the ν in (4.2) follows from (2.8), (2.9) and (3.6).

If $A = \Sigma^{\frac{1}{2}}$, then $\Sigma^{-\frac{1}{2}}a_i = e_i$. In such a case

$$\eta = \beta + \frac{\beta}{p(p+2)} \sum_{i=1}^{m} (\kappa_i - 3).$$

Under the condition in Lemma 4.3, the only possibility for τ, η and ν to be equal is that the second terms on the RHS of (4.1) and (4.2) are zero. If X follows a pseudo elliptical distribution, then $\eta = \nu = \tau$. So the rescaled statistics T_B and T_K originating from elliptical distributions are also valid in the much larger class of pseudo elliptical distributions. If $a_i = \Lambda_i$ and (2.7) holds, then $\nu = \tau$ even if X does not follow a pseudo elliptical distribution.

Lemmas 4.2 and 4.3 lead to the following theorem.

Theorem 4.1. *Let X be given as in Data Model I. Then under Condition* 4.1 *the statistic T_{SB} is asymptotically robust, T_B and T_K are asymptotically robust if X follows a pseudo elliptical distribution, T_{ML} is asymptotically robust only if $\beta = 1$. Furthermore, if* (2.7) *holds and $\Lambda_i = a_i$, T_K is also asymptotically robust.*

Note that even if (2.7) holds with $\Lambda_i = a_i$, and comparing (3.2) with (2.5), the $z_i = r\xi_i$ in (3.2) are not independent as are the z_i in (2.5). So Theorem 4.1 generalizes the validity of T_K to a much larger class of distributions than those stated in Kano (1992). When $\beta = 1$ and $P(r = 1) = 1$, the result on T_{ML} in Theorem 4.1 is essentially the same as that in Amemiya and Anderson (1990), Anderson (1989) and Browne and Shapiro (1988).

Now we turn to Data Model II. To obtain further results, we need more assumptions.

Condition 4.2. For each i, $j = 1, \ldots, p$, $\mathrm{vech}(e_i e_i')$, $\mathrm{vech}(\sigma_i \sigma_i')$, $\mathrm{vech}(e_i \sigma_i')$, $\mathrm{vech}(e_i e_j') \in \mathcal{R}(\dot{\sigma})$.

Condition 4.2 implies ICSF. Similarly as Condition 4.1, Condition 4.2 also depends on the model structure as well as on the data generation method. However, few currently used models satisfy Condition 4.2 with Data Model II. This condition is stated for a general covariance structure, but specific covariance models can be based on a much simpler assumption. When $\Sigma = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{pp})$ for example, the following condition can replace Condition 4.2.

Condition 4.2a. For $i = 1, \ldots, p$, $\mathrm{vech}(E_{ii}) \in \mathcal{R}(\dot{\sigma})$.

Notice that $\partial\Sigma(\theta)/\partial\theta_j$ is symmetric. Applying Lemma 4.1 to the individual terms on the RHS of (3.7), we have the following lemma.

Lemma 4.4. *Let Γ be given as in* (3.7) *and W be given as in* (2.2). *Then under Condition* 4.2, *the $p^* - q$ nonzero eigenvalues of $W\Gamma$ are all equal to* 1.

The following lemma gives the scaling factors τ, η and ν for Data Model II.

Lemma 4.5. *Let X be given as in Data Model II. Then under Condition* 4.2 *the scaling factors for T_{SB}, T_B and T_K are given, respectively, by $\tau = 1$,*

$$\eta = 1 + [(\alpha^{-2} - 1)(10p + 2\sum_{i=1}^{p} \sigma_{ii}\sigma^{ii}) + 12\sum_{i=1}^{p}(\alpha^{-1}\gamma_i - 2\alpha^{-2} + 1)\sigma_{ii}\sigma^{ii}$$

$$
\begin{aligned}
&+3\sum_{i=1}^{p}(\beta_i - 4\alpha^{-1}\gamma_i + 8\alpha^{-2} - 2\alpha^{-4} - 3)(\sigma_{ii}\sigma^{ii})^2 \\
&+(\alpha^{-4} - 2\alpha^{-2} + 1)\sum_{ij}\{5(\sigma_{ij}\sigma^{ij})^2 + \sigma_{ij}^2\sigma^{ii}\sigma^{jj}\}]/\{p(p+2)\},
\end{aligned} \tag{4.3}
$$

where σ^{ij} is an element of the matrix $\Sigma^{-1} = (\sigma^{ij})$, and

$$
\begin{aligned}
\nu = 1 + \{&(\alpha^{-2} - 1)\sum_{i=1}^{p}(P_{1ii}Q_{2ii} + P_{2ii}Q_{1ii} + 10P_{3ii}Q_{3ii}) \\
&+6\sum_{i=1}^{p}(\alpha^{-1}\gamma_i - 2\alpha^{-2} + 1)\sigma_{ii}(P_{1ii}Q_{3ii} + P_{3ii}Q_{1ii}) \\
&+3\sum_{i=1}^{p}(\beta_i - 4\alpha^{-1}\gamma_i + 8\alpha^{-2} - 2\alpha^{-4} - 3)\sigma_{ii}^2 P_{1ii}Q_{1ii} \\
&+(\alpha^{-4} - 2\alpha^{-2} + 1)\sum_{ij}\sigma_{ij}^2(5P_{1ij}Q_{1ij} + P_{1jj}Q_{1ii})\}/(m_1 m_2),
\end{aligned} \tag{4.4}
$$

where

$$
\begin{aligned}
&P_{1ii} = e_i'P(P'\Sigma P)^{-1}P'e_i, \quad P_{1ij} = e_i'P(P'\Sigma P)^{-1}P'e_j, \\
&P_{2ii} = \sigma_i'P(P'\Sigma P)^{-1}P'\sigma_i, \quad P_{3ii} = e_i'P(P'\Sigma P)^{-1}P'\sigma_i, \\
&Q_{1ii} = e_i'Q(Q'\Sigma Q)^{-1}Q'e_i, \quad Q_{1ij} = e_i'Q(Q'\Sigma Q)^{-1}Q'e_j, \\
&Q_{2ii} = \sigma_i'Q(Q'\Sigma Q)^{-1}Q'\sigma_i, \quad Q_{3ii} = e_i'Q(Q'\Sigma Q)^{-1}Q'\sigma_i.
\end{aligned}
$$

Proof. That $\tau = 1$ follows from Lemma 4.4. For p-dimensional vectors b_1, b_2, b_3, b_4, it follows that

$$
\mathrm{tr}[\{D_p'(\Sigma^{-1}\otimes\Sigma^{-1})D_p\}\mathrm{vech}(b_1b_2')\mathrm{vech}(b_3b_4')] = (b_1'\Sigma^{-1}b_3)(b_2'\Sigma^{-1}b_4). \tag{4.5}
$$

Using $\Sigma^{-1}\sigma_i = e_i$, (4.3) follows from (2.4), (3.7) and (4.5). Similarly, the ν in (4.4) follows from (2.8), (2.9), (3.7) and (4.5).

Lemmas 4.4 and 4.5 lead to the following theorem.

Theorem 4.2. *Let X be given as in Data Model II. Then under Condition* 4.2 *the statistics T_{ML} and T_{SB} are asymptotically robust, T_B is asymptotically robust only if the second term on the RHS of* (4.3) *is zero, T_K is asymptotically robust only if the second term on the RHS of* (4.4) *is zero.*

Theorem 4.2 is on the validity of the four test statistics for a general covariance structure. When variables are uncorrelated, the fourth-order moment matrix Γ is given by (3.8). Correspondingly, (4.3) and (4.4) can be much simplified. When $\Sigma = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{pp})$, $\sigma_i = \sigma_{ii} e_i$. It follows from (4.3) that

$$\eta = 1 + \frac{3}{p(p+2)} \sum_{i=1}^{p} (\beta_i - 1). \tag{4.6}$$

Using $\sigma_i = \sigma_{ii} e_i$ and (4.4),

$$\nu = 1 + \frac{3}{m_1 m_2} \sum_{i=1}^{p} (\beta_i - 1) \sigma_{ii}^2 P_{1ii} Q_{1ii}. \tag{4.7}$$

Obviously, $\eta = \nu = 1$ when $\beta_i = 1$. This corresponds to a pseudo normal distribution. We have the following corollary.

Corollary 4.1. *Let $\Sigma = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{pp})$ and suppose a sample is drawn according to Data Model II. Then under Condition 4.2a the statistics T_{ML} and T_{SB} are asymptotically robust, while T_B and T_K are asymptotically robust if the sample is from a pseudo normal distribution.*

In a similar way as the result on T_K in Theorem 4.1, the asymptotic robustness of T_K in Corollary 4.1 is different from that stated by Kano (1992), since the marginal x_i of X may not be independent. When the marginal variables of X are independent, e.g., $G = \mathrm{diag}(\sigma_{11}^{\frac{1}{2}}, \ldots, \sigma_{pp}^{\frac{1}{2}})$ in Data Model II, and (2.7) holds with $\Lambda_i = e_i$, then $\nu = 1$. This is one of the results of Kano (1992). Muirhead and Waternaux (1980) dealt with the uncorrelated variables model when X follows an elliptical distribution. Since even if all the β_i are equal, Data Model II will not generate an elliptical distribution, the result about T_B in Corollary 4.1 is not in conflict with that of Muirhead and Waternaux.

5. Specific Structural Models

In this section, we will consider three popular covariance structure models. One is a confirmatory factor model which is commonly used in psychology, biomedical research, education, and social sciences. The other two models are the classical intra-class model and uncorrelated variables model. Each model can be fitted to a sample generated by either of the data models in Section 3. It will be assumed that the population covariance structure is correctly specified for a given data set.

A confirmatory factor model is given by

$$X = \Lambda f + \varepsilon, \tag{5.1}$$

where Λ is the factor loading matrix, $f = (f_1, \ldots, f_s)'$ is a vector of common factors with $\mathrm{Cov}(f) = \Phi$, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_p)'$ is a vector of unique factors or errors with $\mathrm{Cov}(\varepsilon) = \Psi$. Also, f and ε are uncorrelated. This leads to a covariance structure

$$\Sigma = \Lambda\Phi\Lambda' + \Psi. \tag{5.2}$$

One popular structure for Λ is

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_s \end{pmatrix},$$

where $\lambda_j = (\lambda_{1j}, \ldots, \lambda_{s_j j})'$. That is, each observed variable only depends on one common factor. Φ is an $s \times s$ symmetric matrix, and $\Psi = \mathrm{diag}(\psi_1, \ldots, \psi_p)$ is a diagonal matrix. In order for model (5.2) to be identifiable, it is necessary to fix the scale of each factor f_j. This can be obtained by fixing the last element $\lambda_{s_j j}$ in each λ_j at 1.0. Under these conditions, we have

$$\dot{\Sigma}_{\lambda_{ij}} = (0_{p\times(j-1)}, e_k, 0_{p\times(s-j)})\Phi\Lambda' + \Lambda\Phi(0_{p\times(j-1)}, e_k, 0_{p\times(s-j)})' \tag{5.3a}$$

for $i = 1, \ldots, s_j - 1$, $j = 1, \ldots, s$, where e_k is a p-dimensional unit vector with $k = \sum_{l=1}^{j-1} s_l + i$, $0_{p\times(j-1)}$ is a matrix of 0's;

$$\dot{\Sigma}_{\phi_{ii}} = \Lambda e_i e_i' \Lambda', \quad \dot{\Sigma}_{\phi_{ij}} = \Lambda(e_i e_j' + e_j e_i')\Lambda', \tag{5.3b}$$

for $i, j = 1, \ldots, s$, where e_i and e_j are of s-dimension; and

$$\dot{\Sigma}_{\psi_i} = e_i e_i', \ i = 1, \ldots, p, \tag{5.3c}$$

where e_i is of p-dimension.

The intra-class model is given by

$$\Sigma = \theta_1 I_p + \theta_2 1_p 1_p' \quad \text{and} \quad \dot{\Sigma}_{\theta_1} = I_p, \ \ \dot{\Sigma}_{\theta_2} = 1_p 1_p'. \tag{5.4}$$

The uncorrelated variables model is

$$\Sigma = \mathrm{diag}(\theta_1, \ldots, \theta_p) \quad \text{and} \quad \dot{\Sigma}_{\theta_i} = E_{ii}, \ i = 1, \ldots, p. \tag{5.5}$$

Consider Data Model I first. Suppose we choose the matrix A as

$$A = (a_1, \ldots, a_s, a_{s+1}, \ldots, a_{s+p}) = (\Lambda\Phi^{\frac{1}{2}}, \Psi^{\frac{1}{2}})$$

in the confirmatory factor model. Then it is obvious that $a_i a_i'$ can be expressed as a linear combination of those in (5.3b) for $i = 1$ to s, and a linear combination of those in (5.3c) for $i = s + 1$ to $s + p$. In such a case, the statistic T_{SB} is asymptotically robust regardless of any departure of the observed variables from normality. The statistics T_B and T_K are asymptotically robust if X follows a pseudo elliptical distribution. T_K is also valid if (2.7) holds with $\Lambda_i = a_i$ even if X does not follow a pseudo elliptical distribution. T_{ML} is asymptotically robust only if $\beta = 1$. If we choose $A = (a_1, \ldots, a_p) = \Sigma^{\frac{1}{2}}$, then the a_i will not satisfy Condition 4.1, and none of the four statistics are generally asymptotically robust in such a case. For example, for $p = 15$ with the covariance structure and the population parameters as given in Hu, Bentler and Kano (1992), if we generate ξ_i in (3.2) from standardized independent lognormal variates and $r = \{3/\chi_5^2\}^{\frac{1}{2}}$, the coefficient of variation (CV) of the 87 nonzero eigenvalues τ_i of $W\Gamma$ is $\mathrm{CV}(\tau) \approx .089$.

For the intra-class model (5.4) with $\theta_2 > 0$, if we choose the $p \times (p+1)$ matrix

$$A = (a_1, \ldots, a_{p+1}) = (\theta_1^{\frac{1}{2}} I_p, \theta_2^{\frac{1}{2}} 1_p),$$

then it is obvious that a_i will not satisfy Condition 4.1. In such a case, none of the statistics will be asymptotically $\chi^2_{p^*-2}$. Similarly, if we choose $A = \Sigma^{\frac{1}{2}}$, then Condition 4.1 will not be satisfied, and none of the four statistics will be asymptotically $\chi^2_{p^*-2}$. For example, using $p = 15$ and $\theta_1 = \theta_2 = .5$, the ξ_i in (3.2) from standardized independent lognormal variates and $r = \{3/\chi_5^2\}^{\frac{1}{2}}$, the CV of the 118 nonzero τ_i is $\mathrm{CV}(\tau) \approx 2.38$.

For the uncorrelated variables model, if we choose $A = (a_1, \ldots, a_p)$ with $a_i = \theta_i^{\frac{1}{2}} e_i$, then Condition 4.1 is obviously met. So the statistic T_{SB} is asymptotically robust. The other three statistics will also be valid if the relevant assumptions in Theorem 4.1 are satisfied. If we choose other forms of A that do not satisfy Condition 4.1, then none of the four statistics is asymptotically robust. For example, using $p = 15$, $\theta_1 = \cdots = \theta_{15} = 1$, a random orthogonal matrix A, and the ξ_i in (3.2) from standardized independent lognormal variates and $r = \{3/\chi_5^2\}^{\frac{1}{2}}$, we obtain a $\mathrm{CV}(\tau) \approx 2.24$ of the 105 nonzero eigenvalues τ_i of the $W\Gamma$ matrix.

For Data Model II, the four test statistics will not be asymptotically robust in general for the confirmatory factor model and the intra-class model. For example let $p = 15$. For the factor model with the covariance structure and the population parameters as given in Hu, Bentler and Kano (1992), and for the intra-class model in (5.4) with $\theta_1 = \theta_2 = .5$, we generate r_i from $r_i = \alpha + \epsilon_i$ with $\alpha = .95$ and ϵ_i is given by

$$\epsilon_i = \sqrt{(1-\alpha^2)}(y_i - E(y_i))/\{\mathrm{var}(y_i)\}^{\frac{1}{2}}.$$

For r_1 to r_{15}, we generate each three of the corresponding y_i, $i = 1, \ldots, 15$, respectively, by independent variates of lognormal(0,1), lognormal(0,1.5), lognormal(0,2), and the gamma distributions $\Gamma(1,1)$, and $\Gamma(.1,1)$. This leads to

$\mathrm{CV}(\tau) \approx .014$ for the factor model and $\mathrm{CV}(\tau) \approx 4.44$ for the intra-class model. As for the uncorrelated variables model (5.5), it is obvious that Condition 4.2a is satisfied. So the statistics T_{ML} and T_{SB} are asymptotically robust for the uncorrelated variables model, T_B and T_K are valid when X follows a pseudo normal distribution. If $G = \mathrm{diag}(\sigma_{11}^{\frac{1}{2}}, \ldots, \sigma_{pp}^{\frac{1}{2}})$ in Data Model II and (2.7) holds with $\Lambda_i = e_i$, then T_K is also asymptotically robust even if X does not follow a pseudo normal distribution.

6. Discussion

As one of the major tools for understanding the relationship among multivariate variables and for dimension reduction, covariance structure analysis has been used extensively in many different areas including econometrics, psychology, education, social sciences, biology and medicine. Since commonly used covariance structures are often based on latent variable models, which are basically hypothetical, test statistics play an important role in judging the quality of a proposed model. The normal theory likelihood ratio statistic is the classical test statistic. It is the default option in almost all software packages in this area (e.g., Bentler (1995), Jöreskog and Sörbom (1993)). Unfortunately, this statistic is asymptotically robust only under very special circumstances. Alternative simple statistics such as T_{SB}, T_B, T_K have existed for some time. But for a general covariance structure model, they were believed to be valid only for data from an elliptical distribution. This research generalizes their application to a much wider class of nonnormal distributions than previously known. This is especially true for T_{SB}, which can be applied to a data set whose distribution is far from elliptical. Various simulation results have been reported on the performance of the statistics T_{ML}, T_B and T_{SB}. The results in this paper characterize various mathematical conditions under which these different statistics are valid. One implication of this study is that marginal skewness of a data set will not have much influence on these test statistics when a sample size is relatively large. Another implication is that efforts need not be made on generating variables with specific marginal skewnesses in future simulation studies of these statistics. Since real data in practice can come from an arbitrary mechanism, the third implication is that caution is still needed when using T_{SB} as a fit index of a proposed model, even though previous simulations support this statistic in various ways. For covariance structures generated by latent variable models, the distribution conditions under which the statistic T_K can be applied are almost as wide as those of T_{SB}, as demonstrated in Section 5. But for a general covariance structure, the distributions for which T_K is applicable is still not as wide as that of T_{SB}. Of course, this paper is only concerned with covariance structure analysis.

In situations under which both mean and covariance structures are of interest, the skewnesses of the observed variables will play a role.

Two classes of nonnormal distributions are proposed for the study of test statistics in covariance structure analysis. These represent much wider classes of nonnormal distributions than the class of elliptical symmetric distributions. For example, by letting $m = p$, $r = 1$, $(z_1, \ldots, z_p)' \sim N_p(0, I)$, and $\xi_1 = |z_1|$, $\xi_j = z_j$, $j = 2, \ldots, p$, then the X defined in (3.2) will follow the skew-normal distribution as defined by Azzalini and Valle (1996). Similarly, with proper restrictions, the X defined in (3.4) can also lead to the skew-normal distribution. Since a random vector from either of the two classes can have various marginal skewnesses and kurtoses, these distributions should match more closely with the distribution of a practical data set. Here we have been concerned with up to fourth-order moments of the distributions, since higher-order results were not needed to study the various scaled test statistics. In some applications, the analytical forms of these distributions may be of special interest; this needs further study. Obviously, these distributions can also be used to study the properties of other multivariate methods, especially those in which skewness and kurtosis are involved, e.g., when a mean structure is also of interest as with repeated measures. It also would be interesting to know if some of the results in Muirhead and Waternaux (1980) and Tyler (1983) still hold for the two classes of nonelliptical distributions. These questions are under further investigation.

Now we give an explanation for some results of published simulation studies. In a simulation study with a factor model as in (5.1) and (5.2), using $A = (\Lambda\Phi^{\frac{1}{2}}, \Psi^{\frac{1}{2}})$ is the most natural choice in generating different nonnormal common factors and errors through Data Model I. From the examples in Section 5, this makes T_{SB} valid as a chi-square statistic, regardless of how large the observed marginal skewness and kurtosis may be. In the previously reported Monte-Carlo studies, the CV(τ) has typically not been reported. Even if CV(τ) is not exactly zero, T_{SB} should still perform well if CV(τ) is a small number. This was the case in a simulation study conducted by Yuan and Bentler (1998). Conceivably, many of the previous simulation results on T_{SB} were based on a zero or small CV(τ), leading to a conclusion of robust behavior of this statistic. The situations under which T_B is applicable are very limited. Yet, it is still possible for T_B to be valid for some skewed nonelliptical data. This gives an explanation of some of the simulation results obtained by Browne (1984).

Acknowledgement

The authors would like to thank three anonymous referees whose constructive comments lead to an improved version of this paper. This research was supported by grants DA01070 and DA00017 of the National Institute on Drug Abuse at the U.S. National Institutes of Health.

Appendix

This appendix contains the proofs of Theorems 3.1 and 3.2. This will be facilitated by several lemmas. Let e_i be the ith column unit vector of dimension m, $E_{ij} = e_i e_j'$, and K_m be the commutation matrix as introduced by Magnus and Neudecker (1979) (see also Schott (1997, p.276)). The following lemma is a generalization of the result (i) in Theorem 4.1 of Magnus and Neudecker (1979).

Lemma A.1. *Let $\xi_1, \ldots, \xi_m$ be independent random variables with $E(\xi_i) = 0$, $E(\xi_i^2) = 1$, $E(\xi_i^4) = \kappa_i$, and $\xi = (\xi_1, \ldots, \xi_m)'$. Then*

$$E\{\mathrm{vec}(\xi\xi')\mathrm{vec}'(\xi\xi')\} = K_m + I_{m^2} + \mathrm{vec}(I_m)\mathrm{vec}'(I_m) + \sum_{i=1}^{m}(\kappa_i - 3)\mathrm{vec}(E_{ii})\mathrm{vec}'(E_{ii}).$$

Proof. Since $\mathrm{vec}(\xi\xi') = \xi \otimes \xi$, we have

$$\mathrm{vec}(\xi\xi')\mathrm{vec}'(\xi\xi') = (\xi\xi') \otimes (\xi\xi') = \sum_{i=1}^{m}\sum_{j=1}^{m} E_{ij} \otimes B_{ij}, \tag{A.1}$$

where $B_{ij} = \xi_i\xi_j(\xi\xi')$ is an $m \times m$ matrix. Direct calculation leads to

$$E(B_{ii}) = I_m + (\kappa_i - 1)E_{ii}, \quad \text{and} \quad E(B_{ij}) = E_{ij} + E_{ji},\ i \neq j. \tag{A.2}$$

It follows from (A.2) that

$$E(B_{ij}) = E_{ij} + E_{ji} + \delta_{ij}\{I_m + \frac{1}{2}(\kappa_i - 3)(E_{ij} + E_{ji})\}, \tag{A.3}$$

where δ_{ij} is the Kronecker number with $\delta_{ij} = 1$ if $i = j$ and zero otherwise. Using (A.1) and (A.3), it follows

$$E\{\mathrm{vec}(\xi\xi')\mathrm{vec}'(\xi\xi')\} = \sum_{i=1}^{m}\sum_{j=1}^{m}\{E_{ij}\otimes E_{ij} + E_{ij}\otimes E_{ji}\} + I_m\otimes I_m + \sum_{i=1}^{m}(\kappa_i - 3)(E_{ii}\otimes E_{ii}). \tag{A.4}$$

Since $E_{ii}\otimes E_{ii} = \mathrm{vec}(E_{ii})\mathrm{vec}'(E_{ii})$, using equation (2.11) of Magnus and Neudecker (1979) and the definition for K_m, the lemma follows from (A.4).

When each ξ_i has no excess kurtosis, all the κ_i equal 3, Lemma 1 is still a little more general than the result in (i) of Theorem 4.1 of Magnus and Neudecker (1979), as here the ξ_i can be skewed instead of normal. From Lemma 1, we obtain

$$\mathrm{Cov}\{\mathrm{vec}(\xi\xi')\} = K_m + I_{m^2} + \sum_{i=1}^{m}(\kappa_i - 3)\mathrm{vec}(e_i e_i')\mathrm{vec}'(e_i e_i'). \tag{A.5}$$

Let $Y = A\xi$, $AA' = \Sigma$, and $A = (a_1, \ldots, a_m)$ with a_i being the ith column vector of A. For the vector Y, we have the following lemma.

Lemma A.2. *Let $\xi = (\xi_1, \ldots, \xi_m)'$ satisfy the condition as in Lemma* A.1. *Then*

$$E\{\mathrm{vech}(YY')\mathrm{vech}'(YY)\} = 2D_p^+(\Sigma \otimes \Sigma)D_p^{+'} + \sigma\sigma' + \sum_{i=1}^m (\kappa_i - 3)\mathrm{vech}(a_i a_i')\mathrm{vech}'(a_i a_i').$$

Proof. Since $YY' = A\xi\xi' A'$ and $\mathrm{vech}(YY') = D_p^+ \mathrm{vec}(YY')$, it follows that

$$\mathrm{vech}(YY') = D_p^+(A \otimes A)\mathrm{vec}(\xi\xi'). \tag{A.6}$$

Using Lemma A.1 and (A.6) we have

$$\begin{aligned} E\{\mathrm{vech}(YY')\mathrm{vech}'(YY')\} = D_p^+(A \otimes A)(K_m + I_{m^2})(A' \otimes A')D_p^{+'} + \sigma\sigma' \\ + \sum_{i=1}^m (\kappa_i - 3)\mathrm{vech}(a_i a_i')\mathrm{vech}'(a_i a_i'). \end{aligned} \tag{A.7}$$

Note that since $D_p^+ = (D_p' D_p)^{-1} D_p'$, it follows from Theorem 9 and Theorem 12 of Magnus and Neudecker (1988, pp.47-49) that

$$D_p^+(A \otimes A)(K_m + I_{m^2})(A' \otimes A')D_p^{+'} = 2D_p^+(\Sigma \otimes \Sigma)D_p^{+'}. \tag{A.8}$$

The lemma follows from (A.7) and (A.8).

Proof of Theorem 3.1. Let $Y = A\xi$ and $X = rY$. Since r and Y are independent, we have

$$\Gamma = E(r^4)E\{\mathrm{vech}(YY')\mathrm{vech}'(YY')\} - \sigma\sigma'. \tag{A.9}$$

The theorem follows from (A.9) and Lemma A.2.

The next lemma will be used to prove Theorem 3.2. In the following, e_i will be the ith unit vector of dimension p and $E_{ij} = e_i e_j'$.

Lemma A.3. *Let $X = RA\xi$ be generated through* (3.4)*, and R, A and ξ satisfy the conditions in Data Model II. Then*

$$\begin{aligned} E\{\mathrm{vech}(XX')\mathrm{vech}'(XX')\} = 2D_p^+ E\{(RGR) \otimes (RGR)\}D_p^{+'} \\ + E\{\mathrm{vech}(RGR)\mathrm{vech}'(RGR)\}. \end{aligned} \tag{A.10}$$

Proof. Since $\mathrm{vech}(XX') = D_p^+\{(RA) \otimes (RA)\}\mathrm{vec}(\xi\xi')$ and $\kappa_i = 3$, the lemma follows from using conditional expectation given R, and the results of Theorem 9 and Theorem 12 of Magnus and Neudecker (1988, pp.47-49) together with Lemma A.1.

Proof of Theorem 3.2. Since $E\{\mathrm{vech}(XX')\} = \sigma$ and

$$\Gamma = E\{\mathrm{vech}(XX')\mathrm{vech}'(XX')\} - E\{\mathrm{vech}(XX')\}E\{\mathrm{vech}'(XX')\}, \qquad \text{(A.11)}$$

we only need to calculate the first term on the RHS of (A.11). According to (A.10), we need to calculate $\Pi_1 = E\{(RGR)\otimes(RGR)\}$ and $\Pi_2 = E\{\mathrm{vech}(RGR)$ $\mathrm{vech}'(RGR)\}$. First we work on Π_1. Noting that $RGR = (g_{ij}r_ir_j)$, we have

$$(RGR)\otimes(RGR) = \sum_{i=1}^{p}\sum_{j=1}^{p} g_{ij}E_{ij}\otimes\{r_ir_j(RGR)\}. \qquad \text{(A.12)}$$

Let

$$B_{ij} = r_ir_j(RGR) = G\odot H_{ij}, \qquad \text{(A.13)}$$

where $\odot$ is the notation of a Hadamard product (e.g., Magnus and Neudecker (1988, p.45)), and $H_{ij} = r_ir_j\rho\rho'$ with $\rho = (r_1,\ldots,r_p)'$. Then direct but tedious calculation leads to

$$E(H_{ii}) = \alpha^2 1_p1_p' + (\alpha\gamma_i - \alpha^2)(1_pe_i' + e_i1_p') + (1-\alpha^2)I_p + (\beta_i + 2\alpha^2 - 2\alpha\gamma_i - 1)E_{ii} \qquad \text{(A.14)}$$

and

$$E(H_{ij}) = (\alpha^2 - \alpha^4)(1_pe_i' + 1_pe_j' + e_i1_p' + e_j1_p' + I_p) + (\alpha\gamma_i + 2\alpha^4 - 3\alpha^2)E_{ii} \\ +\alpha^4 1_p1_p' + (\alpha\gamma_j + 2\alpha^4 - 3\alpha^2)E_{jj} + (1 - 2\alpha^2 + \alpha^4)(E_{ij} + E_{ji}). \qquad \text{(A.15)}$$

Let σ_i be the ith column vector of $\Sigma = (\sigma_1,\ldots,\sigma_p)$. Using $G = \alpha^{-2}\Sigma + (1 - \alpha^{-2})D_\Sigma$ and the properties of Hadamard products, it follows from (A.13) and (A.14) that

$$E(B_{ii}) = \Sigma + (\alpha^{-1}\gamma_i - 1)(\sigma_ie_i' + e_i\sigma_i') + (\beta_i - 2\alpha^{-1}\gamma_i + 1)\sigma_{ii}E_{ii}, \qquad \text{(A.16)}$$

and from (A.13) and (A.15) for $i \neq j$, that

$$E(B_{ij}) = \alpha^2\Sigma + (1-\alpha^2)(\sigma_ie_i' + e_i\sigma_i' + \sigma_je_j' + e_j\sigma_j') + (\alpha\gamma_i - 2 + \alpha^2)\sigma_{ii}E_{ii} \\ +(\alpha\gamma_j - 2 + \alpha^2)\sigma_{jj}E_{jj} + (\alpha^{-2} - 2 + \alpha^2)(\sigma_{ij}E_{ij} + \sigma_{ji}E_{ji}). \qquad \text{(A.17)}$$

Noting that $g_{ii} = \sigma_{ii}$ and $g_{ij} = \alpha^{-2}\sigma_{ij}$ for $i \neq j$, it follows from (A.12), (A.13), (A.16) and (A.17) that

$$\Pi_1 = \Sigma\otimes\Sigma + (\alpha^{-2} - 1)\sum_{i=1}^{p}\{\mathrm{vec}(E_{ii})\mathrm{vec}'(\sigma_i\sigma_i') + \mathrm{vec}(\sigma_i\sigma_i')\mathrm{vec}'(E_{ii}) \\ +\mathrm{vec}(e_i\sigma_i')\mathrm{vec}'(\sigma_ie_i') + \mathrm{vec}(\sigma_ie_i')\mathrm{vec}'(e_i\sigma_i')\} \\ +\sum_{i=1}^{p}(\alpha^{-1}\gamma_i - 2\alpha^{-2} + 1)\sigma_{ii}\{\mathrm{vec}(E_{ii})\mathrm{vec}'(e_i\sigma_i') + \mathrm{vec}(E_{ii})\mathrm{vec}'(\sigma_ie_i')$$

$$+\mathrm{vec}(\sigma_i e_i')\mathrm{vec}'(E_{ii})+\mathrm{vec}(e_i\sigma_i')\mathrm{vec}'(E_{ii})\}$$
$$+\sum_{i=1}^{p}(\beta_i-4\alpha^{-1}\gamma_i+8\alpha^{-2}-2\alpha^{-4}-3)\sigma_{ii}^2\mathrm{vec}(E_{ii})\mathrm{vec}'(E_{ii})$$
$$+(\alpha^{-4}-2\alpha^{-2}+1)\sum_{ij}\sigma_{ij}^2\{\mathrm{vec}(E_{ii})\mathrm{vec}'(E_{jj})+\mathrm{vec}(E_{ji})\mathrm{vec}'(E_{ij})\}\qquad\text{(A.18)}$$

Now we turn to Π_2. Since

$$\mathrm{vech}(RGR)\mathrm{vech}'(RGR)=\sum_{i=1}^{p}\sum_{j=1}^{p}g_{ij}\mathrm{vech}(E_{ij})\mathrm{vech}'(B_{ij}),$$

it follows from (A.16) and (A.17) that

$$\Pi_2=\mathrm{vech}(\Sigma)\mathrm{vech}'(\Sigma)+(\alpha^{-2}-1)\sum_{i=1}^{p}\{\mathrm{vech}(e_i\sigma_i')\mathrm{vech}'(e_i\sigma_i')$$
$$+\mathrm{vech}(e_i\sigma_i')\mathrm{vech}'(\sigma_i e_i')+\mathrm{vech}(\sigma_i e_i')\mathrm{vech}'(e_i\sigma_i')+\mathrm{vech}(\sigma_i e_i')\mathrm{vech}'(\sigma_i e_i')\}$$
$$+\sum_{i=1}^{p}(\alpha^{-1}\gamma_i-2\alpha^{-2}+1)\sigma_{ii}\{\mathrm{vech}(E_{ii})\mathrm{vech}'(e_i\sigma_i')+\mathrm{vech}(E_{ii})\mathrm{vech}'(\sigma_i e_i')$$
$$+\mathrm{vech}(\sigma_i e_i')\mathrm{vech}'(E_{ii})+\mathrm{vech}(e_i\sigma_i')\mathrm{vech}'(E_{ii})\}$$
$$+\sum_{i=1}^{p}(\beta_i-4\alpha^{-1}\gamma_i+8\alpha^{-2}-2\alpha^{-4}-3)\sigma_{ii}^2\mathrm{vech}(E_{ii})\mathrm{vech}'(E_{ii})$$
$$+(\alpha^{-4}-2\alpha^{-2}+1)\sum_{ij}\sigma_{ij}^2\{\mathrm{vech}(E_{ij})\mathrm{vech}'(E_{ij})+\mathrm{vech}(E_{ij})\mathrm{vech}'(E_{ji})\}.$$
$$\text{(A.19)}$$

Theorem 3.2 follows from Lemma A.3, (A.11), (A.18) and (A.19).

References

Amemiya, Y. and Anderson, T. W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *Ann. Statist.* **18**, 1453-1463.

Anderson, T. W. (1989). Linear latent variable models and covariance structures. *J. Econometrics* **41**, 91-119.

Anderson, T. W. and Fang, K.-T. (1987). Cochran's theorem for elliptically contoured distributions. *Sankhyā, Ser. A* **49**, 305-315.

Azzalini, A. and Valle, A. D. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715-726.

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

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

Browne, M. W. (1984). Asymptotic distribution-free methods for the analysis of covariance structures. *British J. Math. Statist. Psych.* **37**, 62-83.

Browne, M. W. and Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British J. Math. Statist. Psych.* **41**, 193-208.

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

Fang, K.-T., Kotz, S. and Ng., K. W. (1990). *Symmetric Multivariate and Related Distributions.* Chapman and Hall, London.

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

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

Kano, Y. (1992). Robust statistics for test-of-independence and related structural models. *Statist. Probab. Letters* **15**, 21-26

Kano, Y., Berkane, M. and Bentler, P. M. (1990). Covariance structure analysis with heterogeneous kurtosis parameters. *Biometrika* **77**, 575-585.

Magnus, J. R. and Neudecker, H. (1979). The commutation matrix: Some properties and applications. *Ann. Statist.* **7**, 381-394.

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

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

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

Muirhead, R. J. and Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika* **67**, 31-43.

Olkin, I. (1994). Multivariate non-normal distributions and models of dependency. In *Multivariate Analysis and Its Applications* (Edited by T. W. Anderson, K. T. Fang and I. Olkin), 37-53. Institute of Mathematical Statistics, Hayward, CA.

Satorra, A. and Bentler, P. M. (1986). Some robustness properties of goodness of fit statistics in covariance structure analysis. *American Statistical Association 1986 Proceedings of Business and Economics Sections*, 549-554. American Statistical Association.

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

Satorra, A. and Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Comput. Statist. Data Anal.* **10**, 235-249.

Satorra, A. and Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In *Latent Variables Analysis: Applications for Developmental Research* (Edited by A. von Eye and C. C. Clogg), 399-419. Sage, Thousand Oaks, CA.

Schott, J. (1997). *Matrix Analysis for Statistics.* Wiley, New York.

Shapiro, A. and Browne, M. (1987). Analysis of covariance structures under elliptical distributions. *J. Amer. Statist. Assoc.* **82**, 1092-1097.

Steyn, H. S. (1993). On the problem of more than one kurtosis parameter in multivariate analysis. *J. Mult. Anal.* **44**, 1-22.

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

Yuan, K.-H. and Bentler, P. M. (1997a). Mean and covariance structure analysis: Theoretical and practical improvements. *J. Amer. Statist. Assoc.* **92**, 767-774.

Yuan, K.-H. and Bentler, P. M. (1997b). Generating multivariate distributions with specified marginal skewness and kurtosis. In *SoftStat' 97–Advances in Statistical Software 6* (Edited by W. Bandilla and F. Faulbaum), 385-391. Lucius and Lucius, Stuttgart, Germany.

Yuan, K.-H. and Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British J. Math. Statist. Psych.* **51**, 289-309.

Department of Psychology, University of North Texas, Denton, Texas 76203-1280, U.S.A.

E-mail: kyuan@unt.edu

Departments of Psychology and Statistics, University of California, Box 951563, Los Angeles, CA 90095-1563, U.S.A.

E-mail: bentler@ucla.edu

(Received March 1997; accepted September 1998)