Foldnes, N., & Grønneberg, S. (2017). The asymptotic covariance matrix and its use in simulation studies. *Structural equation modeling: A multidisciplinary Journal, 24*(6), 881-896 DOI: https://doi.org/10.1080/10705511.2017.1341320

The asymptotic covariance matrix and its use in simulation studies

Njål Foldnes and Steffen Grønneberg

Department of Economics

BI Norwegian Business School

Oslo, Norway 0484

Correspondence concerning this article should be addressed to *njal.foldnes@bi.no*

Abstract

The asymptotic performance of SEM tests and standard errors are influenced by two factors: the model and the asymptotic covariance matrix Γ of the sample covariances. Although most simulation studies clearly specify model conditions, specification of Γ is usually limited to values of univariate skewness and kurtosis. We illustrate that marginal skewness and kurtosis are not sufficient to adequately specify a non-normal simulation condition by showing that asymptotic standard errors and test statistics vary substantially among distributions whose skewness and kurtosis are identical. We argue therefore that Γ should be reported when presenting the design of simulation studies. We show how Γ may be exactly calculated under the widely used Vale-Maurelli transform. We suggest to plot the elements of Γ and to report the eigenvalues associated with the test statistic. R code is provided.

Keywords: Monte Carlo, non-normality, Vale-Maurelli, asymptotic covariance matrix, kurtosis, structural equation modeling

The asymptotic covariance matrix and its use in simulation studies

Introduction

The performance of procedures for structural equation modeling (SEM) can often be studied analytically only in the asymptotic case. Therefore, Monte Carlo simulations have for decades been a main source of information about the finite-sample performance of SEM methods, with thousands of papers using random data generation in order to learn how sample size, underlying distribution and level of model misspecification affect estimators and fit statistics. Boomsma (2013) reports that almost one third of all articles published in *Structural Equation Modeling* from 1994 until 2012 were pure simulation studies. The importance of such studies does not seem to diminish. For instance, of the 18 articles in Issues 3 and 4 in Volume 24, 2017 of *Structural Equation Modeling*, 16 contained simulation studies, five of which simulated from continuous non-normal distributions, which is the topic of the present paper.

In the present article we argue that distributional conditions in simulation studies should be more precisely reported than is currently the practice, and we discuss procedures that may help achieve this goal. As an example, consider the well-cited simulation paper by Curran, West, and Finch (1996), whose two non-normal data conditions are reported as skewness 2 and kurtosis 7 for moderate non-normality, and skewness 3 and kurtosis 21 for severe non-normality. Although the authors inform us that data generation was done by the popular Vale-Maurelli (VM) transform (Vale & Maurelli, 1983), and thereby indirectly specify the underlying distribution of the data, there is at present no way to use this information to evaluate the asymptotic performance of SEM procedures. In the present paper we revisit the severe non-normality condition investigated by Curran et al. (1996) and deduce the exact asymptotic rejection rates and standard errors implied by the VM transform. We also show that under a different data generation method, but with the same severe skewness and kurtosis values retained, the rejection rates and standard errors are quite different.

Non-normality in the context of SEM modeling is in essence characterised by the sampling variability of the second-order moments, i.e. the sampling distribution of the sample variances and covariances. In large samples, this moment vector will approximately follow a normal distribution, with a covariance matrix that approaches the so-called asymptotic covariance matrix Γ. Therefore, to more precisely specify the non-normal data distribution, Γ should be reported, or some summary of it. Moreover Γ is a central component of the formulas used to estimate standard errors and compute test statistics in SEM. With the availability of Γ, combined with the model specification, it becomes possible to calculate standard errors and properties of test statistics (e.g, mean value, Type I rejection rates) that are correct asymptotically. This information sheds light on the true degree of non-normality, that is, the exact degree to which the non-normal distribution affects standard errors and test statistics in large samples, and separates this from finite-sample effects.

In the present article we show how Γ may be calculated under the VM transform. As far as we know, such a procedure has not been described before, despite the widespread popularity of the VM transform over three decades. We exemplify the procedure for two concrete models, where we use Γ to calculate asymptotic standard errors and properties of test statistics. We also discuss how the essential aspects of Γ may be reported in simulation studies, using graphs and tables. A final contribution is to contrast the VM transform with the recently proposed independent generator (IG) transform (Foldnes & Olsson, 2016). The IG transform does the same job as the VM transform, e.g., it makes simulation possible from a distribution with prespeficed covariance matrix and marginal skewness and kurtosis values. In addition it accomodates easy calculation of Γ. It is therefore of interest to consider two distributions for which skewnesses and kurtoses are exactly equal, and to demonstrate analytically that they imply different Γ, and therefore different asymptotic behaviour in estimators and test statistics. Such comparisons have hitherto only been empirically using simulations (e.g., Foldnes & Olsson, 2016; Astivia & Zumbo, 2015). Our

findings imply that the widespread practice of specifying only marginal skewness and kurtosis is inadequate, as it leaves out essential components of the non-normality in the simulated data, and how these interact with the model specification.

Theory

A structural equation model posits a covariance structure for an observed p-dimensional vector $Y = (Y_1, \ldots, Y_p)'$. Let s_n denote the $p(p+1)/2$ non-redundant elements of the empirical covariance matrix based on n IID observations with the same distribution as Y. The large-sample behaviour of tests of fit and of standard error estimators are functions of the asymptotic covariance matrix Γ of $\sqrt{n}s_n$, as expounded in the seminal paper by Browne (1984), who treated the estimation of covariance models $\sigma(\theta)$ in the framework of minimal discrepancy functions. These are functions $F = F(s, \sigma)$ that obey the following three conditions: $F(s, \sigma) \geq 0$ for all s, σ; $F(s, \sigma) = 0$ if and only if $s = \sigma$; and F is twice continuously differentiable jointly. An estimator is then obtained as

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}}\, F(s_n, \sigma(\theta)).$$

Note that the widely used normal-theory maximum likelihood (NTML) estimator is such a minimal discrepancy estimator. The minimizer of the above expression, with s replaced by its population counterpart σ_0, is denoted by θ_0. Under the regularity conditions listed by (Browne, 1984), the covariance matrix of the estimated parameters obey

$$\operatorname{cov}(\sqrt{n}\hat{\theta}_n) \longrightarrow \Omega := \{\Delta' V \Delta\}^{-1} \{\Delta' V \Gamma V \Delta\} \{\Delta' V \Delta\}^{-1}, \tag{1}$$

as $n \to \infty$. Here Δ is the $p \times q$ derivative matrix $\partial\sigma(\theta)/\partial\theta'$ and $V = -\frac{1}{2}\frac{\partial^2 F(s,\sigma)}{\partial s \partial \sigma}$, evaluated at θ_0 and σ_0.

A test statistic for global model fit is obtained as $T_n = nF(s, \sigma(\hat{\theta}))$. Under correct model specification and other assumptions presented by Shapiro (1983) and Satorra (1989), the asymptotic distribution of T_n is a mixture of chi-squares:

$$T_n \xrightarrow[n\to\infty]{D} \sum_{j=1}^{d} \lambda_j X_j^2, \qquad X_1, \ldots, X_d \sim N(0,1) \text{ IID}, \tag{2}$$

where $\lambda_1, \ldots, \lambda_d$ are the d non-zero eigenvalues of $U\Gamma$. The matrix U is defined as $U = V - V\Delta \left\{\Delta' V \Delta\right\}^{-1} \Delta' V$.

Although these results are general, we limit ourselves in the present article to the widely used NTML estimator and its associated test statistic T_{ML}. For this estimator, it can be shown that $V = \Gamma_N^{-1}$, where Γ_N is the asymptotic covariance matrix under multivariate normality. This matrix is easily calculated from the population covariance matrix Σ, as $\Gamma_N = 2K_p'(\Sigma \otimes \Sigma)K_p$, where K_p is a pattern matrix (Browne, 1974). The $\otimes$ symbol denotes the Kronecker product. In a given simulation condition, the model is fully specified, so the matrices Σ, V and Δ, which all refer to the underlying model, may be calculated, for instance using the R package lavaan, as demonstrated in Appendix A. The final component needed to use the central results in eqs. (1) and (2) is the asymptotic covariance matrix Γ. We next show how this matrix may be computed under two data simulation techniques, namely the VM and IG transforms.

Calculating the asymptotic covariance matrix

Under multivariate normality there is a closed-form expression for Γ. For non-normal distributions it is harder to obtain Γ. Yuan and Bentler (1999) proposed two classes of non-normal distributions that may be used in simulation studies, and deduced Γ for each of these classes. Foldnes and Olsson (2016) study a subclass of these distributions and demonstrate its usefulness in simulations, referring to the technique as an independent generator approach. However, the traditional workhorse in Monte Carlo evaluation of estimators and tests of fit in covariance structure analysis has been the transform proposed by Vale and Maurelli (1983). The VM method is implemented in almost all current software packages as the default for generating non-normal data. Given the central role of the VM transform for evaluating the finite-sample performance of estimators and test statistics, it therefore seems worthwhile to investigate whether Γ_{VM}, i.e., Γ under the VM transform, may be obtained. Foldnes and Grønneberg (2015) derived the distribution of Y

under VM, and investigated its underlying copula, but did not characterize Γ_{VM}. In fact, we are unaware of any treatment of Γ_{VM} in the literature.

We first shortly discuss the VM transform. In the univariate case, Fleishman's technique for constructing a non-normal random variable Y is based on the stochastic representation

$$Y = a + bX + cX^2 + dX^3,$$

where X is a standard normal variable (Fleishman, 1978). The constants a, b, c and d are chosen to provide Y with pre-specified mean, variance, skewness and kurtosis. The VM method constructs a random vector $Y = (Y_1, \ldots, Y_p)'$ with prespecified univariate skewness and kurtosis values, and a prespecified target covariance matrix Σ, as

$$Y_{VM} = \begin{pmatrix} a_1 + b_1 X_1 + c_1 X_1^2 + d_1 X_1^3 \\ a_2 + b_2 X_2 + c_2 X_2^2 + d_2 X_2^3 \\ \vdots \\ a_p + b_p X_p + c_p X_p^2 + d_p X_p^3 \end{pmatrix}, \tag{3}$$

where $X = (X_1, \ldots, X_p)'$ is a multivariate normal random vector whose elements have unit variance, and whose correlation matrix R is numerically determined so that Y attains its target covariance matrix.

For a general random p-vector Y whose distribution has fourth-order moments, the elements of Γ have the form

$$\Gamma_{ij,kl} = \sigma_{ijkl} - \Sigma_{ij}\Sigma_{kl}, \tag{4}$$

where the fourth-order moments σ_{ijkl} are given by

$$\sigma_{ijkl} = \mathbb{E}(Y_i - \mu_i)(Y_j - \mu_j)(Y_k - \mu_k)(Y_l - \mu_l), \quad \mu_m = \mathbb{E}Y_m,$$

see eq. (2.2) in Browne (1984). Here Σ_{ij} and Σ_{kl} are dictated by the simulation setup from the target covariance matrix Σ. We therefore need to calculate σ_{ijkl} associated with Y_{VM}. We first assume, without loss of generality, that $a_i = -c_i$, so that $\mu_i = a_i + c_i = 0$ for $i = 1, \ldots, p$. Also, to simplify our notation, we replace the coefficients a_i, b_i, c_i and d_i by

$\alpha_{i,1}, \alpha_{i,2}, \alpha_{i,3}$ and $\alpha_{i,4}$, respectively, for $i = 1, \ldots, p$. Then, to determine Γ_{VM} we need to obtain the expectation of

$$\begin{aligned} Y_i Y_j Y_k Y_l &= \left(\sum_{n_i=0}^{3} \alpha_{i,n_i} X_i^{n_i} \right) \left(\sum_{n_j=0}^{3} \alpha_{j,n_j} X_j^{n_j} \right) \left(\sum_{n_k=0}^{3} \alpha_{k,n_k} X_k^{n_k} \right) \left(\sum_{n_l=0}^{3} \alpha_{l,n_l} X_l^{n_l} \right) \\ &= \sum_{n_i=0}^{3} \sum_{n_j=0}^{3} \sum_{n_k=0}^{3} \sum_{n_l=0}^{3} \alpha_{i,n_i} \alpha_{j,n_j} \alpha_{k,n_k} \alpha_{l,n_l} X_i^{n_i} X_j^{n_j} X_k^{n_k} X_l^{n_l}. \end{aligned} \tag{5}$$

Clearly, this amounts to being able to calculate

$$\mathbb{E} X_i^{n_i} X_j^{n_j} X_k^{n_k} X_l^{n_l} \tag{6}$$

for $0 \leq n_i, n_j, n_k, n_l \leq 3$. Note that i, j, k and l are not necessarily distinct, but may overlap, obtaining for instance $\mathbb{E} X_1^{12}$. The moments in eq. (6) can be calculated by the formula of Isserlis (1918), which is based on the intermediate correlations in R. Specifically, if $n_i + n_j + n_k + n_l$ is an odd number, the expectation is zero. Otherwise, $n_i + n_j + n_k + n_l$ is an even number, and we consider all distinct ways of partitioning the set $\{1, 2, \ldots, n_i + n_j + n_k + n_l\}$ into pairs. There are $1 \cdot 3 \cdot 5 \cdot \ldots \cdot (n_i + n_j + n_k + n_l - 1)$ such pair-partitions. Each pair-partition involves $(n_i + n_j + n_k + n_l)/2$ pairs. In each pair-partition we form the product of the corresponding $(n_i + n_j + n_k + n_l)/2$ elements of R. The expectation $\mathbb{E} X_i^{n_i} X_j^{n_j} X_k^{n_k} X_l^{n_l}$ is then the sum of all such products. Consider for instance $\mathbb{E} X_1^3 X_2^2 X_3^3 X_4^2$. Here we have $1 \cdot 3 \cdot 5 \cdot 7 \cdot 9 = 945$ distinct pair-partitions. However,

many of these correspond to the same five correlations from R, so in compact form we get

$$\begin{aligned}\mathbb{E}X_1^3X_2^2X_3^3X_4^2 =& 18r_{1,1}r_{1,2}r_{2,3}r_{3,3}r_{4,4} + 36r_{1,1}r_{1,2}r_{2,3}r_{3,4}^2 + 36r_{1,1}r_{1,2}r_{2,4}r_{3,3}r_{3,4} + \\ & 9r_{1,1}r_{1,3}r_{2,2}r_{3,3}r_{4,4} + 18r_{1,1}r_{1,3}r_{2,2}r_{3,4}^2 + 18r_{1,1}r_{1,3}r_{2,3}^2r_{4,4} + \\ & 72r_{1,1}r_{1,3}r_{2,3}r_{2,4}r_{3,4} + 18r_{1,1}r_{1,3}r_{2,4}^2r_{3,3} + 18r_{1,1}r_{1,4}r_{2,2}r_{3,3}r_{3,4} + \\ & 36r_{1,1}r_{1,4}r_{2,3}^2r_{3,4} + 36r_{1,1}r_{1,4}r_{2,3}r_{2,4}r_{3,3} + 18r_{1,2}^2r_{1,3}r_{3,3}r_{4,4} + \\ & 36r_{1,2}^2r_{1,3}r_{3,4}^2 + 36r_{1,2}^2r_{1,4}r_{3,3}r_{3,4} + 36r_{1,2}r_{1,3}^2r_{2,3}r_{4,4} + \\ & 72r_{1,2}r_{1,3}^2r_{2,4}r_{3,4} + 144r_{1,2}r_{1,3}r_{1,4}r_{2,3}r_{3,4} + 72r_{1,2}r_{1,3}r_{1,4}r_{2,4}r_{3,3} + \\ & 36r_{1,2}r_{1,4}^2r_{2,3}r_{3,3} + 6r_{1,3}^3r_{2,2}r_{4,4} + 12r_{1,3}^3r_{2,4}^2 + \\ & 36r_{1,3}^2r_{1,4}r_{2,2}r_{3,4} + 72r_{1,3}^2r_{1,4}r_{2,3}r_{2,4} + 18r_{1,3}r_{1,4}^2r_{2,2}r_{3,3} + \\ & 36r_{1,3}r_{1,4}^2r_{2,3}^2.\end{aligned}$$

Hence, we calculate each entry $\Gamma_{ij,kl}$ in Γ_{VM} by evaluating the expectation of each monomial by the Isserlis-formula, as exemplified above, and then taking the weighted sum in eq. (5). A more formal description of the proposed procedure is presented in Appendix B. There is a large number of expectations to be calculated as the dimensionality increases. The expectation of the monomials in eq. (6) was therefore computed by the use of an efficient formula proposed by Kan (2008). Assuming that $n = n_i + n_j + n_k + n_l$ is even, the formula is

$$\mathbb{E}X_i^{n_i}X_j^{n_j}X_k^{n_k}X_l^{n_l} = \frac{1}{(\frac{n}{2})!}\sum_{v_i=0}^{n_i}\sum_{v_j=0}^{n_j}\sum_{v_k=0}^{n_k}\sum_{v_l=0}^{n_l}(-1)^{\sum_{i=1}^4 v_i}\binom{n_i}{v_i}\binom{n_j}{v_j}\binom{n_k}{v_k}\binom{n_l}{v_l}\left(\frac{h'Rh}{2}\right)^{\frac{n}{2}}, \quad (7)$$

where $h = [\frac{n_i}{2} - v_i, \frac{n_j}{2} - v_j, \frac{n_k}{2} - v_k, \frac{n_l}{2} - v_l]$. For further implementation details, the reader is referred to the R code in Appendix A.

For the IG approach described by Foldnes and Olsson (2016), obtaining Γ is computationally simpler and faster than the above procedure proposed for VM. The IG transform departs from the exact same specifications as the VM transform: pre-specified univariate skewnesses and kurtoses, and a target covariance matrix. The observed variables are represented as $Y_i = \sum_{j=1}^s a_{ij}X_j$, for $i = 1, \ldots, p$, where the a_{ij} are constant scalars

chosen such that Y has the target covariance matrix. The X_j, $j = 1, \ldots s$, are mutually independent random variables referred to as IG variables. Data simulation is done by drawing random samples for the X_j. Each X_j is specified to follow a univariate distribution, whose skewness and kurtosis is carefully chosen so that Y has the pre-specified skewness and kurtosis in each marginal. The IG transform is more flexible than the VM transform, since the user may specify which distribution, up to the skewness and kurtosis constraints, X_j may be drawn from. Also, the choice of coefficients a_{ij} lends flexibility. For instance, in the case $s = p$ these coefficients are elements of some (not unique) square root matrix $A = (a_{ij})$ of Σ. The independence among the X_j allows for the following closed-form expression for Γ under an IG distribution (Browne & Shapiro, 1988, Theorem 2.1):

$$\Gamma_{IG} = \Gamma_N + K_p' \tilde{A} C \tilde{A}' K_p. \tag{8}$$

Here $\tilde{A}$ is the $p^2 \times s$ matrix whose jth column is $a_j \otimes a_j$, with a_j being the j-th column vector of A. C is the diagonal matrix obtained from the excess kurtosis of the IG variables.

Finally, we remark that both the VM and IG distributions are members of a single distributional class for which Γ is computable. To the best of our knowledge, this new class of non-normal distributions has not been discussed before in the context of VM or IG distributions. We refer the interested reader to Appendix C.

Describing Γ and how it relates to standard errors and test statistics

Having obtained Γ in a specific simulation condition, we next consider how to investigate this large matrix and its impact on SEM procedures. For instance, with 11 indicators, there are $11 \cdot 12/2 = 66$ non-duplicated variances and covariances. Hence Γ will contain no less than $66 \cdot 67/2 = 2211$ non-duplicated elements. We can not therefore simply list all the elements of Γ in a paper, except as a data-file in an online appendix. One solution is to list only the $p(p+1)/2$ diagonal elements of Γ, which conveys information about the large-sample variance of the variances and covariances of the simulated data. Another option is to use graphical tools. We propose to plot each element of Γ against its

corresponding multivariate normal value as a benchmark. That is for each element $\Gamma_{ij,kl}$ of Γ, we form a pair $(\Gamma_{N,ij,kl}, \Gamma_{ij,kl})$, where $\Gamma_{N,ij,kl}$ denotes the corresponding (ij, kl)th element of Γ_N, i.e., the asymptotic covariance matrix of s_n that would be valid if we simulated normal data. Points lying above the line $y = x$ will indicate higher variance and covariance for the simulated condition, relative to the multivariate normal case. The further points are scattered away from the "normal" line $y = x$, the more severe the non-normality is, in terms of the variability of the sample covariance matrix.

The asymptotic covariance matrix Ω of the estimated parameters may be calculated once Γ is available. The other necessary elements in eq. (1) are also obtainable, since we have already specified the simulation condition.

Likewise, the asymptotic behaviour of T_{ML} (and derived test statistics like the Satorra-Bentler scaled statistic) can be exactly calculated from eq. (2), by calculating U in the population and extracting the non-zero eigenvalues of $U\Gamma$. The CDF of the resulting mixture of chi-square variates is obtained by using Imhof's method (Imhof, 1961), as e.g. available in the package CompQuadForm (Duchesne & De Micheaux, 2010). Hence the exact CDF and density curve of, e.g., T_{ML} is available, and we may calculate asymptotically exact Type I error rates for T_{ML} in each simulated condition.

Illustration 1: Moderate non-normality

Consider the structural equation model discussed by Bollen (1989), whose path diagram is sketched in Figure 1. There are 11 observed variables and the model has 35 degrees of freedom. The data-generating model was obtained by fixing all factor loadings and all unique variances to unity, and all five residual covariances to 0.2. In addition the regression of dem65 on dem60 and ind60 was fixed to 0.5 and 0.2, respectively, while the regression of dem60 on ind60 was fixed to 0.8.

Let us assume that a researcher wants to simulate data for this population model, in order to study how non-normality influences NTML estimation. The researcher specifies

moderate non-normality, by requesting skewness equal to one and kurtosis equal to five in all 11 marginal distributions. We calculated Γ_N, Γ_{VM} and Γ_{IG}, the latter from a Cholesky decomposition of Σ. Note that Γ_{VM} and Γ_{IG} are associated with distributions sharing the same covariance matrix and the same marginal skewness and kurtosis. This means that only 11 of the diagonal elements of Γ_{VM} and Γ_{IG} are forced to be equal, namely those corresponding to $\Gamma_{ii,ii}$ for $i = 1, \ldots, 11$. In fact the 2200 other elements of Γ_{VM} and Γ_{IG} are not pairwise equal.

In order to confirm that our proposed procedure for calculating Γ_{VM} was correct, we conducted both visual and formal testing. As a visual confirmation, we simulated 500 samples from the VM transform at sample sizes $n = 10^3, 10^4, 10^5$ and 10^6. For each simulated sample we computed the mean absolute percentage error (MAPE) as a distance measure between the empirical asymptotic covariance matrix $\hat{\Gamma}$ and Γ_{VM}. MAPE is the mean value of $100 \cdot \frac{|\hat{\gamma}-\gamma|}{\gamma}$ where $\hat{\gamma}$ and γ are corresponding elements of $\hat{\Gamma}$ and Γ_{VM}, respectively. With increasing sample size MAPE clearly decreases toward zero, as shown in Figure 2, indicating that $\hat{\Gamma}$ converges toward Γ_{VM} as n increases.

A formal test for the correct specification of Γ_{VM} may be based on the observation that, under the null hypothesis of correct Γ_{VM}, the quadratic form $W = n(s-\sigma)'\Gamma_{VM}^{-1}(s-\sigma)$ will converge in distribution to a chi-square distribution with 66 degrees of freedom. Here s and σ refer to the sample and target covariance matrices, respectively. Whether W has the chi-square distribution may be tested by the Kolmogorov-Smirnov (KS) test. We simulated 10000 samples each of sample size $n = 10^4$. The KS test statistic $D = 0.0077$ corresponds to a p-value of 0.59, which suggests that W in large samples indeed follows the postulated chi-square distribution. We hence have visual and formal support for the correctness of Γ_{VM}.

Note that even for a large sample size of $n = 10^6$, $\hat{\Gamma}$ has considerable variability. Across the 500 samples, each of size $n = 10^6$, the MAPE was 1.6%, with a standard deviation of 0.46%, see Figure 2. This means that the elements in $\hat{\Gamma}$ may differ substantially

from their corresponding population values in Γ_{VM}. Replacing the true Γ_{VM} by some large-sample estimate $\hat{\Gamma}$ is therefore not advisable, given the high degree of variability in $\hat{\Gamma}$ even at large sample sizes. This variability further increases with increasing non-normality in the data. For further analytical investigations on the difficulty of approximating Γ_{VM} using a very large sample, we refer the interested reader to Appendix D.

In the left- and right-hand side panels in Figure 3 we plotted the elements of Γ_{VM} and Γ_{IG}, respectively, against the corresponding elements for the normal case Γ_N. Due to symmetry, in each of the Γ many elements are duplicated. In fact, the number of unique elements in each of Γ_N, Γ_{VM} and Γ_{IG} is respectively equal to 138, 250 and 692. Therefore the scatterplots contain many overplotted points. Nevertheless, the panels indicates a large degree of non-normality, with all elements in both the VM and IG non-normal conditions taking higher values than their corresponding elements for the multivariate normal distribution.

The relation between the two non-normal conditions might also be investigated by plotting VM elements against IG elements, as shown in Figure 4. The plot indicates that the elements of Γ_{VM} are generally larger than the elements of Γ_{IG}, and hence that the VM distribution induces more variability in the sample variances and covariances than does IG.

Having explored the degree of non-normality as reflected by Γ, we next investigate how the non-normality is propagated into the variance of NTML parameter estimates in eq.(1). We calculated the asymptotic covariance matrices under the three distributional conditions of multivariate normality, VM and IG. In practice, only the variance of each parameter estimate are used in SEM analysis, which is located on the diagonal of Ω. In total there are 31 free parameters. In Table 1 the asymptotic variances of $\sqrt{n}\hat{\theta}$ are listed under multivariate normality, VM and IG conditions. It is clear that the NTML parameter estimates have larger variability under non-normality. Also, the asymptotic variances of NTML estimates differ substantially between VM and IG. One extreme case is the unique variance associated with y_8, which has a variance of 4.38 under a normal distribution, 10.29

under the VM distribution, and 34.64 under the IG distribution.

Finally, we consider the asymptotic distribution of T_{ML} under the two non-normal conditions. The non-zero eigenvalues of $U\Gamma_{VM}$ and $U\Gamma_{IG}$ are listed in Table 2. The corresponding density functions for T_{ML} under VM and IG may be constructed from these eigenvalues, using the method of Imhof, and are depicted in Figure 5. Under both VM and IG, it is clear that T_{ML} is inflated relative to the reference χ^2 distribution. Also, the inflation is more pronounced under VM. Asymptotic Type I error rates under VM and IG may also be obtained. Using T_{ML} leads to asymptotic rejection rates under VM and IG of 0.268 and 0.132, respectively.

Illustration 2: Severe non-normality

In this section we revisit the three-factor oblique model used by Curran et al. (1996), under the severe non-normality condition of skewness 3 and kurtosis 21. Each factor has three indicators. Population parameters consisted of factor loadings (fixed to 0.7), uniquenesses (0.51), interfactor correlations (0.3) and factor variances (1).

In the left- and right-hand side panels in Figure 6 we plotted the elements of Γ_{VM} and Γ_{IG}, respectively, against the corresponding elements for the normal case Γ_N. The panels illustrate strong deviation from non-normality, with values in the asymptotic covariance matrix of both VM and IG exceeding the corresponding values under multivariate normality.

We calculated Ω_N, Ω_{VM} and Ω_{IG} from eq. (1), with diagonal elements tabulated in Table 3. The high degree of symmetry in the model, and the homogeneous skewness and kurtosis conditions across the marginals, lead to duplication of elements in Γ_{VM} and to symmetry in Ω_{VM}, while the choice of an asymmetric A yields less symmetry in Γ_{IG} and Ω_{IG}. As was the case for the previous model, both VM and IG induce much larger variability in the NTML estimates compared to underlying multivariate normality. There are also large differences between the variability of specific parameter estimates between IG

and VM. For instance, the residual variances are estimated with very poor precision using NTML under IG. The residual variance associated with x_6, for instance is estimated with a large-sample variance under IG that is 2.5 higher than the large-sample variance under VM, and 17 times higher than the large-sample variance under multivariate normality.

The asymptotic distribution of T_{ML} under multivariate normality, the VM distribution and the IG distribution are depicted in Figure 7. As was the case for Illustration 1, there is a clear inflation of T_{ML} under non-normality, and it is under the VM distribution that T_{ML} performance is most deteriorated, compared to the IG distribution. The Type I error rates of VM and IG are, respectively, 0.489 and 0.108, calculated at the $\alpha = 0.05$ significance level.

Concluding remarks

In a proper reporting of simulation studies the population distribution should be defined explicitly (Boomsma, 2013, p. 523). A strict interpretation of this statement would demand the full cumulative probability distribution to be specified in each experimental condition. However, such a level of precision is hardly feasible or desirable. In current SEM simulation practice often a much lower level of precision is attained, by specifying univariate skewness and kurtosis values. In the present study we demonstrated that this is inadequate, by showing that the performance of SEM procedures vary widely among distributions that match specifications based on skewness and kurtosis. We propose that a reasonable middle-ground between full CDF specification and specification of only univariate skewness and kurtosis may be achieved by reporting the asymptotic covariance matrix Γ, or some important aspects of it. In a sense, Γ provides a fully multivariate measure of skewness and kurtosis by providing all multivariate fourth-order moments. The asymptotic covariance matrix is pivotal in SEM asymptotics, governing both standard errors and fit statistics. We have proposed a procedure to calculate Γ exactly under the popular Vale-Maurelli simulation method.

Even in the case where a researcher is primarily interested in the small-sample performance of estimators and test statistics the information contained in Γ is relevant, since it is essential for understanding the data-generating process that generates the small samples. That is, Γ contains exact population-level information that serves as a useful contrast to the observed finite-sample performance, accentuating the difference betweeen small and large sample sizes, while controlling for the degree of non-normality in the data.

The availability of Γ enables us to separate finite-sample effects from the effects of the underlying non-normality encoded in Γ. This means that analytical investigations, not relying on simulations, may be performed by calculating Γ under various distributions and analyse the effects on standard errors and fit statistics. This comes at the expense that our findings, although not contaminated by the inherent sampling variability of simulation studies, is only valid asymptotically. Γ may also be useful when several models are considered simultaneously, since it provides information about the degree of non-normality that is separate from the model. In such situations, Γ gives a measure of non-normality that is invariant to the details of the models.

A practical use of Γ is to include standard errors and rejection rates obtained from it as the final row when tabulating simulation results. Such tables are abundant in simulation studies, and typically contain rows of increasing sample sizes n. We suggest adding a row denoted by $n = \infty$ at the end of each table. In addition to giving interesting theoretical information, such a row eliminates the need to simulate the very large sample sizes often used to emulate $n = \infty$. The practice of including an unrealistically large sample size is time-consuming, and also yields imprecise information, especially under non-normality. For instance, we saw in Illustration 1 that using $\hat{\Gamma}$ to approximate Γ with a $n = 10^6$ sample on average resulted in a mean absolute percentage error of 1.6%.

The procedures proposed in the present article have several limitations. We first note that Γ is obtainable when data are simulated using either the VM or the IG transforms, or by the more general distribution described in Appendix C. For other non-normal data

simulation methods that control univariate kurtosis and skewness we are unaware of methods for obtaining Γ. A second limitation concerns the long running times of our prototypical implementation for calculating Γ under VM. With more than say, 15-20 observed variables the current implementation has a very long run time. We hope and believe that our code can be improved to substantially reduce running time by taking symmetry considerations into account. For dimensionality above 15-20 the IG approach is still viable, since this method allows for very fast computation of Γ.

Finally, we remark that being able to calculate Γ is a clear advantage for a simulation method due to its relation to central quantities in SEM. The fact that we have identified Γ in the context of VM therefore adds additional value to this simulation method, which now can be used with greater precision. Also, our analysis points to the need to identify calculation methods for Γ also in other non-normal simulation methods.

References

Astivia, O. L. O., & Zumbo, B. D. (2015). A cautionary note on the use of the vale and maurelli method to generate multivariate, nonnormal data for simulation purposes. *Educational and Psychological Measurement*, *75*(4), 541–567.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley. doi: 10.1002/9781118619179

Boomsma, A. (2013). Reporting monte carlo studies in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(3), 518–540.

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

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83.

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*, *41*(2), 193–208.

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*(1), 16-29.

Duchesne, P., & De Micheaux, P. L. (2010). Computing the distribution of quadratic forms: Further comparisons between the liu–tang–zhang approximation and exact methods. *Computational Statistics & Data Analysis*, *54*(4), 858–862.

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

Foldnes, N., & Grønneberg, S. (2015). How general is the Vale–Maurelli simulation approach? *Psychometrika*, *80*(4), 1066–1083.

Foldnes, N., & Olsson, U. H. (2016). A simple simulation technique for nonnormal data

with prespecified skewness, kurtosis, and covariance matrix. *Multivariate behavioral research*, *51*(2-3), 207–219.

Imhof, J. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, *48*(3/4), 419–426.

Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, *12*(1/2), 134–139.

Kan, R. (2008). From moments of sum to moments of product. *Journal of Multivariate Analysis*, *99*(3), 542–554.

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, *54*(1), 131–151.

Shapiro, A. (1983). Asymptotic distribution theory in the analysis of covariance structures. *South African Statistical Journal*, *17*(1), 33–81.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*(3), 465–471.

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*, 831–853.

Parameter	NORM	VM	IG
ind60 =~ x2	3.29	5.77	6.62
ind60 =~ x3	3.29	5.77	5.90
dem60 =~ y2	1.66	3.53	3.02
dem60 =~ y3	1.55	3.41	2.79
dem60 =~ y4	1.66	3.53	2.74
dem65 =~ y6	1.76	3.73	3.11
dem65 =~ y7	1.62	3.55	2.96
dem65 =~ y8	1.76	3.73	2.87
dem60 ~ ind60	3.20	5.34	5.21
dem65 ~ ind60	3.64	4.33	3.82
dem65 ~ dem60	2.22	3.27	2.58
y1 ~~ y5	1.77	2.44	2.00
y2 ~~ y4	2.38	3.85	4.11
y2 ~~ y6	1.39	1.95	1.86
y3 ~~ y7	1.77	2.44	2.91
y4 ~~ y8	1.39	1.95	3.11
y6 ~~ y8	2.55	4.12	5.11
x1 ~~ x1	4.05	9.53	6.84
x2 ~~ x2	4.05	9.53	11.18
x3 ~~ x3	4.05	9.53	18.69
y1 ~~ y1	3.74	9.33	7.70
y2 ~~ y2	4.16	9.94	18.25
y3 ~~ y3	3.74	9.33	28.11
y4 ~~ y4	4.16	9.94	33.80
y5 ~~ y5	3.89	9.57	8.01
y6 ~~ y6	4.38	10.30	19.75
y7 ~~ y7	3.89	9.57	29.41
y8 ~~ y8	4.38	10.30	34.64
ind60 ~~ ind60	8.05	14.90	20.84
dem60 ~~ dem60	6.79	13.39	19.92
dem65 ~~ dem65	5.82	12.16	18.15

Table 1

Illustration 1. Asymptotic variance of $\sqrt{n} \cdot \hat{\theta}$ *for NTML under three distributions: NORM = multivariate normal. VM = Vale-Maurelli. IG = independent generator. Both VM and IG have marginal skewness* 1 *and kurtosis* 5. *=~ : factor loading. ~ : regression coefficient. ~~ : residual variance or covariance, or residual error variance.*

1.60	1.59	1.38	1.34	1.33	1.32	1.29	1.29	1.29	1.29	1.26	1.26
1.26	1.26	1.26	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.23	1.23
1.17	1.16	1.16	1.16	1.16	1.15	1.15	1.15	1.15	1.15	1.15	
2.22	1.99	1.68	1.54	1.49	1.07	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

Table 2

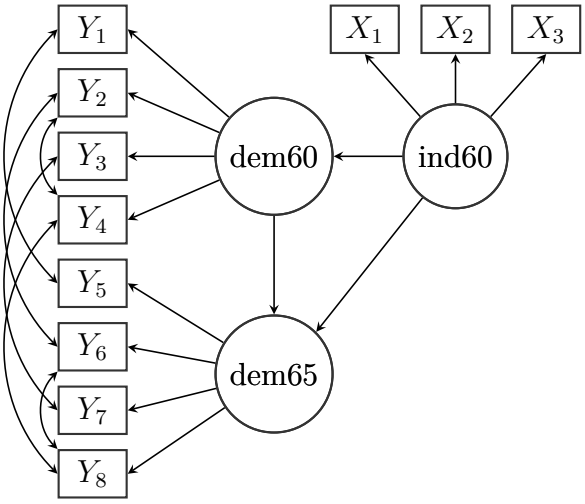
Illustration 1. Non-zero eigenvalues of $U\Gamma_{VM}$ *(upper three rows) and* $U\Gamma_{IG}$ *(lower three rows). Rounded to 2 decimal places.*

Parameter	NORM	VM	IG
F1 =~ x2	1.96	8.45	9.01
F1 =~ x3	1.96	8.45	8.40
F2 =~ x5	1.96	8.45	7.94
F2 =~ x6	1.96	8.45	7.70
F3 =~ x8	1.96	8.45	7.24
F3 =~ x9	1.96	8.45	7.24
x1 ~~ x1	1.22	8.58	4.58
x2 ~~ x2	1.22	8.58	9.71
x3 ~~ x3	1.22	8.58	20.04
x4 ~~ x4	1.22	8.58	4.74
x5 ~~ x5	1.22	8.58	10.62
x6 ~~ x6	1.22	8.58	21.07
x7 ~~ x7	1.22	8.58	4.92
x8 ~~ x8	1.22	8.58	11.28
x9 ~~ x9	1.22	8.58	21.80
F1 ~~ F1	9.09	43.16	65.03
F2 ~~ F2	9.09	43.16	58.76
F3 ~~ F3	9.09	43.16	54.63
F1 ~~ F2	2.14	4.34	4.58
F1 ~~ F3	2.14	4.34	4.54
F2 ~~ F3	2.14	4.34	4.03

Table 3

Illustration 2. Asymptotic variance of $\sqrt{n} \cdot \hat{\theta}$ *for NTML under three distributions: NORM = multivariate normal. VM = Vale-Maurelli. IG = independent generator. Both VM and IG has marginal skewness and kurtosis of 3 and 21, respectively. =~ : factor loading. ~ : regression coefficient. ~~ : residual variance or covariance, or residual error variance.*

Figure 1. Illustration 1. Bollen's political democracy model. dem60: Democracy in 1960. dem65: Democracy in 1965. ind60: Industrialisation in 1960.



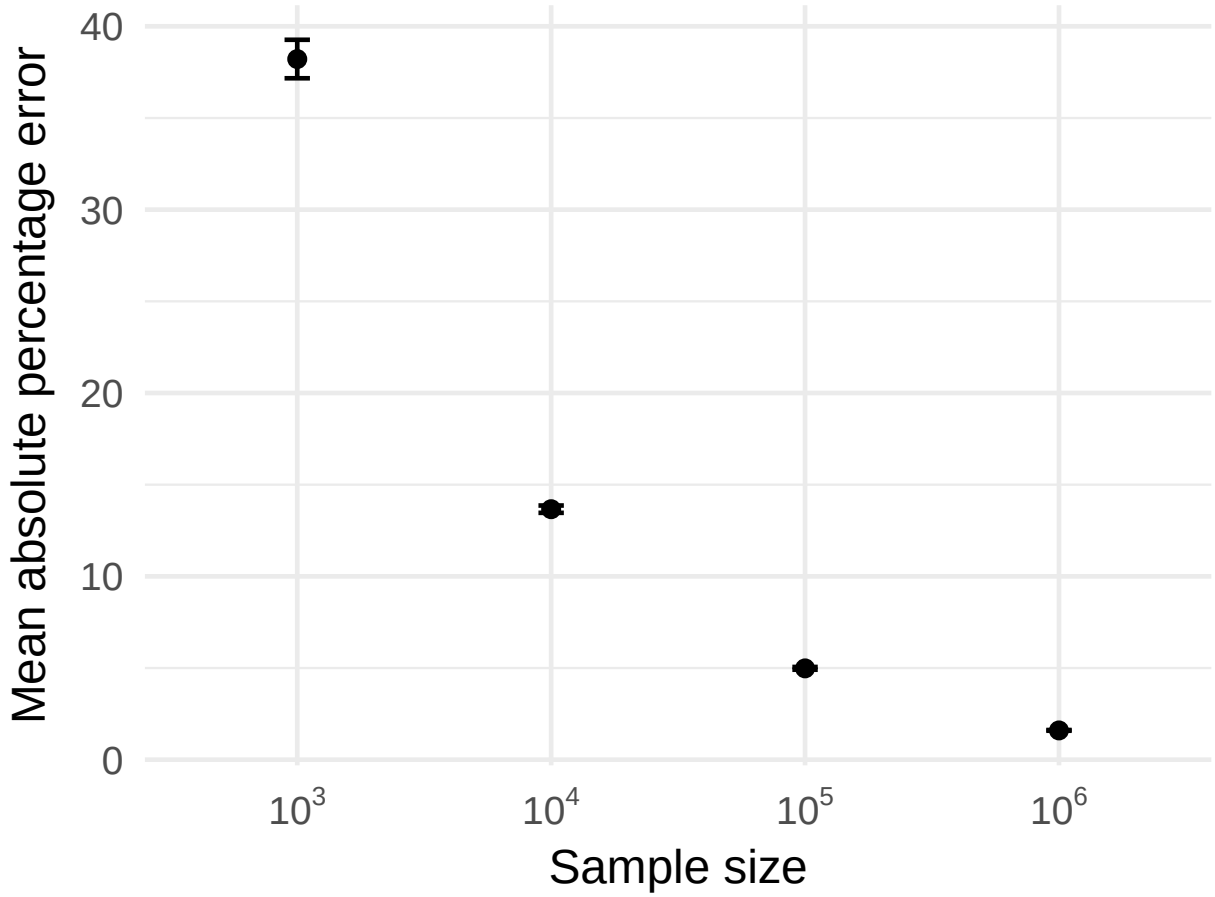


Figure 2. Illustration 1. The mean absolute percentage error between $\hat{\Gamma}$ and Γ_{VM}. Error bars depict 95% confidence intervals.

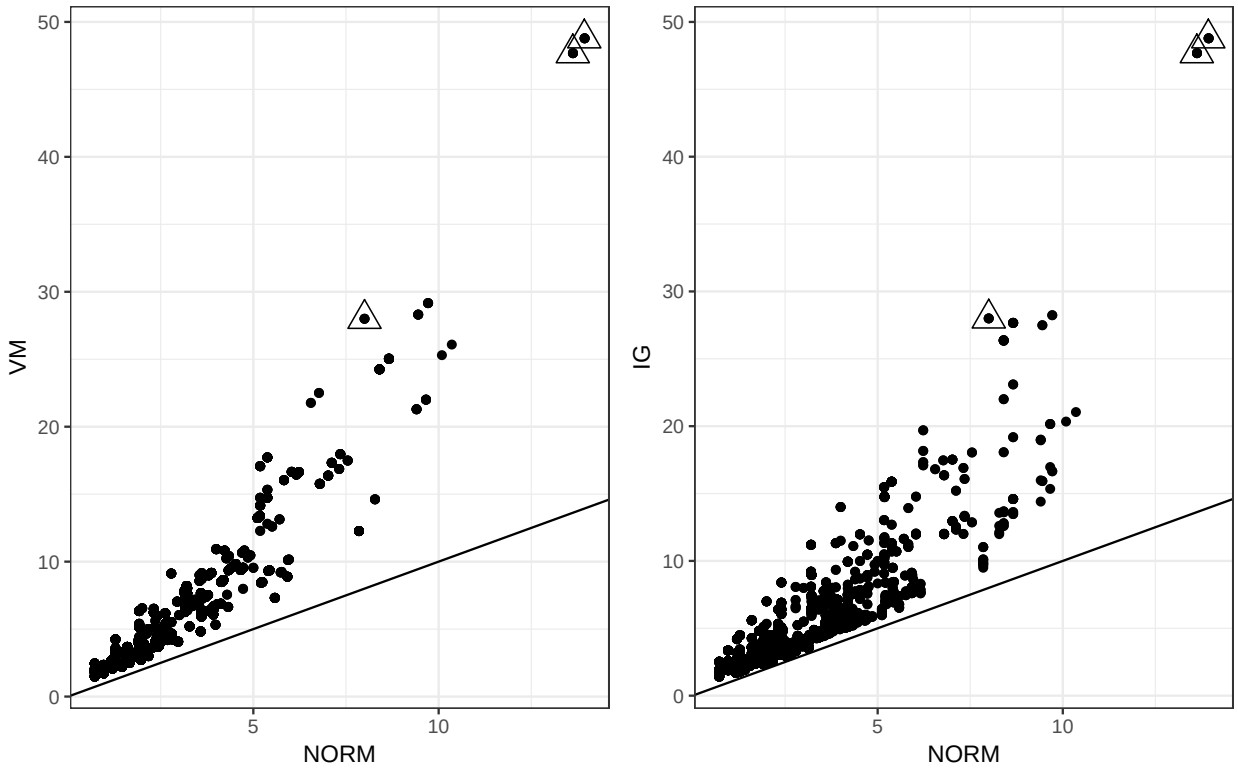


Figure 3. Illustration 1. Left panel: Plotting elements of Γ_{VM} against Γ_N. Right panel: Plotting elements of Γ_{IG} against Γ_N. The straight line represents $y = x$. The three points marked with a $\triangle$ occur in both panels, and correspond to univariate kurtosis conditions satisfied both for VM and IG.

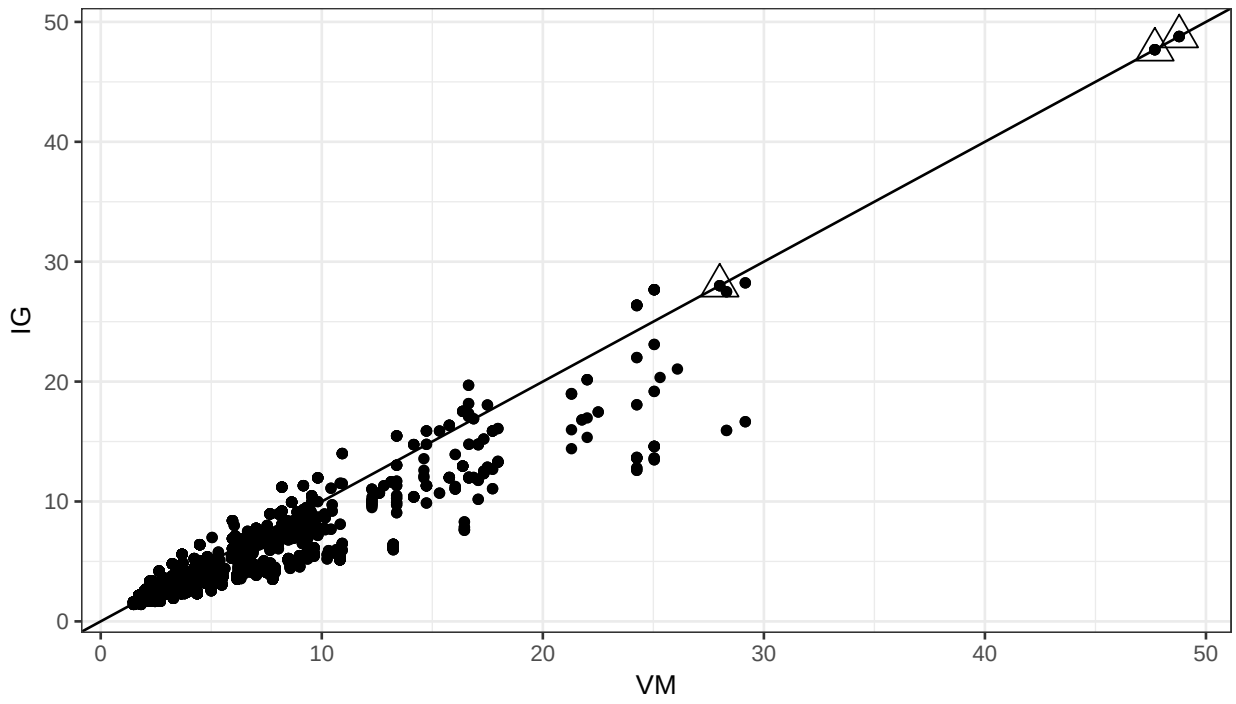


Figure 4. Illustration 1. Plotting elements of Γ_{VM} against Γ_{IG}. The straight line represents $y = x$. The three points marked with a $\triangle$ correspond to univariate kurtosis conditions satisfied both for VM and IG.

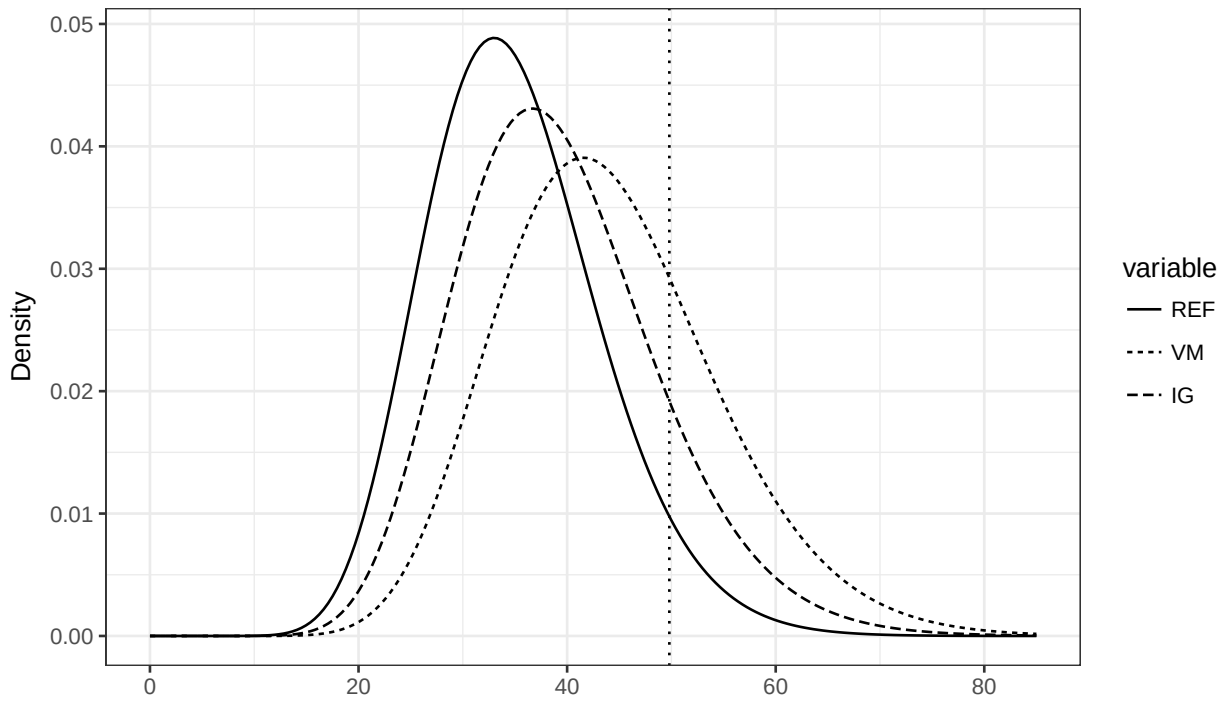


Figure 5. Illustration 1. Density curves of T_{ML} under multivariate normality (REF), VM and IG distributions. The vertical line represents the critical value 49.8 of the chi-square test at the $\alpha = 0.05$ significance level.

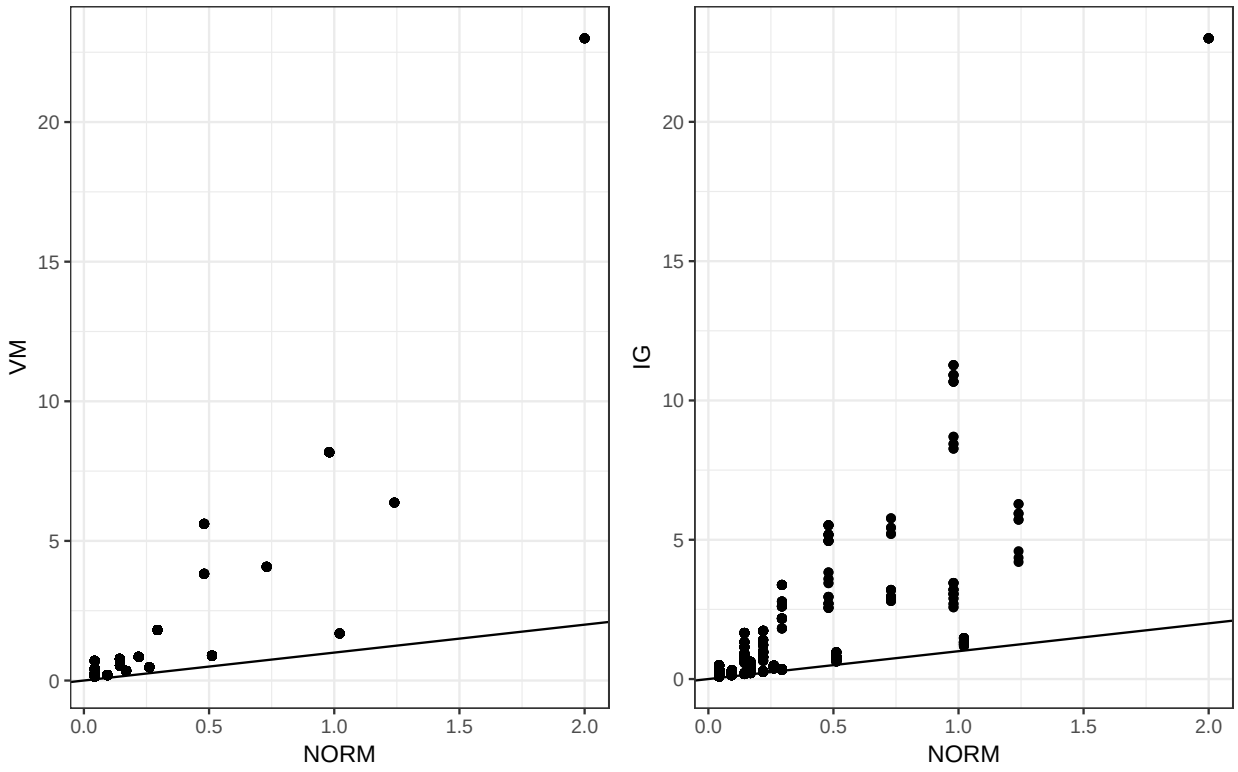


Figure 6. Illustration 2. Left panel: Plotting elements of Γ_{VM} against Γ_N. Right panel: Plotting elements of Γ_{IG} against Γ_N. The straight line represents $y = x$.

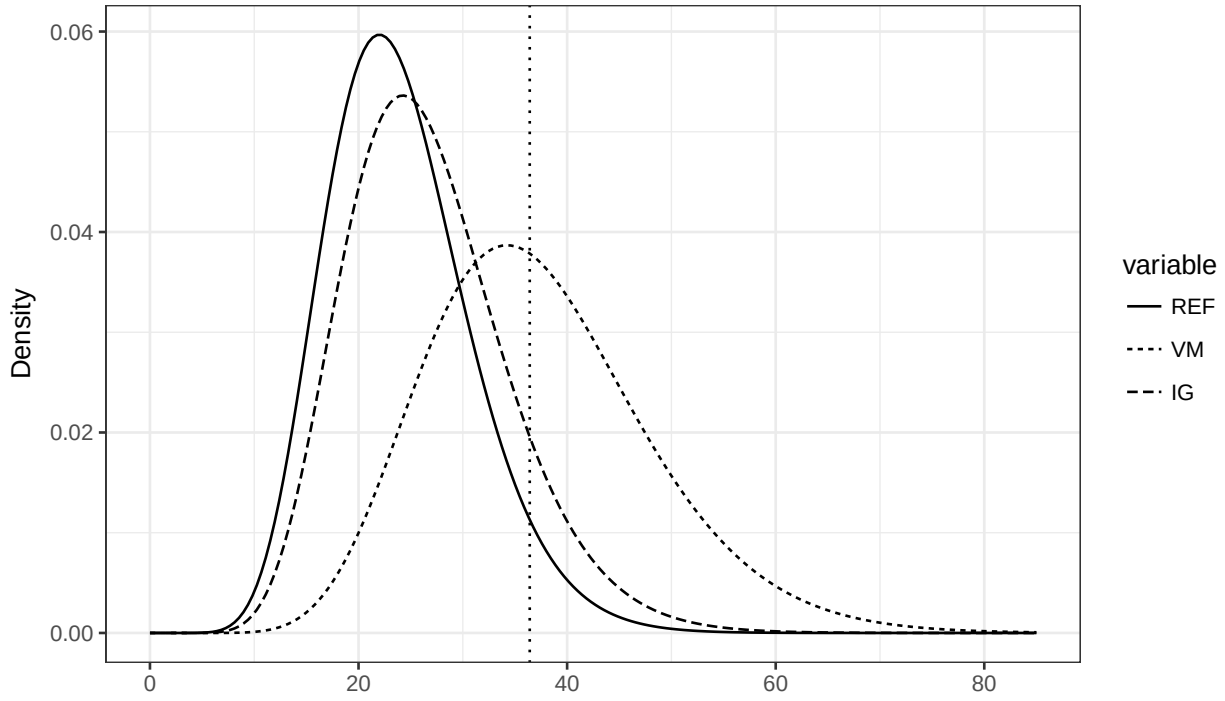


Figure 7. Illustration 2. Density curves of T_{ML} under multivariate normality (REF), VM and IG distributions. The vertical line represents the critical value 36.4 of the chi-square test at the $\alpha = 0.05$ significance level.

Appendix A

R code

```
rm(list = ls())
library(MASS)
library(emulator)#quad.form, may be replaced by better
library(mpoly)
library(CompQuadForm)
library(nleqslv)
library(PearsonDS)
library(psych)
library(lavaan)
library(ggplot2)
library(reshape)
options(scipen = 999)
# HELP FUNCTIONS
getSummand <- function(s,v,sigma){
        h=s/2-v
        res = prod(choose(s,v)) * (quad.form(sigma, h)/2)^(sum(s)/2)
        if(sum(v) %% 2 == 0){
        return(res)
} else {
        return(-res)
}
}
momentCalc <- function(sigma, s) {# proposition 1 in Kan paper.
sSum <- sum(s)
if ((sSum %% 2) != 0) {
return(0)
}
uniqueVcombis = expand.grid(lapply(s, function(x) return(0:x)))
l <- apply(uniqueVcombis,1, getSummand, s=s, sigma=sigma)
return(sum(l)/factorial(sSum/2))
}
# The following function is source code from lavaan package!
getVM <- function(skewness, kurtosis, sigma0){
COR = cov2cor(sigma0)
nvar <- ncol(COR)
# check skewness
if(is.null(skewness)) {
SK <- rep(0, nvar)
} else if(length(skewness) == nvar) {
SK <- skewness
} else if(length(skewness) == 1L) {
SK <- rep(skewness, nvar)
} else {
stop("skewness has wrong length")
}
if(is.null(kurtosis)) {
KU <- rep(0, nvar)
} else if(length(kurtosis) == nvar) {
KU <- kurtosis
} else if(length(kurtosis) == 1L) {
KU <- rep(kurtosis, nvar)
```

```
} else {
stop("kurtosis has wrong length")
}
fleishman1978_abcd <- function(skewness, kurtosis) {
system.function <- function(x, skewness, kurtosis) {
        b.=x[1L]; c.=x[2L]; d.=x[3L]
        eq1 <- b.*b. + 6*b.*d. + 2*c.*c. + 15*d.*d. - 1
        eq2 <- 2*c.*(b.*b. + 24*b.*d. + 105*d.*d. + 2) - skewness
        eq3 <- 24*(b.*d. + c.*c.*(1 + b.*b. + 28*b.*d.) +
        d.*d.*(12 + 48*b.*d. + 141*c.*c. + 225*d.*d.)) - kurtosis
        eq <- c(eq1,eq2,eq3)
        sum(eq*eq) ## SS
}
out <- nlminb(start=c(1,0,0), objective=system.function,
scale=10,
control=list(trace=0),
skewness=skewness, kurtosis=kurtosis)
if(out$convergence != 0) warning("no convergence")
b. <- out$par[1L]; c. <- out$par[2L]; d. <- out$par[3L]; a. <- -c.
if(out$objective > 10^{-5}) {
        warning("!!!Not valid combination of skew/kurtosis")
}
return(c(a.,b.,c.,d.))
}

getICOV <- function(b1, c1, d1, b2, c2, d2, R) {
objectiveFunction <- function(x, b1, c1, d1, b2, c2, d2, R) {
        rho=x[1L]
        eq <- rho*(b1*b2 + 3*b1*d2 + 3*d1*b2 + 9*d1*d2) +
        rho*rho*(2*c1*c2) + rho*rho*rho*(6*d1*d2) - R
        eq*eq
}
out <- nlminb(start=R, objective=objectiveFunction,
scale=10, control=list(trace=0),
b1=b1, c1=c1, d1=d1, b2=b2, c2=c2, d2=d2, R=R)
if(out$convergence != 0) warning("no convergence")
rho <- out$par[1L]
rho
}

# create Fleishman table
FTable <- matrix(0, nvar, 4L)
for(i in 1:nvar) {
FTable[i,] <- fleishman1978_abcd(skewness=SK[i], kurtosis=KU[i])
}
#
ICOR <- diag(nvar)
for(j in 1:(nvar-1L)) {
for(i in (j+1):nvar) {
        if(COR[i,j] == 0) next
        ICOR[i,j] <- ICOR[j,i] <-
        getICOV(FTable[i,2], FTable[i,3], FTable[i,4],
        FTable[j,2], FTable[j,3], FTable[j,4], R=COR[i,j])
}
}
list(FTable*sqrt(diag(sigma0)), ICOR)
```

```
}

#represent monomial as a matrix. index in first column, powers in second. coeff and even powers are NA
strip.mono <- function(mon){
mon=mon[[1]]
mon=mon[-which(names(mon)=="coef")]
if (sum(mon) %% 2 !=0  )
return(NA)
my.mat= matrix(c(as.integer(gsub("X", "", names(mon))), mon), ncol  = 2)

if(nrow(my.mat)==0)
return(NA)#constant

return(matrix(my.mat[order(my.mat[,1]), ], ncol=2))
}

coef.mono <- function(mon){#
mon=mon[[1]]
if(length(mon)==1)
return(NA)#constant
if (sum(mon[-which(names(mon)=="coef")]) %% 2 !=0  )
return(NA)

return(mon[which(names(mon)=="coef")])
}
getGammaV2 <- function(FTable, ICOR, sigma0){# recalculates monomials
d=nrow(FTable)
list.of.polynoms <-  vector("list", d)
for (i in 1:d){
varname <- paste("X", i, sep="")
polynome <- mp(paste(FTable[i,1], "+", FTable[i,2], varname,
"+", FTable[i,3], varname,"^2 +", FTable[i,4], varname,"^3" ))
list.of.polynoms[[i]] <- polynome
}

vech_indx <- NULL
for(i in 1:d) {
for (j in (i:d)) {
        vech_indx <- rbind(vech_indx,c(j,i))
}
}
d.star <- nrow(vech_indx)
smallGamma <- matrix(nrow=d.star, ncol=d.star, 0)# fill in

for (u in (1:d.star)) {#
for (v in (u:d.star)) {
        i <- vech_indx[u,1]; j <- vech_indx[u,2]
        k <- vech_indx[v,1]; l <- vech_indx[v,2]

        polynomial <- list.of.polynoms[[i]]*
        list.of.polynoms[[j]]*
        list.of.polynoms[[k]]*
        list.of.polynoms[[l]]
        monos <- monomials(polynomial)
        res <- NULL
        for(q in 1:length(monos)){
```

```
                mon = monos[[q]][[1]]
                coef <- mon["coef"]
                if(length(mon)==1){
                        res <- c(res, coef)
                        next
                }# special case with only constant. the last monomial?

                vars <- names(mon)
                indx <- NULL
                power <- NULL
                for( name in vars){
                        if (name =="coef")
                        next
                        indx <- c(indx, as.integer(strsplit(name, "X")[[1]][2]))
                        power <- c(power, mon[name])
                }


                res <-c(res, coef*momentCalc(ICOR[indx, indx], power))
        }

        smallGamma[u,v] <- sum(res)-sigma0[i,j]*sigma0[k,l]
        smallGamma[v,u] <- smallGamma[u,v]
}
}
return(smallGamma)
}


# ILLUSTRATION 1
bollen.model <-
"# measurement model
ind60 =~ x1 + x2 + x3
dem60 =~ y1 + y2 + y3 + y4
dem65 =~ y5 + y6 + y7 + y8
# regressions
dem60 ~ start(0.8)*ind60
dem65 ~ start(0.2)*ind60 + start(0.5)*dem60
# residual correlations
y1 ~~ start(0.2)*y5
y2 ~~ start(0.2)*y4 + start(0.2)*y6
y3 ~~ start(0.2)*y7
y4 ~~ start(0.2)*y8
y6 ~~ start(0.2)*y8"

#target covariance matrix
fit= sem(bollen.model, data=NULL)
sigma0 = inspect(fit, "sigma.hat")

# Delta
Delta <- lavaan:::computeDelta(lavmodel = fit@Model)[[1]]

#Gamma normal
GN = 2*lav_matrix_duplication_ginv_pre_post( kronecker(sigma0, sigma0) )

V=solve(GN)
```

```
SW = solve(t(Delta) %*% V %*% Delta) #sandwich element

U= V-V%*% Delta %*% SW %*% t(Delta) %*%V

# VM transform
VM <- getVM(skewness=1, kurtosis=5, sigma0)
Fleishman <- VM[[1]]
IntCorr <- VM[[2]]

# Calculate Gamma VM.
GammaVM <-  getGammaV2(Fleishman,IntCorr, sigma0)

# ASYMPTOTIC COVARIANCE MATRIX OF PARAMETER ESTIMATES:
VMcov = SW %*% t(Delta) %*% V %*% GammaVM %*% V %*% Delta %*% SW

## Eigenvalues
VMeig = Re(eigen(U %*% GammaVM)$values)[1:35]

mycdf <- function(x, eig){
1-imhof(x, eig)$Qq
}

l=200; upper=85; stepsize = upper/l
x=seq(0,upper, length.out=l)
Yvm=sapply(x, mycdf, VMeig)

yvm=Yvm
for (i in 2:200){
yvm[i] = (Yvm[i]-Yvm[i-1])/stepsize
}

#density curve for T_ML
nominal =dchisq(x, df=35)
t= data.frame(x=x, REF=nominal, VM=yvm)

t=melt(t, measure.vars=2:3)
p=ggplot(t, aes(x, value))+geom_line(aes(linetype=variable))+theme_bw()+labs(x=NULL, y="Density")
p+geom_vline(xintercept=qchisq(0.95, df=35), linetype=21)
```

Appendix B

Algorithm for Γ under the VM transform

Algorithm 1 Calculating Γ_{VM}

Require: Target covariance matrix Σ, skewnesses s and kurtoses k

$p \leftarrow$ number of columns in Σ. $p^* \leftarrow p(p+1)/2$

for $i \leftarrow 1, \ldots, p$ **do**

Calculate Fleishman coefficients $\alpha_{i,1}, \alpha_{i,2}, \alpha_{i,3}$ and $\alpha_{i,4}$ to match skewness $s[i]$ and kurtosis $k[i]$

end for

Calculate intermediate correlations in R ▷ by VM method

Construct matrix V ▷ For ordering of elements.

for $i \leftarrow 1, \ldots, d$ **do**

for $j \leftarrow i, \ldots, d$ **do** Add row (j, i) to V

end for

end for ▷ V now has dim $p^* \times 2$

Initialise $\Gamma \leftarrow p^* \times p^*$ matrix

for $u \leftarrow 1, \ldots, p^*$ **do**

for $v \leftarrow u, \ldots, p^*$ **do**

$i \leftarrow V[u, 1]$, $j \leftarrow V[u, 2]$,$k \leftarrow V[v, 1]$,$l \leftarrow V[v, 2]$

polynomial $\leftarrow$ product of Fleishman polynomials i, j, k and l

$g \leftarrow 0$

for monomial in polynomial **do**

$g \leftarrow g + \mathbb{E}$(monomial) ▷ See eq. (7)

end for

$g \leftarrow g - \Sigma[i, j] \cdot \Sigma[j, i]$

$\Gamma[u, v] \leftarrow g$, $\Gamma[v, u] \leftarrow g$,

end for

end for

return Γ

Appendix C

A new class of distributions that embed both VM and IG

Suppose $U = (U_1, \ldots, U_d)'$ contains independent random variables $U_1, \ldots, U_d$. For a $p \times d$ matrix A, let

$$X = AU$$

be a p-dimensional random vector. By definition of matrix multiplication, we have

$$X_j = \sum_{i=1}^{d} a_{j,i} U_i.$$

Now let

$$Y_{VM-IG} = \begin{pmatrix} a_1 + b_1 X_1 + c_1 X_1^2 + d_1 X_1^3 \\ a_2 + b_2 X_2 + c_2 X_2^2 + d_2 X_2^3 \\ \vdots \\ a_p + b_p X_p + c_p X_p^2 + d_p X_p^3 \end{pmatrix}.$$

We are interested in calculating

$$\sigma_{ijkl} = \mathbb{E}(Y_i - \mu_i)(Y_j - \mu_j)(Y_k - \mu_k)(Y_l - \mu_l), \quad \mu_i = \mathbb{E}Y_i,$$

Firstly, let us notice that

$$\mu_j = \mathbb{E}Y_j = a_j + b_j \mathbb{E}X_j + c_j \mathbb{E}X_j^2 + d_j \mathbb{E}X_j^3.$$

Let us focus in on $\mathbb{E}X_j^2$. Since $X_j = \sum_{i=1}^{d} a_{j,i} U_i$, we have

$$\begin{aligned} \mathbb{E}X_j^2 &= \mathbb{E}\left[\left(\sum_{i=1}^{d} a_{j,i} U_i\right)\left(\sum_{k=1}^{d} a_{j,k} U_k\right)\right] \\ &= \mathbb{E}\sum_{i=1}^{d}\sum_{k=1}^{d} a_{j,i} U_i a_{j,k} U_k \\ &= \sum_{i=1}^{d}\sum_{k=1}^{d} a_{j,i} a_{j,k} \mathbb{E}U_i U_k \\ &= \sum_{i=1}^{d}\sum_{k=1}^{d} a_{j,i} a_{j,k} I\{i = j\} \mathbb{E}U_i^2 \\ &= \sum_{i=1}^{d} a_{j,i}^2 \mathbb{E}U_i^2. \end{aligned}$$

This type of argument also applied to σ_{ijkl}, but with more complex sums. For example,

$$\mathbb{E}[(Y_j - \mu_j)^4] = \mathbb{E}[(a_j + b_j X_j + c_j X_j^2 + d_j X_j^3 - \mu_j)^4]$$

can be expanded using again

$$X_j = \sum_{i=1}^{d} a_{j,i} U_i.$$

In general, calculating σ_{ijkl} amounts to calculating weighted sums of expectations of the form

$$\mathbb{E} \prod_{j=1}^{d} U_j^{\alpha_j}$$

for some integers $\alpha_1, \ldots, \alpha_d$. which by independence equals

$$\prod_{j=1}^{d} \mathbb{E} U_j^{\alpha_j}.$$

Assuming these expectations can be calculated (a modest assumption), this means we can calculate Γ associated with Y_{VM-IG}.

Appendix D

On the difficulty of approximating Γ_{VM} from a large simulated sample

In order to demonstrate the large variability inherent in estimating Γ_{VM}, even from a very large sample, let us consider the severe non-normality case in Illustration 2. Here Γ_{VM} is a 45-by-45 matrix, and we focus on a specific diagonal element of Γ_{VM}, namely the case $i = j = k = l = 1$. The argument we present may be extended to any element of Γ_{VM}, albeit with more complex algebraic manipulations. To obtain skewness of 3 and kurtosis of 21 a Fleishman solution is given by

$$Y = -0.2522994 + 0.4186101X - 0.2522994X^2 + 0.1475925X^3,$$

where $X \sim N(0,1)$. We wish to estimate $\sigma = \sigma_{1111}$ based on IID observations Y_i for $i = 1, 2, \ldots, n$. The natural estimator is

$$\hat{\sigma} = n^{-1}\sum_{i=1}^{n} Y_i^4 = n^{-1}\sum_{i=1}^{n}(-0.2522994 + 0.4186101X_i - 0.2522994X_i^2 + 0.1475925X_i^3)^4.$$

By algebraic manipulations, and by repeatedly applying the well-known formula $\mathbb{E}X^{2p} = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2p-1)$, it follows that

$$\begin{aligned}
\text{Var}\,(\hat{\sigma}) &= n^{-1}\text{Var}\,(Y_i^4) = n^{-1}\left(\mathbb{E}Y^8 - (\mathbb{E}Y^4)^2\right) \\
&= n^{-1}\mathbb{E}(-0.2522994 + 0.4186101X_i - 0.2522994X_i^2 + 0.1475925X_i^3)^8 \\
&- n^{-1}\left(\mathbb{E}(-0.2522994 + 0.4186101X_i - 0.2522994X_i^2 + 0.1475925X_i^3)^4\right)^2 \\
&= n^{-1}\left(815106 - 24^2\right) = n^{-1}814530.
\end{aligned}$$

In other words, in order to estimate $\mathbb{E}Y^4 = 24$ so that the variance of the estimator is less than one, a sample size of 814530 is required. Of course, estimating the population parameter $\mathbb{E}Y^4 = 24$ with a variance of one means that the approximation $\hat{\sigma}$ will most likely have poor quality. Moreover, there are more than one thousand elements of Γ_{VM} that must be approximated using sample averages. We may conclude that approximating all these elements to a high level of precision would require a sample size several orders of magnitude above 10^6, which is too large even for modern computers.