Testing structural equation models: the effect of kurtosis

Tron Foss
BI Norwegian Business School

Karl G. Jøreskog
BI Norwegian Business School

Ulf H. Olsson
BI Norwegian Business School

This is the authors' final, accepted and refereed manuscript to the article published in

Computational Statistics and Data Analysis, **55(2011)7: 2263-2275**

DOI: http://dx.doi.org/ 10.1016/j.csda.2011.01.012

Testing Structural Equation Models: The Effect of Kurtosis

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

January 24, 2011

*E-mail address: ulf.h.olsson@bi.no

Abstract

The Satorra Bentler (SB) and the Browne ADF chi-square statistics are used for testing structural equation models with non-normal data. The relationships between the SB and ADF statistics and kurtosis are developed and it is shown that the weighted deviations of the "population" true second-order moments and the fitted second-order moments for these statistics tend to decrease with increasing kurtosis if the model does not hold. The results predict that high kurtosis can lead to loss of power. The results are obtained without simulation.

Keywords

Kurtosis, Scaling correction, ADF,mis-specified

1 Introduction

Structural equation modeling is widely used for studying relationships between observed and unobserved (latent) variables, particularly in the social and behavioral sciences, see *e.g.*, Hershberger (2003).

Various test statistics are used for testing structural equation models. One such test statistic is obtained as n times the minimum of the log-likelihood fit function under multivariate normality, where $N = n + 1$ is the sample size, see e.g.,Jöreskog (1969). Another test statistic is n times the minimum of the generalized least squares (GLS) fit function, see Jöreskog & Goldberger (1972) and Browne (1974).These test statistics are here denoted c_1 and c_2, respectively.

If the model holds and the observed variables have a multivariate normal distribution both c_1 and c_2 have an approximately χ^2_d distribution (d is the degrees of freedom) when n is large.

If the observed variables are non-normal, Satorra & Bentler (1988) proposed another test statistic c_3 (often called the SB rescaled statistic) which is c_1 or c_2 multiplied by a scale factor, often called the Satorra-Bentler scaling correction, which is estimated from the sample and involves an estimate of the asymptotic covariance matrix (ACM) of the sample variances and covariances.

Although the asymptotic distribution of c_3 is not known in general, the asymptotic distribution of c_3 under the null hypothesis, and the χ^2_d distribution agree in mean (Satorra & Bentler (1994)). Still under the null hypothesis and if the distribution of the data is elliptical, Satorra and Bentler (1994, p. 414) conclude "... the scaling correction provides an exact asymptotic chi-square goodness-of-fit statistic." Empirical results suggests that c_3 can also follow a chi -square distribution under certain robustness assumption (See e.g.,Yuan & Bentler, 1998 and Yuan & Bentler, 1999).

The test statistic c_3 is considered as a way of correcting c_1 or c_2 for the effects of non-normality. In fact the Satorra-Bentler correction can be applied to any member of the Swain family (Swain, 1975). See also Satorra (2003, pp.61-62) for a discussion of the application of the scaling correction. In this paper however, c_3 will be the Satorra-Bentler correction applied to c_2 .

Yet another test statistic, the ADF-statistic, here denoted c_4 was proposed by Browne (1984). This is a test statistic valid even under non-normality. Browne (1984) showed that c_4 has an asymptotic χ^2_d distribution under certain standard conditions.

In practice, c_3 is often used as it seems to perform better than c_4 particularly if N is not very large, see e.g., Hu, Bentler, & Kano (1992). Since c_3 and c_4 depends on the ACM and the ACM depends on kurtosis, c_3 and c_4 are affected by kurtosis in the observed variables.

In this paper we develop the relationship between c_3 and c_4 and kurtosis and we show that on average these test statistics tend to decrease with increasing kurtosis. The practical consequence of this is that models that do not hold tend to be accepted by these tests if kurtosis is large. Although the results developed here can be demonstrated by simulating and analyzing random samples, we will use a different approach. Simulation studies depend on rather arbitrary conditions of the design of the simulation and on how random variates are generated. For example, simulation studies depend on specific distributional assumption of the data generating process. By contrast our results are obtained without simulating random variables and they are valid under fairly general conditions.

Curran, West and Finch (1996) presented a simulation study of the SB and ADF test statistics where they concluded: "The most surprising findings are related to the behavior of the SB and ADF test statistics under simultaneous conditions of misspecification and multivariate non-normality (Models 3 and 4). The expected values of these test statistics markedly decreased with increasing non-normality"(Curran, West and Finch, 1996, p.25). Given some assumptions, this paper provides an possible explanation for the seemingly loss of power in such a situation.

2 The Distinction between the The Data Generating Process and The Assumed Model

In this paper we study the behavior of the SB and ADF statistics under the combination of kurtosis and misspecification. To do this we consider the general factor analysis model :

$$\mathbf{x} = \mathbf{\Lambda_x}\boldsymbol{\xi} + \boldsymbol{\delta} \; , \tag{1}$$

where $\mathbf{x}$ is a $k \times 1$ random vector of observables, $\boldsymbol{\xi}(l \times 1)$ and $\boldsymbol{\delta}(k \times 1)$ are uncorrelated random vectors of latent variables with covariance matrices $\mathbf{\Phi}$ and $\mathbf{\Theta}_\delta$, respectively, assumed to be positive definite. The matrix $\mathbf{\Lambda}_x$ is a $(k \times l)$ matrix of unknown factor loadings. We also assume that $E(\boldsymbol{\xi}) = 0$, $E(\boldsymbol{\delta}) = 0$, and $Var(\xi_i) = 1$, $i = 1, 2, ...l$. A method for studying this model is simulation, where one generates "sample data" from a "true confirmatory factor analysis model"

$$\mathbf{x} = \mathbf{\Lambda_x^*}\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{2}$$

where the matrices $\mathbf{\Lambda}^*$, $\mathbf{\Phi}^*$ and $\mathbf{\Theta}_\delta^*$ are fixed at convenient values. The "star" indicate population values. For generating non-normal sample data there are several approaches e.g., see Fleishmann (1978), Vale and Maurelli (1983) and Ramberg et al. (1979) to mention some. Since $\mathbf{\Phi}^*$and $\mathbf{\Theta}_\delta^*$ are positive definite there exists a $(l \times l)$ matrix $\mathbf{T}_1$and a $(k \times k)$ matrix $\mathbf{T}_2$ such that $\mathbf{T}_1\mathbf{T}_1' = \mathbf{\Phi}^*$ and $\mathbf{T}_2\mathbf{T}_2' = \mathbf{\Theta}_\delta^*$. One way to simulate data is to calculate $\mathbf{x} = \mathbf{\Lambda_x^*}\boldsymbol{\xi} + \boldsymbol{\delta}$, where $\boldsymbol{\xi} = \mathbf{T}_1\mathbf{v}_1$ and $\boldsymbol{\delta} = \mathbf{T}_2\mathbf{v}_2$ and the $l \times 1$ vector $\mathbf{v}_1$ and the $k \times 1$ vector $\mathbf{v}_2$ are vectors of independent drawings from a distribution having finite moments up to order four, and with mean vector $\mathbf{0}$ and variance and covariance matrix $\mathbf{I}$. The covariance matrices for $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$ will then be $\mathbf{\Phi}^*$ and $\mathbf{\Theta}_\delta^*$, respectively. The asymptotic covariance matrix (ACM) of the sample variances and covariances will depend on the kurtosis of the elements v_i of the vectors $\mathbf{v}_1$and $\mathbf{v}_2$. The elements v_i can all have different values of kurtosis (see e.g.Mattson 1997). However, we use a different approach than simulation: We *calculate* the asymptotic covariance matrix from the population (true model) instead of generating a large sample and then estimate the asymptotic covariance matrix from this sample. In the following we outline this procedure which is similar to simulation (for very large N). Instead of referring to the term "true model", we will refer to the Data Generating Process (DGP).

Let (2) be the Data Generating Process (DGP). Following the derivations above, we write DGP on a compact form:

Partition $\mathbf{v}' = (\mathbf{v}_1' \; \mathbf{v}_2')$, then DGP can be represented by

$$\mathbf{x} = \mathbf{\Lambda}_x^*\boldsymbol{\xi} + \boldsymbol{\delta} = \begin{pmatrix} \mathbf{\Lambda}_x^*\mathbf{T}_1 & \mathbf{T}_2 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix} = \mathbf{A}\mathbf{v} \; . \tag{3}$$

The elements of $\mathbf{v}$ are independent but this does *not* imply that the elements of $\boldsymbol{\xi} = \mathbf{T}_1\mathbf{v}_1$ are independent. But it does imply that $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$ are independent vectors. In the following we drop the "stars" used to indicate the data generating process when there is no chance of mixing it up with the assumed model (see below). It is only the DGP, where all the parameters are fixed, which is written in the compact form $\mathbf{x} = \mathbf{A}\mathbf{v}$. The assumed model, i.e., the model to be tested where normally there are restrictions on or among the parameters, is not written in the compact form.

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

$$\mathbf{x} = \boldsymbol{\mu} + \sum_{i=1}^{g} \mathbf{A}_i \mathbf{v}_i, \tag{4}$$

where $\boldsymbol{\mu}$ is a constant vector, $\mathbf{A}_i$ is a constant $k \times m_i$ matrix and the $\mathbf{v}_i$ are mutually independent $m_i \times 1$ vector variates for $i = 1, 2, \ldots, g$.

Our DGP is a special case of (4), namely when $\boldsymbol{\mu} = \mathbf{0}$ and each $\mathbf{v}_i$ is a scalar random variable. Then $\mathbf{A}_i$ is a column vector $\mathbf{a}_i$ and (4) can be written

$$\mathbf{x} = \mathbf{A}\mathbf{v} \ , \tag{5}$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots \mathbf{a}_g]$ and $\mathbf{v}$ is a $g \times 1$ (where $g = k + l$) vector of independent random variables having finite moments up to order four. Equation (5) describes the data generating process (DGP), that generates the observables. It is convenient to write DGP as (5) as one can induce non-normality in the x_i-variables by varying the kurtosis of the v_i-variables and calculate the asymptotic covariance matrix as a function of the kurtosis (see equations 10 and 12) .

The assumed model (AM), i.e., the model to be estimated and tested, is different from the DGP. We say that AM holds if AM and DGP are structurally identical i.e., when AM is identically specified as the DGP, but differs from DGP only by the fact that all parameters in DGP are fixed at the "true" values. Otherwise the AM does not hold. In this paper we are interested in the effects of kurtosis on the test statistics that are used for testing the AM. However, instead of analyzing c_3 and c_4 from random samples, we investigate what will happen to $(\frac{c_3}{n})$ and $(\frac{c_4}{n})$ when $n \to \infty$. This is done by studying miss-fit measures of the weighted deviations of the "true" σ_0 and the fitted $\sigma(\theta_0)$ moments , denoted C_3 and C_4 respectively

.

Satorra (1989, 2003) developed a robustness theory for structural equation models where he assumed population drift (see eg., Browne, 1984; Wald, 1943). We do not make this assumption in this paper.

Generally the kurtosis of the observed variables x_i is not identical to the kurtosis of the variables v_i. Even if there is no exact overlap with models given by (5) and the LISREL models, the results derived in this paper should be valid for any structural equation model confined to the class of models given by equation (5) where the elements of $\mathbf{v}$, for the DGP, are independent. This assumption should cover many situations arising in simulation studies e.g,. for a CFA-model given by (3), where $\boldsymbol{\xi} = \mathbf{T}_1\mathbf{v}_1$ and $\boldsymbol{\delta} = \mathbf{T}_2\mathbf{v}_2$ and the vector $\mathbf{v}_1$ and the vector $\mathbf{v}_2$ are vectors of independent drawings from a distribution having finite moments up to order four, and with mean $\mathbf{0}$ and variance $\mathbf{1}$.

In the next section we consider three examples: In example A we study an exploratory factor analysis model as the AM. In examples B and C, AM is a confirmatory factor analysis model. The AM, in both cases, are two different, structurally misspecified versions of DGP.

3 Three Examples

Consider the following three examples, here illustrated with $k = 6$. Section 7 illustrates these examples numerically.

Example A Researcher A is interested in exploratory factor analysis and believes that there are two latent factors. However, he/she realizes that there may be several minor factors affecting the observed variables and these may contribute to minor correlations between the observed variables, see e.g., Tucker, Koopman, & Linn (1969) or MacCallum & Tucker (1991). Let the DGP be of the form (2), where the elements of $\boldsymbol{\xi}$ are independent i.e,. $\boldsymbol{\xi} = \mathbf{v}_1$ and $\boldsymbol{\delta} = \mathbf{B}^*\mathbf{v}_2$, where the elements $\mathbf{v}_2$ are independent. $\mathbf{B}^*$ is not a diagonal matrix. The DGP may then be represented by

$$\mathbf{x} = \boldsymbol{\Lambda}_x^*\mathbf{v}_1 + \mathbf{B}^*\mathbf{v}_2 = (\boldsymbol{\Lambda}_x^* \ \mathbf{B}^*) \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix} = \mathbf{A}\mathbf{v}. \tag{6}$$

The covariance matrix of $\boldsymbol{\delta}$, $\boldsymbol{\Theta}_{\delta}^{*} = \mathbf{B}^{*}\mathbf{B}^{*\prime}$, $\boldsymbol{\Lambda}_{x}^{*}$ consists of the factor loadings of the major factors and $\mathbf{B}^{*}$ consists of the factor loadings of the minor factors.

Then the matrix A may be represented by

$$\mathbf{A} = \begin{pmatrix} \lambda_{11}^{*} & \lambda_{12}^{*} & b_{13}^{*} & b_{14}^{*} & b_{15}^{*} & b_{16}^{*} & b_{17}^{*} & b_{18}^{*} \\ \lambda_{21}^{*} & \lambda_{22}^{*} & b_{23}^{*} & b_{24}^{*} & b_{25}^{*} & b_{26}^{*} & b_{27}^{*} & b_{28}^{*} \\ \lambda_{31}^{*} & \lambda_{32}^{*} & b_{33}^{*} & b_{34}^{*} & b_{35}^{*} & b_{36}^{*} & b_{37}^{*} & b_{38}^{*} \\ \lambda_{41}^{*} & \lambda_{42}^{*} & b_{43}^{*} & b_{44}^{*} & b_{45}^{*} & b_{46}^{*} & b_{47}^{*} & b_{48}^{*} \\ \lambda_{51}^{*} & \lambda_{52}^{*} & b_{53}^{*} & b_{54}^{*} & b_{55}^{*} & b_{56}^{*} & b_{57}^{*} & b_{58}^{*} \\ \lambda_{61}^{*} & \lambda_{62}^{*} & b_{63}^{*} & b_{64}^{*} & b_{65}^{*} & b_{66}^{*} & b_{67}^{*} & b_{68}^{*} \end{pmatrix}, \tag{7}$$

where the λ^{*}'s are factor loadings on the major factors (stars are used to indicate the true, fixed λ's) and the b^{*}'s are factor loadings on the minor factors. The b^{*}'s are small relative to the λ^{*}'s. The AM is the model (1) with

$$\boldsymbol{\Lambda}_{x} = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} & \lambda_{62} \end{pmatrix}, \tag{8}$$

and with the elements of $\boldsymbol{\delta}$ uncorrelated, i.e., with $\boldsymbol{\Theta}_{\delta}$ diagonal. For identification of the AM we fix $\lambda_{12} = 0$. The AM is misspecified because the elements of $\mathbf{B}^{*}\mathbf{v}_{2}$ in DGP are correlated, contrary to what is assumed in exploratory factor analysis where the factors ξ_1 and ξ_2 are supposed to account for the correlations between the x-variables.

Example B Researcher B is interested in confirmatory factor analysis and specifies AM as a model of the form (1) with two correlated factors ξ_1 and ξ_2. Let ϕ be the correlation between ξ_1 and ξ_2. The AM is

$$\boldsymbol{\Lambda}_{x} = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{pmatrix},$$

$$\boldsymbol{\Phi} = \begin{pmatrix} 1 & \\ \phi & 1 \end{pmatrix},$$

$$\boldsymbol{\Theta}_{\delta} = diag(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6).$$

Suppose the true mechanism (DGP) that generates the data is one where $\lambda_{32}^{*} \neq 0$ in $\boldsymbol{\Lambda}_{x}^{*}$. Let

$$\mathbf{T}_1 = \begin{pmatrix} 1 & 0 \\ \phi^{*} & \sqrt{1 - \phi^{*2}} \end{pmatrix},$$

$$\mathbf{T}_2 = diag(\sqrt{\theta_1^*}, \sqrt{\theta_2^*}, \sqrt{\theta_3^*}, \sqrt{\theta_4^*}, \sqrt{\theta_5^*}, \sqrt{\theta_6^*}).$$

Then $\mathbf{T}_1\mathbf{T}_1' = \mathbf{\Phi}^*$, $\mathbf{T}_2\mathbf{T}_2' = \mathbf{\Theta}_\delta^*$, and

$$\mathbf{A} = (\mathbf{\Lambda}_x^* \boldsymbol{T}_1\ \mathbf{T}_2) = \begin{pmatrix} \lambda_{11}^* & 0 & \sqrt{\theta^*}_1 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{21}^* & 0 & 0 & \sqrt{\theta^*}_2 & 0 & 0 & 0 & 0 \\ \lambda_{31}^* + \lambda_{32}^*\phi^* & \lambda_{32}^*\sqrt{1-\phi^{*2}} & 0 & 0 & \sqrt{\theta^*}_3 & 0 & 0 & 0 \\ \lambda_{42}^*\phi^* & \lambda_{42}^*\sqrt{1-\phi^{*2}} & 0 & 0 & 0 & \sqrt{\theta^*}_4 & 0 & 0 \\ \lambda_{52}^*\phi^* & \lambda_{52}^*\sqrt{1-\phi^{*2}} & 0 & 0 & 0 & 0 & \sqrt{\theta^*}_5 & 0 \\ \lambda_{62}^*\phi^* & \lambda_{62}^*\sqrt{1-\phi^{*2}} & 0 & 0 & 0 & 0 & 0 & \sqrt{\theta^*}_6 \end{pmatrix}.$$

In this case, model AM is misspecified because $\lambda_{32}^* \neq 0$. We investigate what happens when λ_{32}^* increases.

Example C Researcher C estimates a model (AM) of the form (1) with one factor. However, the true state of affairs is that there are two factors with correlation $\phi^* \neq 0$ and $\phi^* < 1$. The DGP is the same as in Example B, but with $\lambda_{32}^* = 0$. We investigate what will happen when ϕ^* increases i.e., when the misspecification decreases.

4 The Asymptotic Covariance Matrix

Let $\mathbf{\Sigma}_0$ be the covariance matrix of the data generating process and let $\mathbf{S}$ be a sample covariance matrix estimated from a random sample of $N = n+1$ independent observations of $\mathbf{x}$. 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 $k^\star = \frac{1}{2}k(k+1)$. Similarly, let $\boldsymbol{\sigma}_0$ be a vector of order $k^\star$ of the non-duplicated elements of $\mathbf{\Sigma}_0$. $\mathbf{S}$ converge in probability to $\mathbf{\Sigma}_0$ as $n \to \infty$, *i.e.*, $\mathbf{s} \xrightarrow{p} \boldsymbol{\sigma}_0$. It follows from the multivariate Central Limit Theorem (see *e.g.*, Anderson, 1984, p.81, Theorem 3.4.3) that

$$n^{\frac{1}{2}}(\mathbf{s} - \boldsymbol{\sigma}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Omega}), \tag{9}$$

where $\xrightarrow{d}$ denotes convergence in distribution. Browne & Shapiro (1988, Equation 2.7) give $\mathbf{\Omega}$ for $\mathbf{x}$ in (4) as

$$\mathbf{\Omega} = \mathbf{K}' \left\{ 2(\mathbf{\Sigma_0} \otimes \mathbf{\Sigma_0}) + \sum_{\mathbf{i=1}}^{\mathbf{g}} (\mathbf{A_i} \otimes \mathbf{A_i})\mathbf{C_i}(\mathbf{A_i}' \otimes \mathbf{A_i}') \right\} \mathbf{K}, \tag{10}$$

where $\mathbf{\Sigma_0} = \sum_{\mathbf{i=1}}^{\mathbf{g}} (\mathbf{A_i}\boldsymbol{\phi}_\mathbf{i}\mathbf{A_i}')$; and $\boldsymbol{\phi}_\mathbf{i}$ is the covariance matrix of $\mathbf{v}_i$ and where $\mathbf{K}$ is the matrix $\mathbf{K}_k$ 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 $\mathbf{C}_i$ is the fourth order cumulant matrix of $\mathbf{v}_i$, $i = 1, 2, \ldots, g$, where $g = k + l$.

The mean vector of $\mathbf{x}$ in (5) is $\mathbf{0}$ and, since the elements of $\mathbf{v}$ are independent with unit variances, the covariance matrix of $\mathbf{x}$ is

$$\mathbf{\Sigma}_0 = \sum_{\mathbf{i=1}}^{\mathbf{g}} (\mathbf{a_i a_i'}) = \mathbf{AA}' \; . \tag{11}$$

Let $\mu_{4i} = E(v_i^4)$. The matrix $\mathbf{C}_i$ in (10) is the 1×1 matrix with element $\gamma_{2i} = \mu_{4i} - 3$, the fourth order cumulant or kurtosis of v_i. Then (10) can be written in the following form

$$\mathbf{\Omega} = \mathbf{K}' \left\{ 2(\mathbf{\Sigma}_0 \otimes \mathbf{\Sigma}_0) + \sum_{i=1}^{g} (\mathbf{a}_i \otimes \mathbf{a}_i)(\mathbf{a}_i \otimes \mathbf{a}_i)' \gamma_{2i} \right\} \mathbf{K} \; . \tag{12}$$

Let $\mathbf{G} = [(\mathbf{a}_1 \otimes \mathbf{a}_1), (\mathbf{a}_2 \otimes \mathbf{a}_2), \ldots, (\mathbf{a}_g \otimes \mathbf{a}_g)]$ 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

$$\mathbf{\Omega} = \mathbf{K}'[2(\mathbf{\Sigma}_0 \otimes \mathbf{\Sigma}_0) + \mathbf{GMG}']\mathbf{K} \; . \tag{13}$$

If v_i is normally distributed, then $\mu_{4i} = 3$ and $\gamma_{2i} = 0$. Then the corresponding diagonal element of $\mathbf{M}$ is zero. If v_i is normally distributed *for all* i, then $\mathbf{M} = \mathbf{0}$ so that (13) reduces to

$$\mathbf{\Omega} = \mathbf{K}' 2(\mathbf{\Sigma}_0 \otimes \mathbf{\Sigma}_0)\mathbf{K} \; . \tag{14}$$

It is convenient to use the notation $\mathbf{\Omega}_{\mathrm{NNT}}$ for the matrix in (13) and the notation $\mathbf{\Omega}_{\mathrm{NT}}$ for the matrix in (14). Thus, from (13) it follows that

$$\mathbf{\Omega}_{\mathrm{NNT}} = \mathbf{\Omega}_{\mathrm{NT}} + \mathbf{K}'\mathbf{GMG}'\mathbf{K} \; . \tag{15}$$

A special case of (13) is when all elements of $\mathbf{v}$ have the same kurtosis so that $\gamma_{2i} = \gamma_2$, say, which is the same for all i. Then $\mathbf{M} = \gamma_2 \mathbf{I}$.

Let $\mathbf{W}_{\mathrm{NT}}$ and $\mathbf{W}_{\mathrm{NNT}}$ be consistent estimates of $\mathbf{\Omega}_{\mathrm{NT}}$ and $\mathbf{\Omega}_{\mathrm{NNT}}$, respectively. For example, let the elements of the matrices $\mathbf{W}_{\mathrm{NT}}$ and $\mathbf{W}_{\mathrm{NNT}}$ be

$$w_{ghij}^{\mathrm{NT}} = s_{gi}s_{hj} + s_{gj}s_{hi}, \tag{16}$$

$$w_{ghij}^{\mathrm{NNT}} = m_{ghij} - s_{gh}s_{ij} \; , \tag{17}$$

where

$$m_{ghij} = (1/n) \sum_{a=1}^{N} (z_{ag} - \bar{z}_g)(z_{ah} - \bar{z}_h)(z_{ai} - \bar{z}_i)(z_{aj} - \bar{z}_j) \; . \tag{18}$$

Note that $\mathbf{W}_{\mathrm{NT}}$ and $\mathbf{W}_{\mathrm{NNT}}$ are estimated without the use of the model.

5 Three Test Statistics

Consider a general model $\mathbf{\Sigma}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter vector of order $t < k^*$. The GLS fit function of Jöreskog & Goldberger (1972) is

$$F[\mathbf{S}, \Sigma(\boldsymbol{\theta})] = \frac{1}{2} tr \left\{ \mathbf{S}^{-1} \left[\mathbf{S} - \Sigma(\boldsymbol{\theta}) \right] \right\}^2 . \tag{19}$$

Following Browne (1974), and since

$$\mathbf{W}_{\mathrm{NT}}^{-1} = \frac{1}{2} \left[\mathbf{K}'(\mathbf{S} \otimes \mathbf{S})\mathbf{K} \right]^{-1} = \frac{1}{2} \mathbf{D}'(\mathbf{S}^{-1} \otimes \mathbf{S}^{-1})\mathbf{D} \ ,$$

where $\mathbf{D} = \mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}$, this can also be written

$$F[\mathbf{s}, \boldsymbol{\sigma}(\boldsymbol{\theta})] = \frac{1}{2} \left[\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta}) \right]' \mathbf{D}'(\mathbf{S}^{-1} \otimes \mathbf{S}^{-1})\mathbf{D} \left[\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta}) \right] . \tag{20}$$

The fit function F is to be minimized with respect to the model parameters $\boldsymbol{\theta}$. Let $\widehat{\boldsymbol{\theta}}$ be a minimizer of $F[s, \boldsymbol{\sigma}(\boldsymbol{\theta})]$ and let $\boldsymbol{\theta}_0$ be a minimizer of $F[\boldsymbol{\sigma}_0, \boldsymbol{\sigma}(\boldsymbol{\theta})]$. We assume that $\boldsymbol{\theta}_0$ is unique and, since the model *does not hold*, we have $F[\boldsymbol{\sigma}_0, \boldsymbol{\sigma}(\boldsymbol{\theta}_0)] > 0$.

The test statistic c_2 referred to in the introduction is n times the minimum value of F in (19) or (20). Following Browne (1984), equations 2.20b and 2.20a), this can also be written as

$$c_2 = n(\mathbf{s} - \widehat{\boldsymbol{\sigma}})' \left[\mathbf{W}_{\mathrm{NT}}^{-1} - \mathbf{W}_{\mathrm{NT}}^{-1} \widehat{\boldsymbol{\Delta}} (\widehat{\boldsymbol{\Delta}}' \mathbf{W}_{\mathrm{NT}}^{-1} \widehat{\boldsymbol{\Delta}})^{-1} \widehat{\boldsymbol{\Delta}}' \mathbf{W}_{\mathrm{NT}}^{-1} \right] (\mathbf{s} - \widehat{\boldsymbol{\sigma}}) \tag{21}$$

$$= n(\mathbf{s} - \widehat{\boldsymbol{\sigma}})' \widehat{\boldsymbol{\Delta}}_c (\widehat{\boldsymbol{\Delta}}_c' \mathbf{W}_{\mathrm{NT}} \widehat{\boldsymbol{\Delta}}_c)^{-1} \widehat{\boldsymbol{\Delta}}_c' (\mathbf{s} - \widehat{\boldsymbol{\sigma}}), \tag{22}$$

where $\widehat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}(\widehat{\boldsymbol{\theta}})$, $\widehat{\boldsymbol{\Delta}}_c$ is an orthogonal complement to the matrix $\widehat{\boldsymbol{\Delta}} = \partial \boldsymbol{\sigma} / \partial \boldsymbol{\theta}$ evaluated at $\widehat{\boldsymbol{\theta}}$.

The test statistic c_3 referred to in the introduction is

$$c_3 = \frac{d}{h} c_2, \tag{23}$$

where d is the degrees of freedom and

$$h = tr \left\{ \left[\mathbf{W}_{\mathrm{NT}}^{-1} - \mathbf{W}_{\mathrm{NT}}^{-1} \widehat{\boldsymbol{\Delta}} (\widehat{\boldsymbol{\Delta}}' \mathbf{W}_{\mathrm{NT}}^{-1} \widehat{\boldsymbol{\Delta}})^{-1} \widehat{\boldsymbol{\Delta}}' \mathbf{W}_{\mathrm{NT}}^{-1} \right] \mathbf{W}_{\mathrm{NNT}} \right\} \tag{24}$$

$$= tr\left\{\left[\widehat{\mathbf{\Delta}}_c(\widehat{\mathbf{\Delta}}_c'\mathbf{W}_{\mathrm{NT}}\widehat{\mathbf{\Delta}}_c)^{-1}\widehat{\mathbf{\Delta}}_c'\right]\mathbf{W}_{\mathrm{NNT}}\right\} \tag{25}$$

$$h = tr[(\widehat{\mathbf{\Delta}}_c'\mathbf{W}_{\mathrm{NT}}\widehat{\mathbf{\Delta}}_c)^{-1}(\widehat{\mathbf{\Delta}}_c'\mathbf{W}_{\mathrm{NNT}}\widehat{\mathbf{\Delta}}_c)]. \tag{26}$$

The test statistic c_4 referred to in the introduction is

$$c_4 = n(\mathbf{s} - \widehat{\boldsymbol{\sigma}})'\widehat{\mathbf{\Delta}}_c(\widehat{\mathbf{\Delta}}_c'\mathbf{W}_{\mathrm{NNT}}\widehat{\mathbf{\Delta}}_c)^{-1}\widehat{\mathbf{\Delta}}_c'(\mathbf{s} - \widehat{\boldsymbol{\sigma}}). \tag{27}$$

Still with $\widehat{\boldsymbol{\sigma}}$ evaluated at the GLS estimator $\widehat{\boldsymbol{\theta}}$, it follows from Browne (1984, Proposition 4) that c_4 has an asymptotic χ^2_d distribution if the model holds. This is valid also if $\widehat{\boldsymbol{\sigma}}$ is evaluated at the ML estimator $\widehat{\boldsymbol{\theta}}$.Some computer programs for structural equation modeling (e.g., LISREL) uses $\hat{\boldsymbol{\sigma}}$ instead of $\mathbf{s}$ in (16), where $\hat{\boldsymbol{\sigma}}$ is the vector of the non-duplicated elements of $\mathbf{\Sigma}(\widehat{\boldsymbol{\theta}})$. If the (assumed) model is misspecified, $\mathbf{W}_{\mathrm{NT}}$ is not a consistent estimate of $\mathbf{\Omega}_{\mathrm{NT}}$ but of

$$\mathbf{\Omega}_{\mathrm{NT}}^{AM} = \mathbf{K}'[2(\mathbf{\Sigma}(\boldsymbol{\theta}_0)\otimes\mathbf{\Sigma}(\boldsymbol{\theta}_0)]\mathbf{K}. \tag{28}$$

If the (assumed) model does not hold, $\mathbf{\Omega}_{\mathrm{NT}}^{AM} \neq \mathbf{\Omega}_{\mathrm{NT}}$.

The three test statistics c_2, c_3, and c_4 are all of the form $n\hat{C}$, where $\hat{C}$ converge in probability to a constant C, say. To evaluate C, we replace $\mathbf{s}$ by $\boldsymbol{\sigma}_{\mathbf{0}}$, $\widehat{\boldsymbol{\sigma}}$ by $\boldsymbol{\sigma}(\boldsymbol{\theta}_0)$, and $\widehat{\mathbf{\Delta}}_c$ by $\mathbf{\Delta}_{0c}$, where $\mathbf{\Delta}_{0c}$ is evaluated at $\boldsymbol{\theta}_0$. Furthermore, $\mathbf{W}_{\mathrm{NT}}$ and $\mathbf{W}_{\mathrm{NNT}}$ are replaced by $\mathbf{\Omega}_{\mathrm{NT}}$ and $\mathbf{\Omega}_{\mathrm{NNT}}$. Then we obtain the definitions

$$C_2 = (\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}(\boldsymbol{\theta}_0))'\mathbf{\Delta}_{0c}(\mathbf{\Delta}_{0c}'\mathbf{\Omega}_{\mathrm{NT}}\mathbf{\Delta}_{0c})^{-1}\mathbf{\Delta}_{0c}'(\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}(\boldsymbol{\theta}_0). \tag{29}$$

$$C_3 = \frac{d}{H}C_2. \tag{30}$$

$$H = tr[(\mathbf{\Delta}_{0c}'\mathbf{\Omega}_{\mathrm{NT}}\mathbf{\Delta}_{0c})^{-1}(\mathbf{\Delta}_{0c}'\mathbf{\Omega}_{\mathrm{NNT}}\mathbf{\Delta}_{0c})]. \tag{31}$$

$$C_4 = (\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}(\boldsymbol{\theta}_0))'\mathbf{\Delta}_{0c}(\mathbf{\Delta}_{0c}'\mathbf{\Omega}_{\mathrm{NNT}}\mathbf{\Delta}_{0c})^{-1}\mathbf{\Delta}_{0c}'(\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}(\boldsymbol{\theta}_0). \tag{32}$$

If the model holds, then $\boldsymbol{\sigma}_0 = \boldsymbol{\sigma}(\boldsymbol{\theta}_0)$ and C_2, C_3, and C_4 are all zero. If the model does not hold, then $C_i > 0, i = 1, 2, 3$ and $nC_i \to +\infty$ if $n \to +\infty$. Defining $C_i = p\lim(\frac{c_i}{n})$ requires a less casual definition than the one given here, we therefore define C_i , as mis-fit measures of weighted deviations of the "true" σ_0- and fitted $\sigma(\theta_0)$ moments. Notice that C_i plays the same role as F_0 ($F_0 = \min F[\Sigma_0, \Sigma(\boldsymbol{\theta}_0)]$, the minimum value of the fit function when the model is fitted to the population covariance matrix) does when the "chi-square" is n times the minimum value of a suitable F .

In the following sections we investigate what happens to C_3 and C_4 when the model does not hold and kurtosis increases.

6 The Effect of Kurtosis in DGP

In this section we give a formal proof on how the kurtosis in the asymptotic covariance of the form (10) affect C_3 and C_4.

6.1 The SB-Test

Assuming that $\mathbf{\Delta}_{0c}$ has rank d, we obtain H as

$$H = tr[(\mathbf{\Delta}'_{0c}\mathbf{\Omega}_{\mathrm{NT}}\mathbf{\Delta}_{0c})^{-1}(\mathbf{\Delta}'_{0c}\mathbf{\Omega}_{\mathrm{NNT}}\mathbf{\Delta}_{0c})] \ . \tag{33}$$

The influence of kurtosis on H is only via the diagonal matrix $\mathbf{M}$. All other matrices in (15) are independent of kurtosis. From (15) we have

$$\mathbf{\Delta}'_{0c}\mathbf{\Omega}_{\mathrm{NNT}}\mathbf{\Delta}_{0c} = \mathbf{\Delta}'_{0c}\mathbf{\Omega}_{\mathrm{NT}}\mathbf{\Delta}_{0c} + \mathbf{\Delta}'_{0c}\mathbf{K}'\mathbf{G}\mathbf{M}\mathbf{G}'\mathbf{K}\mathbf{\Delta}_{0c} \ . \tag{34}$$

Hence,

$$(\mathbf{\Delta}'_{0c}\mathbf{\Omega}_{\mathrm{NT}}\mathbf{\Delta}_{0c})^{-1}(\mathbf{\Delta}'_{0c}\mathbf{\Omega}_{\mathrm{NNT}}\mathbf{\Delta}_{0c}) = \mathbf{I}_d + (\mathbf{\Delta}'_{0c}\mathbf{\Omega}_{\mathrm{NT}}\mathbf{\Delta}_{0c})^{-1}\mathbf{P}\mathbf{M}\mathbf{P}' \ , \tag{35}$$

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

$$\mathbf{P} = \mathbf{\Delta}'_{0c}\mathbf{K}'\mathbf{G} \ . \tag{36}$$

Taking the trace of (35), gives

$$H = d + tr(\mathbf{Q}\mathbf{M}) \ , \tag{37}$$

where

$$\mathbf{Q} = \mathbf{P}'(\mathbf{\Delta}'_{0c}\mathbf{\Omega}_{\mathrm{NT}}\mathbf{\Delta}_{0c})^{-1}\mathbf{P} \ . \tag{38}$$

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

$$H = d + \sum_{i=1}^{g} q_{ii}\gamma_{2i} \ . \tag{39}$$

$\mathbf{Q}$ is positive semidefinite and if $\mathbf{Q} \neq \mathbf{0}$, $q_{ii} > 0$ for at least one i. Thus, if $\gamma_{2i} \to \infty$ *for all* i, it follows that $H \to \infty$ and $C_3 \to 0$. If $\gamma_{2i} = \gamma_2$ for *all* i, then

$$H = d + (tr\mathbf{Q})\gamma_2 \tag{40}$$

increases linearly with γ_2. It is also interesting to note that if $\gamma_{2i} < 0$ for all i, then $H < d$ implying $C_3 > C_2$.

The case of $\mathbf{Q} = \mathbf{0}$ will imply $C_3 = C_2$.The fact that $\mathbf{Q} = \mathbf{0}$ can be a consequence of $\mathbf{P} = \mathbf{0}$. From calculation involving some simple examples we have observed the following: If we are in the case of Asymptotic Robustness (AR), (See eg., Satorra, 2003), and the assumed models holds, then $\mathbf{P} = \mathbf{0}$ and hence $\mathbf{Q} = \mathbf{0}$.On the other hand if AR is not present neither $\mathbf{Q}$ or $\mathbf{P}$ is the zero matrix.

Next consider the case when $\mathbf{W}_{\mathrm{NT}}$ is not a consistent estimate of $\mathbf{\Omega}_{\mathrm{NT}}$ but of $\mathbf{\Omega}_{\mathrm{NT}}^{AM}$ in (28). Then

$$\begin{aligned} H &= tr[(\mathbf{\Delta}_{0c}'\mathbf{\Omega}_{\mathrm{NT}}^{AM}\mathbf{\Delta}_{0c})^{-1}(\mathbf{\Delta}_{0c}'\mathbf{\Omega}_{\mathrm{NNT}}\mathbf{\Delta}_{0c})] \\ &= tr[(\mathbf{\Delta}_{0c}'\mathbf{\Omega}_{\mathrm{NT}}^{AM}\mathbf{\Delta}_{0c})^{-1}(\mathbf{\Delta}_{0c}'\mathbf{\Omega}_{\mathrm{NT}}\mathbf{\Delta}_{0c} + \mathbf{\Delta}_{0c}'\mathbf{K}'\mathbf{G}\mathbf{M}\mathbf{G}'\mathbf{K}\mathbf{\Delta}_{0c})] \\ &= m + tr[(\mathbf{\Delta}_{0c}'\mathbf{\Omega}_{\mathrm{NT}}^{AM}\mathbf{\Delta}_{0c})^{-1}\mathbf{P}\mathbf{M}\mathbf{P}'] \ , \end{aligned}$$

where

$$m = tr[(\mathbf{\Delta}_{0c}'\mathbf{\Omega}_{\mathrm{NT}}^{AM}\mathbf{\Delta}_{0c})^{-1}(\mathbf{\Delta}_{0c}'\mathbf{\Omega}_{\mathrm{NT}}\mathbf{\Delta}_{0c})] \ , \tag{41}$$

Then

$$H = m + tr(\mathbf{Q}^{\mathbf{AM}}\mathbf{M}), \tag{42}$$

where

$$\mathbf{Q}^{AM} = \mathbf{P}'(\mathbf{\Delta}_{0c}'\mathbf{\Omega}_{\mathrm{NT}}^{AM}\mathbf{\Delta}_{0c})^{-1}\mathbf{P}. \tag{43}$$

$\mathbf{Q}^{AM}$ is positive semidefinite and if $\mathbf{Q}^{AM} \neq \mathbf{0}$, $q_{ii}^{AM} > 0$ for at least one i. Thus, if $\gamma_{2i} \to \infty$ *for all* i, it follows that $H \to \infty$ and $C_3 \to 0$.

6.2 The ADF-Test

Olsson et al. (2003) showed that F_0 is a non-increasing function of kurtosis when $\widehat{\boldsymbol{\sigma}}$ is evaluated at the WLS estimator $\widehat{\boldsymbol{\theta}}$. The proof presented here is more general since it does not restrict only to the WLS estimator, but include ML, GLS and ULS as well.

For the proof of C_4 we make use of 3 lemmas, they are presented without any proofs since they are only simple extensions of Theorem 23 and 24 in Magnus & Neudecker (1999, p.22).

Lemma 1: Let $\mathbf{A}$ be a positive semidefinite matrix of order $p \times p$ and $\mathbf{B}$ a matrix of order $q \times p$. Then $\mathbf{BAB}'$ is positive semidefinite.

Lemma 2: Let $\mathbf{E}$ and $\mathbf{F}$ be positive semidefinite matrices of order $p \times p$ with $\mathbf{E} \geq \mathbf{F}$ and let $\mathbf{B}$ be a matrix of order $q \times p$. Then $\mathbf{BEB}' \geq \mathbf{BFB}'$.

Lemma 3: Let $\mathbf{A}$ and $\mathbf{B}$ be positive definite matrices of order $p \times p$. If $\mathbf{A} \geq \mathbf{B}$ then $\mathbf{B}^{-1} \geq \mathbf{A}^{-1}$.

Theorem 1: The ADF statistic, here denoted C_4, will either decreases or remain constant when γ_{2i} increases for any i.

Proof: Let $\boldsymbol{\gamma} = (\gamma_{21}, \gamma_{22}, \ldots, \gamma_{2g})'$. Thus, $\boldsymbol{\gamma}$ contains the diagonal elements of $\mathbf{M}$. Consider $C_4(\boldsymbol{\gamma})$ in (32) as a function of $\boldsymbol{\gamma}$ and let $\boldsymbol{\gamma}^{(1)}$ and $\boldsymbol{\gamma}^{(2)}$ be two vectors such that $\gamma_i^{(1)} \leq \gamma_i^{(2)}$, $i = 1, 2, \ldots, g$. We will show that $C_4(\boldsymbol{\gamma}^{(1)}) \geq C_4(\boldsymbol{\gamma}^{(2)})$.

$C_4(\boldsymbol{\gamma})$ in (32) depends on kurtosis only via the matrix $\boldsymbol{\Omega}_{\mathrm{NNT}}$. Olsson *et.al.* (2003, Proposition 1) showed that

$$\boldsymbol{\Omega}_{\mathrm{NNT}}(\boldsymbol{\gamma}^{(1)}) \leq \boldsymbol{\Omega}_{\mathrm{NNT}}(\boldsymbol{\gamma}^{(2)}). \tag{44}$$

From Lemma 2 it follows that

$$\boldsymbol{\Delta}_{0c}'\boldsymbol{\Omega}_{\mathrm{NNT}}(\boldsymbol{\gamma}^{(1)})\boldsymbol{\Delta}_{0c} \leq \boldsymbol{\Delta}_{0c}'\boldsymbol{\Omega}_{\mathrm{NNT}}(\boldsymbol{\gamma}^{(2)})\boldsymbol{\Delta}_{0c}. \tag{45}$$

Then from Lemma 3 we have

$$(\boldsymbol{\Delta}_{0c}'\boldsymbol{\Omega}_{\mathrm{NNT}}(\boldsymbol{\gamma}^{(1)})\boldsymbol{\Delta}_{0c})^{-1} \geq (\boldsymbol{\Delta}_{0c}'\boldsymbol{\Omega}_{\mathrm{NNT}}(\boldsymbol{\gamma}^{(2)})\boldsymbol{\Delta}_{0c})^{-1}. \tag{46}$$

Let $\mathbf{u} = \boldsymbol{\Delta}_{0c}'[\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}(\boldsymbol{\theta}_0)]$. Using **Lemma 2** again shows that

$$\mathbf{u}'(\boldsymbol{\Delta}_{0c}'\boldsymbol{\Omega}_{\mathrm{NNT}}(\boldsymbol{\gamma}^{(1)})\boldsymbol{\Delta}_{0c})^{-1}\mathbf{u} \geq \mathbf{u}'(\boldsymbol{\Delta}_{0c}'\boldsymbol{\Omega}_{\mathrm{NNT}}(\boldsymbol{\gamma}^{(2)})\boldsymbol{\Delta}_{0c})^{-1}\mathbf{u} \ . \tag{47}$$

Hence, $C_4(\boldsymbol{\gamma}^{(1)}) \geq C_4(\boldsymbol{\gamma}^{(2)})$, *i.e.*, C_4 either decreases or remains constant when γ_{2i} increases for any i. The illustrative examples in Section 7 shows that C_4 can decrease with increasing kurtosis.

6.3 Conclusion

We have shown that C_3 decreases towards zero when γ_{2i} increases towards infinity for all i. Note that γ_{2i} increases for all i, but they can increase at different rate. We have also shown that C_4 either decreases or remains constant when γ_{2i} increases for any i. For C_3, this rests on the fact that the scaling correction $\frac{d}{H}$ approaches zero when γ_{2i} increases towards infinity for all i . As noted in the intoduction the Satorra-Bentler correction can be applied to any member of the Swain family (Swain, 1975), C_3 decreases towards zero e.g., for ML, GLS and ULS. These results are valid for any structural equation model as long as the elements of the vector $\mathbf{v}$ for the DGP are independent, conditions that holds in most simulation studies. A practical consequence is that misspecified models *can* be accepted if kurtosis is large. Andreassen, Lorentzen & Olsson (2006) reported a significant drop in the chi-square statistic when they compared the Normal theory chi-square with the ADF- and SB - statistics (ML- Chi-square = 1769.36, SB-Chi-square = 1212.51 and ADF-Chi-square = 518.94). They studied a simplified model (misspecified) of a Satisfaction Model in marketing using a large data set from a satisfaction survey. The number of observed variables in the model was 21, degrees of freedom was 182 and the univariate kurtosis was ranging from -0.5 up to 10.5.

This seemingly low power is not due to the statistics but to their application to misspecified models in combination with data with high kurtosis. We think that researchers should be aware of this.

7 Numerical Examples and illustrations

In this section we illustrate the three examples in Section 3 numerically. Since we are studying C_3 and C_4, the sample size $N = n + 1$ is out of the consideration. But to get LISREL run we have to specify a value for N. In the three following examples N = 101 for convenience. This is arbitrary.

7.1 Example A

For Example A we take $\mathbf{A}$ in (5) as

$$\mathbf{A} = \begin{pmatrix} .9 & .0 & .5 & .2 & .1 & .2 & .2 & .2 \\ .7 & .2 & .2 & .5 & .2 & .2 & .2 & .1 \\ .8 & .2 & .2 & .2 & .5 & .2 & .3 & .2 \\ .3 & .6 & .2 & .1 & .2 & .5 & .2 & .2 \\ .2 & .7 & .2 & .2 & .2 & .2 & .5 & .2 \\ .2 & .6 & .1 & .2 & .2 & .3 & .2 & .5 \end{pmatrix}, \tag{48}$$

and, to begin with, we take v_i to have the kurtosis $\gamma_{2i} = \gamma_2$, the same for all i.

From $\mathbf{A}$ we compute $\mathbf{\Sigma}_0 = \mathbf{AA}'$. Matrix $\mathbf{\Sigma}(\boldsymbol{\theta}_0)$ and $\mathbf{\Delta}_{0c}$ are obtained by fitting the two-factor model to $\mathbf{\Sigma}_0$. The parameter λ_{12} is fixed at zero to make the two-factor model identified.

Using $\mathbf{\Sigma}_0$, $\mathbf{\Sigma}(\boldsymbol{\theta}_0)$, and $\mathbf{\Delta}_{0c}$ all the matrices required to compute C_2, C_3, and C_4 can be computed as functions of γ_2. Table 1 gives the values of C_2, C_3, and C_4 for $\gamma_2 = -2, 0, 20, 30, 50$. In the figures the values can be observed over the interval $0 \le \gamma_2 \le 50$. For convenience, the values of C_2, C_3, and C_4 have been multiplied by 100.

(INSERT TABLE 1 ABOUT HERE)

Table 1 shows that: C_2 does not depend on kurtosis. If $\gamma_2 = 0$, then $C_2 = C_3 = C_4$. If $\gamma_2 < 0$, then $C_2 < C_3$ and $C_2 < C_4$.

If $\gamma_2 > 0$, then $C_2 > C_3$ and $C_2 > C_4$.

Both C_3 and C_4 decreases monotonically with increasing values of γ_2 and C_3 decreases faster than C_4.

These characteristics can also be seen in Figure 1 which shows C_2, C_3, and C_4 as smoothed functions of γ_2 over the interval $0 \le \gamma_2 \le 50$.

(INSERT FIG. 1 ABOUT HERE)

We also consider a case when only one of the v_i have a kurtosis. For example, let $\gamma_{21} \ge 0$ and $\gamma_{2i} = 0$ for $i = 2, 3, \ldots, 8$. The resulting C_3 and C_4 are given in Table 2. As in the previous case both C_3 and C_4 decreases monotonically with increasing values of γ_2, and C_4 appears to decrease slightly faster than C_3. We also observe that the decrease is very small.

(INSERT TABLE 2 ABOUT HERE)

Example B

For Example B we take $\phi^* = 0.6$ so that $\sqrt{1-\phi^{*2}} = 0.8$ and $\mathbf{A}$ in (5) as

$$\mathbf{A} = \begin{pmatrix} .9 & .0 & .5 & .0 & .0 & .0 & .0 & .0 \\ .8 & .0 & .0 & .5 & .0 & .0 & .0 & .0 \\ .7+\phi^*\lambda^*_{32} & \acute{\lambda}^*_{32}\sqrt{1-\phi^{*2}} & .0 & .0 & .5 & .0 & .0 & .0 \\ .6\phi^* & .6\sqrt{1-\phi^{*2}} & .0 & .0 & .0 & .5 & .0 & .0 \\ .7\phi^* & .7\sqrt{1-\phi^{*2}} & .0 & .0 & .0 & .0 & .5 & .0 \\ .8\phi^* & .8\sqrt{1-\phi^{*2}} & .0 & .0 & .0 & .0 & .0 & .5 \end{pmatrix}, \tag{49}$$

where $\lambda^*_{32} = 0.1, 0.3, 0.5$. With $\mathbf{\Sigma}_0 = \mathbf{AA}'$ we fit a two-factor confirmatory factor model with $\lambda_{32} = 0$. Matrices $\mathbf{\Sigma}(\boldsymbol{\theta}_0)$ and $\mathbf{\Delta}_{0c}$ can be obtained as before.

(INSERT TABLE 3 ABOUT HERE)

Table 3 gives values of C_3 for different values of λ^*_{32} and increasing values of γ_2.

It is seen that: C_3 increases with increasing values of λ_{32}.For each value of λ_{32}, C_3 decreases monotonically with increasing values of γ_2.

Table 4 gives values of C_4 for different values of λ_{32} and increasing values of γ_2.

(INSERT TABLE 4 ABOUT HERE)

Again we see that:C_4 increases with increasing values of λ_{32}.For each value of λ_{32}, C_4 decreases monotonically with increasing values of γ_2.

Comparing Tables 3 and 4 it seems that C_4 decreases faster than C_3 with increasing values of γ_2. This holds for all three values of λ_{32}. The same characteristics can be seen in Figure 2 and Figure 3 which give C_3 and C_4, respectively, as smoothed functions.

Since the starting point at $\gamma_2 = 0$ for each value of λ_{32} is the same in Figures 2 and 3 it is clear that C_4 decreases much faster than C_3, and one might think that all three curves goes asymptotically to zero when $\gamma_2 \to +\infty$.

(INSERT FIG 2 AND FIG 3 ABOUT HERE)

7.2 Example C

In example C we take $\mathbf{A}$ as in Example B but with $\lambda^*_{32} = 0$. The DGP is a two-factor model with correlation $\phi^* < 1$. The AM is a one-factor model which is the same as DGP with $\phi^* = 1$. We investigate what happens if ϕ^* increases.

Matrices $\mathbf{\Sigma}_0$, $\mathbf{\Sigma}(\boldsymbol{\theta}_0)$ and $\mathbf{\Delta}_{0c}$ can be obtained as before.

Table 5 and 6 give values of C_3 and C_4 for increasing values of ϕ^* and increasing values of γ_2.

We see that: Both C_3 and C_4 decrease with increasing values of ϕ^*.For each value of ϕ^*, both C_3 and C_4 decreases monotonically with increasing values of γ_2.

(INSERT TABLE 5 AND TABLE 6 ABOUT HERE)

Comparing Tables 5 and 6 it seems that C_4 decreases faster than C_3 with increasing values of γ_2. This holds for all three values of ϕ^*.

Figure 4 shows C_3 as a smoothed function of γ_2 for $\phi^* = 0.5, 0.7, 0.9$ and Figure 5 shows C_4 as a smoothed function of γ_2 for the same value of ϕ^*. It is seen that C_4 decreases much faster than C_3. At $\gamma_2 = 50$, C_4 takes almost the same value for all three values of ϕ^*. For the most severely misspecified model, *i.e.*, when $\phi^* = 0.5$, C_3 drops 61.5% while C_4 drops 93.5% when γ_2 goes from 0 to 50. On the other hand, when $\phi^* = 0.9$, C_3 drops only 27.6% while C_4 drops 77.5% when γ_2 goes from 0 to 50.

(INSERT FIG 4 AND FIG 5 ABOUT HERE)

8 Discussion and Further Research

We have shown that the population value of the scaling correction of the mean corrected SB statistic decreases towards zero with increasing kurtosis. Furthermore, we have shown that C_4 is a non-increasing function (i.e., either decreases or remains constant) of kurtosis. Thus, it is reasonable to conjecture that the test statistics c_3 and c_4 under e.g., an elliptical distribution will loose power as a function of increasing kurtosis in large samples. Our illustrating examples indicate that the decrease is stronger the more misspecified the model is. This holds in all situations. Although the data generating process that we have chosen is similar to procedures used in simulation studies, our results have been obtained without simulations. But, the results are in line with the results in the simulation study reported by Curran, West and Finch (1996), and they are also supported by a simple simulation example in this paper.

It is not unreasonable that the results also are valid for more general processes. Further research should also include situations where $\gamma_{2i} \to \infty$ only for subsets of the vector $\mathbf{v}$. In the examples (A, B and C) we have calculated the matrix $\mathbf{Q}$ (see equations 45 and 46). In example A, where Θ_δ^* is not a diagonal matrix it is hard to see any clear structure in the matrix $\mathbf{Q}$. But in examples B and C, where Θ_δ^* is diagonal, $\mathbf{Q}(8 \times 8)$ is of the form $\mathbf{Q} = \begin{pmatrix} \mathbf{E} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ and $\mathbf{E}$ is (2×2). In e.g., example B when $\lambda_{32}^* = 0.5$, $\mathbf{E} = \begin{pmatrix} 0.00274 & 0.01608 \\ 0.01608 & 0.09445 \end{pmatrix}$. From equation (46) it is relatively easy to observe that C_3 is hardly affected by the kurtosis in position v_1,much more by the kurtosis in position v_2, but nothing from positions v_3 to v_8. It would be of interest to focus on the relationship between the data generating process, the assumed model and the general structure of $\mathbf{Q}$. Referring to Satorra (1989, 2003) and beeing in an asymptotic robustness situation and assuming that AM holds we conjecture that the matrix $\mathbf{P}$ of the form (36) is the zero matrix implying that $\mathbf{Q} = \mathbf{0}$.

9 Acknowledgements

The authors thank the Editor and reviewers for valuable comments in preparing this manuscript.

References

Anderson, T.W. (1984) *An introduction to multivariate statistical analysis*. Second Edition. New York: Wiley.

Andreassen, T.W., Lorentzen, B. & Olsson, U. (2006) The Impact of Non-Normality and Estimation Methods in SEM on Satisfaction Research in Marketing. Quality & Quantity, **40**, 39-58.

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

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. (1987) Robustness of statistical inference in factor analysis and related models. Biometrika,**74**,2,pp.375-84.

Browne, M.W. & Shapiro, A. (1988) Robustness of normal theory methods in the analysis of linear latent variate models. British Journal of Mathematical and Statistical Psychology, **37**, 193–208.

Curran, P.J., 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.

Fleishman, A.I. (1978). A method for simulating non-normal distributions. Psychometrika,43, 521–532.

Hershberger, S.L. (2003) The growth of structural equation modeling. *Structural Equation Modeling*, **10**, 35–46.

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

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

Jöreskog, K.G., and Goldberger, A.S. (1972) Factor analysis by generalized least squares. *Psychometrika*, **37**, 243–250.

Kano, Y. (1992): Robust statistics for test-of-independence and related structural models. *Statistics & Probability Letters*, **15**, 21 - 26.

MacCallum, R.C., & Tucker, L.R. (1991) Representing sources of error in the common factor model: Implications for theory and practice. *Psychological Bulletin*, **109**, 502–511.

Magnus, J.R. & Neudecker, H. (1999) *Matrix differential calculus with applications in statistics and econometrics*. New York: Wliey

Mattson S. (1997). How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research*, **32**, 4, 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.

Ramberg, J.S., Tadikamalla, R.P, Dudewicz, E.J., & Mykytka, E. F. (1979). A probability distribution and its use in fitting data. *Technometrics*, 21, 201-215.

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

Satorra, A. (2003) Power of chi-square goodness-of-fit test in structural equation models: the case of non-normal data, pp. 57 - 68 in *New developments of psychometrics*, H. Yanai, A. Okada, K. Shigemasu, Y. Kano and J.J Meulman (Eds),Springer Verlag, Tokyo.

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.

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

Swain,A.J. (1975). A class of factor analysis estimation procedures with common asymptotic sampling properties. *Psychometrika*, **40**, no.3, 315-335.

Tucker, L.R., Koopman, R.F., & Linn, R.L. (1969) Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, **34**, 421–459

Vale, C. D., & Maurelli, V. A. (1983). Simulating non-normal distributions. *Psychometrika*, 48, 465-471.

Wald, Abraham 1943. "Test of statistical hypotheses concerning several parameters when the number of observation is large." *Transaction of the American Mathematical Society*, 54, 3 ,426–482.

Yuan, K.-H., & Bentler, P.M.(1998). Normal theory based test statistics in structural equation modeling. *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.

γ_2	C_2	C_3	C_4
-2	6.48	7.07	6.92
0	6.48	6.48	6.48
20	6.48	3.55	4.71
30	6.48	2.89	4.26
50	6.48	2.11	3.62

Table 1: Values of C_2 , C_3 , C_4 for increasing values of γ_2
All numbers have been multiplied by 100

γ_{21}	C_3	C_4
-2	6.4824	6.4824
0	6.4821	6.4821
10	6.4806	6.4803
30	6.4777	6.4767
50	6.4748	6.4732

Table 2: Values of C_3 , C_4 for increasing values of γ_{21}
All numbers have been multiplied by 100

	λ_{32}^*		
γ_2	0.1	0.3	0.5
-2	1.18	7.30	13.26
0	1.18	7.20	12.94
10	1.16	6.65	11.53
50	1.10	5.09	8.05

Table 3: Values of C_3 for different values of λ_{32} and increasing values of γ_2
All numbers have been multiplied by 100

	λ_{32}^*		
γ_2	0.1	0.3	0.5
-2	1.20	8.31	16.06
0	1.18	7.20	12.94
10	1.05	4.33	6.56
30	0.87	2.41	3.30
50	0.74	1.67	2.21

Table 4: Values of C_4 for different values of λ_{32} and increasing values of γ_2
All numbers have been multiplied by 100

	ϕ^*		
γ_2	0.5	0.7	0.9
-2	31.81	22.72	7.45
0	29.72	21.73	7.34
10	22.57	17.83	6.82
30	15.20	13.12	5.97
50	11.46	10.38	5.31

Table 5: Values of C_3 for different values of ϕ and increasing values of γ_2
All numbers have been multiplied by 100

	ϕ^*		
γ_2	0.5	0.7	0.9
-2	70.14	35.84	8.51
0	29.78	21.73	7.34
10	7.68	7.32	4.35
30	3.15	3.09	2.39
50	2.01	1.94	1.65

Table 6: Values of C_4 for different values of ϕ and increasing values of γ_2
All numbers have been multiplied by 100

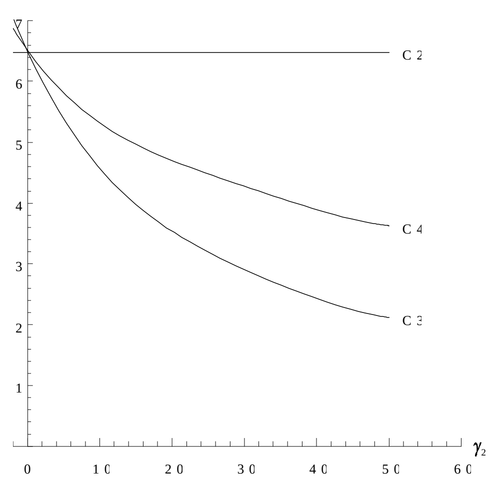

Figure 1: C_2 , C_3 and C_4 as functions of γ_2

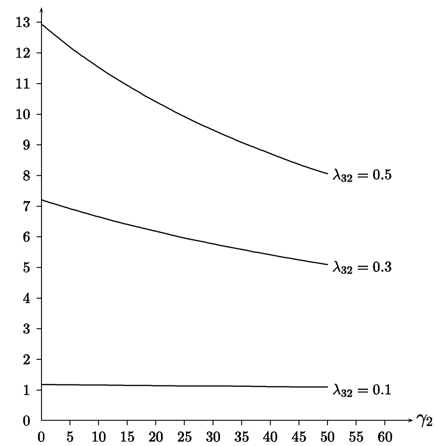

Figure 2: C_3 as a function of γ_2 for $\lambda_{32}^* = 0.1, 0.3, 0.5$

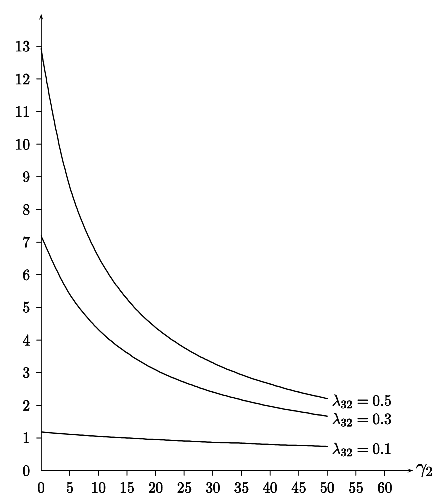

Figure 3: C_4 as a function of γ_2 for $\lambda_{32}^* = 0.1, 0.3, 0.5$

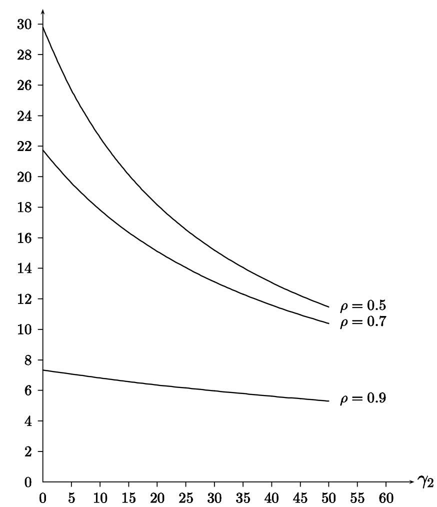

Figure 4: C_3 as a function of γ_2 for $\phi^* = \rho = 0.5, 0.7, 0.9$

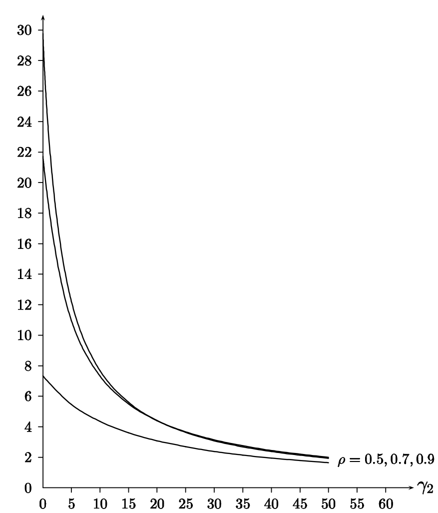

Figure 5: C_4 as a function of γ_2 for $\phi^* = \rho = 0.5, 0.7, 0.9$