Testing Model Fit in Path Models with Dependent Errors Given Non-Normality, Non-Linearity and Hierarchical Data

2 authors, including:

Jacob C Douma
Wageningen University & Research
66 PUBLICATIONS   **2,962** CITATIONS

Testing Model Fit in Path Models with Dependent Errors Given Non-Normality, Non-Linearity and Hierarchical Data

Jacob C. Douma & Bill Shipley

Published online: 04 Oct 2022.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

Testing Model Fit in Path Models with Dependent Errors Given Non-Normality, Non-Linearity and Hierarchical Data

Jacob C. Douma[a] and Bill Shipley[b]

[a]Wageningen University; [b]Université de Sherbrooke

ABSTRACT

We provide a generic method of testing path models that include dependent errors, nonlinear functional relationships and using nonnormal, hierarchically structured data. First, we provide a decomposition of the causal model into smaller, independent sets. These sets can be modeled independently of each other with methods that respect the type of data in these sets. Second, we introduce copulas to model the dependent errors between non-normally distributed variables. Our method yields identical results as classical covariance-based path modelling when meeting its assumptions of linearity and normality, outperforms classical SEM given nonlinear functional relationships, and can easily accommodate any parametric probability function and nonlinear functional relationships.

1. Introduction

Path modeling is an important statistical technique in many domains of science allowing one to model the multivariate dependency between variables based on an *a priori* causal hypothesis (Grace, 2006; Kline, 2005; Shipley, 2016) and to test if the dependencies in observed values in a multivariate observation are consistent with the postulated causal structure. Observed variables can depend on each other in three different ways: (i) an observed variable is directly or indirectly causing another (i.e., causal relationships), (ii) two observed variables are associated because they share common observed causes, and (iii) an observed variable is associated to another but not via the first two ways. The first two dependencies are represented using single-headed arrows ($\rightarrow$) and the latter dependency is represented using bidirected arrows ($\leftrightarrow$). These bi-directed arrows represent so-called correlated errors or free covariances. In this paper, we refer to these as "dependent errors" because the dependencies are not necessarily linear and based on normally distributed variables.

One can choose from two different approaches when fitting path models, each with their pros and cons. First, one can use classical structural equations modelling (SEM) without latent variables (Bollen, 1989) to test path models including correlated errors. Here, the consistency of the data with the hypothesized causal structure is tested by comparing the model-implied covariance matrix to the observed covariance matrix of all variables in the path model via a likelihood ratio test. The model parameters are *simultaneously* optimized such that the model-implied

covariance matrix is as similar as possible the observed covariance matrix while respecting the constraints implied by the hypothesized causal structure and captured by covariance algebra (Bollen, 1989). Classical SEM assumes a multivariate normal distribution with mutually independent observations and linear relationships between the variables. Much effort has gone into relaxing these assumptions and to develop indices that are robust against violations (e.g., Browne, 1984; Kenny & Judd, 1984; Oberski, 2014; Satorra & Bentler, 1994; Wall & Amemiya, 2001). To date, a generic solution to simultaneously deal with non-linearity, non-normality and non-independence of observations is lacking.

The second class of approaches that can be used to fit path models are piecewise structural equations models. Piecewise SEM can easily accommodate non-linear relationships, non-normally distributed variables and data with hierarchical data structures. Key to this approach is that the consistency of the path model to the data is tested through a series of local tests (Shipley, 2009; Shipley & Douma, 2020, 2021). However, this approach can only be applied to path models without dependent errors, i.e., path models that can be represented by a Directed Acyclic Graph (DAGs). Pearl (2009) has shown that the multivariate probability distribution generated by the causal model (topology) of a DAG can be decomposed via the Markov decomposition into the product of a series of univariate probability distributions conditional on their parents irrespective of the functional forms of these univariate probability distributions. This allows one to partition the causal model into sets of variables that can be modeled independently of each other upon conditioning because the parameters in each set can

CONTACT Jacob C. Douma ✉ bob.douma@wur.nl 🖅 Centre for Crop Systems Analysis, Wageningen University, Wageningen, The Netherlands.

then be independently estimated from parameters in other sets. These univariate probability distributions can easily capture nonlinear relationships, non-normally distributed errors and hierarchical designs. The fit of the model to the data can be assessed in two ways. First, one can obtain the likelihood of the full model by summing of the likelihoods of the univariate distributions and compare this to the likelihood of a saturated model using a likelihood ratio test (Shipley & Douma, 2020). When all variables are normally distributed and all relationships are linear, this leads to identical results as classical SEM. Alternatively, one can test a series of independence claims that result from the DAG to obtain a global metric of model fit (Shipley, 2009). The hypothesized model predicts which variables will be independent of each when conditioned on their respective parents if the hypothesized structure is correct. This test applied to DAGs is called the d-sep test, and was recently extended so it can test the structure of path models with dependent errors (i.e., an m-sep test; Shipley & Douma, 2021). To date, no solution is available for estimation of the path coefficients of such a path model when accounting for non-linearity, non-normality and hierarchical designs.

This paper provides two contributions: (i) we provide a convenient decomposition of an acyclic causal graph involving only observed variables but including dependent errors into smaller sets of variables that can be modeled independently of each other, allowing piecewise estimation; (ii) we show that the dependent errors between non-normally distributed variables can be modeled with copulas. The paper presents a proof of concept and identifies practical issues to be solved when doing piecewise estimation of path models with correlated errors.

The paper proceeds as follows. First a decomposition of a path model with dependent errors is given. Second, we explain how to model multivariate distributions through copulas. Third, a likelihood ratio test is proposed to test the consistency of the path model with the data. Fourth, we illustrate our methodology with two case studies. We conclude with a practical step-by-step description of the procedure and identify practical issues when doing piecewise estimation of path models with correlated errors.

2. Decomposition of the Path Model with Correlated Errors into Independent Sets

To model a set of variables that are non-normally distributed or relate non-linearly to each other, it is convenient to reduce the complexity of the multivariate models to be fitted as much as possible. As stated before, reducing the complexity of a DAG is straightforward. The joint probability distribution of n variables that is generated by a DAG can be decomposed through a local Markov decomposition into the product of the n univariate probability distributions (Pearl, 2000, 2009). The likelihood of a variable is fully described by modelling the effect of its causal parents (i.e., its direct causes in the DAG) on the variable. Such a decomposition is not possible for path models with dependent errors because pairs of variables possessing such dependent errors

have at least one common parent that is latent and one cannot condition on such a latent. This next section explains how to obtain an alternative decomposition that is appropriate for acyclic path models with dependent errors and that reduces to Pearl's (Pearl, 2009) decomposition when dependent errors are absent. To explain the decomposition, we first introduce some terminology.

Here, graphs containing a mixture of directed arrows and bi-directed arrows, and having no cyclic relationships, are called mixed acyclic graphs, MAGs (Richardson & Spirtes, 2002; Shipley & Douma, 2021). A MAG consists of vertices (V), and edges (E) that connect the vertices to each other. Once a vertex is associated with a set of observational units then it is also a variable, and we will use "vertex" and "variable" interchangeably. Its edges come in two forms: directed arrows ($\rightarrow$) representing a causal relationship and bidirected arrows ($\leftrightarrow$) representing unresolved causal relationships (dependent errors). MAGs are derived from DAGs with latent variables, in which the bidirected arrows ($v \leftrightarrow v$) represent two vertices in the MAG that are adjacent to each other due to marginalizing over a latent vertex (Pearl, 2009; Richardson & Spirtes 2002; Shipley & Douma, 2021). Marginalizing over a variable (Z) in a multivariate probability distribution P(X, Z) means summing (or integrating) the probabilities of X over all values of Z: $P(X) = \sum_i P(X, Z_i)$. In this paper, we will only consider MAGs obtained after marginalizing over latent vertices and so we will only consider MAGs having either directed or bidirected edges. Further details about the properties of MAGs are given in Appendix S1.

Vertices joined by a bidirected edge ($V_i \leftrightarrow V_j$) mean that V_i does not directly cause V_j, and V_j does not directly cause V_i, but that there is a dependency between V_i and V_j due to a common latent cause of both. A vertex V_i that connects to X with a bidirected arrow ($V_i \leftrightarrow X$) is called the *spouse* of X. Vertices that have an arrow ($V_i \rightarrow X$) pointing into X are called the *parents* of X, and X is their *child*. The *ancestors* of X are all vertices that can reach X by directed paths, including X itself. A *collider vertex* Z along a path between two other vertices is a vertex that has arrows pointing into it from both directions. For a MAG this implies $V_i \rightarrow Z \leftarrow V_j$, $V_i \rightarrow Z \leftrightarrow V_j$, $V_i \leftrightarrow Z \leftarrow V_j$, or $V_i \leftrightarrow Z \leftrightarrow V_j$. A *non-collider vertex* Z along a path between two other vertices is then of the form: $V_i \rightarrow Z \rightarrow V_j$, $V_i \leftarrow Z \leftarrow V_j$, $V_i \leftarrow Z \rightarrow V_j$, $V_i \leftrightarrow Z \rightarrow V_j$, or $V_i \leftarrow Z \leftrightarrow V_j$. Given a MAG, a path p between two vertices is *m-connected* (i.e., information flows between the two vertices) given a set of vertices **Z** if (Richardson & Spirtes, 2002):

i.   every non-collider on the path is not in **Z**, and
ii. every collider on the path is in **Z** or has an ancestor in **Z**.

If every path between the two vertices is not m-connected then the two vertices are *m-separated*; otherwise the two vertices are *m-connected*. Richardson and Spirtes (2002) has proven that m-separation between the two vertices in the graph implies conditional independence of the two variables in the resulting probability distribution. To partition the MAG into
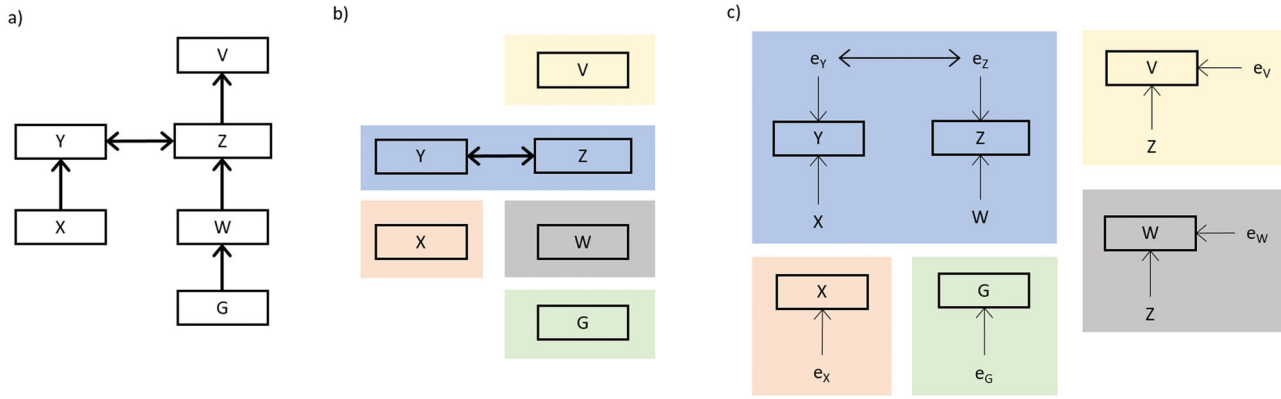
Figure 1. A hypothetical path model with directional paths (arrows) and correlated errors (bidirected arrows) (a), and unique district sets of variables (colors/shaded boxes) whose loglikelihoods can be summed (b), and the univariate and bivariate regressions to be fitted (c).

sets that can be modeled independently of each other we use the m-separation criterion to find subsets of observed variables that are independent of each other given the appropriate conditioning sets. We call a set of variables that cannot be decomposed any further a district set, **D**.

The district sets of a MAG are defined as follows (Evans & Richardson, 2014). For a graph G with vertices **V**, form the *induced bidirected graph* G' by removing all unidirectional arrows between pairs of vertices while maintaining all bidirected arrows. All V_j that can be reached from V_i (including V_i) in G' by traversing over the paths in G' while ignoring the directions of the arrowheads define the *district set* D_i of V_i. There are as many district sets as there are vertices in the MAG, but these district sets are not necessarily unique; two district sets, D_i and D_j are not unique if $D_i = D_j$. Let the set of unique district sets in G' be $D = \{D_1, \ldots, D_k\}$.

A graph G will consist of k≤ n unique district sets, where n is the total number of vertices in G. If the MAG is also a DAG, then each vertex forms a unique district set. All unique district sets that contain only one vertex can be modeled with a univariate distribution and are independent of all other unique district sets when conditioned on its respective parents. When a unique district set contains multiple vertices, a multivariate distribution is needed that accounts for the interdependencies between the vertices. Each unique district set is independent of all other unique district sets when its vertices are conditioned on their respective exogenous parents. The *exogenous parents* of a unique district i, $epa(D_i)$, are the set of parents of any vertex in district set i that are not, themselves, members of district set i. Appendix S1 gives the proof and application of this claim, examples, and an R-script with the factorization algorithm.

Finally, the joint probability distribution over the entire set of variables in a MAG can be decomposed into the product of a series of probability distributions involving only variables in each unique district as follows:

$$p(V_i, \ldots, V_n) = \prod_{i=1}^{k} p_i(D_i | epa(D_i)) \qquad (1)$$

In the special case in which the MAG does not contain any bidirected edges (↔), there are as many unique districts

as there are variables, each unique district is $D_i = \{V_i\}$, every parent of a vertex is an external parent, and the decomposition reduces to the Markov decomposition given by Pearl (Pearl, 2009):

$$p(V_i, \ldots, V_n) = \prod_{i=1}^{n} p_i(V_i | pa(V_i)) \qquad (2)$$

The above decomposition maybe useful to speed up the fitting of path models with dependent errors, as the resulting subgraphs each have fewer parameters than the total graph, and can be fitted and optimized independently of each other. This may lead to faster parameter optimization (Bellman, 1957), and it could be useful for optimizing parts of the graph with respect to the functional forms chosen or the parametric distributions, without having to worry that this affects other parts of the graph. Finally, the decomposition holds for any graph written as a MAG and is therefore applicable to classical path models as well as Bayesian belief networks.

We present two case studies to illustrate the application of the algorithm. Figure 1 shows an arbitrary example of a MAG and the resulting district sets. Y and Z will be modeled using a multivariate distribution conditioned on their respective external parents (X and W respectively). The other variables can be modeled with univariate distributions conditioned on their external parents. Likewise, Figure 2 shows a more complex example of a MAG taken from the literature (Scherber et al. 2010) containing six district sets, of which two district sets each contain three variables.

Our piecewise model M contains a vector of parameters $\theta = \{\theta_1, \ldots, \theta_p\}$. This model gives the multivariate probability distribution, or probability density function, $p_M(V | \theta)$, of this set of random variables $V = \{V_1, \ldots V_n\}$. The likelihood of this model for θ given the data (**V**) is $\mathcal{L}_M(\theta | V)$ where $\mathcal{L}_M(\theta | V) = p_M(V | \theta)$. If the values of the parameters θ are chosen to maximize this likelihood, then this gives the maximum likelihood $\mathcal{L}_M(\hat{\theta} | V)$. Given Equation (1), the log likelihood of the model M is equal to the sum of the log likelihoods of each of the k unique district sets; in this equation the notation θ_i and V_i refer to all of the parameters and variables involved in the ith submodel.
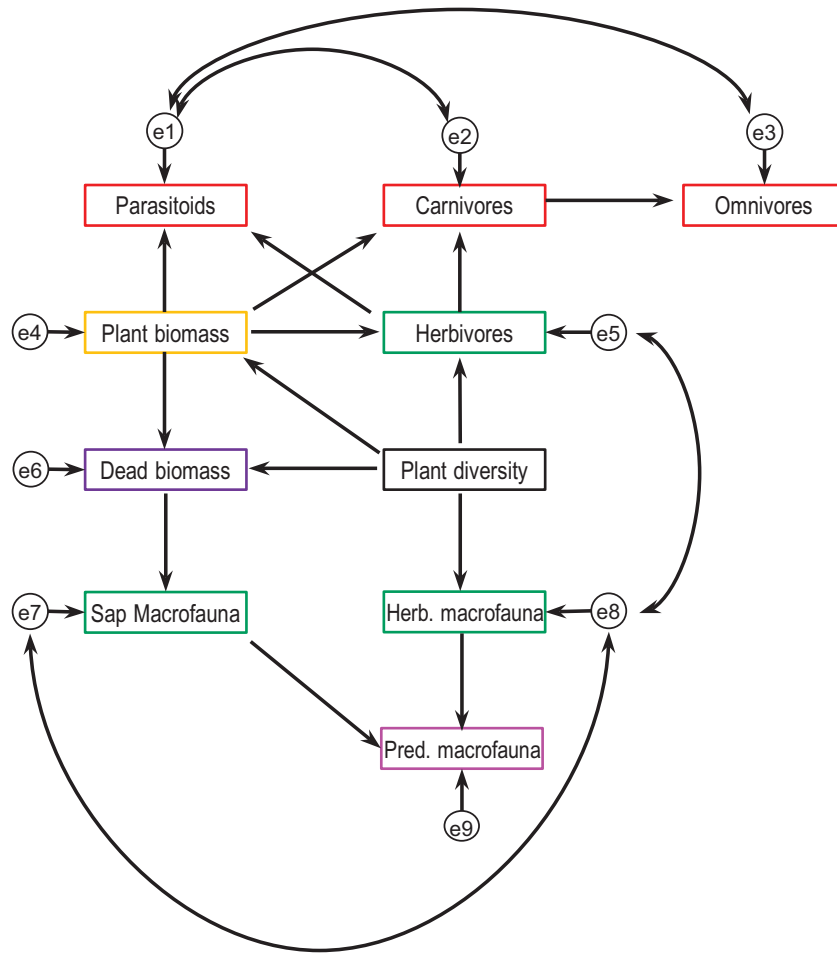
Figure 2. Path diagram representing the relationships between organism from different trophic levels ultimately as a function of plant diversity. Solid one-directional arrows represent causal relationships. Double-headed arrows represent dependent errors between variables. The path diagram is partitioned into district sets, and are represented by different colours. Full details of the study can be found in Scherber et al. (2010).

$$\log(\mathcal{L}_M(\boldsymbol{\theta}_M|\mathbf{V})) = \log\left(\prod_{i=1}^{k} \mathcal{L}_i(\boldsymbol{\theta}_i\mathbf{V}_i)\right) = \sum_{i=1}^{k} \log(\mathcal{L}_i(\boldsymbol{\theta}_i\mathbf{V}_i)) \tag{3}$$

3. Copulas Explained

The n vertices that reside in one unique district set must be modeled as a joint probability distribution. However, complex multivariate distributions arise when variables are non-normally distributed. Although well-defined multivariate distributions are sometimes available, this often not the case. When this occurs, copulas can be used. A copula decomposes the joint distribution of n variables into the product of the n marginal distributions of the variables plus a copula that describes the dependence between the n variables (Sklar's theorem, Sklar, 1959). A copula is a multivariate cumulative distribution function with uniform margins, i.e., on the interval [0, 1]. Appendix S2 provides a more detailed description of copulas and an illustration in R of a gaussian copula. The probability distribution of any random variable can be made uniform by feeding it into the cumulative distribution function from which it was generated (a probability integral transform). By doing so, information

about the characteristics of this distribution is lost (such as its moments) and what remains is only the dependence structure between the variables on $[0, 1]^n$. As such, the copula can link arbitrarily distributed variables and specifies the degree of monotonic dependence (see Figure 3 for a conceptual illustration). If a variable is endogenous and therefore has parents then its cumulative distribution can be obtained after regressing the variable on its parents, and use these regression parameters to feed into a cumulative distribution function where its parameters are the same as obtained in the regression (Hofert et al., 2018).

3.1. Two Important Technical Notes

Although Sklar's theorem of decomposing the joint probability distribution holds for any joint distribution, the copula function is not always differentiable or cannot always be conveniently estimated. Differentiating the copula (which is a cumulative distribution) into a probability density function is convenient because then it can be used for maximum likelihood estimation of its parameters (Hofert et al., 2018). Therefore, in practice, a limited number of conveniently defined copulas are frequently used (Hofert et al., 2018). These copulas do not necessarily exactly decompose the
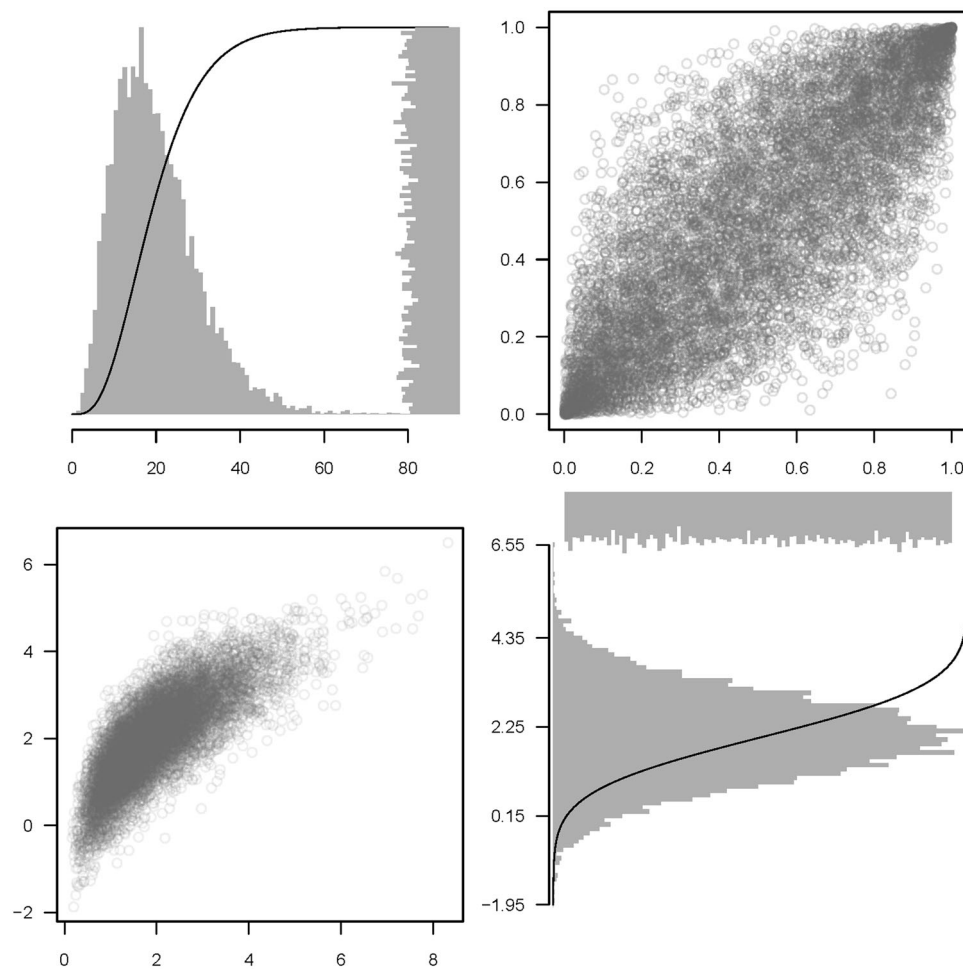
Figure 3. Conceptual illustration of the copula. The bottom left graph shows two correlated variables: a gamma distributed variable (histogram in top left) and a normally distributed variable (histogram in bottom right) on their original scales. After feeding the marginal distributions into their respective cumulative distributions (solid black lines), two correlated uniform distributions are left (vertical grey bars). The resulting two uniform distributions have the same rank correlation (top right) as the untransformed variables. The resulting two uniform distributions are modeled with a copula (a Gaussian in this example). After Meucci (2011).

joint likelihood of n variables, but rather approximate it; like the normal distribution is often used to approximate normal-looking data, and the beta distribution is used to model data between 0 and 1.

In Figure 4 different copulas are shown that have different properties. For example, the Gumbel copula allows for strong right tail dependence, the Clayton copula has strong left tail dependence, and the Gaussian copula has symmetric tail dependence. One can choose between the different copulas, based on the observed properties of the data, or by considerations of model fit (Genest & Favre, 2007). Only the gaussian copula can be easily extended to >2 dimensions, and we therefore recommend it as the preferred choice. If other copulas are chosen, they can be made multidimensional through a graphical method called vines. See Appendix S2 for more details on vine copulas (section 'multivariate copulas') and Appendix S3 for a script in R to make a visual assessment of tail dependence.

The margins of the copula function are assumed to be uniform and continuous on the interval [0, 1]. If the cumulative distribution functions that are used to transform a variable to uniform margins are continuous, the copula is uniquely defined. However, when the margins are defined as count distributions (e.g., a Poisson distribution), the copula is not unique because the probability integral transform returns a countable number of values on the interval [0, 1]. This may lead to biased estimates of the copula dependence parameter (Genest & Nešlehová, 2007; Trivedi & Zimmer, 2017). In practice, for Poisson and negative binomial distributed variables, this does not lead to problems except under certain conditions. Trivedi and Zimmer (2017) showed that the dependence parameter between two Poisson distributed variables can be quite accurately estimated when their means are larger than 1.05. When these variables are functions of covariates the problem virtually diminishes and means of at least 0.15 can be estimated accurately. This is because the copula function will have predicted means as input, and these are continuous. We are not aware of studies that have investigated the identification issue for dependent Bernoulli or binomial variables in practice, but we refer to Masarotto and Varin (2012) who apply copula regressions to discrete data, including binomial data.
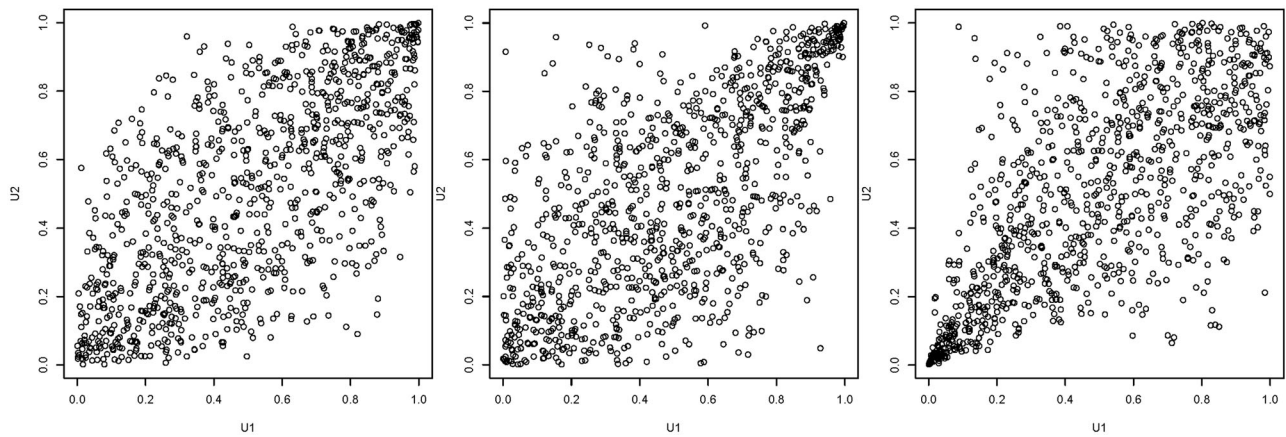
Figure 4. Simulated dependence between two variables U1 and U2 assuming a Normal, Gumbel and Clayton copula respectively. The latter two show stronger dependence at the upper and lower tail respectively.

4. Maximum Likelihood Estimation of a Copula

Although the copula approach allows one to separate the dependence between variables from their marginal distributions, the dependence parameters cannot be estimated independently of the parameters of the marginals, because the parameter estimates of the marginal distribution affect the parameter estimates of the copulas (Appendix S2). For this reason, it is necessary to simultaneously optimize both the marginal distributions and their associated copulas.

When the number of variables in the path model becomes large, or the model becomes complex for other reasons, simultaneous optimization of the marginals and the copula parameters through maximum likelihood estimation may fail (Joe & Xu, 1996; Joe, 2005). Currently it is an open question how many variables a copula regression can contain while still being able to numerically optimize it (but see Oh & Patton, 2017). If the optimization fails, it is proposed to estimate the marginals and the copulas in a two stage procedure (called inference from margins; IFM). First, the parameters of the marginal distributions are estimated with maximum likelihood. Thereafter, the copula parameters are estimated with the uniformized margins given the maximum likelihood parameters of the marginals. After having these estimates one can proceed in two ways. Preferably one uses the estimates obtained from the two-stage estimation as starting values for the simultaneous optimization. In case that this fails, one uses IFM without the simultaneous optimization afterwards. The advantage of IFM is that one only needs univariate models for the margins. For example, one could fit (generalized) linear mixed effect models on the margins and estimate the correlation between variables with a copula afterwards (Ferreira et al., 2019).

The ease of the IFM approach comes at the cost of the precision, thus the standard error, of the estimated parameters. IFM has been shown to produce parameter estimates that converge to the true parameter value with increasing sample size (consistency, Joe & Xu, 1996; Ko & Hjort, 2019). Practical guidelines for when IFM can be safely used at relatively low sample sizes (50-200) are, to the best of our knowledge, missing but the simulation study described below gives some insights. Furthermore, the average

deviation between the estimated parameter value and the true parameter value, i.e., parameter bias, was found to be low for large sample sizes (i.e., asymptotic efficiency), but usually the deviations were larger than for simultaneous estimation (Joe, 2005) and the asymptotic efficiency was found to be low when dependence between variables was strong (Joe, 2005). A final drawback of the IFM approach is that the standard error of the copula parameter will be underestimated and thus cannot be used for inference on the significance of the copula parameters (Hofert et al., 2018). Bootstrap estimates can be used to create standard errors for the copula parameters (Ferreira et al., 2019).

5. Testing the Consistency of the Path Model to the Data

To test whether the patterns of dependence that are observed between the variables are consistent with the patterns of dependence that are implied by the hypothesized path model, and to test whether any deviations between these can reasonably be ascribed only to random sampling variation, the likelihood ratio test (LRT, Wilks, 1938) or the m-sep test (Shipley & Douma, 2021) can be used. For the LRT, if the hypothesized causal structure has generated the data, then twice the difference between the likelihood of the hypothesized and a saturated model will asymptotically follow a χ^2 distribution with the degrees of freedom equal to the difference in number of free parameters between the hypothesized and the saturated model (Wilks, 1938). A saturated model is a path model with no missing arrows between the variables, hence saturated. See Shipley and Douma (2020) for a more detailed description of how to define the saturated model for DAGs.

Defining a saturated model for non-normally distributed variables and for non-linear relationships between variables is less straightforward than in classical SEM. In classical SEM one assumes that variables are linearly related to each other, and that variables are normally distributed. Given these assumptions, a path model consisting of free covariances between all variables—which is the saturated model in classical SEM—is always the model with the highest

Table 1. Loglikelihood of the path model presented in Figure 1 (generating equations in Table S4.1 in Appendix S4) obtained from classical SEM, the piecewise copula and the method of inference from margins (IFM).

Variable i	Classical SEM	Piecewise copula	Piecewise copula IFM
G	–	−132.45	−132.45
X	–	−138.161	−138.161
W	–	−139.488	−139.488
Y,Z	–	−255.87	−145.859
			−146.667
			35.458
V	–	−145.81	−145.81
Full model	−811.779	$\Sigma_i = $ −811.779	$\Sigma_i = $ −812.972
Saturated model	−808.889	−808.889	−808.908

The piecewise copula method calculates a loglikelihood for each set. The method of inference from margins calculates a loglikelihood for each variable and the copula. The loglikelihood of the saturated model is given in the bottom row and cannot be decomposed further.

likelihood and returns the same likelihood as a saturated model with only directed paths. This is because a multivariate normal distribution is closed under conditioning (Devroye & Lehn, 2010); i.e., the distribution resulting after conditioning on a variable is still normal. As a result, it does not matter in terms of likelihood whether one fits a bivariate normal, or the product of two normals with one normal conditioned on the other (Devroye & Lehn, 2010). However, for non-normal variables, or for variables that relate non-linearly to each other, this cannot be done. Therefore, a saturated model with only dependent errors is not necessarily the model with the highest likelihood. Thus, when using non-linear relationships, or employing generalized linear models (GLMs) that use a link function other than the identity link one cannot simply use a saturated model that assumes dependent errors between all variables.

We therefore propose to define the saturated model as a path model equal to the hypothesized model plus bidirected arrows joining pairs of variables that are not joined by an edge (directed or bidirected arrows) in the hypothesized model. Systematic deviations in the maximum likelihood of the hypothesized path model to the saturated model will be due to misspecifications of the hypothesized model due to missing paths. Of course, as is always the case in any statistical test, a lack of significant deviation of the hypothesized model to the saturated model does not 'prove' that the hypothesized model captures the right causal structure, only that the data do not provide sufficient evidence to reject the model (see simulation study below).

The alternative to the method described here, i.e., comparing the maximum likelihood of the hypothesized model to the maximum likelihood of a saturated model, is to use the m-sep test described in Shipley and Douma (2021). The m-test does not require a saturated model, which is an advantage when fitting more complicated path models, but tests only the causal structure of the model, not details about the functional form of the cause—effect links.

5.1. Case Study 1: A Simulation Study with a Numerical Example

To illustrate the fitting of a path model with dependent errors, we generated 100 observations following the path model in Figure 1. The generating equations for each variable are listed in Table S4.1 in Appendix S4 (R-code is shown in Appendix S5). Three estimation techniques were used: (1) the piecewise copula method; (2) inference from margins (estimating the margins for all variables, followed by the estimation of the copula for Y and Z); (3) classical maximum likelihood SEM (that estimates the model implied variance-covariance matrix and the free model parameters simultaneously). The latter is possible since we assumed that all variables were normally distributed and related linearly to each other. Next, we fitted the saturated model and the baseline model to the data. The saturated model and the baseline are equivalent for linear path models with multivariate normal distributions.

The piecewise copula method, using a gaussian copula to model correlated errors, produced the same likelihood for the hypothesized and the saturated model as did classical SEM, in contrast to the estimates by the method of inference from margins (Table 1). As a result, the chi-square test to test for difference between the hypothesized and saturated model were equivalent for the piecewise copula method and the classical SEM, but not for the method of inference from margins. However, all methods showed that the tested causal structure was consistent with the data according to the likelihood ratio test ($X^2_{ML} = 5.782$, df $= 10$, p-value $= 0.833$ for classical SEM and the copula method and $X^2_{ML} = 8.128$, df $= 10$, p-value $= 0.616$ for the method of inference from margins). When, instead, Z→Y is incorrectly used rather than Z↔Y, all three methods strongly rejected this model. Again, classical SEM and the copula method produced the same results according to the likelihood ratio test ($X^2_{ML} = 70.21$, df $= 10$, p-value $= 4.04e-11$).

Next, we simulated from the same causal model (Figure 1), but variables X, V, Z and Y were respectively lognormally, Poisson, beta and gamma distributed and we assumed a non-linear relationship (exponential) between W and the gamma distributed variable Z, a Michaelis-Menten relationship between X and Y (Y = aX/(b + X) and a Ricker function between V and Z (V = a*Z*(exp(−b*Z)), Table S5.3 in Appendix S5 for the generating equations). All other variables were normally distributed, and a Gaussian copula was used to model the dependent error between Y and Z. To test if the hypothesized model was consistent with the data, we compared the log-likelihood of the hypothesized model with the saturated model (model #4 in Table 2) following the procedure outlined in the section above and in Appendix S4. We also tested two alternative models that contained errors. First, we replaced the dependent error (Y↔Z) by a path from Z to Y (Y←Z). This model required its own saturated model (model #5). Second, we removed the dependent error between Z and Y. This second model has the same saturated model as our hypothesized model. Table 2 reports the results of the overall model statistics. The hypothesized model was consistent with the data as evidenced by a non-significant Chi-square test using a saturated model ($X^2_{ML} = 2.55$, df $= 10$, p-value $= 0.99$). The path model with an incorrect path Z→Y produced a p-value of 0.30 ($X^2_{ML} = 11.84$, df $= 10$, p-value $= 0.30$). The model with

Table 2. Model statistics of path model presented in Figure 1. Model statistics based on 100 observations generated according to the equations shown in Table S4.3 Appendix S4.

No.		LL	AIC	BIC	LRT
1.	Hypothesized model	−1127.96	**2283.93**	**2320.4**	Model 1 vs 5: $X^2_{ML} = 2.55$, df $=10$, p-value $= 0.99$
2.	Alternative model 1	−1156.55	2347.10	2391.39	Model 2 vs 6: $X^2_{ML} = 11.84$, df $=10$, p-value $= 0.30$
3.	Alternative model 2	−1164.57	2359.15	2398.23	Model 3 vs. 5: $X^2_{ML} = 75.77$, df $=11$, p-value $= 9.61\text{e-}12$
4.	Saturated model 1	−1126.69	2293.37	2345.47	
5.	Saturated model 2	−1150.63	2357.26	2430.21	

Loglikelihood (LL), Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC) and the Likelihood ratio test (LRT). The likelihood ratio test compares the hypothesized model to a saturated path model. Each hypothesized model has its own saturated model as explained in the main text; hypothesized model (#1) is compared to saturated model 1 (#4), and alternative model 1 to saturated model 2 (#5). The first alternative model (#3) had a directed path from Z to Y instead of a correlated error between Z and Y. The second alternative model (#3) had no edge between Z and Y. The latter model has the same saturated model as model #1. Bold values highlight the most parsimonious model according to AIC and BIC.

an incorrect missing edge between Z and Y was clearly rejected ($X^2_{ML} = 75.77$, df $=11$, p-value $= 9.61\text{e-}12$).

To understand the behaviour and compare the power of the saturated model to detect misspecifications in the path model we compared our proposed saturated model approach to existing approaches. As result, we tested and compared seven different approaches: (1) A saturated model as used in classical SEM; (2) as 1 but with Satorra-Bentler as robust test statistic; (3) as 1 but with mean-variance scaling as robust test statistic; (4) a saturated model using *bidirected* arrows between each pair of variables that do not have a path between them in the hypothesized path model by copulas. See Appendix S4 Figure S4.2 for an illustration.; (5) a saturated model using *bidirected* arrows between each pair of variables that do not have a path between them in the hypothesized path model by copulas and using IFM; (6) A saturated model using *directed* arrows between each pair of variables that do not have a path between them in the hypothesized path model, following. See Appendix S4 Figure S3 for an illustration; (7) the m-sep test, following Shipley (2009) and Shipley and Douma (2021) for details.

With each method, the fit of three different hypothesized path models to the data were tested: (1) a path model corresponding to the path model that was actually used to generate the data; (2) a path model similar to Figure 1 but with a directed path from Z to Y instead of a bidirected arrow; (3) a path model with a missing path between Z and Y.

The simulated data sets were always generated according to the path model shown in Figure 1 and had different sample sizes (50, 100, 200, 500, 1,000) per simulated data set. Additionally, three different sets of generating equations were used to generate the data, varying in complexity. First, we simulated all data assuming only linear relationships and normally distributed variables (Table S4.1 in Appendix S4) because we expect the results of our approach to be identical to the classical SEM. Second, we relaxed the assumptions of linearity and made three out of four relationship non-linear (Table S4.2 in Appendix S4). Third, we introduced both non-linear and non-normally distributed variables (Table S4.3 in Appendix S4). Each simulation was repeated 2000 times to obtain the rejection rates.

From the simulation study we conclude that when variables are normally distributed and have linear relationships with their parents, all approaches tested converge to the 5% rejection rate, with exception of the saturated model with inference from margins (#5). When using IFM rejection rates are 1.5-2 times higher compared to the simultaneous

estimation. The saturated model approach proposed in our study gives identical results as classical sem. When relaxing the assumptions of normality and linearity, the classical sem approach gives substantially higher rejection rates than the theoretical 5% ($>51\%$) with the robust estimators performing better compared to the standard chi-square estimators ($>41\%$) but still showing substantial type I errors. In contrast, the saturated model approach with copula (#4) and the m-sep test behave comparably when assessing the fit of hypothesized path models to the data, and converge at large sample sizes at a rejection of 5% when the hypothesized model is correct, and higher rejection rates when the hypothesized model is incorrect. The saturated model approach with directed paths did not perform as well as the approach with bidirected arrows, and the rejection rates depended on the direction in which the arrows were specified. When the path model does not conform with how the data was generated then the rejection rates went up in all methods tested. Overall, the copula method and the m-sep test performed best across cases containing non-linear relationships and nonnormal relationships, and have to be preferred over the saturated model made from adding directed paths to pairs of variables that do not yet have an edge between them.

All simulations were performed in R. The classical SEM model was fitted with the R-package lavaan (Rosseel, 2012).

5.2. Case Study 2: Effect of Plant Diversity on Multitrophic Interactions

The second case study illustrates our approach on data collected in a long term ecosystem experiment (Jena, Germany, Scherber et al. 2010). The aim of this experiment is to explore the long-term effects of plant diversity on ecosystem processes. Different numbers of perennial plant species were sown in different plots, and in each plot the abundance of different organismal groups from different trophic levels were measured. The interdependency between the different organismal groups was modeled through a structural equation model with only measured variables. In the original analysis, all variables were log transformed and scaled to [0,1] prior to analysis, and linearity and normality was assumed. However, some variables representing abundance were measured on a count scale while others were measured on a continuous scale, and variation was heterogeneous across the predictors. Additionally, some relationships were non-linear.

Table 3. Non-linear equations with MLE parameters estimates associated to variables in the path model presented in Figure 2.

Variable	Equation
Parasitoids (P)*	NB (μ = exp($4.18 + 7.67e\text{-}04$*H + $1.30e\text{-}03$*PB), k = 1.54)
Carnivores (C)*	G (μ = 2.37*PB$^{0.13}$ + 0.27*H$^{0.28}$, s = 4.90)
Omnivores (O)*	G (μ = exp($2.92 + 2.94e\text{-}03$*C), s = 7.75)
Plant biomass (PB)	G (μ = 171.87*PD$^{0.28}$, s = 57.57)
Herbivores (H)**	G (μ = 165.18*PB$^{0.16}$ + 78.00*PD$^{0.56}$, s = 71.67)
Dead biomass (D)	G (μ = 1.64*PB$^{0.30}$ + 1.22*PD$^{0.22}$, s = 3.95)
Sap. macrofauna (S)**	NB (μ = exp($1.77 + 1.50e\text{-}2$*D), k = 1.94)
Herb. macrofauna (HB)**	NB (μ = 3.15 + PD$^{0.24}$, k = 1.07)
Pred. macrofauna (PB)	NB (μ = 1.28*SB$^{0.75}$ + 4.41*HB$^{0.54}$, k = 1.12)
Plant diversity (PD)	NB (exp(2.42), k = 0.75)
e1 ∼∼ e2*	**0.40**
e1 ∼∼ e3*	0.14
e5 ∼∼ e8**	0.20
e7 ∼∼ e9**	**0.32**

Different distribution functions were used depending on the variable (G = Gamma distribution, NB = Negative binomial distribution). Four correlated errors were fitted, represented in the table by e ∼∼ e. Bold values represent MLE estimates that are significantly different from zero. *, and ** in the first column ("Variable") represent respectively two district sets that were simultaneously fitted with the regressions and a gaussian copula linking the variables through dependent error.

We used the same path diagram that was found to be consistent with data in the original analysis for our analysis and fitted it with piecewise SEM while accounting for non-linearity and non-normality. First, for each variable univariate regressions were fitted with probability distributions depending on the type of data (negative binomial distribution for counts, and a Gamma distribution for continuous variables as we observed increased variance with increased values of that variable), and different non-linear functions were fitted and compared. The functional forms that lead to the lowest AIC were used in the full path model. Second, the factorization algorithm was applied to identify the districts sets. If variables were connected to each other through dependent errors a gaussian copula was used. The consistency of the path model to the data was tested both using the saturated model approach and the m-sep test. Through the partitioning to district sets, six parameter sets containing respectively 14, 12, 6, 5, 3 and 2 parameters were fitted instead of fitting 42 parameters at once. The whole fitting procedure is described in detail in Appendix S7. The resulting path diagram is shown in Figure 2, and the regression equations are shown in Table 3. The path model was consistent with the data, both according to the likelihood ratio test ($X^2 = 2.21$, df = 27, p-value = >0.99) as well as the m-sep test (C = 51.27, $k = 54$, p-value = 0.58).

6. Proposed Procedure and Practical Recommendations

Below we outline a step-by-step description of the procedure to fit path models with dependent errors using the approach outlined in this paper and the R functions that could be used in each step. Appendix S6 applies each step below to case study 2.

Step 1: Define a path model containing directed and/or bi-directed edges. One can make a DAG or MAG with the R-package 'ggm' (Marchetti, 2006).

Step 2: Decide which test to use to test the consistency of the model with the data. One can either use the m-sep test

(move to step 3) or the saturated model approach (move to step 4).

Step 3: Obtain the independence claims resulting from the hypothesized path model. Next, test each independence claim, collect p-values and test whether the data is consistent with the hypothesize causal structure (see Shipley, 2009; Shipley & Douma, 2021 for a detailed description of the procedure). If the hypothesized structure is consistent with the data, move to step 4, otherwise go to step 1 and revise the hypothesized path model. To obtain the independence claims of a MAG one can use the functions "basiSet.mag" from the R-package CauseAndCorrelation (https://github.com/BillShipley/CauseAndCorrelation).

Step 4: Partition the MAG into district sets. Here on can use the function "districtSet" from the package CauseAndCorrelation. The subgraphs to be fitted including the exterior parents can be obtained through the function "districtGraph" from the package CauseAndCorrelation. Move to Step 5.

Step 5: If a district set contains only one variable one can proceed with fitting a univariate regression with the parents being the independent variables. The choices of the distributional forms the functional relationships follow the same decisions as for generalized (non-)linear (mixed effect) modelling, i.e., one can have a priori reasons to choose certain parametric distributions or functional forms, or one chooses distributions and functional forms that lead to a good fit (Bolker et al., 2009; McCullagh & Nelder, 1989; Zuur et al., 2009). Non-parametric functions can be used as long as they can be optimised in the maximum likelihood framework. See Bolker, 2008 for a good guide on which functional forms and parametric distributions to choose.

If a district set contains multiple variables, the recommended procedure would be to first fit the univariate regressions of the variables in the district (based on the same considerations as given above), choose a copula function to link the variables. Different copula functions can be chosen. See Hofert et al. (2018) for examples and Appendix S3 for how to assess the degree of tail dependence. Calculate the correlation of a pair (i.e., the residuals

of the) variables using a rank correlation. Construct a likelihood function that contains the variables in the district and the dependent errors through the copula (See Appendix S2 for an example). Use the estimates of the univariate regressions, and the dependence parameters as starting values for the optimisation. Providing good starting values is particularly important when the district set contains a large number of variables. This may happen when one assumes that exogenous variables are dependent on each other. Copulas can be fit using various packages in R (Hofert et al., 2020; Nagler et al., 2021; Nagler & Vatter, 2022) , but to date, to the best of our knowledge, no convenient R packages exist that can flexibly fit copula regressions.

Repeat the above for all district sets and finally sum all the loglikelihoods of the district sets.

Step 6: If one has chosen to compare the fit of the hypothesized model to the data using the saturated model approach proceed as follows. Combine the likelihood function of all districts sets to one likelihood function and estimate in the same optimisation step the dependencies between variables that do not have a directed path or bidirected path between. The number of parameters that are additionally estimated with a hypothetical model with n variables are (n*n-n)/2-p-z with p the number of directed paths and z the number of dependent errors. Obtain the loglikelihood of the saturated model. Use the likelihood ratio test to compare the logLikelihood of the hypothesized with the log Likelihood of the saturated model.

Step 7: If the hypothesized model appears to be consistent with the data, report the path coefficients, dependent errors, their standard errors and significance.

7. Discussion

This paper has made two contributions: Firstly, we give a Markov decomposition of the multivariate probability distribution and associated likelihood that is generated by path models with dependent errors into sets of variables (unique districts) that can be modeled independently of each other. Secondly, we used copulas to model the dependent errors between non-normally distributed variables. Together, these advances provide a generic solution to causal modelling of observed variables that are not normally distributed, are not linearly related to each other, or for which observations are not mutually independent. Importantly, it allows the user to tailor the probability distributions and functional forms according to their hypothesis or data.

Piecewise estimation is not common practice in path modelling, even though it has advantages compared to simultaneous estimation. Factorizing the graphs into independent subsets is frequently done in the domain of Bayesian networks represented as DAGs (e.g., Liu et al., 2012; Pearl, 2009). Evans and Richardson (2014) have provided an algorithm for discrete data that partitions MAGs into even smaller subsets than the method that we proposed, but this partitioning cannot be applied to continuous distributions. Decomposing the path model into unique districts has two

advantages: First, it may ease model fitting since the complexity of the full path model is reduced to subgraphs of smaller size. For example, Table 2 provides a separation of the full model likelihood into the likelihood of each unique district set, which gives insight in the contribution of each variable to the full model likelihood. Second, it may improve the estimation of the path coefficients and the dependent errors because the appropriate estimation methods can be used for each unique district set. Also, each unique district set can be optimized independent of other district sets by comparing different functional forms and probability distributions.

To test the consistency of the hypothesized path model with the data we have introduced the saturated model. The saturated model was constructed by adding dependent errors (i.e., bidirected arrows) to each pair of variables in the hypothesized path model that are not already joined by an edge. We compared the performance of this method to existing approaches. The saturated model approach generated as expected identical results as with classical SEM. The copula saturated model method and the m-sep test performed best across cases containing non-linear relationships and nonnormal relationships, and have to be preferred over the saturated model made from adding directed paths to pairs of variables that do not yet have an edge between them. The difference in rejection rate between these saturated models diminishes when the non-normal distributions become more symmetric (see Appendix S4).

Despite that the copula method obtains slightly better rejection rates compared to the m-sep test (at least, in the simulations used here), the m-sep method is more convenient to use. The m-separation claims of the m-sep test only involve univariate distributions, and only if the path model is not rejected does one proceed to model the functional links between the variables including the fitting the copulas. This two-step approach may be a particularly attractive option when fitting mixed effect models that would require a saturated model that may be difficult to fit. In contrast, the saturated model approach requires that all variables are modeled simultaneously which may involve a lot of parameters. The latter may cause optimization problems when the saturated model contains a large number of parameters. Application of this approach to real data is needed to identify the conditions when the saturated model can be fitted anymore.

The disadvantage of the m-sep test is that it is sometimes more cumbersome to include non-linear relationships in the independence claims. Both the m-sep test and the saturated model approach make assumptions on the functional form of the relationships. This is comparable to when applying classical SEM that assumes linearity to data that contains non-linear relationships.

The use of copulas to model dependent errors has three main advantages. First, as indicated, it allows one to model the dependence between variables that are not-normally distributed. This may be especially relevant for distributions with strong kurtosis (Hoogland & Boomsma 1998). Second, since copulas have probability density functions, this allows

maximum likelihood estimation of the copula parameter, and hence a likelihood can be calculated for the full model. As a result, the chi-square test and model selection techniques such as AIC and BIC can be used to test for differences in and parsimony of the causal structure and specification of the path model. Third, one can choose among copulas to be able to account for types of dependency between variables, e.g., to account for stronger left or right tail dependence (Figure 4).

Copulas have been used frequently in the literature, in particular in the domain of economics (e.g., Anderson et al., 2016; Genest & Favre, 2007; Popovic et al., 2018). However, there is currently a lack of practical guidance on which copula to choose, under which conditions copulas can be applied to discrete data, when inference from margins can safely be used, and when numerical optimisation of copula (regression) is not possible anymore. Our simulations showed that IFM is inferior to simultaneous optimization and suggests that the latter should be preferred. We expect that practical experience and more extensive simulations studies may identify cases when copulas have added benefit over classical SEM approaches with robust statistics. Additionally, copulas have not often been applied in a mixed model framework, allowing for random effects (but see Ferreira et al., 2019; Zhang et al., 2020 and Appendix S7). More research is needed to identify the maximum complexity that can be fitted within the copula framework in terms of the number of dependent variables and their random effects. Even though canned copula solutions to model dependent (mixed effect) regressions do not exist yet, the rapid developments we have seen in fitting complex univariate mixed effect models make us optimistic that these will become available in near future.

To the best of our knowledge, we are the first to use copulas to model dependent errors in a SEM without latents. However, copulas have also been used to include latent variables and fit measurement models, i.e., models with latent and non-normally distributed indicator variables (Brechmann & Joe, 2014; Krupskii & Joe, 2013; Murray et al., 2013). So far, applications of copulas in SEM with latent variables have been scarce (but see Braeken et al., 2013). However, our work together with the cited work, may offer a generic solution to relaxing the assumptions of classical SEM.

Acknowledgements

Funding

ORCID

Jacob C. Douma ⓘ http://orcid.org/0000-0002-8779-838X
Bill Shipley ⓘ http://orcid.org/0000-0002-7026-3880

References

Anderson, M. J., de Valpine, P., Punnett, A., & Miller, A. E. (2019). A pathway for multivariate analysis of ecological communities using copulas. *Ecology and Evolution*, 9, 3276–3294. https://doi.org/10.1002/ece3.4948.

Bellman, R. E. (1957). *Dynamic programming*. Princeton University Press.

Bolker, B. M. (2008). *Ecological models and data in R*. Princeton University Press.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24, 127–135. https://doi.org/10.1016/j.tree.2008.10.008.

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley and Sons.

Braeken, J., Kuppens, P., Boeck, P. D., & Tuerlinckx, F. (2013). Contextualized Personality Questionnaires: A case for copulas in structural equation models for categorical data. *Multivariate Behavioral Research*, 48, 845–870. https://doi.org/10.1080/00273171.2013.827965.

Brechmann, E. C., & Joe, H. (2014). Parsimonious parameterization of correlation matrices using truncated vines and factor analysis. *Computational Statistics & Data Analysis*, 77, 233–251. https://doi.org/10.1016/j.csda.2014.03.002

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62–83. https://doi.org/10.1111/j.2044-8317.1984.tb00789.x

Devroye, L., & Lehn, J. (2010). *Recent developments in applied probability and statistics: dedicated to the memory of Jürgen Lehn*. Physica Springer.

Evans, R. J., & Richardson, T. S. (2014). Markovian acyclic directed mixed graphs for discrete data. *The Annals of Statistics*, 42, 1452–1482. 1431. https://doi.org/10.1214/14-AOS1206

Ferreira, p H., Fiaccone, r L., Lordelo, j S., Sena, S O. L., & Duran, v R. (2019). Bivariate copula-based linear mixed-effects models: An application to longitudinal child growth data. *TEMA*, 20, 37–59. https://doi.org/10.5540/tema.2019.020.01.037

Genest, C., & Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12, 347–368. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347)

Genest, C., & Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bulletin*, 37, 475–515. https://doi.org/10.1017/S0515036100014963

Grace, J. B. (2006). *Structural equation modeling and natural systems*. Cambridge University Press. doi: https://doi.org/10.1017/CBO9780511617799

Hofert, M., Kojadinovic, I., Mächler, M., & Yan, J. (2018). *Elements of copula modeling with R*. Springer International Publishing. https://doi.org/10.1007/978-3-319-89635-9

Hofert, M., Kojadinovic, I., Maechler, M., & Yan, J. (2020). *Copula: Multivariate dependence with copulas*. https://cran.r-project.org/web/packages/copula/copula.pdf

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, 26, 329–367. https://doi.org/10.1177/0049124198026003003

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94, 401–419. https://doi.org/10.1016/j.jmva.2004.06.003

Joe, H., & Xu, J. J. (1996). *The estimation method of inference functions for margins for multivariate models*. University of British Columbia Library. https://doi.org/10.14288/1.0225985

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, 96, 201–210. https://doi.org/10.1037/0033-2909.96.1.201

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). The Guilford Press. https://doi.org/10.1080/10705511.2012.687667

Ko, V., & Hjort, N. L. (2019). Model robust inference with two-stage maximum likelihood estimation for copulas. *Journal of Multivariate Analysis*, 171, 362–381. https://doi.org/10.1016/j.jmva.2019.01.004

Krupskii, P., & Joe, H. (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120, 85–101. https://doi.org/10.1016/j.jmva.2013.05.001

Liu, Z., Malone, B., & Yuan, C. (2012). Empirical evaluation of scoring functions for Bayesian network model selection. *BMC Bioinformatics*, 13, S14. https://doi.org/10.1186/1471-2105-13-S15-S14

Marchetti, G. M. (2006). Independencies Induced from a graphical Markov model after marginalization and conditioning: The R package ggm. *Journal of Statistical Software*, 15, 1–15. https://doi.org/10.18637/jss.v015.i06

Masarotto, G., & Varin, C. (2012). Gaussian copula marginal regression. *Electronic Journal of Statistics*, 6, 1517–1549.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). Chapman and Hall.

Meucci, A. (2011). A short, comprehensive, practical guide to copulas. *Risk Professional*. www.garp.org.

Murray, J. S., Dunson, D. B., Carin, L., & Lucas, J. E. (2013). Bayesian Gaussian copula factor models for mixed data. *Journal of the American Statistical Association*, 108, 656–665. https://doi.org/10.1080/01621459.2012.762328.

Nagler, T., Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., & Erhardt, T. (2021). *VineCopula: Statistical inference of vine copulas.* https://cran.r-project.org/web/packages/VineCopula/VineCopula.pdf

Nagler, T., & Vatter, T. (2022). *rvinecopulib: High performance algorithms for vine copula modeling.* https://cran.r-project.org/web/packages/rvinecopulib/rvinecopulib.pdf

Oberski, D. (2014). lavaan.survey: An R package for complex survey analysis of structural equation models. *Journal of Statistical Software*, 57, 27. https://doi.org/10.18637/jss.v057.i01

Oh, D. H., & Patton, A. J. (2017). Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics*, 35, 139–154. https://doi.org/10.1080/07350015.2015.1062384

Pearl, J. (2000). *Causality*. Cambridge University Press.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511803161.

Popovic, G. C., Hui, F. K. C., & Warton, D. I. (2018). A general algorithm for covariance modeling of discrete data. *Journal of Multivariate Analysis*, 165, 86–100. https://doi.org/10.1016/j.jmva.2017.12.002

Richardson, T., & Spirtes, P. (2002). Ancestral graph Markov models. *The Annals of Statistics*, 30, 962–1030.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. https://doi.org/10.18637/jss.v048.i02

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage Publications, Inc.

Scherber, C., Eisenhauer, N., Weisser, W. W., Schmid, B., Voigt, W., Fischer, M., … Tscharntke, T. (2010). Bottom-up effects of plant diversity on multitrophic interactions in a biodiversity experiment. *Nature*, 468, 553–556. https://doi.org/10.1038/nature09492.

Shipley, B. (2009). Confirmatory path analysis in a generalized multilevel context. *Ecology*, 90, 363–368. https://doi.org/10.1890/08-1034.1.

Shipley, B. (2016). Cause and correlation in biology: A user's guide to path analysis. *Structural Equations and Causal Inference with R.* Cambridge University Press, Cambridge. https://doi.org/10.1017/cbo9781139979573

Shipley, B., & Douma, J. C. (2020). Generalized AIC and chi-squared statistics for path models consistent with directed acyclic graphs. *Ecology*, 101, e02960. https://doi.org/10.1002/ecy.2960.

Shipley, B., & Douma, J. C. (2021). Testing piecewise structural equations models in the presence of latent variables and including correlated errors. *Structural Equation Modeling*, 28, 582–589. https://doi.org/10.1080/10705511.2020.1871355

Sklar, A. (1959). *Fonctions de répartition à n dimensions et leurs marges.* Publications de l'Institut Statistique de l'Université de Paris.

Trivedi, P. K., & Zimmer, D. (2017). A note on identification of bivariate copulas for discrete count data. *Econometrics*, 5, 10. https://doi.org/10.3390/econometrics5010010

Wall, M. M., & Amemiya, Y. (2001). Generalized appended product indicator procedure for nonlinear structural equation analysis. *Journal of Educational and Behavioral Statistics*, 26, 1–29. https://doi.org/10.3102/10769986026001001

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9, 60–62. https://doi.org/10.1214/aoms/1177732360

Zhang, W., Zhang, M., & Chen, Y. (2020). A copula-based GLMM model for multivariate longitudinal data with mixed-types of responses. *Sankhya B*, 82, 353–379. https://doi.org/10.1007/s13571-019-00197-8

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effect models and extensions in ecology with R.* Springer. https://doi.org/10.1007/978-0-387-87458-6