Testing Structural Equation Models: The Effect of Kurtosis

Tron Foss, Karl G Jöreskog & Ulf H Olsson
Norwegian School of Management

October 18, 2006

Abstract

Various chi-square statistics are used for testing structural equation models. A commonly used chi-square is the Satorra-Bentler (SB) statistic. This is a normal-theory chi-square corrected for non-normality in the observed variables. The relationship between the SB statistic and kurtosis is developed and it is shown that the SB statistic tends to decrease with increasing kurtosis if the model does not hold. It is shown that the population value of the correction factor for the SB statistic decreases toward zero with increasing kurtosis.

Various chi-square statistics are used for testing structural equation models. If the model is fitted by the maximum likelihood (ML) method, one such chi-square statistic is obtained as n times the minimum of the ML fit function, where n is the sample size. An asymptotically equivalent chi-square statistic is n times the minimum value of a quadratic form using a weight matrix evaluated under multivariate normality of the observed variables. Following the notation in Jöreskog, Sörbom, Du Toit, & Du Toit (2003), these chi-square statistics are denoted c_1 and c_2, respectively. They are valid under multivariate normality of the observed variables and if the model holds.

If the observed variables are non-normal, Satorra & Bentler (1988) proposed another chi-square statistic c_3 (often called the SB statistic) which is c_1 or c_2 multiplied by a scale factor which is estimated from the sample and involves an estimate of the asymptotic covariance matrix (ACM) of the sample variances and covariances. Satorra & Bentler (1988) show that under multivariate non-normality, c_1 and c_2 have asymptotically a distribution which is a linear combination of chi-squares with one degree of freedom. The scale factor is estimated such that c_3 has an asymtotically correct mean even though it does not have an asymptotic chi-square distribution. In practice, c_3 is conceived of as a way of correcting c_1 or c_2 for the effects of non-normality and c_3 is often used as it performs better than the ADF test proposed by Browne (1984), particularly if n is not very large, see e.g., Hu, Bentler, & Kano (1992).

In this paper we develop the relationship between c_3 and kurtosis and we show that the c_3 statistic tends to decrease with increasing kurtosis. The practical consequence of this is that models that *do not hold* tend to be accepted by the chi-square test if kurtosis is large. Thus, this test have low power for detecting misspecified models. Although the results developed here can be demonstrated by simulating and analyzing random samples, we will use a different approach.

Curran, West and Finch (1996) presented a simulation study of these chi-square statitics and concluded "The most surprising findings are related to the behavior of the SB and ADF test statistics under simultanous conditions of misspecification and multivariate nonnormality (Models 3 and 4). The expected values of these test statistics markedly decreased with increasing nonnormality. That is, all else being equal, the SB and ADF test statistics were less likely to dectect a specification error given increasing departures from a multivariate normal distribution. The more severe the nonnormality, the greater the corresponding loss of power. This result was unexpected, and we are not aware of any previous discussions of this finding" (Curran, West and Finch, 1996, p.25). This paper provides an explanation for their results.

The Asymptotic Covariance Matrix under Non-Normality

Browne & Shapiro (1988) considered the following general structure for an observable $k \times 1$ random vector z:

$$z = \mu + \sum_{i=i}^{g} A_i v_i, \tag{1}$$

where μ is a constant vector, A_i is a constant $k \times m_i$ matrix and the v_i are independent $m_i \times 1$ vector variates for $i = 1, 2, \ldots, g$.

Let $\mathbf{S}$ be a sample covariance matrix estimated from a random sample of n observations of z, and let $\mathbf{s} = (s_{11}, s_{21}, s_{22}, \ldots, s_{kk})'$ be a vector of order $\frac{1}{2}k(k+1) \times 1$ of the non-duplicated elements of $\mathbf{S}$. Let $k^\star = \frac{1}{2}k(k+1)$. We assume that $\mathbf{S}$ converge in probability to Σ_0 as $n \to \infty$.

It follows from the multivariate central limit theorem that

$$n^{\frac{1}{2}}(\mathbf{s} - \sigma_0) \xrightarrow{d} N(0, \Omega) \ , \tag{2}$$

where $\xrightarrow{d}$ denotes convergence in distribution.

Browne & Shapiro (1988, Equation 2.7) give Ω as

$$\Omega = K' \left\{ 2(\Sigma_0 \otimes \Sigma_0) + \sum_{i=1}^{g} (A_i \otimes A_i) C_i (A_i' \otimes A_i') \right\} K \tag{3}$$

where K is the matrix K_k of order $k^2 \times k^\star$ defined in Browne (1974, Section 2) or in Browne (1984, Section 4), and $\otimes$ denotes the Kronecker product. The matrix C_i is the fourth order cumulant matrix of v_i, $i = 1, 2, \ldots, g$.

For our purpose it is sufficient to consider a special case of (1), namely when each v_i is a scalar random variable. Then A_i is a column vector a_i and (1) can be written

$$z = Av \ , \tag{4}$$

where A is a matrix of order $k \times g$ and v is a $g \times 1$ vector of independent random variables having moments up to order four. For convenience, we assume that v_i has mean 0 and variance 1.

The rationale for (4) is as follows (see also Olsson, Foss & Troye, 2003 and Mattson, 1997). Suppose we are given a structural equation model (SEM) with specified parameter matrices, for example, the factor analysis model

$$z = \Lambda \xi + \delta \ , \tag{5}$$

where $\xi(q \times 1)$ and $\delta(k \times 1)$ are independent with covariance matrices Φ and Θ_δ respectively, both assumed to be positive definite. Since Φ is positive definite there exist a $q \times q$ matrix T_1 such that $T_1 T_1' = \Phi$. Then we can write $\xi = T_1 v_1$, where we assume that the elements of $v_1(q \times 1)$ are independent. In the same manner we can write $\delta = T_2 v_2$, where the elements of v_2 of order $k \times 1$ are independent and $T_2(k \times k)$ is a non-singular matrix such that $T_2 T_2' = \Theta_\delta$. Assuming that v_1 and v_2 are independent it follows that ξ and δ are independent and that

$$z = \Lambda \xi + \delta \ = (\Lambda T_1 \ T_2) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = Av \ , \tag{6}$$

as in (4). The Appendix shows that more general SEM models can also be constructed from a set of independent random variables v.

Note that v is not a vector of latent variables (not a ξ-vector) but rather a vector of random drawings from a given distribution. We have assumed the v's are independent. But this does not imply that the ξ's ($\xi = T_1 v_1$) are independent; they will in general be correlated. But it does imply that ξ and δ are independent vectors as required.

The mean vector of z is 0 and the covariance matrix is

$$\Sigma_0 = AA' \ . \tag{7}$$

Let $\mu_{4i} = E(v_i^4)$. The matrix C_i in (3) is the 1×1 matrix $\gamma_{2i} = \mu_{4i} - 3$, the fourth order cumulant or excess kurtosis of v_i. Then (3) can be written in the following form

$$\Omega = K' \left\{ 2(\Sigma \otimes \Sigma) + \sum_{i=1}^{g} (a_i \otimes a_i)(a_i \otimes a_i)' \gamma_{2i} \right\} K \ . \tag{8}$$

Let $\mathbf{G} = [(a_1 \otimes a_1)(a_2 \otimes a_2) \ldots (a_g \otimes a_g)]$ and let $\mathbf{M} = diag(\gamma_{21}, \gamma_{22}, \ldots, \gamma_{2g})$. $\mathbf{G}$ is of order $k^2 \times g$ and $\mathbf{M}$ is of order $g \times g$. Then

$$\Omega = \mathbf{K}'[2(\Sigma_0 \otimes \Sigma_0) + \mathbf{GMG}']\mathbf{K} \; . \tag{9}$$

If v_i is normally distributed, then $\mu_{4i} = 3$ and $\gamma_{2i} = 0$. Then the corresponding diagonal element of $\mathbf{M}$ is zero and the corresponding column of $\mathbf{G}$ can be eliminated. If v_i is normally distributed *for all* i, then v has a multivariate normal distribution and $\mathbf{M} = 0$ so that the second term in (9) vanishes. It is convenient to use the notation Ω_{nnt} for the matrix in (9) and the notation Ω_{nt} for the matrix $\mathbf{K}'[2(\Sigma_0 \otimes \Sigma_0)]\mathbf{K}$. Thus, from (9) it follows that

$$\Omega_{\text{nnt}} = \Omega_{\text{nt}} + \mathbf{K}'\mathbf{GMG}'\mathbf{K} \; . \tag{10}$$

A special case of (9) is when all elements of v have the same kurtosis so that $\gamma_{2i} = \gamma_2$, say, which is the same for all i. Then $\mathbf{M} = \gamma_2 \mathbf{I}$. Note that we made no assumption about third order moments of the elements of v. The assumption of an elliptic distribution of z with homogeneous kurtosis parameters was considered by several authors, see, e.g., Browne (1984), and Browne & Shapiro (1987). A general case with heterogenous kurtosis parameters is considered by Kano, Bercane & Bentler (1990).

The Effect of Kurtosis

Consider a general model $\Sigma(\theta)$, where θ is a parameter vector of order t, and a general fit function of the form $F[\mathbf{s}, \sigma(\theta)]$ satisfying the conditions of Browne (1984) and Satorra (1989). For example, F can be

$$F[\mathbf{s}, \sigma(\theta)] = [\mathbf{s} - \sigma(\theta)]'\mathbf{V}[\mathbf{s} - \sigma(\theta)] \tag{11}$$

where $\mathbf{V}$ is either a fixed positive definite matrix or a random matrix converging in probability to a positive definite matrix $\bar{\mathbf{V}}$. This covers all fit functions available in computer softwares for estimating structural equation models. F is to be minimized with respect to the model parameters θ. Let $\hat{\theta}$ be the minimizer of $F[\mathbf{s}, \sigma(\theta)]$ and let θ_0 be a unique minimizer of $F[\sigma_0, \sigma(\theta)]$. We assume that the model *does not hold* so that $F[\sigma_0, \sigma(\theta_0)] > 0$. Browne & Shapiro (1988, equation 3.9, see also Satorra (1989, p. 135) considered a different case, the model of parametric drift, where $\sigma_0 = \sigma(\theta_0) + n^{-1/2}\mu$, where μ is a constant vector. This implies that the model holds in the population so that $F[\sigma_0, \sigma(\theta_0)] = 0$, contrary to our assumption.

Using the notation in Jöreskog & Sörbom (1999), the chi-square statistics c_2 and c_3 discussed initially are defined as

$$c_2 = n(\mathbf{s} - \hat{\sigma})'\Delta_c(\Delta_c'\mathbf{W}_{\text{nt}}\Delta_c)^{-1}\Delta_c'(\mathbf{s} - \hat{\sigma}) \tag{12}$$

$$c_3 = \frac{d}{h} c_2 \; , \tag{13}$$

where d is the degrees of freedom and

$$h = tr[(\Delta_c'\mathbf{W}_{\text{nt}}\Delta_c)^{-1}(\Delta_c'\mathbf{W}_{\text{nnt}}\Delta_c)] \; . \tag{14}$$

Here $\hat{\sigma} = \sigma(\hat{\theta})$, $\mathbf{W}_{\text{nt}}$ and $\mathbf{W}_{\text{nnt}}$ are consistent estimates of Ω_{nt} and Ω_{nnt}, respectively, and Δ_c is an orthogonal complement to the matrix $\Delta = \partial\sigma/\partial\theta$ evaluated at $\hat{\theta}$. The matrix Δ is of order $k^\star \times t$ with $t < k^\star$ and the matrix Δ_c is of order $k^\star \times d$, where $d = k^\star - t$ is the degrees of freedom of the model.

Since we assume that the model *does not hold*, the matrices $\mathbf{W}_{\mathrm{nt}}$ and $\mathbf{W}_{\mathrm{nnt}}$ are estimated without regard to the model. For example, take the elements of $\mathbf{W}_{\mathrm{nt}}$ and $\mathbf{W}_{\mathrm{nnt}}$ as

$$w_{ghij}^{\mathrm{nt}} = s_{gi}s_{hj} + s_{gj}s_{hi} \tag{15}$$

$$w_{ghij}^{\mathrm{nnt}} = m_{ghij} - s_{gh}s_{ij} \ , \tag{16}$$

where

$$m_{ghij} = (1/n)\sum_{a=1}^{n}(z_{ag}-\bar{z}_g)(z_{ah}-\bar{z}_h)(z_{ai}-\bar{z}_i)(z_{aj}-\bar{z}_j) \ . \tag{17}$$

Most computer programs for structural equation modeling assume that the model holds and therefore use $\hat{\sigma}_{ij} = \sigma_{ij}(\hat{\theta})$ instead of s_{ij} in (15).

The correction factor in (13) is d/h, where d is a constant and h is a random variable converging in probability to H, say. To obtain H we replace $\mathbf{s}$ by σ_0, $\hat{\sigma}$ by $\sigma(\theta_0)$, and Δ_c by Δ_{0c}, where Δ_{0c} is evaluated at θ_0. Furthermore, $\mathbf{W}_{\mathrm{nt}}$ and $\mathbf{W}_{\mathrm{nnt}}$ are replaced by Ω_{nt} and Ω_{nnt}. We assume that Δ_{0c} has rank d. Then

$$H = tr[(\Delta_{0c}'\Omega_{\mathrm{nt}}\Delta_{0c})^{-1}(\Delta_{0c}'\Omega_{\mathrm{nnt}}\Delta_{0c})] \ . \tag{18}$$

The influence of kurtosis on H is only via the diagonal matrix $\mathbf{M}$. All other matrices in (10) are independent of kurtosis. From (10) we have

$$\Delta_{0c}'\Omega_{\mathrm{nnt}}\Delta_{0c} = \Delta_{0c}'\Omega_{\mathrm{nt}}\Delta_{0c} + \Delta_{0c}'\mathbf{K}'\mathbf{G}\mathbf{M}\mathbf{G}'\mathbf{K}\Delta_{0c} \ . \tag{19}$$

Hence,

$$(\Delta_{0c}'\Omega_{\mathrm{nt}}\Delta_{0c})^{-1}(\Delta_{0c}'\Omega_{\mathrm{nnt}}\Delta_{0c}) = \mathbf{I}_d + (\Delta_{0c}'\Omega_{\mathrm{nt}}\Delta_{0c})^{-1}\mathbf{P}\mathbf{M}\mathbf{P}' \ , \tag{20}$$

where $\mathbf{I}_d$ is the identity matrix of order d and

$$\mathbf{P} = \Delta_{0c}'\mathbf{K}'\mathbf{G} \ . \tag{21}$$

Taking the trace of (20), gives

$$H = d + tr(\mathbf{QM}) \ , \tag{22}$$

where

$$\mathbf{Q} = \mathbf{P}'(\Delta_{0c}'\Omega_{\mathrm{nt}}\Delta_{0c})^{-1}\mathbf{P} \ . \tag{23}$$

$\mathbf{Q}$ is symmetric and of order $g \times g$. Since $\mathbf{M}$ is diagonal,

$$H = d + \sum_{i}^{g} q_{ii}\gamma_{2i} \ . \tag{24}$$

$\mathbf{Q}$ is positive semidefinite and if $\mathbf{Q} \neq 0$, $q_{ii} > 0$ for at least one i. Thus, if $\gamma_{2i} \to \infty$ for all i, it follows that $H \to \infty$. If $\gamma_{2i} = \gamma_2$ for *all* i, then

$$H = d + (tr\mathbf{Q})\gamma_2 \ . \tag{25}$$

increases linearly with γ_2. Thus, the correction factor d/h tends to be small in large samples if kurtosis is large.

References

Browne, M.W. (1974) Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, **8**, 1–24.

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

Browne, M.W. & Shapiro, A. (1987) Adjustments for kurtosis in factor analysis with elliptically distributed errors. *Journal of the Royal Statistical Society. Series B (Methodological)*, **49**, 346–352.

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

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

Jöreskog, K.G. & Sörbom, D. (1999) LISREL 8*: User's Reference Guide*. Second Edition. Lincolnwood, IL: Scientific Software International.

Jöreskog, K.G., Sörbom, D., Du Toit, S., & Du Toit, M. (2003) LISREL 8*: New Statistical Features*. Third printing with revisions. Lincolnwood, IL: Scientific Software International.

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

Mattsson, S. (1997) How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research*, **32**, 355–373.

Olsson, U.H., Foss, T., & Troye, S.V. (2003) Does the ADF fit function decrease when the kurtosis increases. *British Journal of Mathematical and Statistical Psychology*, **56**, 289–303.

Satorra, A. (1989) Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, **54**, 131–151.

Satorra, A., & Bentler, P.M. (1988) Scaling corrections for chi-square statistics in covariance structure analysis. Proceedings of the Business and Economic Statistics Section of the American Statistical Association, 308–313.

Appendix

In a general SEM model we need to construct independent random vectors $\xi(n \times 1)$, $\varsigma(m \times 1)$, $\varepsilon(p \times 1)$, and $\delta(q \times 1)$ with covariance matrices $\Phi, \Psi, \Theta_\varepsilon, \Theta_\delta$, respectively, satisfying

$$\eta = B\eta + \Gamma\xi + \varsigma \tag{26}$$

$$y = \Lambda_y \eta + \varepsilon \tag{27}$$

$$x = \Lambda_x \xi + \delta \tag{28}$$

where $B, \Gamma, \Lambda_y, \Lambda_x$ are coefficient matrices with $I - B$ non-singular. For further explication of this model, see Jöreskog & Sörbom (1999).

Solving (29) for η gives

$$\eta = (I - B)^{-1}\Gamma\xi + (I - B)^{-1}\varsigma, \tag{29}$$

and substituting this into (30) gives

$$y = \Lambda_y(I - B)^{-1}\Gamma\xi + \Lambda_y(I - B)^{-1}\varsigma + \varepsilon, \tag{30}$$

Combining (33) and (31) gives

$$\begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} \Lambda_y(I - B)^{-1}\Gamma & \Lambda_y(I - B)^{-1} & I & 0 \\ \Lambda_x & 0 & 0 & I \end{pmatrix} \begin{pmatrix} \xi \\ \varsigma \\ \varepsilon \\ \delta \end{pmatrix}$$

Take $k = p{+}q$ and $g = n{+}m{+}p + q$ in (4) and $v' = (v_1', v_2', v_3', v_4')$,where the subvectors are of orders n, m, p and q, respectively. Choose square non-singular matrices T_1, T_2, T_3 *and* T_4, such that $\Phi = T_1T_1',\ \Psi = T_2T_2',\ \Theta_\varepsilon = T_3T_3'$ *and* $\Theta_\delta = T_4T_4'$ and choose $\xi = T_1v_1,\ \varsigma = T_2v_2,\ \varepsilon = T_3v_3$ *and* $\delta = T_4v_4$. Then

$$\begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} \Lambda_y(I - B)^{-1}\Gamma T_1 & \Lambda_y(I - B)^{-1}T_2 & T_3 & 0 \\ \Lambda_x T_1 & 0 & 0 & T_4 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix}$$

which is of the form (4).