A simple simulation technique for nonnormal data with prespecified
skewness, kurtosis, and covariance matrix

Njål Foldnes
BI Norwegian Business School

Ulf Henning Olsson
BI Norwegian Business School

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

Publisher's version available at

http://www.tandfonline.com/doi/full/10.1080/00273171.2015.1133274

A simple simulation technique for non-normal data with pre-specified skewness, kurtosis

and covariance matrix

Njål Foldnes and Ulf Henning Olsson

BI Norwegian Business School

Author Note

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

Abstract

We present and investigate a simple way to generate non-normal 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 is shown to have a non-Gaussian copula. Analytically, we obtain asymptotic robustness conditions for the IG distribution. Empirically, we show 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 non-normal data generation. We provide an implementation of the IG transform in the R environment.

Keywords: Simulation, Non-normal multivariate data, Kurtosis, Asymptotic Robustness, Structural equation modeling

A simple simulation technique for non-normal data with pre-specified skewness, kurtosis and covariance matrix

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 non-normal distributions (Micceri, 1989). It is therefore important to evaluate normality-based ML under realistic sample sizes and non-normal 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 non-normal distributions. The tradition in SEM simulation is to set up a non-normal 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 non-normal 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. So by specifying the covariance matrix, and also, 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 paper 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 is generated by the simulation technique proposed in this article. For instance, 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 non-normality. However such a conclusion is contingent on the special kind of non-normality offered by the VM transform. A less optimistic evaluation of the appropriateness of T_{ML} under non-normality 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 non-normality 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:

Multivariate distribution = Univariate marginal distributions + Copula.

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 instance, the copula associated with multivariate normality is defined as follows. Given a random multivariate normal vector $\boldsymbol{X} = (X_1, \ldots, X_d)'$ we may define a new vector $\boldsymbol{U}_X = (\phi(\frac{X_1-\mu_1}{\sigma_1}), \ldots, \phi(\frac{X_d-\mu_d}{\sigma_d}))'$, obtained from $\boldsymbol{X}$ by applying the standard normal cumulative distribution function ϕ to each standardised univariate marginal in $\boldsymbol{X}$. The copula of $\boldsymbol{X}$, known as the normal copula, is the cumulative distribution function of the vector $\boldsymbol{U}_X$. This procedure may be applied to any non-normal random vector $\boldsymbol{Y}$, whose copula is the distribution of the vector $\boldsymbol{U}_Y = (F_1(Y_1), \ldots, F_d(Y_d))'$, 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 both all univariate marginal distributions are normally distributed, and the associated copula is normal as defined above. Violation of the normality assumption can occur either if at least one univariate marginal distribution is non-normal, or if the copula is non-normal, or both. For instance, 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 non-normal 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 non-normal univariate observed variables, but with a normal copula. In this latter case, the distribution will appear highly non-normal 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 pre-specified univariate skewness and kurtosis, and a pre-specified covariance matrix. See Curran, West, and Finch (1996) for an example of an extensive Monte Carlo paper which 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 aspects of the VM-transform, i.e., its copula, is very closely related to the multivariate normal copula. Moreover, attempts to generalise 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 non-normal. 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 non-normal data with exact control of the covariance matrix and univariate skewness and kurtosis, and with truly non-normal copulas. The new method hence provides an alternative to the VM transform, and gives rise to a different set of non-normal distributions than 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 $\boldsymbol{Y}$ of p

observed variables. The transform represents the observed variables as $Y_i = \sum_{j=1}^{s} a_{ij} X_j$, for

$i = 1, \ldots, p$, where the a_{ij} are constant scalars and the X_j, $j = 1, \ldots s$, are mutually

independent random variables referred to as *independent generator* (IG) variables. In

practice, the vector $\boldsymbol{Y}$ 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 $\boldsymbol{Y}$. We say that any $\boldsymbol{Y}$ that can be represented stochastically in this way

has an IG distribution. These distributions are convenient for simulating multivariate

non-normal 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 $\boldsymbol{Y}$. 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 non-normal

distribution whose covariance matrix equals the implied covariance matrix. The task of

non-normal 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, e.g.,

repeated-measures designs (Molenaar & Rovine, 1998; Berkovits, Hancock, & Nevitt,

2000), multilevel random coefficients models (Rovine & Molenaar, 2000) and ANCOVA

(Headrick & Sawilowsky, 2000). Evaluating such methods under non-normal 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 where only the third- and

fourth-order moments are pre-specified for each univariate marginal distribution.

The goals of this article are threefold. First, to introduce and illustrate a new simulation technique based on IG variables. The second goal is to demonstrate the need for non-normal 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 non-normality, even for fixed levels of univariate skewness and kurtosis. In a real-world setting, we do not know the exact nature of our non-normal data, so we should be careful to include different kinds of non-normality into 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 non-normal IG distributions. We also demonstrate that the IG distribution has a non-normal copula, in contrast to the VM distribution, again showing that the IG transform produces conditions that may complement the VM transform.

This article is organised as follows. In the next section we present the IG transform, summarised in a five-step algorithm. This is followed by a section where 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 non-technical 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. The last section discusses limitations of the IG transform and gives 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

This paper investigates distributions obtained from linear combinations of independent non-normal variables:

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}, \tag{1}$$

where $\boldsymbol{X}$ is a column s-vector of generator variables, while $\boldsymbol{Y}$ is a column p-vector of generated variables. A researcher may form $\boldsymbol{Y}$ from the observed variables in the model under scrutiny. The topic of the current section is how we may choose the $\boldsymbol{A}$ 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 $j = 1, \ldots, s$, contained in the vector $\boldsymbol{X}$, are assumed to be mutually independent, and $\boldsymbol{A}$ 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 $\boldsymbol{\Sigma}$ for $\boldsymbol{Y}$, we employ some matrix $\boldsymbol{A}$ such that

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

This will ensure, given the unit variance and mutual independence of the IG variables, that $\boldsymbol{Y}$ has the target covariance $\boldsymbol{\Sigma}$. In a straightforward application of eq. (1) one can set $s = p$ and choose $\boldsymbol{A}$ as a square root matrix of $\boldsymbol{\Sigma}$. 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 $\boldsymbol{A}$ can be constructed from the model at hand as follows. A SEM model can be formulated (Bollen, 1989) as

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \qquad \boldsymbol{Y}_1 = \boldsymbol{\Lambda}_1\boldsymbol{\eta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{Y}_2 = \boldsymbol{\Lambda}_2\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{2}$$

where $\boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ are latent vector variates and the observed variate is $\boldsymbol{Y} = (\boldsymbol{Y}_1', \boldsymbol{Y}_2')'$. The covariance matrices of $\boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ are denoted by $\boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Theta}_\epsilon$ and $\boldsymbol{\Theta}_\delta$, respectively, and

are assumed to be positive definite. The matrices $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ contain regression coefficients, while $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ contain factor loadings. Note that (2) contains the multivariate regression model as a submodel, when the $\boldsymbol{Y}_i$ are measured without errors and $\boldsymbol{B} = \boldsymbol{0}$.

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

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

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

After $\boldsymbol{A}$ in eq. (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 α_V of a random variable V with expectation μ_V and standard deviation σ_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 (see Westfall (2014) for a recent discussion of the interpretation of kurtosis in terms of tail extremity). Kurtosis involves fourth-order moments. The relative, or excess, kurtosis β_V of V 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:

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

Hence, the skewness of the Y_i depends linearly on the skewnesses in the generator variables. Given user-specified skewness values $\alpha_{Y_i}, i = 1, \ldots, p$, there are p equations of type (4). Together they constitute a linear system in p equations with $s \geq p$ unknowns, namely the $\alpha_{X_j}, j = 1, \ldots, s$. In general such a system is consistent, and can be easily solved for the α_{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:

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

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

We can now summarise the IG transform simulation method:

1. The user specifies a target covariance matrix $\boldsymbol{\Sigma}$, and target skewness (α_{Y_i}) and kurtosis (β_{Y_i}) for each univariate marginal $Y_1, \ldots, Y_p$.

2. A matrix $\boldsymbol{A}$ is determined such that $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{\Sigma}$.

3. The systems of linear equations (4) and (5) are solved for α_{X_j} and β_{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 α_{X_j} and kurtosis β_{X_j}.

5. For a given sample size, draw independently random realisations of $X_j, i = 1 \ldots, s$. Apply the linear transformation (1) to obtain a random sample from $\boldsymbol{Y}$.

The IG transform above stochastically defines a large class of distributions, where steps 2, 3 and 4 can be executed in various ways. In step 2 many $\boldsymbol{A}$ are possible. If $s > p$ there might be several possible solutions for α_{X_j} and β_{X_j} in step 3. In step 4 there are different candidate distributions for X_j with given skewness and kurtosis, for instance 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 which 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. 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 dataset (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 $\boldsymbol{\Sigma} = \hat{\boldsymbol{\Sigma}}$:

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

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 $Y_1, Y_2, \ldots, Y_{11}$: $(0, 0, 0, 0, 1, 1, 1, 1, 2, 2, 2)$, and the following excess

kurtosis values: $(1, 1, 1, 1, 3, 3, 3, 3, 7, 7, 7)$. For instance, for Y_5 we desire a distribution with skewness 1 and excess kurtosis 3.

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

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

The third step involves solving the equations (4) and (5) for skewness α_{X_j} and kurtosis β_{X_j}, respectively. Solving a linear system of equations is a routine task performed by any mathematical software package. The solutions for skewness α_{X_j}, $j = 1, \ldots, 11$, 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 $(1, 2.18, 2.46, 5.81, 17.36, 14.06, 18.79, 29.51, 12.44, 96.07, 56.42)$. So for instance, 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 which have the required skewness and kurtosis. For instance, 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 instance 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 instance, 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 & KlÃ¼ßner, 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 eleven samples are combined as columns to form a $N \times 11$ matrix $\boldsymbol{D}$. The final simulated dataset is then obtained as the $N \times 11$ data matrix $\boldsymbol{D} \cdot \boldsymbol{A}'$, where each row represents a single draw from the random vector $\boldsymbol{Y}$.

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 non-normal 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 non-normal. It is well known that under some specific combinations of non-normal 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 characterised in a body of literature referred to as asymptotic robustness (AR) (e.g., Shapiro, 1987; Browne & Shapiro, 1988; Amemiya & Anderson, 1990; Satorra & Bentler, 1990). If the IG transform is to be used to realistically assess T_{ML}, for instance 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 non-null, 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. Interestingly, 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, then 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 non-normal distributions is presently unknown.

It is also important to establish that the IG transform is able to generate data with a non-normal 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 non-normal 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 non-normal 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

non-optimal conditions should avoid AR. So what we need is a necessary condition for AR.

By violating this condition in the simulation set-up, 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 ξ and three observed variables Y_1, Y_2, Y_3, with $var(\xi) = 1$ 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 ψ_i^2 is the variance of the measurement residual corresponding to Y_j.

Hence the free parameters in Model S are λ, ψ_1^2 and ψ_3^2. These parameters constitute the

parameter vector $\boldsymbol{\theta} = (\lambda, \psi_1^2, \psi_3^2)'$. The model-implied covariance matrix of Model S is

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

This matrix is symmetric, and contains redundant elements. For a symmetric $p \times p$-matrix

$\boldsymbol{G}$, let $vech(\boldsymbol{G})$ represent the column vector consisting of the $p^* = p(p+1)/2$

non-duplicated elements of $\boldsymbol{G}$, that is, the diagonal and supra-diagonal elements. So we

can represent $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ by the column vector

$$
\boldsymbol{\sigma}(\boldsymbol{\theta}) = vech(\boldsymbol{\Sigma}(\boldsymbol{\theta})) = (\lambda^2 + \psi_1^2, \lambda^2, \lambda^2 + \psi_1^2, \lambda^2, \lambda^2, \lambda^2 + \psi_3^2)',
$$

containing the $p^* = 3 \cdot 4/2 = 6$ unique elements of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. The same procedure may be

applied to the sample covariance matrix $\boldsymbol{S}$, which is then more economically represented as

$\boldsymbol{s} = vech(\boldsymbol{S})$. 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 $\boldsymbol{\Delta}(\boldsymbol{\theta}) = \partial \boldsymbol{\sigma}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'$, where each entry of $\boldsymbol{\sigma}(\boldsymbol{\theta})$ is partially differentiated with respect to each parameter in $\boldsymbol{\theta}$. For model S we have

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

Next we consider a central element in SEM, namely the sampling distribution of the covariances, i.e., the distribution of $\boldsymbol{s}$. The asymptotic covariance matrix of $\sqrt{n}\boldsymbol{s}$ is denoted by $\boldsymbol{\Gamma}$, 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 λ, ψ_1^2 and ψ_3^2 to 1, such that the target covariance matrix is $\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$. Then we set $\boldsymbol{A} = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & \sqrt{3/2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{6} & 2/\sqrt{3} \end{pmatrix}$ where $\boldsymbol{A}\boldsymbol{A}' = \Sigma$. There is a general formula for $\boldsymbol{\Gamma}$ when data comes from an IG transform (Browne & Shapiro, 1988, Theorem 2.1):

$$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_N + \boldsymbol{K}_p'\tilde{\boldsymbol{A}}\boldsymbol{C}\tilde{\boldsymbol{A}}'\boldsymbol{K}_p. \tag{6}$$

Here $\tilde{\boldsymbol{A}}$ is the $p^2 \times s$ matrix whose jth column is $\boldsymbol{a}_j \otimes \boldsymbol{a}_j$, with $\boldsymbol{a}_j$ being the j-th column vector of $\boldsymbol{A}$. The $\otimes$ symbol denotes the Kronecker product. Then $\boldsymbol{a}_j \otimes \boldsymbol{a}_j$ consists of stacking p copies of $\boldsymbol{a}_j$, where the ith copy is being multiplied by a_{ij}, for $i = 1, \ldots, p$. For

Model S, with the $\boldsymbol{A}$ given above, this gives

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

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

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

Finally, eq. (6) contains $\boldsymbol{\Gamma}_N$, defined as the asymptotic covariance matrix of $\sqrt{n}\boldsymbol{s}$ for multivariate normal data, which is given by $\boldsymbol{\Gamma}_N = 2\boldsymbol{K}_p'(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\boldsymbol{K}_p$. Returning to Model S, with fixed parameter values λ, ψ_1^2 and ψ_3^2 all equal to 1, and with excess kurtosis in the IG variables denoted by $\beta_i = \beta_{X_i}$ for $i = 1, 2, 3$, eq. (6) gives:

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

Setting $\beta_i = 0$ for $i = 1, 2, 3$ in the above expression recovers $\boldsymbol{\Gamma}_N$. Recall that the elements of $\boldsymbol{s}$ are ordered as $s_{11}, s_{12}, s_{22}, s_{13}, s_{23}, s_{33}$. So, for instance, 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 $\hat{\boldsymbol{\theta}}$ minimises

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

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

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

where the χ_1^2 are mutually independent chi-squares with one degree of freedom and $\xrightarrow{d}$ denotes convergence in distribution. The γ_j are the non-zero eigenvalues of $\boldsymbol{U}\boldsymbol{\Gamma}$ where $\boldsymbol{U} = \boldsymbol{\Gamma}_N^{-1} - \boldsymbol{\Gamma}_N^{-1}\boldsymbol{\Delta}\left\{\boldsymbol{\Delta}'\boldsymbol{\Gamma}_N^{-1}\boldsymbol{\Delta}\right\}^{-1}\boldsymbol{\Delta}'\boldsymbol{\Gamma}_N^{-1}$. 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 as follows: $\beta_1 = 2, \beta_2 = 4$ and $\beta_3 = 6$, the non-zero eigenvalues of $\boldsymbol{U}\boldsymbol{\Gamma}$ 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 $\boldsymbol{\Sigma}$, 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 above 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 $c = d/trace(\boldsymbol{U\Gamma})$, 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 $\boldsymbol{D}$ such that

$$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_N + \boldsymbol{\Delta D}' + \boldsymbol{D\Delta}'. \tag{8}$$

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

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

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

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

Model S one orthogonal complement is given by

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

So that $\boldsymbol{P} = \begin{pmatrix} 0 & 0 & 0 \\ 1/2 & -1/2 & 0 \\ 3/2 & -3/2 & 0 \end{pmatrix}$, which contains non-zero elements.

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

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

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

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 $\boldsymbol{P} = 0$ is actually equivalent to AR.

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

Proof. By AR, it follows from equations (6) and (8) that some $\boldsymbol{D}$ exists such that $\boldsymbol{\Delta D'} + \boldsymbol{D\Delta'} = \boldsymbol{K'_p \tilde{A} C \tilde{A}' K_p}$ and premultiplication with $\boldsymbol{\Delta'_c}$ and postmultiplication by $\boldsymbol{\Delta_c}$ gives the required result. Now assume that $\beta_{X_j} > 0$ for all j, i.e. that the diagonal elements of $\boldsymbol{C}$ are all positive. The diagonal elements of $\boldsymbol{PCP'}$ are quadratic forms $\boldsymbol{p_j C p'_j}$ where $\boldsymbol{p_j}$ is the jth row of $\boldsymbol{P}$. Now $\boldsymbol{p_j C p'_j} = 0$ implies that $\boldsymbol{p_j} = \boldsymbol{0}$ 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 γ_j is equal to d. In general this does not imply that we are in an AR situation, with $\gamma_j = 1$ for all $j = 1, \ldots, d$. 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 $\boldsymbol{P} = \boldsymbol{0}$. Under an IG distribution we have from (6)

$$trace(\boldsymbol{U\Gamma}) = trace((\boldsymbol{\Delta'_c \Gamma_N \Delta_c})^{-1}(\boldsymbol{\Delta'_c \Gamma \Delta_c})) = trace(\boldsymbol{I}) + tr(\boldsymbol{QC}),$$

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

Assume $c = 1$. Then $tr(\boldsymbol{QC}) = 0$. Define $\boldsymbol{M} = (\boldsymbol{\Delta'_c \Gamma_N \Delta_c})^{-1}$. Then

$$trace(\boldsymbol{QC}) = trace(\boldsymbol{P'MPC}) = trace(\boldsymbol{\tilde{P}\tilde{P}'M}) = trace(\boldsymbol{\tilde{P}'M\tilde{P}}) = trace(\boldsymbol{\breve{P}'\breve{P}})$$

where $\tilde{\boldsymbol{P}} = \boldsymbol{P}\boldsymbol{C}^{1/2}$ and $\check{\boldsymbol{P}} = \tilde{\boldsymbol{P}}\boldsymbol{M}^{1/2}$. Since $tr(\boldsymbol{Q}\boldsymbol{C}) = 0$, it follows that $\check{\boldsymbol{P}} = \boldsymbol{0}$, which is equivalent to $\tilde{\boldsymbol{P}}\boldsymbol{M}^{1/2} = \boldsymbol{P}\boldsymbol{C}^{1/2}\boldsymbol{M}^{1/2} = \boldsymbol{0}$. Since $\boldsymbol{C}$ and $\boldsymbol{M}$ are non-singular, it follows that $\boldsymbol{P} = \boldsymbol{0}$, which implies AR, by Proposition 1. $\qquad\square$

We have seen that when all IG variables have more, or all have less, kurtosis than a normal variable, then $\boldsymbol{P}\boldsymbol{C}\boldsymbol{P}' = \boldsymbol{0}$ and $\boldsymbol{P} = \boldsymbol{0}$ are equivalent conditions, both again equivalent to AR. A natural follow-up question is whether $\boldsymbol{P} = \boldsymbol{0}$ 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 $\boldsymbol{P} = \begin{pmatrix} 0 & 0 & 0 \\ 1/2 & -1/2 & 0 \\ 3/2 & -3/2 & 0 \end{pmatrix} \neq \boldsymbol{0}$. 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 $\boldsymbol{D}' = \frac{1}{12}\begin{pmatrix} 8 & 3 & -1 & 3 & 0 & 1 \\ 8 & 2 & -10 & 2 & -4 & 0 \\ -6 & -2 & 0 & -2 & 0 & 0 \end{pmatrix}$, then $\boldsymbol{K}_3'\tilde{\boldsymbol{A}}\boldsymbol{C}\tilde{\boldsymbol{A}}'\boldsymbol{K}_3 = \boldsymbol{\Delta}\boldsymbol{D}' + \boldsymbol{D}\boldsymbol{\Delta}'$. Hence, in this case $\boldsymbol{P}$ is not zero, but if $\beta_1 = 1, \beta_2 = -1$ and $\beta_3 = 1/16$, then AR holds. So $\boldsymbol{P} = \boldsymbol{0}$ 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 $F_i(y) = P(Y_i \leq y)$, $i = 1, \ldots, p$. Sklar (1959) showed that a full description of the random vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_p)'$ may be uniquely obtained by separating the marginal distributions F_i, from the dependence structure, i.e. the copula. More precisely, the probability integral transform applied to each component of $\boldsymbol{Y}$ gives the random vector $(U_1, \ldots, U_p) = (F_1(Y_1), \ldots, F_p(Y_p))$, with uniform marginal distributions. The copula C of $\boldsymbol{Y}$ is the joint cumulative distribution function of $(U_1, \cdots, U_p)$. So 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 non-normal nature of the univariate marginal distributions, appears to be highly non-normal. 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. Intuitively this is not surprising, given that the Fleishman polynomials in the VM transform do not contain interactions between the normally distributed generator variables $\boldsymbol{Z}$ for $\boldsymbol{Y}$. 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 $\boldsymbol{Z} = (Z_1, \ldots, Z_p)'$ is multivariate normal, and the VM output $\boldsymbol{Y} = (Y_1, \ldots, Y_p)'$ inherits some basic properties of $\boldsymbol{Z}$.

Compare this to the IG transform in eq. (1), where each Y_i is a function of potentially all the generator variables X_j. Although eq. (1) is a simple way to obtain dependence between the Y_i, i.e. by assuming linearity and independence, the resulting distribution for $\boldsymbol{Y}$ has a non-normal 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 $\boldsymbol{A}$ is non-singular. Suppose that the probability density function (pdf) of X_1 is non-zero everywhere, while the pdf of X_2 is non-zero only in an interval $[b, \infty)$. For instance might X_1 be normally distributed and X_2 a scaled χ^2. Fix Y_1 to any value y_1. If we assume that both a_{11} and a_{21} are non-zero, it follows from

$$y_1 \;\; = a_{11}x_1 + a_{12}x_2$$

$$y_2 \;\; = a_{21}x_1 + a_{22}x_2$$

that

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

Since $\boldsymbol{A}$ is invertible we have $\left(a_{22} - \frac{a_{21}a_{21}}{a_{11}}\right) \neq 0$, and suppose without loss of generality that

this expression is positive. Then, given $Y_1 = y_1$, the conditional density of Y_2 is non-zero only in $[\frac{a_{21}}{a_{11}}y_1 + (a_{22} - \frac{a_{12}a_{21}}{a_{11}})b, \infty)$. So 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 non-zero and since the pdf of X_1 is non-zero everywhere, it follows that the pdf of Y_2 is non-zero 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 $\boldsymbol{Y}$ has $c_{12}(F_1(y_1), F_2(y_2)) = 0$ for some y_1, y_2. However, since the bivariate normal pdf is non-zero everywhere, the bivariate normal copula is non-zero for all interior points of the unit square, so c_{12} can not 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 non-normal data. Our goal is to illustrate the IG transform in a specific example, and also 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 non-normality conditions, we employ a simple model, given in Figure 2, with two factors ξ_1 and ξ_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 ξ_1 and ξ_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

$$
\mathbf{\Sigma} = \begin{pmatrix} 1.40 & 0.80 & 0.20 & 0.16 \\ 0.80 & 1.04 & 0.16 & 0.13 \\ 0.20 & 0.16 & 1.40 & 0.80 \\ 0.16 & 0.13 & 0.80 & 1.04 \end{pmatrix}, \qquad \mathbf{A} = \begin{pmatrix} 1.183 & 0 & 0 & 0 \\ 0.676 & 0.763 & 0 & 0 \\ 0.169 & 0.060 & 1.170 & 0 \\ 0.135 & 0.048 & 0.662 & 0.762 \end{pmatrix}.
$$

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

The Monte Carlo design include the following factors

- Two sample sizes, $n = 100$ and $n = 500$

- Two levels of non-normality.

    - *Moderate* non-normality has skewness α_{Y_i} equal to $0, 0, 1, 1$ and kurtosis β_{Y_i} equal to $1, 1, 3, 3$ in Y_1, Y_2, Y_3 and Y_4, respectively.

    - *Severe* non-normality has skewness α_{Y_i} equal to $2, 2, 3, 3$ and kurtosis β_{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 non-normality in the Y_i variables, $i = 1, \ldots, 4$, we calculated skewness and kurtosis values in the generator variables $X_1, \ldots, X_4$. 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 $X_1, \ldots, X_4$ are given in Table 1. Each generator variable X_j was then simulated independently, with zero mean, unit variance, zero 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 give 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 2000 replications in each cell, so that the maximal standard error for the rejection rates are $\sqrt{0.5^2/2000} = 0.011$. 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 that of a χ^2 distribution with seven degrees of freedom. So the asymptotic mean and variance is 7 and 14, respectively. As expected, due to excess kurtosis, T_{ML} is inflated, with rejection rates larger than the nominal levels. The non-normality correction in T_{SB} only partly ameliorates the inflation. Importantly, in all conditions the behaviour of of both T_{ML} and T_{SB} are 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 instance, in the $n = 500$ condition with severe non-normality, 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 transform 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 non-normal 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 non-normality 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, say, 10, it may be necessary for some IG variables to have a kurtosis of, say, 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. Hovever, these are linear systems of equations, where a solution does not exist only if the coefficient matrix is singular, in which case the matrix $\boldsymbol{A}$ 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 $\boldsymbol{A}$. Whether this poses a serious

limitation for the IG transform is a topic for future research.

Note that while the VM transform defines one single 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 does only partially specify the multivariate distribution. The IG class of distributions, for given covariance and univariate moments, is spanned by two independent choices. Firstly, the choice of $\boldsymbol{A}$, among the many different square and non-square matrices whose product $\boldsymbol{A}\boldsymbol{A}'$ equals the target covariance matrix. Each choice of $\boldsymbol{A}$ leads to a different multivariate distribution. Secondly, the choice of distributions for the IG variables. There are many univariate distributions with a given skewness and kurtosis, and a researcher may use Fleishman polynomials, Pearson distributions or Johnson distributions. All of these will result in different distributions for the IG transform. Finally, we remark that the IG transform can easily be extended to higher order moments, e.g. 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 non-normal 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 than 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 non-normal 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 non-normality. 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 robustess of normal-theory SEM inference.

ACKNOWLEDGMENTS

References

Amemiya, Y., & Anderson, T. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics*, 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*, 1-27. doi: 10.1177/0013164414548894

Becker, M., & KlÃűßner, 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.* 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), 1066–18. 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

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

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

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Addison-Wesley Pub. Co.

Mair, P., Satorra, A., & Bentler, P. M. (2012). Generating Nonnormal Multivariate Data Using Copulas: Applications to SEM. *Multivariate Behavioral Research*, *47*(4), 547–565. doi: 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

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

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

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

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. In *Proceedings of the american statistical association* (p. 308-313). Alexandria, VA. doi: 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

Shapiro, A. (1987). Robustness properties of the mdf analysis of moment structures. *South African Statistical Journal*, *21*, 39-62.

Sklar, M. (1959). *Fonctions de rÃľpartition à n dimensions et leurs marges*. UniversitÃľ Paris 8.

Vale, C., & Maurelli, V. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*(465–471). doi: 10.1007/BF02293687

Vuolo, M. (2015, June). Copula Models for Sociology: Measures of Dependence and Probabilities for Joint Distributions. *Sociological Methods & Research*, 1–45. doi: 10.1177/0049124115584477

Westfall, P. H. (2014). Kurtosis as peakedness, 1905–2014. rip. *The American Statistician*, *68*(3), 191–195. doi: 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

Table 1

Skewness α_{X_j} and kurtosis β_{X_j} in the generator variables $X_1, \ldots, X_4$.

Non-normality		X_1	X_2	X_3	X_4
Moderate	α_{X_j}	0	0	1.035	1.716
	β_{X_j}	1	2.569	3.142	7.820
Severe	α_{X_j}	2	3.378	3.100	5.140
	β_{X_j}	5	12.843	15.711	39.098

Table 2

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

| Non-normality | 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

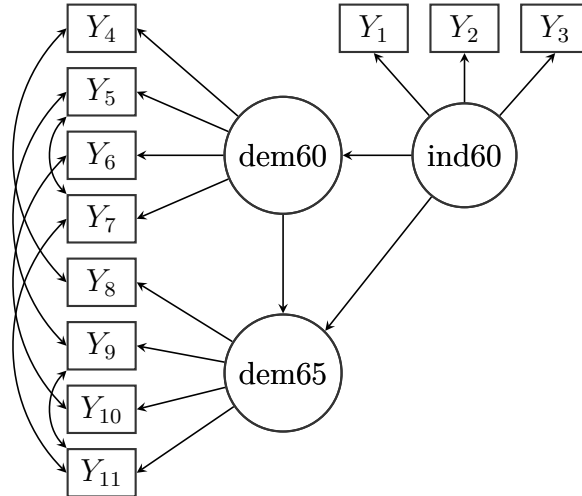*Figure 1*. Bollen's political democracy model. dem60: Democracy in 1960. dem65: Democracy in 1965. ind60: Industrialisation in 1960.

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