




## A Simple Simulation Technique for Nonnormal Data with Prespecified Skewness, Kurtosis, and Covariance Matrix

Njål Foldnes & Ulf Henning Olsson


**To cite this article:** Njål Foldnes & Ulf Henning Olsson (2016) A Simple Simulation Technique for Nonnormal Data with Prespecified Skewness, Kurtosis, and Covariance Matrix, *Multivariate Behavioral Research*, 51:2-3, 207-219, DOI: [10.1080/00273171.2015.1133274](https://doi.org/10.1080/00273171.2015.1133274)

**To link to this article:** <https://doi.org/10.1080/00273171.2015.1133274>

 View supplementary material 

 Published online: 25 Mar 2016.

 Submit your article to this journal 

 Article views: 1121

 View related articles 

 View Crossmark data 

 Citing articles: 13 View citing articles 

## A Simple Simulation Technique for Nonnormal Data with Prespecified Skewness, Kurtosis, and Covariance Matrix

Njål Foldnes and Ulf Henning Olsson

BI Norwegian Business School

### ABSTRACT

We present and investigate a simple way to generate nonnormal data using linear combinations of independent generator (IG) variables. The simulated data have prespecified univariate skewness and kurtosis and a given covariance matrix. In contrast to the widely used Vale-Maurelli (VM) transform, the obtained data are shown to have a non-Gaussian copula. We analytically obtain asymptotic robustness conditions for the IG distribution. We show empirically that popular test statistics in covariance analysis tend to reject true models more often under the IG transform than under the VM transform. This implies that overly optimistic evaluations of estimators and fit statistics in covariance structure analysis may be tempered by including the IG transform for nonnormal data generation. We provide an implementation of the IG transform in the R environment.

### KEYWORDS

Simulation; nonnormal multivariate data; kurtosis; asymptotic robustness; structural equation modeling

### Introduction

The use of Monte Carlo simulation to empirically assess estimators and goodness-of-fit statistics is widespread in structural equation modeling (SEM) research. The most common estimation method is normal-theory maximum likelihood (ML), which is the default choice in SEM software. Normal-theory ML also provides an associated goodness-of-fit test, which under ideal conditions of correct model specification, multivariate normality, and large sample size follows a chi-square distribution. However, in most real-world situations, data are drawn from nonnormal distributions (Micceri, 1989). It is therefore important to evaluate normality-based ML under realistic sample sizes and nonnormal distributions. In fact, for any estimation technique to be evaluated, normal-theory based or not, we need to generate data with methods that can provide many different kinds of multivariate nonnormal distributions. The tradition in SEM simulation is to set up a nonnormal data-generating process by specifying a given covariance matrix and given skewness and kurtosis in the univariate marginal distributions. This is natural since SEM is based on analysis of the covariance structure. However, in the nonnormal case, there are many distributions that have a given covariance matrix. Similarly, the skewness and kurtosis of a univariate distribution do not fully define a univariate distribution. Thus by specifying the covariance matrix, and, for each observed variable, the skewness and kurtosis values, there are potentially

many data-generating processes that can attain the target covariance matrix, skewnesses, and kurtosis. The data-generating process of Vale-Maurelli (Vale & Maurelli, 1983) has been the default choice for researchers conducting SEM simulation studies. However, the Vale-Maurelli (VM) transform represents only one of many data-generating processes that attain the prespecified covariance structure and univariate skewness and kurtosis of the observed variables.

The aim of the current article is to introduce a new simulation technique that generates data that are different in important ways from data obtained with the VM transform. We demonstrate that using the VM transform may result in a too-optimistic evaluation of normal-theory-based test statistics in SEM. A more moderate evaluation of the performance of these popular test statistics is obtained when data are generated by the simulation technique proposed in this article. For example, we show that at fixed levels of skewness and kurtosis in the observed variables, the performance of the ML chi-square goodness-of-fit test  $T_{ML}$  rejects true models more often when data are simulated with our proposed method compared to when data are generated with the VM transform. Hence, a researcher using only the VM transform may be led to conclude that  $T_{ML}$  performs quite well under nonnormality. However, such a conclusion is contingent on the special kind of nonnormality offered by the VM transform. A less optimistic evaluation of the appropriateness of  $T_{ML}$  under nonnormality may have resulted had

the data been generated by the new method instead of the VM transform. Researchers can more broadly evaluate the finite-sample performance of statistical techniques under nonnormality by employing this new simulation technique in addition to the VM transform.

Let us remark that although convenient, the traditional SEM approach of specifying the covariance matrix and some finite-order moments of the univariate marginal distributions does not fully identify a multivariate distribution. There is a more general concept of dependency among random variables than the covariance matrix; namely, the *copula*. The importance of the copula stems from the following unique decomposition of any continuous multivariate distribution into two parts:

$$\begin{aligned} &\text{Multivariate distribution} \\ &= \text{Univariate marginal distributions} + \text{Copula.} \end{aligned}$$

The first part is a specification of the univariate marginal distributions of the multivariate distribution; that is, the distribution of each observed variable. The second part pertains to the dependence structure among these observed variables. This structure may be modeled by the copula of the distribution, which describes the dependency among the observed variables. For example, the copula associated with multivariate normality is defined as follows. Given a random multivariate normal vector  $\mathbf{X} = (X_1, \dots, X_d)'$ , we may define a new vector  $\mathbf{U}_X = (\phi(\frac{X_1 - \mu_1}{\sigma_1}), \dots, \phi(\frac{X_d - \mu_d}{\sigma_d}))'$  obtained from  $\mathbf{X}$  by applying the standard normal cumulative distribution function  $\phi$  to each standardized univariate marginal in  $\mathbf{X}$ . The copula of  $\mathbf{X}$ , known as the normal copula, is the cumulative distribution function of the vector  $\mathbf{U}_X$ . This procedure may be applied to any nonnormal random vector  $\mathbf{Y}$ , whose copula is the distribution of the vector  $\mathbf{U}_Y = (F_1(Y_1), \dots, F_d(Y_d))'$ , where we have applied the cumulative distribution function  $F_i(y_i) = P(Y_i \leq y_i)$  to each  $Y_i$ . See Vuolo (2015) for an introduction to copulas in sociology.

Note that multivariate normality occurs only when all univariate marginal distributions are normally distributed and the associated copula is normal as defined above. Violation of the normality assumption can occur if at least one univariate marginal distribution is nonnormal, if the copula is nonnormal, or both. For example, it may be the case that all univariate marginal distributions are normally distributed but that the copula is not normal, which will result in a nonnormal multivariate distribution with normal univariate marginal distributions. A practitioner investigating the observed variables one by one will in such a case get the false impression that the data are normally distributed. Conversely, it is possible to have

nonnormal univariate observed variables but with a normal copula. In this latter case, the distribution will appear highly nonnormal when investigating each observed variable, but its dependence structure—that is, its copula—is identical to the multivariate normal case.

In fact, this latter situation closely describes the VM transform, which is implemented in software packages like Mplus (Muthén & Muthén, 2012), EQS (Bentler, 2006) and lavaan (Rosseel, 2012). The VM transform delivers distributions with prespecified univariate skewness and kurtosis and a prespecified covariance matrix. See Curran, West, and Finch (1996) for an example of an extensive Monte Carlo study that uses this method. In the VM transform, one samples from an underlying multivariate normal distribution, whose covariance matrix is found with a numerical routine. Then for each marginal variable  $Z$ , a resulting observed variable  $Y$  is generated from  $Z$  by applying a Fleishman polynomial; that is, a third or higher order polynomial whose coefficients are determined to match the target skewness and kurtosis values (Fleishman, 1978). There has recently been some critique of the VM transform, both empirically (Astivia & Zumbo, 2014) and theoretically (Foldnes & Grønneberg, 2015). In the latter work, it is shown that the truly multivariate aspect of the VM transform—that is, its copula—is very closely related to the multivariate normal copula. Moreover, attempts to generalize the VM transform by including Fleishman polynomials of higher degrees (Headrick, 2002) do not transcend this limitation. The copula-based simulation approach proposed by Mair, Satorra, and Bentler (2012) offers an alternative to the VM transform, where the resulting copula is nonnormal. However, controlling the univariate marginal distributions is not possible with their method.

In the present article, we present and investigate a new method for simulating nonnormal data with exact control of the covariance matrix and univariate skewness and kurtosis and with truly nonnormal copulas. The new method hence provides an alternative to the VM transform and gives rise to a different set of nonnormal distributions from those offered by the VM transform. We illustrate that the proposed simulation method produces distributions under which normal-theory ML inference performs differently than it does under the data generated by the VM transform, in spite of both distributions sharing the same covariance matrix and the same univariate skewness and kurtosis.

We propose a new transform that generates random samples from a vector  $\mathbf{Y}$  of  $p$  observed variables. The transform represents the observed variables as  $Y_i = \sum_{j=1}^s a_{ij}X_j$ , for  $i = 1, \dots, p$ , where the  $a_{ij}$  are constant scalars and the  $X_j$ ,  $j = 1, \dots, s$ , are mutually independent random variables referred to as *independent*

generator (IG) variables. In practice, the vector  $\mathbf{Y}$  will contain all observed variables in a SEM model that we wish to generate data for, while the  $X_j$  variables may be unrelated to the model, propagating randomness into  $\mathbf{Y}$ . We say that any  $\mathbf{Y}$  that can be represented stochastically in this way has an IG distribution. These distributions are convenient for simulating multivariate nonnormal data because it is straightforward to independently draw random samples from the univariate distributions of the  $X_j$  and then use the linear combinations to generate data for  $\mathbf{Y}$ . The algebraic simplicity of the stochastic representation also makes it possible to develop analytical results. Foldnes, Olsson, and Foss (2012) used IG distributions to theoretically study the loss of power of robust test statistics with increasing kurtosis.

To investigate whether normal-theory ML inference is robust to various violations of the normality assumption, researchers need to generate data that depart from the normal case in significant ways. In SEM simulation, a researcher with a target model in mind usually specifies the population parameters of the model and calculates the implied covariance matrix of the observed data. Then simulations are executed from a nonnormal distribution whose covariance matrix equals the implied covariance matrix. The task of nonnormal data generation with given covariance matrix is important in related fields of statistics as well. Many methods rely on some form of covariance structure analysis; for example, repeated-measures designs (Berkovits, Hancock, & Nevitt, 2000; Molenaar & Rovine, 1998), multilevel random coefficients models (Rovine & Molenaar, 2000), and analysis of covariance (ANCOVA) (Headrick & Sawilowsky, 2000). Evaluating such methods under nonnormal data hence requires a simulation from an underlying distribution whose covariance matrix equals a given target matrix and whose univariate marginal distributions are partially or fully specified. In the present article, we consider the traditional case in which only the third- and fourth-order moments are prespecified for each univariate marginal distribution.

The goals of this article are threefold. The first goal is to introduce and illustrate a new simulation technique based on IG variables. The second goal is to demonstrate the need for nonnormal data simulation techniques that differ from the often used VM transform. We do this by Monte Carlo, showing that popular SEM fit statistics are sensitive to the underlying type of nonnormality, even for fixed levels of univariate skewness and kurtosis. In a real-world setting, we do not know the exact nature of our nonnormal data, so we should be careful to include different kinds of nonnormality in our simulation conditions. Our third goal is to present analytical results for IG distributions. We establish exact conditions for the validity of normal-theory based test statistics under nonnormal IG

distributions. We also demonstrate that the IG distribution has a nonnormal copula, in contrast to the VM distribution, again showing that the IG transform produces conditions that may complement the VM transform.

This article is organized as follows. In the next section we present the IG transform, summarized in a five-step algorithm. This is followed by a section in which the five-step algorithm is illustrated on an empirical example. We then present a technical section developing analytical results for the IG transform, beginning with a nontechnical summary of these results. This is followed by a section with a Monte Carlo study that evaluates the rejection rates of two popular test statistics under data produced by both the IG and the VM transform. In the last section, we discuss limitations of the IG transform and give conclusive remarks. Sample code written in the R environment is available as online supplementary material for methodologists that wish to implement the IG transform.

## The IG transform

We investigate distributions obtained from linear combinations of independent nonnormal variables:

$$\mathbf{Y} = \mathbf{A}\mathbf{X}, \quad (1)$$

where  $\mathbf{X}$  is a column  $s$ -vector of generator variables, while  $\mathbf{Y}$  is a column  $p$ -vector of generated variables. A researcher may form  $\mathbf{Y}$  from the observed variables in the model under scrutiny. The topic of the current section is how we may choose the  $\mathbf{A}$  matrix and distributions for the  $X_j$  so that the observed variables have the desired covariance matrix, skewness, and kurtosis. The IG variables  $X_j$  for  $j = 1, \dots, s$ , contained in the vector  $\mathbf{X}$ , are assumed to be mutually independent, and  $\mathbf{A}$  is a  $p \times s$  ( $s \geq p$ ) matrix of constant coefficients. We assume without loss of generality that  $X_j$  have zero mean and unit variance for all  $j$ . In the following, we refer to data generation by way of (1) as the IG transform.

For a given target covariance matrix  $\Sigma$  for  $\mathbf{Y}$ , we employ some matrix  $\mathbf{A}$  such that

$$\mathbf{A}\mathbf{A}' = \Sigma.$$

This will ensure, given the unit variance and mutual independence of the IG variables, that  $\mathbf{Y}$  has the target covariance  $\Sigma$ . In a straightforward application of Equation (1), one can set  $s = p$  and choose  $\mathbf{A}$  as a square root matrix of  $\Sigma$ . However, a researcher investigating a particular model might want to investigate how skewness and kurtosis in various parts of the model influence estimation and goodness-of-fit statistics. Then a more elaborate  $\mathbf{A}$  can be constructed from the model at hand as follows. A SEM



model can be formulated (Bollen, 1989) as

$$\begin{aligned}\eta &= \mathbf{B}\eta + \mathbf{\Gamma}\xi + \zeta, & Y_1 &= \mathbf{\Lambda}_1\eta + \epsilon, \\ Y_2 &= \mathbf{\Lambda}_2\xi + \delta,\end{aligned}\quad (2)$$

where  $\xi$ ,  $\zeta$ ,  $\epsilon$ , and  $\delta$  are latent vector variates and the observed variate is  $\mathbf{Y} = (Y'_1, Y'_2)'$ . The covariance matrices of  $\xi$ ,  $\zeta$ ,  $\epsilon$ , and  $\delta$  are denoted by  $\Phi$ ,  $\Psi$ ,  $\Theta_\epsilon$ , and  $\Theta_\delta$ , respectively, and are assumed to be positive definite. The matrices  $\mathbf{B}$  and  $\mathbf{\Gamma}$  contain regression coefficients, while  $\mathbf{\Lambda}_1$  and  $\mathbf{\Lambda}_2$  contain factor loadings. Note that (2) contains the multivariate regression model as a submodel, when the  $Y_i$  are measured without errors and  $\mathbf{B} = \mathbf{0}$ .

We decompose these matrices by  $\Phi = \mathbf{A}_1\mathbf{A}'_1$ ,  $\Psi = \mathbf{A}_2\mathbf{A}'_2$ ,  $\Theta_\epsilon = \mathbf{A}_3\mathbf{A}'_3$ , and  $\Theta_\delta = \mathbf{A}_4\mathbf{A}'_4$  where the  $\mathbf{A}_i$  are square root matrices. We set  $\xi = \mathbf{A}_1\mathbf{X}_1$ ,  $\zeta = \mathbf{A}_2\mathbf{X}_2$ ,  $\epsilon = \mathbf{A}_3\mathbf{X}_3$ , and  $\delta = \mathbf{A}_4\mathbf{X}_4$  where the variables contained in the vector  $\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2, \mathbf{X}'_3, \mathbf{X}'_4)'$  are mutually independent, each with variance one and zero mean. Consequently we can cast (2) into the form of (1):

$$\mathbf{Y} = \begin{pmatrix} \mathbf{\Lambda}_1(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}\mathbf{A}_1 & \mathbf{\Lambda}_1(\mathbf{I} - \mathbf{B})^{-1}\mathbf{A}_2 & \mathbf{A}_3 & \mathbf{0} \\ \mathbf{\Lambda}_2\mathbf{A}_1 & \mathbf{0} & \mathbf{0} & \mathbf{A}_4 \end{pmatrix} \times \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \\ \mathbf{X}_4 \end{pmatrix}. \quad (3)$$

Thus, we can control skewness and kurtosis in each random vector  $\xi$ ,  $\zeta$ ,  $\epsilon$ , and  $\delta$  of the proposed model by applying the IG transform to each of these. The distribution of the observed vector  $\mathbf{Y}$  will then still be an IG distribution.

After  $\mathbf{A}$  in Equation (1) has been determined, we wish to find suitable distributions  $F_j$  for the  $X_j$  such that each  $Y_i$  has a given skewness and kurtosis. The skewness  $\alpha_V$  of a random variable  $V$  with expectation  $\mu_V$  and standard deviation  $\sigma_V$  is a measure of asymmetry and involves the third-order moment:  $\alpha_V = \mathbb{E}[(V - \mu_V)/\sigma_V]^3$ . Kurtosis is a measure of the propensity of a distribution to produce outliers (for a recent discussion of the interpretation of kurtosis in terms of tail extremity, see Westfall, 2014). Kurtosis involves fourth-order moments. The relative, or excess, kurtosis  $\beta_V$  of  $V$  is defined as  $\beta_V = \mathbb{E}[(V - \mu_V)^4]/\sigma_V^4 - 3$ .

The issues of skewness and kurtosis can be handled separately. For skewness, the pairwise independence between the  $X_j$  makes it a rather uncomplicated exercise in algebra to obtain an expression for the skewness of  $Y_i$ :

$$\alpha_{Y_i} = \frac{a_{i1}^3\alpha_{X_1} + a_{i2}^3\alpha_{X_2} + \cdots + a_{is}^3\alpha_{X_s}}{(a_{i1}^2 + a_{i2}^2 + \cdots + a_{is}^2)^{3/2}}. \quad (4)$$

Hence, the skewness of the  $Y_i$  depends linearly on the skewnesses in the generator variables. Given user-specified skewness values  $\alpha_{Y_i}$ ,  $i = 1, \dots, p$ , there are  $p$

equations of type (4). Together they constitute a linear system in  $p$  equations with  $s \geq p$  unknowns; namely, the  $\alpha_{X_j}$ ,  $j = 1, \dots, s$ . In general, such a system is consistent and can be easily solved for the  $\alpha_{X_j}$  by using statistical or mathematical software like R, Mathematica, or Matlab.

An expression similar to (4) might be obtained for kurtosis, again relying on the mutual independence and unit variance of the IG variables. By some straightforward algebra, we obtain

$$\beta_{Y_i} = \frac{a_{i1}^4\beta_{X_1} + a_{i2}^4\beta_{X_2} + \cdots + a_{is}^4\beta_{X_s}}{(a_{i1}^2 + a_{i2}^2 + \cdots + a_{is}^2)^2}. \quad (5)$$

With given user-specified kurtosis values  $\beta_{Y_i}$ , we again end up with a system of  $p$  linear equations with more unknowns ( $\beta_{X_j}$ ,  $j = 1, \dots, s$ ) than equations. And again, we can solve this system by using routines in a software package like R.

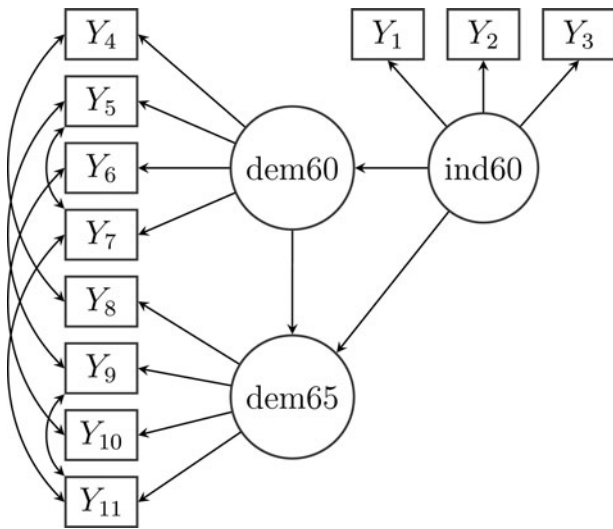
We can now summarize the IG transform simulation method:

1. The user specifies a target covariance matrix  $\Sigma$  and target skewness ( $\alpha_{Y_i}$ ) and kurtosis ( $\beta_{Y_i}$ ) for each univariate marginal  $Y_1, \dots, Y_p$ .
2. A matrix  $\mathbf{A}$  is determined such that  $\mathbf{A}\mathbf{A}' = \Sigma$ .
3. The systems of linear equations (4) and (5) are solved for  $\alpha_{X_j}$  and  $\beta_{X_j}$ , respectively.
4. For each IG variable  $X_j$ , a univariate distribution  $F_j$  is specified such that  $X_j$  has zero expectation, unit variance, skewness  $\alpha_{X_j}$ , and kurtosis  $\beta_{X_j}$ .
5. For a given sample size, draw independently random realizations of  $X_j$ ,  $j = 1, \dots, s$ . Apply the linear transformation (1) to obtain a random sample from  $\mathbf{Y}$ .

The IG transform described stochastically defines a large class of distributions, where steps 2, 3, and 4 can be executed in various ways. In step 2, many  $\mathbf{A}$  are possible. If  $s > p$ , there might be several possible solutions for  $\alpha_{X_j}$  and  $\beta_{X_j}$  in step 3. In step 4, there are different candidate distributions for  $X_j$  with given skewness and kurtosis, for example through Fleishman polynomials, the Johnson family of distributions (Johnson, 1949) or the Pearson family of distributions (Pearson, 1895). However, the focus in the present article is to introduce and illustrate a new simulation method that matches target skewness, kurtosis, and covariance and is fast and easy to simulate from. The flexibility contained in the IG class of distributions may be a topic of future research.

### Illustration: Bollen's political democracy model

In this section, we demonstrate how to simulate data for the well-known political democracy model for developing countries, described in the textbook by Bollen (1989). The associated R code is available as supplementary material.



**Figure 1.** Bollen's political democracy model. dem60: democracy in 1960. dem65: democracy in 1965. ind60: industrialization in 1960.

The model contains four measures of political democracy measured twice (1960 and 1965) and three measures of industrialization measured once (1960) and is depicted in Figure 1.

Suppose we want to generate data from a population that fits perfectly with the model. In our case, we have access to the original data set (Bollen, 1989, p. 428, Table 9.4) that contains  $n = 75$  developing countries, so we can use the model-implied covariance matrix as the target matrix  $\Sigma = \hat{\Sigma}$ :

$$\Sigma = \begin{pmatrix} 0.53 & & & & & & & & & & \\ 0.98 & 2.25 & & & & & & & & & \\ 0.82 & 1.78 & 1.95 & & & & & & & & \\ 0.67 & 1.45 & 1.21 & 6.83 & & & & & & & \\ 0.84 & 1.82 & 1.52 & 6.21 & 15.18 & & & & & & \\ 0.70 & 1.53 & 1.28 & 5.23 & 6.57 & 10.6 & & & & & \\ 0.84 & 1.83 & 1.53 & 6.25 & 9.17 & 6.61 & 11.05 & & & & \\ 0.81 & 1.77 & 1.48 & 5.14 & 5.68 & 4.78 & 5.72 & 6.77 & & & \\ 0.96 & 2.1 & 1.75 & 5.36 & 8.89 & 5.67 & 6.78 & 5.24 & 11.17 & & \\ 1.04 & 2.27 & 1.89 & 5.78 & 7.27 & 6.91 & 7.31 & 5.66 & 6.71 & 10.67 & \\ 1.03 & 2.25 & 1.87 & 5.72 & 7.19 & 6.05 & 7.58 & 5.6 & 7.99 & 7.16 & 10.34 \end{pmatrix}.$$

In addition to specifying the target covariance matrix, the researcher also specifies skewness and kurtosis for each of the 11 observed variables. We specify the following skewness values for  $Y_1, Y_2, \dots, Y_{11}$ : (0, 0, 0, 0, 1, 1, 1, 1, 2, 2, 2). We specify the following excess kurtosis values: (1, 1, 1, 1, 3, 3, 3, 3, 7, 7, 7). For example, for  $Y_5$  we desire a distribution with skewness 1 and excess kurtosis 3.

The second step is to determine a matrix  $A$  such that  $AA' = \Sigma$ . This is done in mathematical or statistical software like R. In this illustration, we take  $A$  to be the lower-triangular Choleski factor of  $\Sigma$ :

$$A = \begin{pmatrix} 2.61 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 2.38 & 3.09 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 2.00 & 0.59 & 2.50 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 2.39 & 1.13 & 0.47 & 1.96 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 1.97 & 0.33 & 0.26 & 0.27 & 1.63 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 2.05 & 1.30 & 0.32 & 0.13 & 0.41 & 2.23 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 2.21 & 0.65 & 0.84 & 0.46 & 0.46 & 0.36 & 2.02 & 0.00 & 0.00 & 0.00 & 0.00 \\ 2.19 & 0.64 & 0.52 & 0.70 & 0.47 & 0.99 & 0.28 & 1.76 & 0.00 & 0.00 & 0.00 \\ 0.25 & 0.07 & 0.06 & 0.06 & 0.16 & 0.11 & 0.12 & 0.07 & 0.63 & 0.00 & 0.00 \\ 0.55 & 0.16 & 0.13 & 0.13 & 0.34 & 0.25 & 0.26 & 0.16 & 1.09 & 0.65 & 0.00 \\ 0.46 & 0.14 & 0.11 & 0.11 & 0.29 & 0.21 & 0.21 & 0.14 & 0.91 & 0.39 & 0.73 \end{pmatrix}.$$

The third step involves solving Equations (4) and (5) for skewness  $\alpha_{X_j}$  and kurtosis  $\beta_{X_j}$ , respectively. Solving a linear system of equations is a routine task performed by any mathematical software package. The solutions for skewness  $\alpha_{X_j}$ ,  $j = 1, \dots, 11$ , in the IG variables are (0, 0, 0, 0, 4.07, 3.33, 4.14, 5.44, 2.99, 9.32, 6.47). For excess kurtosis, the solutions are (1, 2.18, 2.46, 5.81, 17.36, 14.06, 18.79, 29.51, 12.44, 96.07, 56.42). For example, the IG variable  $X_5$  should have skewness 4.07 and excess kurtosis 17.36.

The fourth step involves finding univariate distributions for the IG variables  $X_j$  that have the required skewness and kurtosis. For example, we must define a distribution  $F_5$  such that the skewness and kurtosis of  $X_5$  equal 4.07 and 17.36, respectively. There are different ways of doing this; one could, for example, use Fleishman polynomials as done in the VM transform. However, for this illustration we use the Pearson distribution system, which allows the user to specify mean, variance, skewness, and kurtosis in order to obtain a corresponding Pearson distribution. For example,  $F_5$  is a member of the Type I Pearson family of distributions, which are transformations of Beta distributions, so that the probability density of  $X_5$  is  $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{x-\lambda}{s}\right)^{a-1} \left(1 - \frac{x-\lambda}{s}\right)^{b-1}$ . To attain zero mean, unit variance, skewness 4.07, and kurtosis 17.36, we use R (Becker & Klößner, 2013) to calibrate the parameters:  $a = 0.05$ ,  $b = 1.07$ ,  $\lambda = -0.32$ , and  $s = 6.99$ .

The fifth and last step is to draw independently for each  $X_j$  random samples of a given sample size  $N$ . These 11 samples are combined as columns to form an  $N \times 11$  matrix  $D$ . The final simulated data set is then obtained as the  $N \times 11$  data matrix  $D \cdot A'$ , where each row represents a single draw from the random vector  $Y$ .

## Analytical investigations of IG distributions

The material contained in this section is quite mathematical in nature and may be of interest primarily to technical readers. Applied researchers may wish to skip the technicalities. For those readers, we provide the following summary of our analytical investigations.

### Summary of analytical results

An important use of nonnormal data generation in SEM is to evaluate the finite-sample performance of estimators and fit statistics under a variety of distributional conditions. Normal-theory-based ML is the most popular estimation method, which also provides an associated fit statistic,  $T_{ML}$ . An important issue is whether  $T_{ML}$  can still be trusted when data are nonnormal. It is well known that under some specific combinations of nonnormal data and model conditions,  $T_{ML}$  retains its asymptotic chi-square distribution. These combinations involve certain exact relationships between the multivariate fourth-order moments of the data and the derivatives of the model-implied covariance matrix and are characterized in a body of literature referred to as asymptotic robustness (AR; e.g., Amemiya & Anderson, 1990; Browne & Shapiro, 1988; Satorra & Bentler, 1990; Shapiro, 1987). If the IG transform is to be used to realistically assess  $T_{ML}$ , for example in terms of Type I error rates, it is important that AR does not hold. In this section, we therefore investigate under what conditions the IG transform will violate AR. With the results obtained in this section, we can affirm whether a given IG transform in conjunction with a given model will violate AR. This is done by examining an explicit matrix that must be zero in all elements for AR to hold. A researcher employing the IG method for data generation could check whether this matrix is nonnull and hence be confident that the simulated data will distort the distribution of the ML chi-square fit statistic. We are able to deduce this result because the IG transform is analytically tractable due to the independence of the IG variables. We also establish in the context of IG distributions that whenever the well-known Satorra-Bentler correction is equal to one—that is, when there is no correction—AR must hold, and hence both the ML chi-square and the Satorra-Bentler chi-square are to be trusted under reasonably large sample sizes. Whether this result extends to other classes of nonnormal distributions is presently unknown.

It is also important to establish that the IG transform is able to generate data with a nonnormal copula so that the resulting IG distribution departs from normality both in terms of each observed variable and in terms of their

dependence structure. In the second subsection, we show that IG distributions, in contrast to those obtained with the VM transform, in general have a nonnormal copula. Hence, the IG transform may serve to modify previous performance evaluations of estimators and fit statistics in SEM that were based on the VM transform, or other nonnormal data simulation techniques based on the normal copula, like the NORTA method (Cario & Nelson, 1997) and its extensions (Kowalchuk & Headrick, 2010).

Applied researchers may skip the technicalities contained in the next two subsections without loss of continuity and proceed to the next section, where the IG transform is compared to the VM transform in a Monte Carlo study.

### Asymptotic robustness

IG distributions belong to a class of distributions referred to as Data Model 1 by Yuan and Bentler (1999), where AR conditions were derived for four test statistics. These results were based on a crucial assumption (Yuan & Bentler, 1999, Condition 4.1) relating the data generation method and the model structure. This condition is sufficient for AR to hold. However, a researcher seeking to evaluate normal-theory-based statistics under nonoptimal conditions should avoid AR. Thus, what we need is a necessary condition for AR. By violating this condition in the simulation setup, the researcher will know that AR is violated.

To develop AR conditions under (1), we need to define some central vectors and matrices and operations upon them. For illustrative purposes, we exemplify in the following these definitions on a simple model, referred to as Model S. Model S has one factor  $\xi$  and three observed variables  $Y_1, Y_2, Y_3$ , with  $\text{var}(\xi) = 1$  for identification. We constrain all three loading coefficients to be equal:  $\lambda_i = \lambda$  for  $i = 1, 2, 3$ . As is common in parallel measures (Lord & Novick, 1968), we also constrain some error variances to be equal:  $\psi_1^2 = \psi_2^2$ , where  $\psi_i^2$  is the variance of the measurement residual corresponding to  $Y_i$ . Hence, the free parameters in Model S are  $\lambda$ ,  $\psi_1^2$ , and  $\psi_3^2$ . These parameters constitute the parameter vector  $\theta = (\lambda, \psi_1^2, \psi_3^2)'$ . The model-implied covariance matrix of Model S is

$$\Sigma(\theta) = \begin{pmatrix} \lambda^2 + \psi_1^2 & \lambda^2 & \lambda^2 \\ \lambda^2 & \lambda^2 + \psi_1^2 & \lambda^2 \\ \lambda^2 & \lambda^2 & \lambda^2 + \psi_3^2 \end{pmatrix}.$$

This matrix is symmetric and contains redundant elements. For a symmetric  $p \times p$  matrix  $\mathbf{G}$ , let  $\text{vech}(\mathbf{G})$  represent the column vector consisting of the  $p^* = p(p+1)/2$  nonduplicated elements of  $\mathbf{G}$ ; that is,

the elements along and below the diagonal. Thus, we can represent  $\Sigma(\theta)$  by the column vector

$$\begin{aligned}\sigma(\theta) &= \text{vech}(\Sigma(\theta)) \\ &= (\lambda^2 + \psi_1^2, \lambda^2, \lambda^2 + \psi_1^2, \lambda^2, \lambda^2 + \psi_3^2)',\end{aligned}$$

containing the  $p^* = 3 \cdot 4/2 = 6$  unique elements of  $\Sigma(\theta)$ . The same procedure may be applied to the sample covariance matrix  $S$ , which is then more economically represented as  $s = \text{vech}(S)$ . We let  $p$  denote the number of observed variables and  $q$  the number of free parameters. An important matrix is the Jacobian matrix of derivatives  $\Delta(\theta) = \partial\sigma(\theta)/\partial\theta'$ , where each entry of  $\sigma(\theta)$  is partially differentiated with respect to each parameter in  $\theta$ . For model S we have

$$\Delta(\theta) = \begin{pmatrix} 2\lambda & 1 & 0 \\ 2\lambda & 0 & 0 \\ 2\lambda & 1 & 0 \\ 2\lambda & 0 & 0 \\ 2\lambda & 0 & 0 \\ 2\lambda & 0 & 1 \end{pmatrix}.$$

Next we consider a central element in SEM: the sampling distribution of the covariances; that is, the distribution of  $s$ . The asymptotic covariance matrix of  $\sqrt{ns}$  is denoted by  $\Gamma$  and should not be confused with the regression coefficient matrix in Equation (2). To exemplify, let us consider an IG distribution tailored to simulate data for Model S. We fix all the parameter values  $\lambda$ ,  $\psi_1^2$ , and  $\psi_3^2$  to 1, such that the target covariance matrix is  $\Sigma = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$ . Then we set  $A = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & \sqrt{3}/2 & 0 \\ 1/\sqrt{2} & 1/\sqrt{6} & 2/\sqrt{3} \end{pmatrix}$  where  $AA' = \Sigma$ . There is a general formula for  $\Gamma$  when data come from an IG transform (Browne & Shapiro, 1988, Theorem 2.1):

$$\Gamma = \Gamma_N + K'_p \tilde{A} \tilde{C} \tilde{A}' K_p. \quad (6)$$

Here  $\tilde{A}$  is the  $p^2 \times s$  matrix whose  $j$ th column is  $a_j \otimes a_j$ , with  $a_j$  being the  $j$ th column vector of  $A$ . The  $\otimes$  symbol denotes the Kronecker product. Then  $a_j \otimes a_j$  consists of stacking  $p$  copies of  $a_j$ , where the  $i$ th copy is being multiplied by  $a_{ij}$ , for  $i = 1, \dots, p$ . For Model S, with the  $A$  given above, this gives

$$\tilde{A} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1/2 & 3/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ 1/2 & 1/6 & 4/3 \end{pmatrix}.$$

$C$  is the  $s \times s$  diagonal matrix that has the excess kurtosis of the IG variables on the diagonal, and zeros elsewhere. The constant  $p^2 \times p^*$  matrix  $K_p$  consists of elements with values in  $\{0, \frac{1}{2}, 1\}$  and is defined in, for example, Browne (1974, Section 2). It has the property that  $\text{vech}(G) = K'_p \text{vec}(G)$  where  $\text{vec}(G)$  is the vector formed by stacking the columns of  $G$ . For Model S,  $p = 3$ , and

$$K'_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Finally, Equation (6) contains  $\Gamma_N$ , defined as the asymptotic covariance matrix of  $\sqrt{ns}$  for multivariate normal data, which is given by  $\Gamma_N = 2K'_p(\Sigma \otimes \Sigma)K_p$ . Returning to Model S, with fixed parameter values  $\lambda$ ,  $\psi_1^2$ , and  $\psi_3^2$  all equal to 1, and with excess kurtosis in the IG variables denoted by  $\beta_i = \beta_{Y_i}$  for  $i = 1, 2, 3$ , Equation (6) gives

$$\Gamma = \begin{pmatrix} 8 + 4\beta_1 & & & & & & & & \\ 4 + 2\beta_1 & 5 + \beta_1 & & & & & & & \\ 2 + \beta_1 & 4 + \frac{\beta_1}{2} & 8 + \frac{\beta_1 + 9\beta_2}{4} & & & & & & \\ 4 + 2\beta_1 & 3 + \beta_1 & 2 + \frac{\beta_1}{2} & 5 + \beta_1 & & & & & \\ 2 + \beta_1 & 3 + \frac{\beta_1}{2} & 4 + \frac{\beta_1 + 3\beta_2}{4} & 3 + \frac{\beta_1}{2} & 5 + \frac{\beta_1 + \beta_2}{4} & & & & \\ 2 + \beta_1 & 2 + \frac{\beta_1}{2} & 2 + \frac{\beta_1 + \beta_2}{4} & 4 + \frac{\beta_1}{2} & 4 + \frac{\beta_1 + \beta_2}{4} & 8 + \frac{9\beta_1 + \beta_2 + 64\beta_3}{36} & & & \end{pmatrix}.$$

Setting  $\beta_i = 0$  for  $i = 1, 2, 3$  in the previous expression recovers  $\Gamma_N$ . Recall that the elements of  $s$  are ordered as  $s_{11}, s_{12}, s_{22}, s_{13}, s_{23}, s_{33}$ . Thus, for example, the covariance between the sample covariance of  $Y_2$  and  $Y_3$  and the variance of  $Y_2$  is approximately equal to  $4 + (\beta_1 + 3\beta_2)/4$  divided by the sample size  $n$ , provided the sample is large.

The normal-theory ML estimate  $\hat{\theta}$  minimizes

$$F_{ML}(\theta) = \log |\Sigma(\theta)| + \text{tr}(\Sigma^{-1}(\theta)) - \log |S| - p.$$

With multivariate normality  $T_{ML} = (n-1) \cdot F_{ML}(\hat{\theta})$  is asymptotically distributed as a chi-square with  $d = p^* - q$  degrees of freedom. With nonnormal data,  $T_{ML}$  is however asymptotically distributed as a mixture of chi-squares:

$$T_{ML} \xrightarrow{d} \sum_{j=1}^d \gamma_j \chi_1^2, \quad (7)$$

where the  $\chi_1^2$  are mutually independent chi-squares with one degree of freedom, and  $\xrightarrow{d}$  denotes convergence in distribution. The  $\gamma_j$  are the nonzero eigenvalues of  $U\Gamma$  where  $U = \Gamma_N^{-1} - \Gamma_N^{-1} \Delta \{\Delta' \Gamma_N^{-1} \Delta\}^{-1} \Delta' \Gamma_N^{-1}$ . AR is hence equivalent to  $\gamma_j = 1$ ,  $j = 1, \dots, d$ . Returning to Model



S, with the three parameters fixed to one and with excess kurtosis in the IG variables  $\beta_1 = 2$ ,  $\beta_2 = 4$ , and  $\beta_3 = 6$ , the nonzero eigenvalues of  $\mathbf{U}\mathbf{\Gamma}$  are  $\gamma_1 = 2.25$ ,  $\gamma_2 = 1$ , and  $\gamma_3 = 1$ . Hence, for large samples produced by the specified IG transform,  $T_{ML}$  obtained from testing Model S will have an asymptotic mean of 4.25, which is larger than the mean  $d = 3$  of the reference chi-square distribution. Hence, although Model S has perfect fit to the target covariance matrix  $\mathbf{\Sigma}$ , the  $T_{ML}$ -based test will reject Model S too often. In fact, at the nominal significance level 0.05, model S will have a rejection rate of about 0.14 in large samples. The fact that  $T_{ML}$  tends to inflate with increasing kurtosis is well known. However, as demonstrated with Model S, working with IG distributions allows for an exact calculation of the asymptotic distribution of  $T_{ML}$ . One way to remedy the inflation of  $T_{ML}$  is to scale by a factor  $c = d/\text{trace}(\mathbf{U}\mathbf{\Gamma})$ , as proposed by Satorra and Bentler (1988). Then the asymptotic mean of  $T_{SB} = c \cdot T_{ML}$  equals the degrees of freedom  $d$ .

Shapiro (1987) showed that AR holds if and only if there exists some  $p^* \times q$  matrix  $\mathbf{D}$  such that

$$\mathbf{\Gamma} = \mathbf{\Gamma}_N + \mathbf{\Delta}\mathbf{D}' + \mathbf{D}\mathbf{\Delta}'. \quad (8)$$

Moreover, under the slightly less general representation  $\mathbf{\Gamma} = \mathbf{\Gamma}_N + \mathbf{\Delta}\mathbf{E}\mathbf{\Delta}'$  for some symmetric matrix  $\mathbf{E}$ , normal-theory-based estimators like ML are asymptotically efficient within the class of minimum discrepancy function (MDF) estimators (Shapiro, 1987, Corollary 5.3). An MDF estimator  $\hat{\boldsymbol{\theta}}$  is defined as the minimizer of  $F(\mathbf{S}, \mathbf{\Sigma}(\boldsymbol{\theta}))$  where the function  $F$  has the following three properties:  $F(\mathbf{E}, \mathbf{G}) \geq 0$  for all  $\mathbf{E}$  and  $\mathbf{G}$ ;  $F(\mathbf{E}, \mathbf{G}) = 0$  if and only if  $\mathbf{E} = \mathbf{G}$ ;  $F$  is twice continuously differentiable jointly in  $\mathbf{E}$  and  $\mathbf{G}$  (Browne, 1982).

In Yuan and Bentler (1999), a crucial assumption, therein referred to as Condition 4.1, is that  $\text{vech}(\mathbf{a}_i\mathbf{a}_i')$  is contained in the column space of  $\mathbf{\Delta}$  for all  $i$ . We reformulate this condition as  $\mathbf{P} = \mathbf{0}$ , where

$$\mathbf{P} \equiv \mathbf{\Delta}_c' \mathbf{K}_p' \tilde{\mathbf{A}}.$$

Here,  $\mathbf{\Delta}_c$  is an orthogonal complement of  $\mathbf{\Delta}(\boldsymbol{\theta})$ ; that is, a  $p^* \times (p^* - \text{rank}(\mathbf{\Delta}))$  matrix of full column rank where each column in  $\mathbf{\Delta}_c$  is orthogonal to each column in  $\mathbf{\Delta}(\boldsymbol{\theta})$ . For Model S, one orthogonal complement is given by

$$\mathbf{\Delta}_c = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

so that  $\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 \\ 1/2 & -1/2 & 0 \\ 3/2 & -3/2 & 0 \end{pmatrix}$ , which contains nonzero elements.

Yuan and Bentler (1999) showed that  $\mathbf{P} = \mathbf{0}$  is a sufficient condition for AR. Our first proposition gives a simple alternative proof that  $\mathbf{P} = \mathbf{0}$  entails AR. In addition, we show that the condition  $\mathbf{P} = \mathbf{0}$  implies that normal-theory estimators are asymptotically efficient.

**Proposition 1.** *Consider the covariance model  $\mathbf{\Sigma}(\boldsymbol{\theta})$ , and assume that it is correctly specified for an underlying IG distribution of Type (1). Then, if  $\mathbf{P} = \mathbf{0}$ , asymptotic robustness holds. Also, if  $\mathbf{P} = \mathbf{0}$ , then normal theory estimators (e.g., ML, GLS) are asymptotically efficient within the class of minimum discrepancy function estimators.*

**Proof.**  $\mathbf{\Delta}_c' \mathbf{K}_p' \tilde{\mathbf{A}} = \mathbf{0}$  implies that the column space of  $\mathbf{K}_p' \tilde{\mathbf{A}}$  is a subset of the column space of  $\mathbf{\Delta}$ . Hence there exists a matrix  $\mathbf{L}$  such that  $\mathbf{K}_p' \tilde{\mathbf{A}} = \mathbf{\Delta}\mathbf{L}$ . Substitution in the second term on the right-hand side of (6) gives  $\mathbf{K}_p' \tilde{\mathbf{A}} \mathbf{C} \tilde{\mathbf{A}}' \mathbf{K}_p = \mathbf{\Delta} \mathbf{L} \mathbf{C} \mathbf{L}' \mathbf{\Delta}'$ , and  $\mathbf{\Gamma} = \mathbf{\Gamma}_N + \mathbf{\Delta} \mathbf{E} \mathbf{\Delta}'$  where  $\mathbf{E} = \mathbf{L} \mathbf{C} \mathbf{L}'$ .  $\square$

A necessary condition for AR was not given in Yuan and Bentler (1999). The following proposition identifies a necessary condition. It also states that if all the  $X_j$  have more, or all have less, kurtosis than the normal distribution, then  $\mathbf{P} = \mathbf{0}$  is actually equivalent to AR.

**Proposition 2.** *Suppose that asymptotic robustness holds when testing a model  $\mathbf{\Sigma}(\boldsymbol{\theta})$  against an underlying IG distribution defined by (1). Then  $\mathbf{PCP}' = \mathbf{0}$ . Moreover, if either  $\beta_{X_j} > 0$  for all  $j$  or  $\beta_{X_j} < 0$  for all  $j$ , then  $\mathbf{P} = \mathbf{0}$ .*

**Proof.** By AR, it follows from Equations (6) and (8) that some  $\mathbf{D}$  exists such that  $\mathbf{\Delta}\mathbf{D}' + \mathbf{D}\mathbf{\Delta}' = \mathbf{K}_p' \tilde{\mathbf{A}} \mathbf{C} \tilde{\mathbf{A}}' \mathbf{K}_p$  and premultiplication with  $\mathbf{\Delta}_c'$  and postmultiplication by  $\mathbf{\Delta}_c$  gives the required result. Now assume that  $\beta_{X_j} > 0$  for all  $j$ ; that is, that the diagonal elements of  $\mathbf{C}$  are all positive. The diagonal elements of  $\mathbf{PCP}'$  are quadratic forms  $\mathbf{p}_j \mathbf{C} \mathbf{p}_j'$  where  $\mathbf{p}_j$  is the  $j$ th row of  $\mathbf{P}$ . Now  $\mathbf{p}_j \mathbf{C} \mathbf{p}_j' = 0$  implies that  $\mathbf{p}_j = \mathbf{0}$  for all  $j$ . The same argument holds for the situation with  $\beta_{X_j} < 0$  for all  $j$ .  $\square$

As a corollary, we obtain an interesting observation. In general, if the Satorra-Bentler correction  $c$  is inactive—that is, if  $c = 1$ —we know that the sum of the eigenvalues  $\gamma_j$  is equal to  $d$ . In general, this does not imply that we are in an AR situation, with  $\gamma_j = 1$  for all  $j = 1, \dots, d$ . The following corollary shows, however, that under an IG distribution with excess kurtosis in the IG variables,  $c = 1$  is indeed equivalent to AR.

**Corollary 3.** *Consider an IG distribution (1) where either  $\beta_{X_j} > 0$  for all  $j$  or  $\beta_{X_j} < 0$  for all  $j$ . Then the Satorra-Bentler correction  $c$  equals one if and only if AR holds.*

**Proof.** Assume AR holds. Then from the previous proposition it follows that  $\mathbf{P} = \mathbf{0}$ . Under an IG distribution we

have from (6)

$$\begin{aligned} \text{trace}(\mathbf{U}\mathbf{\Gamma}) &= \text{trace}((\mathbf{\Delta}'_c \mathbf{\Gamma}_N \mathbf{\Delta}_c)^{-1} (\mathbf{\Delta}'_c \mathbf{\Gamma} \mathbf{\Delta}_c)) \\ &= \text{trace}(\mathbf{I}) + \text{tr}(\mathbf{Q}\mathbf{C}), \end{aligned}$$

where  $\mathbf{Q} = \mathbf{P}'(\mathbf{\Delta}'_c \mathbf{\Gamma}_N \mathbf{\Delta}_c)^{-1} \mathbf{P}$ . Hence from  $\mathbf{Q} = \mathbf{0}$  it follows that  $\text{trace}(\mathbf{U}\mathbf{\Gamma}) = d$  so that  $c = d/d = 1$ .

Assume  $c = 1$ . Then  $\text{tr}(\mathbf{Q}\mathbf{C}) = 0$ . Define  $\mathbf{M} = (\mathbf{\Delta}'_c \mathbf{\Gamma}_N \mathbf{\Delta}_c)^{-1}$ . Then

$$\begin{aligned} \text{trace}(\mathbf{Q}\mathbf{C}) &= \text{trace}(\mathbf{P}'\mathbf{M}\mathbf{P}\mathbf{C}) = \text{trace}(\tilde{\mathbf{P}}\tilde{\mathbf{P}}'\mathbf{M}) \\ &= \text{trace}(\tilde{\mathbf{P}}'\mathbf{M}\tilde{\mathbf{P}}) = \text{trace}(\tilde{\mathbf{P}}'\tilde{\mathbf{P}}), \end{aligned}$$

where  $\tilde{\mathbf{P}} = \mathbf{P}\mathbf{C}^{1/2}$  and  $\tilde{\mathbf{P}}' = \tilde{\mathbf{P}}\mathbf{M}^{1/2}$ . Since  $\text{tr}(\mathbf{Q}\mathbf{C}) = 0$ , it follows that  $\tilde{\mathbf{P}} = \mathbf{0}$ , which is equivalent to  $\tilde{\mathbf{P}}\mathbf{M}^{1/2} = \mathbf{P}\mathbf{C}^{1/2}\mathbf{M}^{1/2} = \mathbf{0}$ . Since  $\mathbf{C}$  and  $\mathbf{M}$  are nonsingular, it follows that  $\mathbf{P} = \mathbf{0}$ , which implies AR, by Proposition 1.  $\square$

We have seen that when all IG variables have more, or all have less, kurtosis than a normal variable, then  $\mathbf{P}\mathbf{C}\mathbf{P}' = \mathbf{0}$  and  $\mathbf{P} = \mathbf{0}$  are equivalent conditions, both again equivalent to AR. A natural follow-up question is whether  $\mathbf{P} = \mathbf{0}$  is also necessary for AR in the general case where some IG variables have more kurtosis, and some less, than the normal distribution. In the following, we use Model S to construct a counterexample. For Model S, we noted earlier that  $\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 \\ 1/2 & -1/2 & 0 \\ 3/2 & -3/2 & 0 \end{pmatrix} \neq \mathbf{0}$ . If we set the excess kurtosis in the IG variables to  $\beta_1 = 1$ ,  $\beta_2 = -1$ , and  $\beta_3 = \frac{1}{16}$  and define  $\mathbf{D}' = \frac{1}{12} \begin{pmatrix} 8 & 3 & -1 & 3 & 0 & 1 \\ 8 & 2 & -10 & 2 & -4 & 0 \\ -6 & -2 & 0 & -2 & 0 & 0 \end{pmatrix}$ , then  $\mathbf{K}'_3 \tilde{\mathbf{A}} \mathbf{C} \tilde{\mathbf{A}}' \mathbf{K}_3 = \mathbf{\Delta} \mathbf{D}' + \mathbf{D} \mathbf{\Delta}'$ . Hence, in this case  $\mathbf{P}$  is not zero, but if  $\beta_1 = 1$ ,  $\beta_2 = -1$ , and  $\beta_3 = 1/16$ , then AR holds. Thus  $\mathbf{P} = \mathbf{0}$  implies AR only in the case where all IG variables have positive, or all have negative, excess kurtosis.

### The copula of the IG transform

Consider a continuous multivariate distribution, with univariate marginal distributions  $F_i(y) = P(Y_i \leq y)$ ,  $i = 1, \dots, p$ . Sklar (1959) showed that a full description of the random vector  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)'$  may be uniquely obtained by separating the marginal distributions  $F_i$  from the dependence structure; that is, the copula. More precisely, the probability integral transform applied to each component of  $\mathbf{Y}$  gives the random vector  $(U_1, \dots, U_p) = (F_1(Y_1), \dots, F_p(Y_p))$ , with uniform marginal distributions. The copula  $C$  of  $\mathbf{Y}$  is the joint cumulative distribution function of  $(U_1, \dots, U_p)$ . Thus a copula is a cumulative distribution function with uniform marginal distributions. We refer to Joe (2014) for a thorough review of copula theory.

Similar to the IG transform, under the VM transform the user may specify kurtosis and skewness for each univariate marginal, together with a covariance matrix. The resulting distribution, given the nonnormal nature of the univariate marginal distributions, appears to be highly nonnormal. It has however been shown (Foldnes & Grønneberg, 2015) that the dependence structure of the VM transform is closely related to that of a multivariate normal distribution. This is not surprising, given that the Fleishman polynomials in the VM transform do not contain interactions between the normally distributed generator variables  $\mathbf{Z}$  for  $\mathbf{Y}$ . That is, in the VM transform each output variable  $Y_i$  is obtained as a polynomial in a single corresponding normally distributed variable  $Z_i$ . The generator vector  $\mathbf{Z} = (Z_1, \dots, Z_p)'$  is multivariate normal, and the VM output  $\mathbf{Y} = (Y_1, \dots, Y_p)'$  inherits some basic properties of  $\mathbf{Z}$ .

Compare this to the IG transform in Equation (1), where each  $Y_i$  is a function of potentially all the generator variables  $X_j$ . Although Equation (1) is a simple way to obtain dependence between the  $Y_i$ —that is, by assuming linearity and independence—the resulting distribution for  $\mathbf{Y}$  has a nonnormal copula. To the best of our knowledge there exists no general results on the copula obtained from the IG transform. However, as shown in the following, this copula is not in general normal. Consider the two-dimensional case  $p = s = 2$  and assume that  $\mathbf{A}$  is nonsingular. Suppose that the probability density function (pdf) of  $X_1$  is nonzero everywhere, while the pdf of  $X_2$  is nonzero only in an interval  $[b, \infty)$ . For example, might  $X_1$  be normally distributed and  $X_2$  a scaled  $\chi^2$ . Fix  $Y_1$  to any value  $y_1$ . If we assume that both  $a_{11}$  and  $a_{21}$  are nonzero, it follows from

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 \\ y_2 &= a_{21}x_1 + a_{22}x_2 \end{aligned}$$

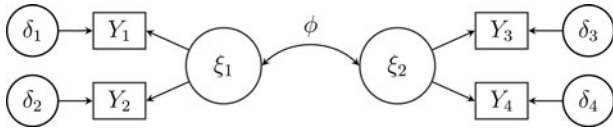
that

$$y_2 = \frac{a_{21}}{a_{11}}y_1 + (a_{22} - \frac{a_{12}a_{21}}{a_{11}})x_2.$$

Since  $\mathbf{A}$  is invertible, we have  $(a_{22} - \frac{a_{12}a_{21}}{a_{11}}) \neq 0$ , and suppose without loss of generality that this expression is positive. Then, given  $Y_1 = y_1$ , the conditional density of  $Y_2$  is nonzero only in  $[\frac{a_{21}}{a_{11}}y_1 + (a_{22} - \frac{a_{12}a_{21}}{a_{11}})b, \infty)$ . Thus there exists  $y_1, y_2$  such that  $f_{2|1}(y_2|y_1) = 0$ , where  $f_{2|1}$  denotes the conditional density of  $Y_2$  given  $Y_1 = y_1$ . Furthermore, since  $a_{21}$  is nonzero and since the pdf of  $X_1$  is nonzero everywhere, it follows that the pdf of  $Y_2$  is nonzero everywhere:  $f_2(y_2) > 0$  for all  $y_2$ . Then since

$$f_{2|1}(y_2|y_1) = f_2(y_2) \cdot c_{12}(F_1(y_1), F_2(y_2)),$$

it follows that the copula density  $c_{12}$  of  $\mathbf{Y}$  has  $c_{12}(F_1(y_1), F_2(y_2)) = 0$  for some  $y_1, y_2$ . However, since



**Figure 2.** Model S: A simple two-factor model.

the bivariate normal pdf is nonzero everywhere, the bivariate normal copula is nonzero for all interior points of the unit square, so  $c_{12}$  cannot be a normal copula.

### Monte Carlo illustration

In this section, we investigate empirically the performance of the two most popular fit statistics under the IG and VM simulation methods. We choose to focus on test statistics since model fit evaluation is a central task in SEM. Specifically, we consider the performance of  $T_{ML}$ , which is the default statistic in most SEM software packages, and  $T_{SB}$ , which is the most widely used fit statistic for nonnormal data. Our goal is to illustrate the IG transform in a specific example and to show that it produces distributions that differ from the VM transform to the extent that important SEM statistics are affected.

Since our purpose here is not to evaluate the IG method over a span of models, sample sizes, and nonnormality conditions, we employ a simple model, given in Figure 2, with two factors  $\xi_1$  and  $\xi_2$ , each with two indicators.

The population model has factor loadings equal to 1, 0.8, 1, 0.8 for  $Y_1, Y_2, Y_3$ , and  $Y_4$ , respectively. Both factor variances are equal to one, and the covariance between  $\xi_1$  and  $\xi_2$  is  $\phi = 0.2$ . The residual variances of  $Y_1, Y_2, Y_3$ , and  $Y_4$  are all equal to 0.4. The model has only three free parameters; namely, the elements of the covariance matrix  $\text{Cov}(\xi_1, \xi_2)$ . All factor loadings and residual variances are fixed in the model to their population counterparts, so there are seven degrees of freedom. The resulting covariance matrix of the observables, and a corresponding triangular square root matrix, are given by

$$\Sigma = \begin{pmatrix} 1.40 & 0.80 & 0.20 & 0.16 \\ 0.80 & 1.04 & 0.16 & 0.13 \\ 0.20 & 0.16 & 1.40 & 0.80 \\ 0.16 & 0.13 & 0.80 & 1.04 \end{pmatrix},$$

$$A = \begin{pmatrix} 1.183 & 0 & 0 & 0 \\ 0.676 & 0.763 & 0 & 0 \\ 0.169 & 0.060 & 1.170 & 0 \\ 0.135 & 0.048 & 0.662 & 0.762 \end{pmatrix}.$$

We remark that other square root matrices of  $\Sigma$  could have been used, for example a symmetric  $A$ , but for the sake of simplicity we do not proceed to study these.

The Monte Carlo design includes the following factors:

**Table 1.** Skewness  $\alpha_{X_j}$  and kurtosis  $\beta_{X_j}$  in the generator variables  $X_1, \dots, X_4$ .

Nonnormality		$X_1$	$X_2$	$X_3$	$X_4$
Moderate	$\alpha_{X_j}$	0	0	1.035	1.716
	$\beta_{X_j}$	1	2.569	3.142	7.820
Severe	$\alpha_{X_j}$	2	3.378	3.100	5.140
	$\beta_{X_j}$	5	12.843	15.711	39.098

- Two sample sizes,  $n = 100$  and  $n = 500$
- Two levels of nonnormality
  - *Moderate* nonnormality has skewness  $\alpha_{Y_i}$  equal to 0, 0, 1, 1 and kurtosis  $\beta_{Y_i}$  equal to 1, 1, 3, 3 in  $Y_1, Y_2, Y_3$ , and  $Y_4$ , respectively.
  - *Severe* nonnormality has skewness  $\alpha_{Y_i}$  equal to 2, 2, 3, 3 and kurtosis  $\beta_{Y_i}$  equal to 5, 5, 15, 15 in  $Y_1, Y_2, Y_3$ , and  $Y_4$ , respectively.
- Two simulation methods: VM and IG

For the two levels of nonnormality in the  $Y_i$  variables,  $i = 1, \dots, 4$ , we calculated skewness and kurtosis values in the generator variables  $X_1, \dots, X_4$ . This was done with a numerical routine for the equation systems (4) and (5), as shown in the supplementary online material. The resulting skewness and kurtosis values for  $X_1, \dots, X_4$  are given in Table 1. Each generator variable  $X_j$  was then simulated independently, with zero mean, unit variance, skewness, and excess kurtosis as specified in Table 1. To this end, we employed the Pearson family of distributions (Pearson, 1895), which allows the user to specify the first four moments. We note that other kinds of distributions that match these moments could have been used and that the sensitivity of the IG transform to variation in the univariate distributions of the IG variables is a direction for future research.

Results are given in Table 2. In each condition, we report for  $T_{ML}$  and  $T_{SB}$  the mean, variance, and rejection rate at the  $\alpha = 0.05$  level. There are 2,000 replications in each

**Table 2.** Mean, variance, and percentage rejection rates for two test statistics.

Nonnormality	$n$	Method	$T_{ML}$			$T_{SB}$		
			Mean	Var	RR	Mean	Var	RR
Moderate	100	VM	10.1	32.8	20.5	8.0	28.6	11.2
	100	IG	11.3	135.8	23.9	8.3	43.0	12.6
	500	VM	10.2	33.3	21.9	7.2	17.5	7.0
	500	IG	11.8	73.5	29.1	7.3	24.2	9.0
Severe	100	VM	22.1	268.5	66.5	9.5	69.6	17.9
	100	IG	32.2	700.3	74.6	13.1	382.2	25.6
	500	VM	23.4	305.0	67.4	7.5	30.1	9.1
	500	IG	33.8	939.9	74.5	8.8	111.6	15.0

Note.  $T_{ML}$ : normal-theory-based likelihood ratio test statistic.  $T_{SB}$ : Satorra and Bentler mean-adjusted test statistic. VM: data obtained with VM transform. IG: data obtained with IG transform. RR: rejection rate. Var: variance.

cell, so that the maximal standard error for the rejection rates is  $\sqrt{0.5^2/2000} = 0.011$ . The main observation is that fixing the covariance matrix and both univariate skewness and kurtosis still leaves a lot of flexibility that will influence the behaviour of  $T_{ML}$  and  $T_{SB}$ . Both statistics differ in mean, variance, and rejection rates between the VM and IG transforms. Under normal data, the asymptotic mean and variance of these statistics equal those of a  $\chi^2$  distribution with seven degrees of freedom. Thus, the asymptotic mean and variance are 7 and 14, respectively. As expected, due to excess kurtosis,  $T_{ML}$  is inflated, with rejection rates larger than the nominal levels. The nonnormality correction in  $T_{SB}$  only partly ameliorates the inflation. In all conditions, the behavior of both  $T_{ML}$  and  $T_{SB}$  is sensitive to the underlying transform. The mean and variance are higher under IG than VM, resulting in consistently higher rejection rates for data generated with the IG transform compared to the VM transform. Hence, it seems that data obtained from the IG transform are farther removed from normality than data obtained with the VM transform, even when controlling for covariance matrix and skewness and kurtosis in the marginal distributions. For example, in the  $n = 500$  condition with severe nonnormality, one gets a much more positive impression of the performance of  $T_{SB}$  under the VM transform than under the IG transform (rejection rates are 9.1% and 15.0%, respectively). Since the difference between the IG and VM transforms lies solely in the copula, we have empirically demonstrated that the underlying copula of a distribution can significantly affect SEM goodness-of-fit testing.

## Discussion

### Limitations of the IG transform

It is likely that the class of IG distributions contains a quite limited subset of the class of all nonnormal multivariate distributions. The question whether a random vector is generated by some mechanism based on linear combinations of mutually independent random constituents is hard to verify. Hence, there may be types of multivariate nonnormality that may occur in practice but are not obtainable through the IG transform.

A more practical and critical limitation is the application of the IG transform for large dimensionality. If  $d$  is large, say  $d > 50$ , the observed variables will be linear combinations of a large set of independent generator variables, and the resulting sum will tend to be normally distributed, due to the central limit theorem. To counteract this convergence to normality, the IG variables must possess extreme skewness and kurtosis. In theory this is not a problem, but in practice it is our experience that it

may be difficult to simulate from univariate distributions with such extreme third- and fourth-order moments. For example, with a target kurtosis of 10, it may be necessary for some IG variables to have a kurtosis of 100. We may speculate that available statistical software routines will not be fully able to generate data with such a high kurtosis. This concern is a topic for further research.

One might wonder whether the two systems of Equations (4) and (5) can be solved for skewness and kurtosis values, respectively. However, these are linear systems of equations, where a solution does not exist only if the coefficient matrix is singular, in which case the matrix  $A$  may be replaced to obtain a solution. Given solutions for skewness and kurtosis in (4) and (5), another problem may arise. Not all combinations of univariate skewness and kurtosis are feasible, and we may end up in such a situation for one of the observed variables. In such a case, one could again replace  $A$ . Whether this poses a serious limitation for the IG transform is a topic for future research.

Note that while the VM transform defines one specific distribution for a given target covariance matrix, and given univariate skewness and kurtosis, the IG transform defines a much larger class of distributions. This generality reflects the fact that specifying the covariance and third- and fourth-order moments of the marginal distributions only partially specifies the multivariate distribution. The IG class of distributions, for given covariance and univariate moments, is spanned by two independent choices. First, the choice of  $A$ , among the many different square and nonsquare matrices whose product  $AA'$  equals the target covariance matrix. Each  $A$  implies a particular multivariate distribution, so that two different  $A$ s will result in two different distributions. Second, each set of distributions for the IG variables implies a particular multivariate distribution. There are many univariate distributions for a given skewness and kurtosis, and a researcher may use Fleishman polynomials, Pearson distributions, or Johnson distributions. Choosing two different sets of IG distributions will result in two different distributions. Finally, we remark that the IG transform can easily be extended to higher order moments—for example, fifth-order—which would involve solving systems of linear equations of the the same kind as Equations (4) and (5).

## Conclusion

We have studied the use of independent generator variables for simulating nonnormal data with a prescribed covariance matrix and skewness and kurtosis in each univariate marginal. We have shown analytically and empirically that the IG transform generates data with different multivariate characteristics from the VM transform,



even when the two transforms share the same covariance structure and univariate skewness/kurtosis. While the VM transform results in data whose copula is closely related to the normal case, the IG transform produces data with a nonnormal copula. A simple Monte Carlo study demonstrated that two popular SEM goodness-of-fit statistics perform worse with data from the IG transform than when data are generated with the VM transform. This demonstrates that the use of the IG transform may offer new insights on the robustness of SEM inference to violations of the normality assumption, relative to those gained through the use of the VM transform.

The proposed IG transform may be applied in simulation studies designed to test the sensitivity of normal-theory-based estimators and test statistics like  $T_{ML}$  to conditions of nonnormality. A researcher could use the IG transform for data generation and make certain that the necessary condition in Proposition 2 is violated. This will most likely prevent an overly optimistic evaluation of the distributional robustness of normal-theory SEM inference.

## Article information

**Conflict of Interest Disclosures:** Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

**Ethical Principles:** The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

**Funding:** This work was not funded.

**Role of the Funders/Sponsors:** None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

**Acknowledgments:** The authors thank Randi Hammer-vold, Stefan Schaubert, and Albert Satorra for their comments on prior versions of this manuscript. The ideas and opinions expressed herein are those of the authors

alone, and endorsement by the authors' institution is not intended and should not be inferred.

## References

- Amemiya, Y., & Anderson, T. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics*, 18(3), 1453–1463. doi:10.1214/aos/1176347760
- Astivia, O. L. O., & Zumbo, B. D. (2014). 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, 541–567. doi:10.1177/0013164414548894
- Becker, M., & Klößner, S. (2013). PearsonDS: Pearson distribution system [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=PearsonDS> (R package version 0.97)
- Bentler, P. M. (2006). *Eqs 6 structural equations program manual*. Encino, CA: Multivariate Software.
- Berkovits, I., Hancock, G. R., & Nevitt, J. (2000). Bootstrap resampling approaches for repeated measure designs: Relative robustness to sphericity and normality violations. *Educational and Psychological Measurement*, 60(6), 877–892. doi:10.1177/00131640021970961
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley. doi:10.1002/9781118619179
- Browne, M. (1974). Generalized least-squares estimators in the analysis of covariance structures. *South African Statistical Journal*, 8, 1–24. doi:10.1002/j.2333-8504.1973.tb00197.x
- Browne, M. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge, UK: Cambridge University Press. doi:10.1017/cbo9780511897375.003
- Browne, M., & Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variable models. *British Journal of Mathematical and Statistical Psychology*, 41, 193–208. doi:10.1111/j.2044-8317.1988.tb00896.x
- Cario, M., & Nelson, B. (1997). Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. *Northwestern University, IEMS Technical Report*, 50, 100–150.
- 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. doi:10.1037/1082-989X.1.1.16
- Fleishman, A. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43, 521–532. doi:10.1007/BF02293811
- Foldnes, N., & Grønneberg, S. (2015). How general is the Vale–Maurelli simulation approach? *Psychometrika*, 80(4), 1066–1018. doi:10.1007/s11336-014-9414-0
- Foldnes, N., Olsson, U. H., & Foss, T. (2012). The effect of kurtosis on the power of two test statistics in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, 65(1), 1–18. doi:10.1111/j.2044-8317.2010.02010.x
- Headrick, T. C. (2002). Fast fifth-order polynomial transforms for generating univariate and multivariate nonnormal distributions. *Computational Statistics & Data Analysis*, 40(4), 685–711. doi:10.1016/S0167-9473(02)00072-5

- Headrick, T. C., & Sawilowsky, S. S. (2000). Properties of the rank transformation in factorial analysis of covariance. *Communications in Statistics-Simulation and Computation*, 29(4), 1059–1087. doi:[10.1080/03610910008813654](https://doi.org/10.1080/03610910008813654)
- Joe, H. (2014). *Dependence modeling with copulas*. Boca Raton, FL: CRC Press.
- Johnson, N. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36, 149–176. doi:[10.1093/biomet/36.1-2.149](https://doi.org/10.1093/biomet/36.1-2.149)
- Kowalchuk, R. K., & Headrick, T. C. (2010). Simulating multivariate g-and-h distributions. *British Journal of Mathematical and Statistical Psychology*, 63(1), 63–74. doi:[10.1348/000711009X423067](https://doi.org/10.1348/000711009X423067)
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Pub. Co.
- Mair, P., Satorra, A., & Bentler, P. M. (2012). Generating non-normal multivariate data using copulas: Applications to SEM. *Multivariate Behavioral Research*, 47(4), 547–565. doi:[10.1080/00273171.2012.692629](https://doi.org/10.1080/00273171.2012.692629)
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156–166. doi:[10.1037/0033-2909.105.1.156](https://doi.org/10.1037/0033-2909.105.1.156)
- Molenaar, P. C. M., & Rovine, M. J. (1998). A nonstandard method for estimating a linear growth model in LISREL. *International Journal of Behavioral Development*, 22(3), 453–473. doi:[10.1080/016502598384225](https://doi.org/10.1080/016502598384225)
- Muthén, B., & Muthén, L. (2012). *Mplus version 7: User's guide*. Los Angeles, CA: Muthén & Muthén.
- Pearson, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London. A*, 343–414. doi:[10.1098/rsta.1894.0003](https://doi.org/10.1098/rsta.1894.0003)
- Rosseel, Y. (2012). lavaan: Latent variable analysis [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=lavaan> (R package version 0.4-12)
- Rovine, M. J., & Molenaar, P. C. M. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research*, 35(1), 51–88. doi:[10.1207/S15327906MBR350103](https://doi.org/10.1207/S15327906MBR350103)
- Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. In *Proceedings of the American Statistical Association* (pp. 308–313). Alexandria, VA: American Statistical Association. doi:[10.1007/BF02296192](https://doi.org/10.1007/BF02296192)
- Satorra, A., & Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics & Data Analysis*, 10(3), 235–249. doi:[10.1016/0167-9473\(90\)90004-2](https://doi.org/10.1016/0167-9473(90)90004-2)
- Shapiro, A. (1987). Robustness properties of the mdf analysis of moment structures. *South African Statistical Journal*, 21, 39–62.
- Sklar, M. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Paris: Université Paris 8.
- Vale, C., & Maurelli, V. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48, 465–471. doi:[10.1007/BF02293687](https://doi.org/10.1007/BF02293687)
- Vuolo, M. (2015). Copula models for sociology: Measures of dependence and probabilities for joint distributions. *Sociological Methods & Research*, 1–45. doi:[10.1177/0049124115584477](https://doi.org/10.1177/0049124115584477)
- Westfall, P. H. (2014). Kurtosis as peakedness, 1905–2014. R.I.P. *The American Statistician*, 68(3), 191–195. doi:[10.1080/00031305.2014.917055](https://doi.org/10.1080/00031305.2014.917055)
- 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, 831–854. doi:[10.1111/j.2044-8317.1998.tb00682.x](https://doi.org/10.1111/j.2044-8317.1998.tb00682.x)