# How general is the Vale-Maurelli simulation approach?

Njål Foldnes
BI Norwegian Business School

Steffen Grønneberg
BI Norwegian Business School

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

*Psychometrika,* 80(2015)4:1066-1083

Publisher's version available at http://dx.doi.org/10.1007/s11336-014-9414-0

# HOW GENERAL IS THE VALE-MAURELLI SIMULATION APPROACH?

NJÅL FOLDNES AND STEFFEN GRØNNEBERG

Abstract. The Vale-Maurelli (VM) approach to generating non-normal multivariate data involves the use of Fleishman polynomials applied to an underlying Gaussian random vector. This method has been extensively used in Monte Carlo studies during the last three decades to investigate the finite-sample performance of estimators under non-Gaussian conditions. The validity of conclusions drawn from these studies clearly depends on the range of distributions obtainable with the VM method. We deduce the distribution and the copula for a vector generated by a generalized VM transformation, and show that it is fundamentally linked to the underlying Gaussian distribution and copula. In the process we derive the distribution of the Fleishman polynomial in full generality. While data generated with the VM approach appears to be highly non-normal, its truly multivariate properties are close to the Gaussian case. A Monte Carlo study illustrates that generating data with a different copula than that implied by the VM approach severely weakens the performance of normal-theory based ML estimates.

Contents

## 1. Introduction and summary

Multivariate data simulation is the main tool in evaluating the finite-sample performance of estimation methods and goodness-of-fit measures in structural equation modeling (SEM). A popular and seemingly rather general data generation technique is the use of Fleishman polynomials (Fleishman, 1978) applied to an underlying multivariate normal vector, as presented by Vale & Maurelli (1983). Here, a Gaussian random vector $Z$ is transformed to a non-Gaussian random vector $X$ through applying third degree polynomials to each of its marginals. The polynomials are chosen so that $X$ has prespecified univariate moments up to fourth order and the covariance matrix of $Z$ is thereafter chosen so that $X$ has a prespecified covariance structure. The Vale-Maurelli (VM) approach is implemented in popular software packages like Mplus (Muthén & Muthén, 2010), EQS (Bentler, 2006), Lisrel (Jöreskog & Sorbom, 2006) and lavaan (Rosseel, 2012). Over a span of several decades, Monte Carlo studies have employed the VM technique for data generation (e.g., Bentler & Tong, 2013; Fouladi, 2000; Curran et al., 1996).

In this paper we investigate the dependence structure of a generalization of the VM transformation. The generalization consists in allowing each marginal transform to be any function whatever, not just a low-degree polynomial, and we also allow the generator vector $Z$ to be non-Gaussian. We point out some fundamental limitations of these type of transformations through deriving a precise expression of the resulting copula of the non-Gaussian distribution. These limitations are essentially due to the lack of interaction terms in the transformation from $Z$ to $X$. The copula – or dependence properties – of the simulated vector is found to be closely linked with the copula of $Z$. In the classical VM framework, were $Z$ is Gaussian, this means that the truly multivariate properties of the random vectors generated by VM approach is either very close to or equal to the Gaussian case, even though it would seem that the resulting highly kurtotic random variables are very far away from the Gaussian case. It follows that evaluating the finite-sample behaviour of estimators and goodness-of-fit statistics with data generated by the VM approach may give biased conclusions favoring methods performing well with observations from Gaussian copulas. This is a conceptually serious issue, as most non-likelihood based methods are designed to work well irrespective of the underlying distribution – except for certain moment properties such as the covariance structure.

Tadikamalla (1980) criticized the Fleishman transform for having an unknown distribution and quantile function, and Kotz et al. (2000, Chap. 44.10) similarly criticized the VM transform for not having a known distribution. We derive the full VM distribution, which unfortunately is computationally difficult to use, except in cases where the Fleishman transforms are mostly monotonous. We also produce a concrete formula for the cumulative distribution function of the Fleishman approach. While numerical methods are needed for its inversion to produce

the quantile function, this inversion is well-behaved by the strict monotonousness of the cumulative distribution function. The computational complexity of calculating quantiles of the Fleishman transform is therefore admittedly more complex than merely applying a formula, but is well within reach for modern computers. To a large degree, this counters the critique of the Fleishman transform raised by Tadikamalla (1980) .

An alternative method for simulating data with pre-specified covariance matrix from Gaussian variables is the NORTA method of Cario & Nelson (1997). Although NORTA has the advantage of allowing completely arbitrary marginals while the classical VM transform only controls lower-order moments, it has not to our knowledge been employed in the SEM literature. Also the NORTA method yields data whose full distribution is easily calculated and understood. However, NORTA shares the limitation of the classical VM transform that its dependence structure is strongly tied to the Gaussian case.

Our basic perspective for studying multivariate dependence in a random vector is the perspective of Joe (1997): We will study multivariate dependence through the copula. Since the concept of a copula is fundamental to both our arguments concerning the limitations of the VM approach, we start our paper with a brief introductory section concerning copulas focused towards moment based models, such as covariance structure models. The concept of tail dependence is introduced, and we explain how properties of moment based estimators depend not just on the marginal distribution of the data, but also on the copula of the data. In Section 3, we deduce the distribution and copula of the generalized VM transform and provide a basic investigation of its properties. Section 3.3 investigates some consequences for the classical VM method. We next illustrate by simulation how the finite-sample behavior of factor model estimates are affected by changing the copula of the data in Section 4. A concluding discussion is given in Section 5.

## 2. The concept of dependence and copulas

SEM and covariance structure analysis are usually seen as moment based methods, where dependency among random variables is typically modeled not by a full specification of the distribution of the data, but merely by the covariance matrix $\Sigma$ of a random vector $X$ induced by relationships between postulated latent variables. As is well known, if $X$ is Gaussian, the dependence structure is indeed completely described by $\Sigma$. However, in the non-Gaussian case two multivariate distributions may differ fundamentally in terms of dependence structure, but still have the same covariance matrix $\Sigma$. By dependence structure, we mean a copula, i.e. a function that binds together marginal distributions to obtain a joint distribution. In psychometric and SEM literature, the use of copulas is relatively limited. Braeken et al. (2007) used copulas for modeling residual dependencies in Rasch models.

Mair et al. (2012) propose to use copulas to generate data with a pre-specified covariance matrix, offering an alternative to the VM method. Their copula-based method has the disadvantage relative to the VM method that it does not control the marginals. That is, an exact specification of univariate skewness and kurtosis allowed for in the VM approach is not possible with the copula-based approach proposed by Mair et al. (2012). However, their approach can generate data with a wide scope of copulas, and the method does give some control of the marginal properties of the generated random vector.

2.1. **Copula Theory.** We refer to the monograph by Joe (1997) for a thorough review of copula theory, and limit ourselves here to some basic definitions. We will focus exclusively on the case when the marginal distributions $H_i(x) = P(X_i \leq x)$, $i = 1, \ldots, d$ are continuous. In this case, Sklar (1959) noticed that a full description of a random vector $X = (X_1, X_2, \ldots, X_d)^t$ may be uniquely obtained by separating the marginal distributions $H_i(x) = P(X_i \leq x)$, $i = 1, \ldots, d$, from the dependence structure, i.e. the copula. More precisely, the integral transform applied to each component of $X$ gives the random vector $(U_1, \ldots, U_d) = (H_1(X_1), \ldots, H_d(X_d))$, with uniform margins. The copula $C$ of $X$ is the joint cumulative distribution function of $(U_1, \cdots, U_d)$. So a copula is a cumulative distribution function with uniform marginals. Sklar's theorem states that, for any $d$-dimensional cdf $H$ with marginals $H_1, \ldots, H_d$, there exists a unique copula function $C$ such that $H(x_1, \ldots, x_d) = C(H_1(x_1), \ldots, H_d(x_d))$, and conversely, given a copula $C$ and any marginals $H_1, \ldots, H_d$, the function

$$H(x) = C(H_1(x_1), \ldots, H_d(x_d))$$

is a cumulative distribution function.

To measure the dependence in a pair $(X_i, X_j)$ rank-based parameters like Kendall's tau and Spearman's rho are functions only of the copula of $(X_i, X_j)$, and not their marginals. This is unlike the more familiar correlation coefficient of Pearson, which depends upon both marginal distribution and the copula. To illustrate the limitation of the VM approach we will investigate another dependence measure that is a copula property, namely tail dependence. Upper tail dependence in $(X_i, X_j)$ is the probability that $X_i$ exceeds its $q$-quantile, given that $X_j$ exceeds its own $q$-quantile. More precisely, the upper tail dependence parameter $\lambda_u$, if it exists, is defined as the limit of this conditional probability as $q \to 1^-$. The lower tail dependence parameter $\lambda_l$ is motivated similarly. Mathematically, upper and lower tail dependence can be expressed in terms of the copula $C$ of $(X_i, X_j)$ as

$$\lambda_u = \lim_{q \to 1^-} \frac{1 - 2q + C(q, q)}{1 - q}, \qquad \lambda_l = \lim_{q \to 0^+} \frac{C(q, q)}{q}$$

provided the limits exist. When $(X_i, X_j)$ has a bivariate Gaussian distribution both lower and upper tail dependence are equal to zero. We will show that this property

is carried over to the VM transformed vector, which exemplifies the restrictions of the types of non-normality achievable under the VM method.

2.2. **The role of copulas in SEM.** Covariance models only specify a parametric structure of second order moments of the data, and estimation techniques aim at estimating these parameters without assuming anything else of the observations except the correctness of these moment assumptions. In SEM, all standard estimators in use are minimum distance (MD) estimators that match model implied moments with empirical moments. The finite sample distribution of these estimators is a function of the full joint distribution of the data. As the copula point of view is simply a re-phrasing of the joint distribution into a univariate part and a dependency part, this obviously means that finite sample behaviour is a function of the copula and the univariate marginals.

We are unaware of any uniform bounds on the behaviour of MD estimates in the whole class of all distributions pertaining to the correct covariance structure. The traditional technique for assessing finite sample behaviour is to simulate data from a particular class of distributions. It is then hoped that the estimation behaviour under the chosen class of distributions reflects real world data sets.

The selection of simulation distributions is typically based on two considerations. Firstly, one tries to emulate distributional aspects of real world datasets. Secondly, one can use asymptotic theory to motivate which design factors to vary. As we will shortly summarise, in asymptotic theory, the fourth order moments is the most important aspect of the distribution of the estimates, and most simulation studies have focused on varying these.

Both considerations are related to copulas. The distributional aspects that are emulated are typically only univariate, or – rarely – bivariate. However, the full multivariate distribution will play a role, and the type of higher dimensional copulas that can be expected in social science data seems to be unknown. This perspective could lead to new and more realistic classes of distributions to simulate from, and to the identification of other distributional aspects that have important influence on the finite sample behaviour of MD estimators.

In large samples, mathematical theory shows exactly what part of the data-generating mechanism influences the distribution of parameter estimates. SEM estimation involves MD estimators that fulfil the conditions of Browne (1984). This implies that MD estimators are asymptotically normal and with covariance matrix of the vectorized parameters as given in eq. (2.12a) in Browne (1984), namely

$$(\Delta_0' V \Delta_0)^{-1} \Delta_0' V \Gamma V \Delta_0 (\Delta_0' V \Delta_0)^{-1} = A \Gamma A'. \tag{1}$$

Here, the matrix constituents of $A$ depend solely on the chosen estimator or the model. The matrix $\Gamma$, however, depends on the distribution of the data. $\Gamma$ is the asymptotic covariance matrix of the vectorized sample covariance matrix $s$:

$$\sqrt{n}(s-\sigma) \xrightarrow{L} N(0,\Gamma).$$

The elements of $\Gamma$ have the form

$$\Gamma_{ij,kl} = \sigma_{ijkl} - \Sigma_{ij}\Sigma_{kl}, \tag{2}$$

where the fourth-order moments $\sigma_{ijkl}$ are given by

$$\sigma_{ijkl} = \mathbb{E}(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)(X_l-\mu_l), \quad \mu_i = \mathbb{E}X_i,$$

see eq. (2.2) in Browne (1984). When $\Sigma$ is pre-specified, as is always the case in simulation studies, the asymptotic behaviour of the MD estimator is entirely determined by $\sigma_{ijkl}$, which depends on the multivariate aspects of the data, i.e. the copula. More precisely, assume for simplicity that all variables have zero mean, then we have that

$$\sigma_{ijkl} = \int_{\mathbb{R}^4} x_i x_j x_k x_l \mathrm{d}H_{ijkl}(x_i,x_j,x_k,x_l),$$

where $H_{ijkl}$ is the cumulative distribution function of $(X_i,X_j,X_k,X_l)$. With a change of variables this can be reformulated in terms of the copula, as

$$\sigma_{ijkl} = \int_{[0,1]^4} H_i^{-1}(u_i)H_j^{-1}(u_j)H_k^{-1}(u_k)H_l^{-1}(u_l)\mathrm{d}C_{ijkl}(u_i,u_j,u_k,u_l),$$

where $C_{ijkl}$ is the copula of $(X_i,X_j,X_k,X_l)$ and $H_i^{-1}$ is the quantile function of $X_i$. This shows how $\Gamma$ is a function of both the univariate marginals and the underlying copula $C$ of $X$.

The popularity of the VM approach is due to its ability to control univariate kurtosis, motivated by the above form of $\Gamma$. However, as is clearly seen from the above display, the VM approach only specifies the special case where $i=j=k=l$, but offers no control in the general case. This means that even though only the first four (multivariate) moments of the data matters for the large sample behaviour of MD estimators, the classical VM method, and its generalisation defined in the next section, is not general enough to replicate all possible values of $\Gamma$.

## 3. The multivariate distribution of the generalized Vale-Maurelli transform

Suppose given a random vector $Z = (Z_1, Z_2, \ldots, Z_d)^t$ with joint distribution

$$F(z_1,\ldots,z_d) = P(Z_1 \leq z_1, \ldots, Z_d \leq z_d)$$

and marginal distributions

$$F_i(z_i) = P(Z_i \leq z_i).$$

We assume that $Z$ is a continuous random vector, and our canonical example is the multivariate Gaussian distribution, denoted $Z \sim N_d(\mu,\Sigma)$. Consider the variable

$X = (X_1, X_2, \ldots, X_d)^t$, where $X = p(Z)$, in which $p : \mathbb{R}^d \mapsto \mathbb{R}^d$ is defined through continuous functions $p_1, p_2, \ldots, p_d$ where $p_i : \mathbb{R} \mapsto \mathbb{R}$, and

$$p(z_1, z_2, \ldots, z_d) = (p_1(z_1), \ldots, p_d(z_d))^t .$$

That is, the $k$'th coordinate of $p(z)$ only depends on $z_k$.

Let us denote the cumulative distribution function of $X$ by

$$H(x_1, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d) = P(p_1(Z_1) \leq x_1, \ldots, p_d(Z_d) \leq x_d)$$

and the $i$'th marginal distribution of $X$ by

$$H_i(x_i) = P(X_i \leq x_i) = P(p_i(Z_i) \leq x_i).$$

The transformation $p(Z)$ does not induce any interactions between the elements of $Z$. When the $p_i$ are polynomial functions, this is the VM transform. Hence, we will call this class of transformations the generalized VM approach.

Note that the choice $p_i(z) = G_i^{-1}(F_i(z))$ where $G_i$ is a continuous CDF induces $H_i = G_i$. Hence, the generalized VM transformation supports any marginal distribution of $X$. This follows because of the well-known fact that $F_i(Z_i) \sim U[0, 1]$, and that $G_i^{-1}(U) \sim G_i$ when $G_i$ is a CDF and $U \sim U[0, 1]$. Hence, the generalized VM transformation not only supports any univariate skewness and kurtosis, but any marginal distribution whatever.

We will show that despite this seeming generality, the fact that $p(Z)$ does not induce any interactions between the elements of $Z$ implies that the generalized VM approach has a copula which is closely linked to the copula of the generator variable $Z$.

3.1. **The univariate distribution.** Let $d = 1$ and suppose that $Z$ is a random variable with a continuous cumulative distribution function. Note that $Z$ may be any random variable, not necessarily normally distributed. Let us consider the transformed variable

$$X = p(Z)$$

for some continuous function $p : \mathbb{R} \mapsto \mathbb{R}$. Our canonical example of the kinds of transformations we will focus on is the polynomial case of Fleishman (1978). Here, $Z$ is standard normal variable and $p$ is a third-degree polynomial.

For analytical tractability the transformation $p$ has been previously restricted to be strictly monotone (Headrick, 2007; Headrick & Pant, 2012). Tadikamalla (1980) noted that the exact distribution of $X$ was not known, even for third-degree polynomials. The following result provides the closed form expression of the density of $X$ under fairly general assumptions on $p$, including non-monotonous polynomials of any degree. While this result is surely not new, we have been unable to find a reference for it in the literature. As the multivariate results in Section 3.2 will use the technical details of the following univariate proof, we feel justified in providing a full technical proof for this fundamental result.

We will need the following notation. Let $\{A_i\}_{i=0,\dots,n}$ be the maximal intervals of $\mathbb{R}$ such that that $p(z)$ is strictly monotonous for $z \in A_i$. We denote

$$A_i = (a_i, a_{i+1}], \quad \text{for } i = 0, 1, \dots, n,$$

where $a_0 = -\infty$ and $a_{n+1} = +\infty$ so that $A_n = (a_n, \infty)$. We assume that the minimum possible partition is chosen. If $n$ is allowed to be infinite, this can always be arranged for any continuous function $p$. While our proof technique applies also when $n$ is infinite, we focus attention to the case when $n$ is finite, which means that $p$ is non-monotonic only in a finite region.

We will make the following assumption in the rest of the paper.

**Assumption 1.** *We assume that the cdf $F$ of $Z$ and $p$ are continuous, $n$ is finite and that $p$ is not constant on any open interval.*

In the traditional VM method this assumption is fulfilled, since the $p$'s are polynomials and $Z$ is multivariate Gaussian. The techniques of this paper can be extended to cases where $p$ has jump discontinuities and is constant on intervals. We will not deal with these cases in the present paper in order to reduce the technical level of our presentation, as well as limiting our treatment to cases similar to the traditional VM method.

Partition the indices $\{0, 1, \dots, n\}$ into two subsets $\mathcal{I}$ and $\mathcal{D}$ such that $i$ belongs to $\mathcal{I}$ ($\mathcal{D}$) if $p(z)$ increases (decreases) on $A_i$. For each $A_i$ let $p_i(z)$ denote the function $p(z)$ restricted to $A_i$. As each $p_i$ is monotonous, it is injective on $A_i$ and has an inverse function we will denote by $p_i^{-1}$. Let us extend the domain of $p_i^{-1}$ to all of $\mathbb{R}$ through the re-definition

$$p_i^{-1}(x) = \begin{cases} p_i^{-1}(x), & x \in p(A_i) \\ a_{i+1}, & x > p(a_{i+1}) \text{ and } i \in \mathcal{I} \\ a_i, & x > p(a_i) \text{ and } i \in \mathcal{D} \\ a_i, & x < p(a_i) \text{ and } i \in \mathcal{I} \\ a_{i+1}, & x < p(a_{i+1}) \text{ and } i \in \mathcal{D} \end{cases}$$

where $p(A_i) = \{p(z) : z \in A_i\}$ is the image-set of $A_i$. An illustration of these definitions is given in Figure 1. Also, let us introduce $\delta : \{0, \dots, n\} \mapsto \{0, 1\}$ by

$$\delta(i) = I\{i \in \mathcal{D}\},$$

so that $\delta(i)$ is the indicator function for the event that $p(z)$ is decreasing in $A_i$.

**Proposition 1.** *Under Assumption 1, the cumulative distribution function of $X$ is*

$$H(x) = \delta(n) + \sum_{i=0}^{n} (-1)^{\delta(i)} F(p_i^{-1}(x)).$$

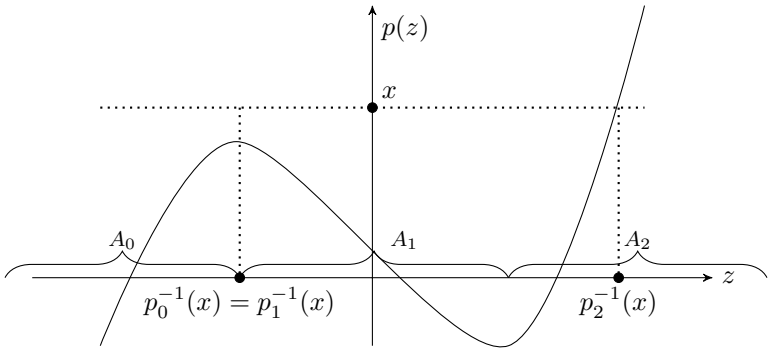


FIGURE 1. Illustration for the description of $p_i^{-1}(x)$

*If $Z$ has a density $f(x)$ with respect to Lebesgue measure, the density of $X$ with respect to Lebesgue measure is*

$$h(x) = I\{x \in Im(p)\} \sum_{i=0}^{n} (-1)^{\delta(i)} \frac{f\left(p_i^{-1}(x)\right)}{p'(p_i^{-1}(x))} I\{x \in A_i\}.$$

*where $Im(p) := \{p(z) : z \in \mathbb{R}\}$ is the image of $p$. Note that this density is only defined on $x \in \mathbb{R} \setminus \{a_1, a_2, \ldots, a_n\}$, whose complement has Lebesgue-measure zero.*

*Proof.* Let us first note that $P(X \leq x)$ is either zero or one if $x \notin \{p(z) : z \in \mathbb{R}\}$. Assume that this is not the case. We have

$$\begin{aligned} P(X \leq x) = P(p(Z) \leq x) &= \sum_{k=0}^{n} P\left(\{p(Z) \leq x\} \cap \{Z \in A_k\}\right) \\ &= \sum_{i \in \mathcal{I}} P\left(\{Z \leq p_i^{-1}(x)\} \cap \{Z \in A_i\}\right) + \sum_{j \in \mathcal{D}} P\left(\{Z \geq p_j^{-1}(x)\} \cap \{Z \in A_j\}\right) \\ &= \sum_{i \in \mathcal{I}} P\left(a_i \leq Z \leq p_i^{-1}(x)\right) + \sum_{j \in \mathcal{D}} P\left(p_j^{-1}(x) \leq Z \leq a_{j+1}\right), \end{aligned}$$

where the last line uses the extended definition of $p_i^{-1}$. We hence have that

$$H(x) = \sum_{i \in \mathcal{I}} \left[F\left(p_i^{-1}(x)\right) - F(a_i)\right] + \sum_{j \in \mathcal{D}} \left[F(a_{j+1}) - F\left(p_j^{-1}(x)\right)\right].$$

Note that we have hitherto not used the assumption that $n$ is finite. If $n = \infty$, the above display is clearly still valid. However, this case will not interest us in the remainder of our investigation.

Because the function must alternate between increasing and decreasing, for each $i$ except $i = 0$ and $i = n + 1$, the first sum in the above display includes $-F(a_i)$ while the second sum includes $F(a_i)$. Hence, similar to telescoping sums, these terms cancel out. Note that if $0 \in \mathcal{D}$, then $a_0$ is not in any of the sums in the above display. Similarly, if $n \in \mathcal{I}$, then $a_{n+1}$ is not included in the above expression.

Hence,

$$H(x) = (1 - \delta(0))\, F(a_0) + \delta(n) F(a_{n+1}) + \sum_{i=0}^{n} (-1)^{\delta(i)} F(p_i^{-1}(x)).$$

As we always have $a_0 = -\infty$ and $a_{n+1} = \infty$, we can conclude that

$$H(x) = (1 - \delta(0))\, F(-\infty) + \delta(n) F(\infty) + \sum_{i=0}^{n} (-1)^{\delta(i)} F(p_i^{-1}(x)) \tag{3}$$

$$= \delta(n) + \sum_{i=0}^{n} (-1)^{\delta(i)} F(p_i^{-1}(x)). \tag{4}$$

The density is found through identifying $H'(x)$, and equals

$$\begin{aligned} h(x) &= \sum_{i=0}^{n} (-1)^{\delta(i)} I\{x \in A_i\} f\left(p_i^{-1}(x)\right) \frac{\mathrm{d}p_i^{-1}(x)}{\mathrm{d}x} \\ &= \sum_{i=0}^{n} (-1)^{\delta(i)} I\{x \in A_j\} \frac{f\left(p_j^{-1}(x)\right)}{p'(p_i^{-1}(x))}, \end{aligned}$$

when $x \in Im(p)$. □

3.2. **The multivariate distribution.** We will now provide an expansion of the cdf and copula of $X$ in terms of the cdf and copula of $Z$. While the copula of $X$ is in general not identical to that of $Z$, the copula is a particular type of linear combination of the copula of $Z$. An immediate conclusion is that the copula of $X$ is fairly restrictively connected to the copula of $Z$.

In order to understand the copula of $X$, we need to extend the notation of Section 3.1 to take care of each coordinate $p_k$ in the mapping

$$X = p(Z) = (p_1(Z_1), \ldots, p_d(Z_d))^t.$$

For each $1 \leq k \leq d$, let $\{A_i^k\}_{i=0,\ldots,n_k}$ be the maximal intervals of $\mathbb{R}$ such that $p_k(z_k)$ is strictly monotonous for $z_k \in A_i^k$. We denote

$$A_i^k = (a_i^k, a_{i+1}^k], \quad \text{for } i = 0, 1, \ldots, n_k,$$

where $a_0^k = -\infty$ and $a_{n_k+1}^k = +\infty$ so that $A_{n_k}^k = (a_{n_k}^k, \infty)$. For each $A_i^k$ let $p_{k,i}(z)$ denote the function $p_k(z)$ restricted to $A_i^k$. As each $p_{k,i}$ is monotonous, it is injective on $A_i^k$ and has an inverse function denoted by $p_{k,i}^{-1}$. Furthermore, the definitions of $\mathcal{I}$ and $\mathcal{D}$ are naturally extended for each $k$ to subsets of $\{0, \ldots, n_k\}$ denoted by $\mathcal{I}_k$ and $\mathcal{D}_k$, respectively. The function $\delta(i)$ is also extended to $\delta_k : \{0, \ldots, n_k\} \mapsto \{0, 1\}$, so that $\delta_k(i)$ is the indicator function for the event that $p_k(z_k)$ is decreasing on $A_i^k$.

Finally, let us extend the domain of $p_{k,i}^{-1}$ to all of $\mathbb{R}$ through

$$p_{k,i}^{-1}(x) = \begin{cases} p_{k,i}^{-1}(x), & x \in p_k(A_i^k) \\ a_{i+1}^k, & x > p(a_{i+1}^k) \text{ and } i \in \mathcal{I}_k \\ a_i^k, & x > p(a_i^k) \text{ and } i \in \mathcal{D}_k \\ a_i^k, & x < p(a_i^k) \text{ and } i \in \mathcal{I}_k \\ a_{i+1}^k, & x < p(a_{i+1}^k) \text{ and } i \in \mathcal{D}_k. \end{cases}$$

We now extend Proposition 1 to the multivariate case. Note that, on the basis of this result, one can clearly derive the density of $X$. However, we will not introduce the necessary notation to calculate this density, and will be content to present the cumulative distribution function

$$H(x) = H(x_1, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d)$$

and the copula

$$C(u) = C_X(u) = C(u_1, \ldots, u_d) = P(H_1(X_1) \leq u_1, \ldots, H_1(X_1) \leq u_1)$$

of $X = p(Z)$. We start with the cumulative distribution function.w

**Theorem 1.** *If Assumption 1 is valid for each marginal, we have that*

$$H(x) = \sum_{J \subseteq \{1,2,\ldots,d\}} \left( \prod_{k \in J^C} \delta_k(n_k) \right) \sum_{\substack{0 \leq i_j \leq n_j \\ j \in J}} (-1)^{\sum_{l \in J} \delta_l(i_l)} P\left( \bigcap_{l \in J} \{ Z_l \leq p_{l,i_l}^{-1}(x_{i_l}) \} \right),$$

*where the outer sum is over all unique subsets of $\{1, 2, \ldots, d\}$, including the empty set, $J^C$ denotes the complement of $J$ with respect to $\{1, 2, \ldots, d\}$ and the product operator $\prod$ equals 1 when indexed by the empty set. We here also use the convention that the sum over the empty set equals one.*

Before we give the proof, we would like to comment upon how to read sums of the form

$$\sum_{\substack{0 \leq i_j \leq n_j \\ j \in J}} (-1)^{\sum_{l \in J} \delta_l(i_l)} P\left( \bigcap_{l \in J} \{ Z_l \leq p_{l,i_l}^{-1}(x_{i_l}) \} \right).$$

Suppose $J = \{1, 4, 6\}$. Then this sum equals

$$\begin{aligned} \sum_{0 \leq i_1 \leq n_1} \sum_{0 \leq i_4 \leq n_4} \sum_{0 \leq i_6 \leq n_6} & (-1)^{\sum_{l \in \{1,4,6\}} \delta_l(i_l)} P\left( \bigcap_{l \in \{1,4,6\}} \{ Z_l \leq p_{l,i_l}^{-1}(x_{i_l}) \} \right) \\ &= \sum_{0 \leq i_1 \leq n_1} \sum_{0 \leq i_4 \leq n_4} \sum_{0 \leq i_6 \leq n_6} (-1)^{\delta_1(i_1) + \delta_4(i_4) + \delta_6(i_6)} \\ &\qquad P\left( Z_1 \leq p_{1,i_1}^{-1}(x_{i_1}), Z_4 \leq p_{4,i_4}^{-1}(x_{i_4}), Z_6 \leq p_{6,i_6}^{-1}(x_{i_6}), \right). \end{aligned}$$

That is, the sum in the statement of the result expands to a total of $2^d \prod_{k=1}^d n_k$ terms, many of which will be zero.

*Proof.* Let us denote $L_i = \{p_i(Z_i) \leq x_i\}$, so that

$$H(x_1, x_2, \ldots, x_d) = P\left(\bigcap_{i=1}^{d} L_i\right) = P\left(\left[\bigcap_{i=1}^{d-1} L_i\right] \bigcap L_d\right).$$

All the steps in the proof of Proposition 1 can be applied to the probability of $L_d$ – when one only considers the events where $\cap_{i=1}^{d-1} L_i$ occurs – except the step going from eq. (3) to eq. (4). We have

$$H(x) = \delta_d(n_d) P\left(\bigcap_{i=1}^{d-1} L_i\right) + \sum_{i_d=0}^{n_d} (-1)^{\delta_d(i_d)} P\left(\bigcap_{i=1}^{d-1} L_i \cap \{Z_d < p_{d,i_d}^{-1}(x_d)\right).$$

The same expansion, restricted to events where $\{Z_d < p_{d,i_d}^{-1}(x_d)\}$ occurs, results in

$$\begin{aligned}
H(x) =& \delta_d(n_d)\delta_{d-1}(n_{d-1}) P\left(\bigcap_{i=1}^{d-2} L_i\right) \\
&+ \delta_d(n_d) \sum_{i_{d-1}=0}^{n_{d-1}} (-1)^{\delta_{d-1}(i_{d-1})} P\left(\bigcap_{i=1}^{d-2} L_i \cap \{Z_{d-1} < p_{d-1,i_{d-1}}^{-1}(x_{d-1})\right) \\
&+ \delta_{d-1}(n_{d-1}) \sum_{i_d=0}^{n_d} (-1)^{\delta_d(i_d)} P\left(\bigcap_{i=1}^{d-2} L_i \cap \{Z_d < p_{d,i_d}^{-1}(x_d)\right) \\
&+ \sum_{i_{d-1}=0}^{n_{d-1}} (-1)^{\delta_{d-1}(i_{d-1})} \sum_{i_d=0}^{n_d} (-1)^{\delta_d(i_d)} P\Bigg(\bigcap_{i=1}^{d-2} L_i \cap \{Z_{d-1} < p_{d-1,i_{d-1}}^{-1}(x_{d-1})\} \\
&\cap \{Z_d < p_{d,i_d}^{-1}(x_d)\}\Bigg).
\end{aligned}$$

This procedure can be iterated until we get the stated formula. To see this, first notice that the above expansion includes the four possible subsets of $\{d-1, d\}$. Then notice that the numbering of the marginals is irrelevant for the iteration argument: For example,

$$P\left(\bigcap_{i=1}^{d-2} L_i \cap \{Z_{d-1} < p_{d-1,i_{d-1}}^{-1}(x_{d-1})\right)$$

has the same structure as the original problem, but with $d-1$ marginals. □

An immediate consequence of Theorem 1 is that the copula of $X$ is

$$\begin{aligned}
C(u) = \sum_{J \subseteq \{1,2,\ldots,d\}} & \left(\prod_{k \in J^C} \delta_k(n_k)\right) \\
& \sum_{\substack{0 \leq i_j \leq n_j \\ j \in J}} (-1)^{\sum_{l \in J} \delta_l(i_l)} P\left(\bigcap_{l \in J} \{U_l \leq F_l(p_{l,i_l}^{-1}(H_l^{-1}(u_{i_l})))\}\right).
\end{aligned} \tag{5}$$

We note that when all $p_k$ are strictly monotonously increasing, i.e. when $\delta_k(n_k) = \delta_k(0) = 0$ for all $k$, the copula of $X$ is equal to that of $Z$, as is well known.

Since $F_l$, $p_{l,i_l}^{-1}$ and $H_l^{-1}$ are all monotonic functions, the above display shows the copula of $X$ is a linear combination of monotonic transformations of marginal copulas of $Z$. Note that both $F_l$ and $H_l^{-1}$ are strictly increasing, while $p_{l,i_l}^{-1}$ is strictly monotonous in $p_l(A_{l,i_l})$ and constant elsewhere. While the $p$-transformation will change the copula of $X$, it will only do so in a limited sense.

As a consequence of eq. (5), we note that several quantities of interest, such as $\sigma_{ijkl}$, the central constituent appearing in eq. (2) for $\Gamma$, and many measures of multivariate dependence of the random vector $X$, are of the form

$$\tau_h := \int_{[0,1]^d} h(u)\,\mathrm{d}C_X(u)$$

where $C_X$ is the copula of $X$. For example, both Kendall's tau and Spearman's rho can be written in this form (see Section 2.1.9 of Joe, 1997). The linearity of the Lebesgue-Stieltjes integral in the integrand means that eq. (5) provides a way to calculate such integrals through finite sums.

The use of eq. (5) in computations is complicated by its inclusion of $H_k^{-1}$. As $H_k$ is a sum of functions, there is no generally available expression for $H_k^{-1}$ except in special regions. In these regions, where all $p_k$ have inverses, we will now identify the copula of $X$. Let us define $\mathcal{M}$ to be the subset of $\mathbb{R}^d$ with elements $x = (x_1, \ldots, x_d)$ such that for all $k = 1, \ldots, d$ the equation $p_k(z_k) = x_k$ has an unique solution in some neighbourhood of $x_k$. Also define

$$\mathcal{M}^u = \{(H_1(x_1)\ldots, H_d(x_d)) : (x_1, \ldots, x_d) \in \mathcal{M}\} \subseteq [0,1]^d.$$

Define $\Delta_k(x_k)$ to be equal to 0 if $p_k$ is increasing at $z_k$ where $z_k$ is the unique root of $p_k(z_k) = x_k$, and 1 otherwise. Similarly, define $\Delta_k^u(u_k) = \Delta_k(H_k^{-1}(u_k)))$.

**Proposition 2.** *Assume that Assumption 1 holds for all $d$ marginals. Then we have the following property of $C_X$, the copula of $X = p(Z)$, described in terms of the copula of $Z$, i.e. the distribution of $U = (U_1, \ldots, U_d) = (F_1(Z_1), \ldots, F_d(Z_d))$. For $u \in \mathcal{M}^u$, we have that*

$$C_X(u_1, u_2, \ldots, u_d) = P\left(\bigcap_{k=1}^{d} \{\Delta_k^u(u_k) + (-1)^{\Delta_k^u(u_k)} U_k \leq u_k\}\right).$$

*If, in addition each $p_k$ has image $Im(p_k) = \mathbb{R}$, then the complement of $\mathcal{M}$ has finite Lebesgue-measure.*

*Proof.* Suppose that $x \in \mathcal{M}$. Then for any $k$

$$H_k(x_k) = P(p_k(Z_k) \leq x_k) = P((-1)^{\Delta_k(x_k)} Z_k \leq (-1)^{\Delta_k(x_k)} p_k^{-1}(x_k)).$$

Let $u = (u_1, \dots, u_d) \in \mathcal{M}^u$. Then for any $k \in \{1, \dots, d\}$ we can write the inverse function as

$$\begin{aligned} H_k^{-1}(u_k) &= \begin{cases} p_k(F_k^{-1}(u_k)), & \text{for } \Delta_k^u(u_k) = 0 \\ p_k(F_k^{-1}(1-u_k)), & \text{for } \Delta_k^u(u_k) = 1 \end{cases} \\ &= p_k(F_k^{-1}(\Delta_k^u(u_k) + (-1)^{\Delta_k^u(u_k)} u_k)). \end{aligned}$$

Let $K_k(u_k) := \{H_k(X_k) \leq u_k\}$ and recall that

$$K_k(u_k) = \{X_k \leq H_k^{-1}(u_k)\} = \{p_k(Z_k) \leq H_k^{-1}(u_k)\}.$$

Since $u \in \mathcal{M}^u$ we have

$$\begin{aligned} K_k(u_k) &= \{p_k(Z_k) \leq p_k(F_k^{-1}(\Delta_k^u(u_k) + (-1)^{\Delta_k^u(u_k)} u_k))\} \\ &= \{(-1)^{\Delta_k^u(u_k)} Z_k \leq (-1)^{\Delta_k^u(u_k)} F_k^{-1}(\Delta_k^u(u_k) + (-1)^{\Delta_k^u(u_k)} u_k)\} \\ &= \{\Delta_k^u(u_k) + (-1)^{\Delta_k^u(u_k)} U_k \leq u_k\}, \end{aligned}$$

where $U_k = F_k(Z_k)$. The result follows, as $C_X(u) = P(\cap_{k=1}^d K_k(u_k))$. In the case that each $Im(p_k) = \mathbb{R}$, the $p_k$ are unbounded as $x_k \to \pm\infty$. Hence $\mathcal{M}^C$ is a bounded region and has finite measure.

□

Given a $d$-dimensional variable $U = (U_1, \dots, U_d)$ distributed according to a copula $C$, there are $2^d - 1$ so-called associated copulas (Section 1.6 Joe, 1997). These are the distributions of random vectors of the form

$$(q_1 + (-1)^{q_1} U_1, \dots, q_d + (-1)^{q_d} U_d), \qquad \text{where } q_1, \dots, q_d \in \{0, 1\}.$$

Proposition 2 says that the copula of $X$ evaluated in $\mathcal{M}^u$ is equal to an associated copula of $Z$. Hence, in $\mathcal{M}^u$, the generalized VM transform can only change the copula of $Z$ in the same way as multiplying the elements of $Z$ by either 1 or $-1$.

While we could in theory re-phrase Proposition 2 in terms of the CDFs of these associated copulas, this is only practical in lower-dimensional cases. For simplicity, we will only give this more direct expression for $d = 2$ and when $u \in \mathcal{M}^u$.

**Corollary 1.** *Assume that $d = 2$ and that Assumption 1 holds for both marginals. Then, for $u \in \mathcal{M}^u$, we have that*

$$C_X(u_1, u_2) = \begin{cases} C_Z(u_1, u_2), & \text{for } \Delta_1^u(u_1) = 0, \Delta_2^u(u_2) = 0 \\ u_1 + u_2 - 1 + C_Z(1-u_1, 1-u_2), & \text{for } \Delta_1^u(u_1) = 1, \Delta_2^u(u_2) = 1 \\ u_1 - C_Z(u_1, 1-u_2), & \text{for } \Delta_1^u(u_1) = 0, \Delta_2^u(u_2) = 1 \\ u_2 - C_Z(1-u_1, u_2), & \text{for } \Delta_1^u(u_1) = 1, \Delta_2^u(u_2) = 0 \end{cases}.$$

*Proof.* As mentioned in Section 1.6 in Joe (1997), for $(U_1, U_2) \sim C_Z$, the three associated copulas of $Z$ are the distributions of

$$U_{(1)} = (1 - U_1, 1 - U_2), \quad U_{(2)} = (U_1, 1 - U_2), \quad U_{(3)} = (1 - U_1, U_2),$$

and are given by

$$C_{(1)}(u_1, u_2) = u_1 + u_2 - 1 + C_Z(1 - u_1, 1 - u_2)$$
$$C_{(2)}(u_1, u_2) = u_1 - C_Z(u_1, 1 - u_2)$$
$$C_{(3)}(u_1, u_2) = u_2 - C_Z(1 - u_1, u_2)$$

respectively. The result then follows directly from Proposition 2.

□

3.3. **Limitations of the classical Vale-Maurelli approach.** The classical VM method uses the Gaussian distribution for $Z$, and third-degree polynomials for $p_k$. As long as the third-order coefficient of $p_k$ is non-zero, the image set of $p_k$ is equal to $\mathbb{R}$. This means that points sufficiently close to the edge set of $[0, 1]^d$ are contained in $\mathcal{M}^u$. It follows immediately from Corollary 1 that the copula of the classical VM transform is equal to an associated copula of the Gaussian copula in the tail region. Such an associated copula equals the copula of a random vector obtained by multiplying some elements of $Z$ by $-1$. Clearly, such a random vector is itself Gaussian. As no Gaussian copula has tail dependence this immediately implies that the classical VM transform $X$ has zero tail dependence. Note that this argument is also valid for odd-degree polynomials of higher order than three.

That the Gaussian copula has no tail dependence is a striking feature, since many popular copulas exhibit tail dependence. Most approximations in statistics are based on the central limit theorem, whose convergence speed decreases as we move further away from the Gaussian case. This means that the classical VM transform $X$ inherits a Gaussian-like property, indicating that simulation studies based on the VM approach might give overly optimistic impressions of finite-sample properties of estimators with non-Gaussian data.

## 4. A Monte Carlo illustration

Through our general analytical results in the previous sections, the limitation implicit in the VM approach should be clear. Our simulation study therefore focuses on a simple and clear-cut case in contrast to performing a large simulation study for many SEM setups. We illustrate how the finite sample behaviour of the popular normal theory based maximum likelihood (ML) estimator is affected by the underlying copula in a simple confirmatory factor model. The model, given in Figure 2, has two factors $\xi_1$ and $\xi_2$, each with two indicators. In our model formulation there is only one free parameter to estimate, namely $\phi = \mathrm{Cov}(\xi_1, \xi_2)$. The other parameters are constants: the variances of $\xi_1$ and $\xi_2$ are fixed to one, the four factor loadings are fixed to one, and the residual variances $\mathrm{Var}(\delta_i)$, for $i = 1, \ldots, 4$, are fixed to 0.5.

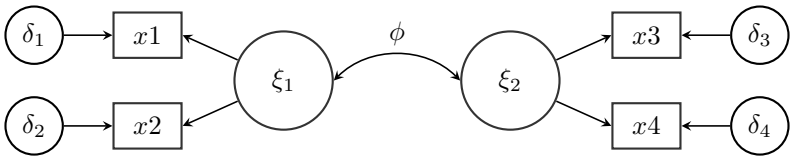


FIGURE 2. A two-factor model.

4.1. **Data generation.** Simulation of data proceeds according to these fixed parameters, so that the model is correctly specified. We generate independent measurement errors $\delta_i$ from the Gaussian distribution $N(0, 0.5)$ while the vector $(\xi_1, \xi_2)$ is simulated from distributions with varying copulas. These copulas generate various degrees of dependence between $\xi_1$ and $\xi_2$. However, as next described, the data-generating process ensures that $\mathrm{Cov}\,(\xi_1, \xi_2) = 0$, i.e. there is no correlation between $\xi_1$ and $\xi_2$.

Let $Q_0 = 2Z^3 - 3Z^2$ where $Z \sim N(0, 1)$, which is the result of the non-monotonous third degree polynomial transformation $p(z) = 2z^3 - 3z^2$. The skewness and kurtosis of $Q_0$ are, respectively, $\beta_1 = -5$ and $\beta_2 = 54.8$. The function $p(z)$ is increasing for $z \leq 0$, decreasing for $0 < z \leq 1$ and increasing for $z > 1$.For $x > p(0) = 0$ and $x < p(1) = 2 - 3 = -1$, there is only one real root. In the region $-1 \leq x \leq 0$, we have three real roots.

The CDF of $Q_0$ is, according to Proposition 1,

$$H(x) = \Phi\left(p_0^{-1}(x)\right) - \Phi\left(p_1^{-1}(x)\right) + \Phi\left(p_2^{-1}(x)\right),$$

where $\Phi$ is the cdf for the standard normal distribution. In order to describe $p_i^{-1}$, we need $A_0 = (-\infty, 0], \quad A_1 = (0, 1], \quad A_2 = (1, \infty)$, and the function

$$W(x) = \sqrt[3]{2\sqrt{x^2 + x} + 2x + 1}.$$

Recall the extended definition of $p_i^{-1}$. We have that for $x_i \in p(A_i) = \{p(z) : z \in A_i\}$,

$$\begin{aligned}
p_0^{-1}(x_0) &= -\frac{1 - i\sqrt{3}}{4} W(x_0) - \frac{1 + i\sqrt{3}}{4W(x_0)} + \frac{1}{2} \\
p_1^{-1}(x_1) &= -\frac{1 + i\sqrt{3}}{4} W(x_1) - \frac{1 - i\sqrt{3}}{4W(x_1)} + \frac{1}{2} \\
p_2^{-1}(x_2) &= \frac{1}{2} W(x_2) + \frac{1}{2W(x_2)} + \frac{1}{2}.
\end{aligned}$$

Now, let $(U_1, U_2)$ have the so-called Joe-copula with dependence parameter $\theta$ as its joint distribution. The CDF is given by

$$C_\theta(u_1, u_2) = 1 - [(1 - u_1)^\theta + (1 - u_2)^\theta - (1 - u_1)^\theta (1 - u_2)^\theta]^{1/\theta}$$

for a $\theta \geq 1$. Then

$$Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \begin{pmatrix} H^{-1}(U_1) \\ H^{-1}(U_2) \end{pmatrix}$$

has the exact same marginal distributions as $Q_0$ for both marginals, and $Q$ has the same copula as $(U_1, U_2)$ because $H^{-1}$ is a strictly monotonous transformation. In order to generate non-correlated $\xi_1$ and $\xi_2$ based on $Q$, we need to calculate the covariance of $Q_1$ and $Q_2$. Because the marginals of $Q$ are identical, we thus need to calculate the moments $\mathbb{E}Q_1, \mathbb{E}Q_1^2$ and $\mathbb{E}Q_1Q_2$. Using standard properties of the Normal distribution, we see that $\mathbb{E}Q_1 = -3$ and $\mathbb{E}Q_1^2 = 87$. As a function of $\theta$, the expectation $\mathbb{E}Q_1Q_2$ equals

$$\int_{[0,1]^d} H^{-1}(u)H^{-1}(v)c(u, v; \theta)\,\mathrm{d}u\,\mathrm{d}v \tag{6}$$

expressed through the copula density

$$c(u, v; \theta) = (\bar{u}^\theta + \bar{v}^\theta - \bar{u}^\theta\bar{v}^\delta)^{-2+1/\theta}\,\bar{u}^{\theta-1}\bar{v}^{\theta-1}[\theta - 1 + \bar{u}^\theta + \bar{v}^\theta - \bar{u}^\theta\bar{v}^\theta],$$

where $\bar{u} = 1-u, \bar{v} = 1-v$. Note that because $\theta = 1$ corresponds to the independence copula, we know that in this case $\mathbb{E}Q_1Q_2 = (\mathbb{E}Q_1)(\mathbb{E}Q_2) = 9$, from normal theory. For the other values of $\theta$ employed in this study, $\mathbb{E}Q_1Q_2$ and similar higher moments were calculated to a high level of precision by numerical integration of integrals of type (6). Finally, to ensure that the covariance $\phi$ between $\xi_1$ and $\xi_2$ is zero, we define

$$\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \mathrm{Cov}\,(Q_1, Q_2)^{-\frac{1}{2}} \cdot \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}.$$

For the special case $\theta = 1$ we have that $U_1$ and $U_2$ are independent, which implies that $Q_1$ and $Q_2$ are independent. It follows that $\mathrm{Cov}\,(Q_1, Q_2)$, and hence also $\mathrm{Cov}\,(Q_1, Q_2)^{-\frac{1}{2}}$, are diagonal matrices. Therefore, for $\theta = 1$, the multiplication with $\mathrm{Cov}\,(Q_1, Q_2)^{-\frac{1}{2}}$ does not introduce any cross terms. This implies that for $\theta = 1$, $\xi_1$ and $\xi_2$ are produced by a classical VM transform, i.e. from an underlying Gaussian copula. When $\theta > 1$, however, the distribution of $(\xi_1, \xi_2)$ is a linear combination of a generalized VM transform with a non-Gaussian underlying copula.

4.2. **Design of study.** Our design conditions are as follows. We consider three sample sizes, $n = 50, 200$ and 500. We consider four underlying Joe copulas, defined by $\theta = 1, 1.5, 5$ and 10. These copulas generate various degrees of upper tail dependence $\lambda_u$ and heterogeneous kurtosis in $(\xi_1, \xi_2)$. The Joe copula has zero lower tail dependence, while the upper tail dependence is given by Joe (1997, p. 35) as $\lambda_u = 2 - 2^{1/\theta}$. In Table 1 we tabulate $\lambda_u$ and Mardia's (Mardia, 1970) multivariate kurtosis statistic $\beta_2$ for $(Q_1, Q_2)$. The values of $\beta_2$ were calculated by numerical integration, and also apply to $(\xi_1, \xi_2)$. This is because $\beta_2$ is invariant under affine transformations, and $(\xi_1, \xi_2)$ is a linear transformation of $(Q_1, Q_2)$. This linear transformation unfortunately distorts $\lambda_u$ for $(\xi_1, \xi_2)$ in a non-trivial way, and $\lambda_u$

was only calculated for $(Q_1, Q_2)$. Note that under multivariate normality $\lambda_u = 0$ and $\beta_2 = 8$. Table 1 clearly indicates that as $\theta$ increases, the dependence structure of the resulting distribution becomes increasingly non-Gaussian.

| $\theta$ | 1 | 1.5 | 5 | 10 |
|---|---|---|---|---|
| $\lambda_u$ | 0 | 0.41 | 0.85 | 0.93 |
| $\beta_2$ | 111.5 | 115.4 | 132.5 | 156.4 |

TABLE 1. Upper tail dependence and heterogeneous kurtosis for $(Q_1, Q_2)$

In each of the resulting 12 design cells we run 2000 replications. Simulation and estimation was obtained through the use of the R (R Core Team, 2013) packages copula (Hofert et al., 2013) and lavaan (Rosseel, 2012), respectively.

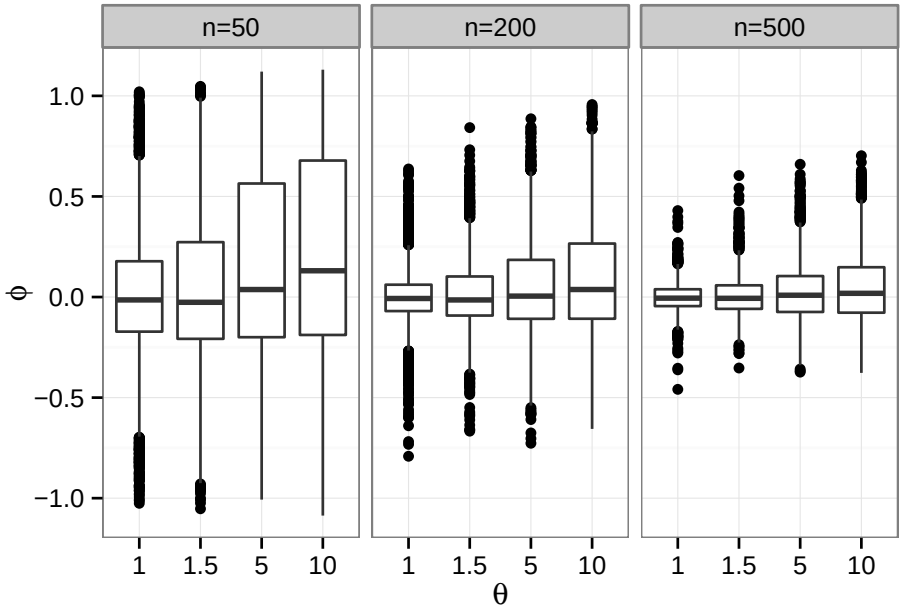


FIGURE 3. The estimation of $\phi$ for varying dependence parameters $\theta$ and sample sizes $n$.

4.3. **Results and remarks.** The results are visualised in Figure 3 as boxplots. Clearly, the true population parameter value $\phi = 0$ is best estimated when $\theta = 1$, across all sample sizes. That is, the ML estimator performs best under the classical VM transform. With increasing levels of $\theta$, both bias and standard errors become larger. This shows that when the underlying copula introduces more and more dependence between $\xi_1$ and $\xi_2$, which nevertheless remain uncorrelated, the ML

estimation of $\phi$ in finite samples becomes more biased and less efficient. Predictably, as the ML estimator is consistent, parameter bias decreases with increasing $n$. Note however that the standard error, even for large sample sizes, is clearly affected by the underlying copula, in accordance with the asymptotic equation (1). Even for the case $\theta = 1.5$, which in practice is barely distinguishable from the VM case using standard tools such as scatter plots, there is an increased variability in the estimation. In fact, the empirical standard error increased by about 20 % for sample sizes $n = 200$ and $n = 500$ when moving from $\theta = 1$ to $\theta = 1.5$. We conclude that the underlying copula of $(\xi_1, \xi_2)$ has a marked effect on the quality of the ML estimator.

As a final note, one might wonder whether the deteriorating performance of the ML estimator with increasing $\theta$ is solely explained by the increase in multivariate kurtosis $\beta_2$ that accompanies increasing values of $\theta$. To investigate this issue we compared ML estimation in the following two distributions. The first distribution is based on uncorrelated $\xi_1$ and $\xi_2$ generated by a VM transform such that skewness and kurtosis in $\xi_i$ equals $-5$ and 77.2, respectively. The resulting bivariate kurtosis parameter is then equal to (Mardia, 1970, eq. 3.9) $\beta_2 = 2(1 + 77.2) = 156.4$. This value of $\beta_2$ is equal to the distribution of $(\xi_1, \xi_2)$ obtained from the Joe copula with $\theta = 10$, see Table 1. Hence we consider two distributions, one obtained from Vale-Maurelli generation of $(\xi_1, \xi_2)$, and the other obtained from using the Joe copula with $\theta = 10$ to generate $(Q_1, Q_2)$ and then multiplying by $\mathrm{Cov}\,(Q_1, Q_2)^{-\frac{1}{2}}$. So although the marginal kurtosis in the VM distribution is higher than in the Joe distribution, the bivariate kurtosis $\beta_2$ equals 156.4 in both distributions. Monte Carlo simulation from these two distributions, with 2000 replications in each case, and with sample sizes $n = 50, 200, 500$ yields the results presented in Table 2. Clearly, the ML estimator is much better behaved under data generation with VM than with Joe. In accord with the information in Figure 3, and even when we control for bivariate kurtosis, the bias and standard errors become inflated under a copula that departs from the Gaussian-like copula obtained from the VM transform.

| | $n = 50$ | | $n = 200$ | | $n = 500$ | |
|---|---|---|---|---|---|---|
| | $\widehat{\phi}$ | S.E. | $\widehat{\phi}$ | S.E. | $\widehat{\phi}$ | S.E. |
| VM | 0.001 | 0.443 | 0.001 | 0.185 | -0.003 | 0.088 |
| Joe | 0.181 | 0.525 | 0.090 | 0.282 | 0.042 | 0.171 |

TABLE 2. Mean and standard error (S.E.) of $\widehat{\phi}$ under a VM and a Joe distribution with equal $\beta_2$.

## 5. Conclusion

Simulation and distributional assumptions are in some sense two sides of the same coin. If one has a distribution that one cannot simulate from, it is difficult to claim that one has fully understood it. And similarly, a simulation method that one does not know the precise distribution of cannot be said to be fully understood. The main mathematical contribution of this paper is to derive the distribution of a generalized version of the Vale-Maurelli technique.

This mathematical result has several important practical consequences. In moment-based models, such as SEM, simulation studies are important in assessing the finite sample behavior of estimators. To get a realistic picture of this behaviour, it is important to be able to generate data that deviates from Normality in a variety of ways. In this paper, we have worked with a generalized version of the VM approach of generating non-Normal observations with a given covariance structure. The method enables the researcher to specify marginal skewness and excess kurtosis, which are two of the univariate features most in contrast with Normality. Inspection of marginal distributions will therefore clearly give the impression that the resulting data-set is far from Normal. However, marginal distributions are only a small portion of the full distribution, and one of the main conclusions in our paper is that the truly multivariate aspects of data generation using the VM approach is exactly equal to the Normal model, except in regions where the Fleishman polynomials are non-monotonous. Hence, evaluating the robustness of Gaussian ML estimation with the VM method may give a false impression that this popular estimator is more well-behaved under even severely non-Normal data than it really is. Gaussian ML may seem to outperform other estimators that are not based on the Normality assumption, such as unweighted least squares or Brownes asymptotically distribution-free estimator, due to the use of a simulation method that only appears to generate data far from the Normal case. In conclusion, this accentuates the need for more general and flexible simulation methods that incorporates the copula perspective to better evaluate the behaviour of SEM estimators.

## References

Bentler, P. (2006). Eqs 6 structural equations program manual.

Bentler, P. M. & Tong, X. (2013). Evaluation of a New Mean Scaled and Moment Adjusted Test Statistic for SEM . *Structural Equation Modeling: A Multidisciplinary Journal* **20**, 148–156.

Braeken, J., Tuerlinckx, F. & Boeck, P. (2007). Copula Functions for Residual Dependency. *Psychometrika* **72**, 393–411.

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

Cario, M. C. & Nelson, B. L. (1997). Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Tech. rep., Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

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

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

Fouladi, R. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Structural Equation Modeling: A Multidisciplinary Journal* **7**, 356–410.

Headrick, T. (2007). *Real data analysis*, chap. Methods for simulating real world data for the psycho-educational sciences. Information Age Publishing, pp. 79–97.

Headrick, T. C. & Pant, M. D. (2012). Simulating non-normal distributions with specified L-moments and L-correlations. *Statistica Neerlandica* **66**, 422–441.

Hofert, M., Kojadinovic, I., Maechler, M. & Yan, J. (2013). *copula: Multivariate Dependence with Copulas*. R package version 0.999-7.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*, vol. 73. Chapman & Hall/CRC.

Jöreskog, K. & Sorbom, D. (2006). Lisrel version 8.8. lincolnwood, il: Scientific software international.

Kotz, S., Balakrishnan, N. & Johnson, N. L. (2000). *Continuous multivariate distributions*, vol. 1. John Wiley & Sons.

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

Mardia, K. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519–530.

Muthén, L. & Muthén, B. (2010). Mplus software (version 6.1).

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rosseel, Y. (2012). lavaan: An r package for structural equation modeling. *Journal of Statistical Software* **48**, 1–36.

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

Tadikamalla, P. R. (1980). On simulating non-normal distributions. *Psychometrika* **45**, 273–279.

Vale, C. & Maurelli, V. (1983). Simulating nonnormal distributions. *Psychometrika* **48**.

BI Norwegian Business School