Parsimonious parameterization of correlation matrices using truncated vines and factor analysis

Eike C. Brechmann[a], Harry Joe[b]

[a]*Corresponding author. Center for Mathematical Sciences, Technische Universität München, Boltzmannstr. 3, 85747 Garching, Germany. E-mail: eike.brechmann@tum.de. Phone: +49 89 289 17425. Fax: +49 89 289 17435.*
[b]*Department of Statistics, University of British Columbia, Canada.*

Abstract

Both in classical multivariate analysis and in modern copula modeling, correlation matrices are a central concept of dependence modeling using multivariate normal distributions and copulas. Since the number of correlation parameters quadratically increases with the number of variables, parsimonious parameterizations of large correlation matrices in terms of $\mathcal{O}(d)$ parameters are important. While factor analysis is commonly used for this purpose, the use of vines is an attractive alternative: Vines are graphical models based on a sequence of trees, and are based on the decomposition of a correlation matrix in terms of algebraically independent correlations and partial correlations. By limiting the number of trees, with so-called truncation, parsimonious parameterizations of correlation matrices may be found. Moreover, truncated vines and factor models may be joined to define a combined model, with individual benefits from each of the two approaches. The different parameterizations and how they are estimated for data are discussed. In particular, spanning tree algorithms for truncated vines and a modified EM algorithm for the combined factor-vine model are proposed and evaluated in a simulation study. Three applications to psychometric and finance data sets illustrate the different parsimonious models.

Keywords: multivariate normal, partial correlations, regular vines, Markov trees, EM algorithm

1. Introduction

The vine graphical model or pair-copula construction has been popular in the copula literature in recent years in order to use high-dimensional multivariate non-normal models. These models depend on a sequence of bivariate or pair-copulas, some of which are applied to univariate margins and most of which are applied to univariate conditional distributions. When all of the pair-copulas are bivariate normal, the resulting multivariate vine copula is multivariate normal or Gaussian. Vines have been important for copula construction but they also provide an alternative way of view Gaussian dependence. For the case of bivariate normal copulas applied to univariate conditional distributions, the parameters are interpreted as partial correlations. For a vine, the correlation matrix of a d-variate normal distribution is parametrized in terms of $d-1$ correlations and $(d-1)(d-2)/2$ partial correlations that are algebraically independent. There are many such partial correlation representations of a correlation matrix (see Kurowicka and Cooke (2003) and Lewandowski et al. (2009)) and a vine is a graphical model for these correlations and partial correlations summarized in a multiple tree structure with $d-1$ trees (Bedford and Cooke (2002) and Kurowicka and Joe (2011)). The parametrization with algebraic independence of the parameters is a key to the extension to non-Gaussian dependence.

For high-dimensional applications with the number of variables d being large, it is common in classical multivariate statistics to use a parsimonious correlation structure with $\mathcal{O}(d)$ parameters instead of $d(d-1)/2$ correlations. One such approach is factor analysis with p latent variables, where observed variables are conditionally independent given the latent variables. Factor analysis with p factors, generally with $p \ll d$, leads to one such structured correlation matrix with the number of identifiable parameters equal to $pd - p(p-1)/2$. Another approach is through truncated vines with the vine is truncated after ℓ trees with $1 \leq \ell < d-1$; the result is called an ℓ-truncated vine. If $\ell = 1$, then the result is a Markov tree dependence structure where two variables not connected by an edge are conditionally independent given the variables in the tree between them. However Markov tree dependence may be too simple to explain all of the dependence in d variables, and 2-truncated and 3-truncated partial correlation vines would be considered as parsimonious models with $2d-3$ and $3d-6$ parameters, respectively.

These two approaches have been considered in the copula literature. Krupskii and Joe (2013) have introduced the copula extension of factor mod-

els based on truncated vines rooted at latent variables, Brechmann et al. (2012) study copula models based on truncated vines and is the second paper with methodology on truncated vines for copulas after Kurowicka (2011). Truncated vine structures are relatively new in the multivariate literature as an approach to get a parsimonious dependence structure for high dimensions. With Gaussian pair-copulas, truncated partial correlation vines have not been fully exploited for applications except the special case of autoregressive (in time).

The main new contribution in this paper is the study of another parsimonious dependence structure that combines factor models and truncated vines. We call this a combined factor-vine model. To avoid being too general, we focus only on Gaussian dependence here and consider the copula extension in subsequent research. The simplest interpretation of these structures is that there is conditional dependence given latent variables. This can be especially useful and interpretable for applications when there is a latent variable that can explain most but not all of the dependence in the observed variables. Rather than adding latent factors, a truncated vine structure is assumed on the residuals conditional on the latent variable.

The estimation of a combined factor-vine model is non-trivial because of the latent variables and because the best truncated vine for the residual conditional dependence is to be determined. Methods of factor analysis do not apply and estimation techniques for vines can not be used directly. We show that estimation can proceed with a modified expectation-maximization (EM) algorithm, making use of minimum spanning trees for the truncated vine part of the model.

After full specification of the truncated vines, with or without latent variables, they can be written as structural equation models (SEMs), for which there are graphical representations through path diagrams to show conditional independence or relationships given latent variables. However not all SEMs have the form of truncated vines. We view this subclass of SEMs as (a) having representations in terms of partial correlations that are algebraically independent, and (b) being amenable to optimization via spanning tree algorithms in the sense of being able to find a 'good' structure without a complete specification in advance. In the SEM literature (e.g., Bollen (1989), Steiger (2001)), often the structural relations are assumed to come from some psychometric theory. Bollen and Long (1992) mention that one could go through a cycle of model specification and respecification. Our view is that there might be latent variables that can explain most of the depen-

dence, but there might be 'residual' dependence conditioned on the latent variables that can be parsimoniously explained by a truncated vine, more precisely, using our newly proposed combined factor-vine model—in contrast to factor models, where the variables are conditionally independent given the latent variables.

An example of 'residual' dependence in the item response theory literature is given in Braeken et al. (2007). In their model, there is one latent variable to explain most of the dependence, and then conditionally there is exchangeable dependence of each of several subgroups of items. We consider more general 'residual' dependence, because depending on the context, there might not be obvious subgroups of variables that are approximately exchangeable. Within a vine specification, no such restrictions need to be made.

An alternative approach to specify parsimonious correlation matrices is via setting entries of the inverse correlation matrix to zero (see, e.g., Dempster (1972) and Whittaker (2008)) and this leads to another graphical model. The non-diagonal entries are proportional to negative partial correlations of two variables given all the others. In contrast to vine-based models, this however does not lead to a parameterization in terms of algebraically independent parameters. Moreover, if variables are added to or deleted from a data set, the set of relevant partial correlations represented by inverse correlation matrix changes. This is not the case in a vine. In fact, a truncated partial correlation vine also implies zeros in the inverse correlation matrix, since higher order partial correlations are set to zero. For factor models, the common dependence on latent variables implies that two observed variables are conditionally dependent given other observed variables and the inverse correlation matrices does not have zeros and might not have small non-diagonal values. If the context suggest dependence through latent variables and the inverse correlation matrix is not "sparse", the combined factor-vine approach can yield flexible and yet parsimonious parameterizations of correlation matrices in terms of algebraically independent parameters.

The organization of the rest of this paper is as follows. In Section 2, a structural form is given for multivariate normal models for d variables with correlation structure having $\mathcal{O}(d)$ parameters. Notation is introduced for Markov trees, truncated vines, factor models and the novel combined factor-vine models. Section 3 has the general form of the log-likelihood for vine and factor-vine models, and discusses spanning tree algorithms for identifying the best spanning trees and doing the optimization. Section 4 has a small simulation study to show that the algorithms do well in finding the vine

structure. Applications to psychometric and finance data are given in Section 5. Section 6 discusses some future research.

Throughout, we are assuming that the sample size is large enough relative to the number of variables so that the correlation matrix can be well estimated and various structural correlation forms can be distinguished. When the sample size is not large enough or even smaller than the number of variables, and if there is some idea of a parsimonious structured correlation matrix such as considered in this paper, then shrinkage towards the structured matrix can proceed as in Daniels and Kass (2001).

2. Correlation matrix parameterizations

In this section, we introduce different multivariate normal models that have a structured correlation matrix with $\mathcal{O}(d)$ dependence parameters.

Let $\boldsymbol{Z} = (Z_1, \ldots, Z_d)'$, $d \geq 3$, be the vector of observed random variables, where each component is without loss of generality standard normal, that is, $Z_j \sim N(0,1)$ for all $j = 1, \ldots, d$. Further, let $\boldsymbol{V} = (V_1, \ldots, V_p)'$ be a vector of p unobserved random variables, which are also standard normal, that is, $V_j \sim N(0,1)$ for all $j = 1, \ldots, p$. The random vector $(\boldsymbol{Z}', \boldsymbol{V}')'$ is assumed to be jointly multivariate normal.

We assume the linear representation:

$$Z_j = \boldsymbol{\varphi}_j' \boldsymbol{Z} + \boldsymbol{\delta}_j' \boldsymbol{V} + \psi_j \varepsilon_j, \quad j = 1, \ldots, d, \tag{2.1}$$

where $\varepsilon_j \sim N(0,1)$ is independent of Z_i, $i \neq j$, and $\boldsymbol{V}$, the ε_js are mutually independent, and ψ_j is chosen such that $\mathrm{Var}(Z_j) = 1$. The variables in $\boldsymbol{V}$ are latent factors of $\boldsymbol{Z}$ and have loading vector $\boldsymbol{\delta}_j \in \mathbb{R}^p$. The vector $\boldsymbol{\varphi}_j = (\varphi_{j1}, \ldots, \varphi_{jd})' \in \mathbb{R}^d$, on the other hand, specifies between-variable dependence of the components of $\boldsymbol{Z}$. We assume that $\varphi_{jj} = 0$, and also the vectors $\boldsymbol{\varphi}_j$, $j = 1, \ldots, d$, have to be chosen such that the model is well-defined.

In other words, we assume that Z_j is explained by its relationship to other components of $\boldsymbol{Z}$ as well as to a set of common latent factors $\boldsymbol{V}$. The idiosyncratic variance is given through the ε_js. In the following three subsections we explain how specific choices of $\boldsymbol{\varphi}_j$ and $\boldsymbol{\delta}_j$ lead to parsimonious and well-defined parameterizations of the correlation matrix $\boldsymbol{\Sigma} = (\rho_{ij})_{i,j=1,\ldots,d}$ of $\boldsymbol{Z}$.

As a summary of the parsimony, we want $\mathcal{O}(d)$ non-zero parameters among the $\boldsymbol{\varphi}_j$ and $\boldsymbol{\delta}_j$ for $j = 1, \ldots, d$. For example, for a 1-truncated vine structure in (2.1), there is at most one non-zero element of $\boldsymbol{\varphi}_j$ for each j, and for a 2-truncated vine structure, there are at most two non-zero elements of $\boldsymbol{\varphi}_j$ for each j. See (2.8) and (2.3) for examples for how (2.1) is written for 1-truncated and 2-truncated vines, respectively. If the truncated vine is not completely specified a priori, the positions of the non-zero coefficients are not fixed in advance. A parsimonious model, with the truncated vine form to be determined, has not been previously studied.

A reason for assuming a truncated vine structure to determine the $\boldsymbol{\varphi}_j$ is that this permits a representation (or reparametrization) in terms of correlations and partial correlations that are algebraically independent. For notation, $\rho_{jk;m_1,\ldots,m_M}$ is the partial correlation of the variables j, k given or conditioned on the variables $m_1, \ldots, m_M$. The partial correlation with no conditioning variables is a correlation. With one conditioning variable, $\rho_{jk;m} = (\rho_{jk} - \rho_{jm}\rho_{km})/\sqrt{(1-\rho_{jm}^2)(1-\rho_{km}^2)}$. There are recursive equations for higher order partial correlations in terms of those of one fewer conditioning variable.

2.1. Markov trees and truncated vines

We first consider models without latent variables, that is, where $\boldsymbol{\delta}_j = \mathbf{0}$ for all $j = 1, \ldots, d$.

Markov trees

A Markov tree is a simple choice to build a valid multivariate model in terms of $\boldsymbol{\varphi}_j$. In a Markov tree on d variables, $d-1$ of the $d(d-1)/2$ possible pairs are identified as the edges of a tree, an acyclic connected undirected graph, with d nodes corresponding to the variables. An exemplary Markov tree for five variables is shown in Figure 1. Its edges are $\{1,2\}$, $\{2,3\}$, $\{2,4\}$ and $\{1,5\}$.

Without loss of generality let the pairs in the Markov tree be $\{1,2\}$ and $\{j, k(j)\}$ for $j = 3, \ldots, d$, where $1 \le k(j) < j$ (in Figure 1: $k(3) = 2$, $k(4) = 2$, $k(5) = 1$). Model (2.1) becomes

$$\begin{aligned} Z_1 &= \psi_1 \varepsilon_1, \\ Z_2 &= \varphi_2 Z_1 + \psi_2 \varepsilon_2, \\ Z_j &= \varphi_j Z_{k(j)} + \psi_j \varepsilon_j, \quad j = 3, \ldots, d, \end{aligned} \tag{2.2}$$

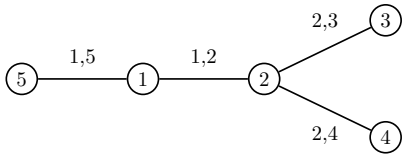

Figure 1: A five-dimensional Markov tree with edge indices.

where $\varphi_1 = 0$ and $\psi_j = \sqrt{1-\varphi_j^2}$ for all $j = 1, \ldots, d$. Clearly, it holds that $\varphi_j = \rho_{j,k(j)}$. For those pairs that are not selected in the Markov tree, conditional independence is assumed. For instance, if $k(3) = 2$ as in Figure 1, then $\varphi_3 = \rho_{32}$ and it assumed that $\rho_{31;2} = 0$, so that ρ_{31} is modeled as the product of ρ_{21} and ρ_{32}; and similarly for all other pairs by recursively exploiting the relationship

$$\rho_{j\ell} = \mathrm{Cov}(Z_j, Z_\ell) = \mathrm{Cov}(\varphi_j Z_{k(j)} + \psi_j \varepsilon_j, Z_\ell) = \varphi_j \rho_{k(j),\ell} = \rho_{j,k(j)} \rho_{k(j),\ell},$$

for $\ell \notin \{j, k(j)\}$.

In summary, a Markov tree model parameterizes the correlation matrix of $\boldsymbol{Z}$ in terms of $d-1$ parameters $\varphi_j,\ j = 2, \ldots, d$, or equivalently ρ_{12} and $\rho_{j,k(j)},\ j = 3, \ldots, d$, that is, we have a parameterization $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\varphi_2, \ldots, \varphi_d) = \boldsymbol{\Sigma}(\rho_{12}, \{\rho_{j,k(j)} : j = 3, \ldots, d\})$.

Truncated vines

To overcome the strict assumption of conditional independence of pairs that are not selected in the Markov tree, Bedford and Cooke (2002) proposed the graphical model which they called a vine. A vine is a sequence of linked trees, where the first tree is a Markov tree and subsequent trees specify conditional dependencies. In order to obtain a valid probability distribution from this graphical model, the notion of a d-dimensional regular vine has been defined in terms of $d-1$ trees $T_1, \ldots, T_{d-1}$ as follows.

(i) T_1 has nodes $N_1 = \{1, \ldots, d\}$ and edges E_1.
(ii) Tree T_i has nodes $N_i = E_{i-1}$ for $i = 1, \ldots, d-1$.
(iii) Two edges in tree T_i are joined in tree T_{i+1} only if they share a common node in tree T_i (proximity condition).

A five-dimensional example is shown in Figure 2. The proximity condition may be easily verified in this example. For instance, in tree T_2 one finds the

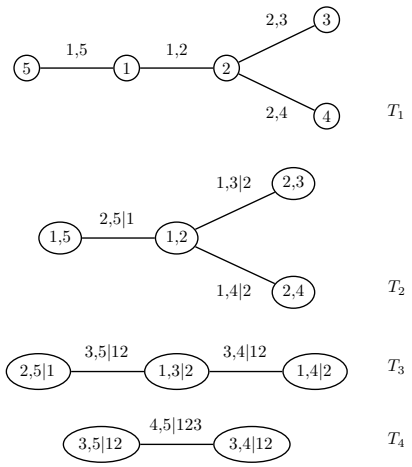

Figure 2: A five-dimensional regular vine with edge indices.

edge $\{\{1,5\},\{1,2\}\} =: \{2,5|1\}$ which satisfies the proximity condition, since $\{1,5\}$ and $\{1,2\}$ share node 1 in tree T_1.

Vines are now widely used in the copula literature to build dependence models with flexible bivariate components (see Kurowicka and Joe (2011) for a recent overview). In the context of multivariate normal distributions it has been shown in Kurowicka and Cooke (2003) that when each of the $d(d-1)/2$ edges of a regular vine is identified by a (partial) correlation in $(-1,1)$, then the resulting correlation matrix from this parameterization is positive definite. Without loss of generality let $\{1,2\}$ and $\{j,k_1(j)\},\ j=3,\ldots,d$ be the pairs identified by T_1, where $k_1(j) < j$. Then T_2 identifies pairs $\{\{j,k_1(j)\},\{k_2(j),k_1(j)\}\} =: \{j,k_2(j)|k_1(j)\}$ for $j=3,\ldots,d$ with $k_2(j) < j$, which correspond to partial correlations $\rho_{j,k_2(j);k_1(j)}$ (in Figure 2: $k_1(3)=2$, $k_2(3)=1$, $k_1(4)=2$, $k_2(4)=1$, $k_1(5)=1$, $k_2(5)=2$). This can be iterated up to T_{d-1}. For $i=1,\ldots,d-1$, in tree T_i, there are $i-1$ conditioning variables for each edge, and there are $d-i+1$ nodes and $d-i$ edges. The total number of edges over all trees is $d(d-1)/2$ and each pair (j,k) appears exactly once as conditioned variables (and the partial correlation has form $\rho_{j,k;m_1,\ldots,m_{i-1}}$ if j,k are paired conditioned variables in tree i). If the sequence is stopped at a specific level $\ell \leq d-1$, the vine is called ℓ-truncated and remaining partial correlations are assumed to be 0 (see Brechmann et al. (2012)).

For the case of a 2-truncated vine in the form of Model (2.1):

$$\begin{aligned} Z_1 &= \psi_1 \varepsilon_1, \\ Z_2 &= \varphi_{21} Z_1 + \psi_2 \varepsilon_2, \\ Z_j &= \varphi_{j1} Z_{k_1(j)} + \varphi_{j2} Z_{k_2(j)} + \psi_j \varepsilon_j, \quad j = 3, \ldots, d, \end{aligned} \tag{2.3}$$

where $\varphi_{11} = \varphi_{12} = \varphi_{22} = 0$ and $\psi_j = \sqrt{1 - \varphi_{j1}^2 - \varphi_{j2}^2 - 2\rho_{k_1(j),k_2(j)}\varphi_{j1}\varphi_{j2}}$. The extension to the ℓ-level case is straightforward, while the Markov tree model (2.2) corresponds to a 1-truncated vine.

The parameters φ_{j1} and φ_{j2} can be parameterized in terms of correlations $\rho_{j,k_1(j)}$ and partial correlations $\rho_{j,k_2(j);k_1(j)}$. For this we temporarily abbreviate $k_1(j)$ to k_1 and $k_2(j)$ to k_2 in order to improve readability. It holds that

$$\begin{aligned} &\rho_{jk_1} = \mathrm{Cov}(Z_j, Z_{k_1}) = \varphi_{j1} + \varphi_{j2}\rho_{k_1k_2}, \\ &1 = \mathrm{Var}(Z_j) = \varphi_{j1}^2 + \varphi_{j2}^2 + 2\varphi_{j1}\varphi_{j2}\rho_{k_1k_2} + \psi_j^2, \\ &\rho_{jk_2;k_1}\sqrt{(1-\rho_{jk_1}^2)(1-\rho_{k_1k_2}^2)} = \mathrm{Cov}(Z_j, Z_{k_2}|Z_{k_1}) = \varphi_{j2}\mathrm{Var}(Z_{k_2}|Z_{k_1}) = \varphi_{j2}(1-\rho_{k_1k_2}^2), \end{aligned}$$

where the right-hand sides follow by plugging in the model equations (2.3). Solving for $\varphi_{j1}, \varphi_{j2}$ and ψ_j leads to:

$$\begin{aligned} \varphi_{j2} &= \rho_{jk_2;k_1}\sqrt{(1-\rho_{jk_1}^2)/(1-\rho_{k_1k_2}^2)}\,, \\ \varphi_{j1} &= \rho_{jk_1} - \varphi_{j2}\rho_{k_1k_2}, \\ \psi_j^2 &= 1 - \varphi_{j1}^2 - \varphi_{j2}^2 - 2\varphi_{j1}\varphi_{j2}\rho_{k_1k_2} = (1-\rho_{jk_1}^2)(1-\rho_{jk_2;k_1}^2). \end{aligned} \tag{2.4}$$

Note that this includes autoregressive time series models of order 2 as a special case: For $j = 3, k_1 = 2$ and $k_2 = 1$ as well as $\rho_{12} = \rho_{23} = \rho$ and $\rho_{13;2} = \alpha$, it follows that $\varphi_{32} = \alpha$, $\varphi_{31} = \rho(1-\alpha)$ and $\psi_3 = (1-\rho^2)(1-\alpha^2)$, which matches results in books on times series analysis (see, e.g., Box et al. (2008)).

Similar to the Markov tree model (2.2), the full correlation matrix can be conveniently computed from this parameterization. While the $\rho_{jk_1(j)}$s are directly given from the model, the $\rho_{jk_2(j)}$s are obtained using

$$\rho_{jk_2(j)} = \mathrm{Cov}(Z_j, Z_{k_2(j)}) = \varphi_{j1}\rho_{k_1(j)k_2(j)} + \varphi_{j2}$$

and all other correlations by exploiting

$$\rho_{j\ell} = \mathrm{Cov}(Z_j, Z_\ell) = \varphi_{j1}\rho_{k_1(j)\ell} + \varphi_{j2}\rho_{k_2(j)\ell}, \quad \text{for } \ell \notin \{j, k_1(j), k_2(j)\}. \tag{2.5}$$

For instance, for $j = 3$ in Figure 2, we have $k_1(3) = 2$ and $k_2(3) = 1$, so that $\rho_{31} = \varphi_{31}\rho_{21} + \varphi_{32}$ and $\rho_{3\ell} = \varphi_{31}\rho_{2\ell} + \varphi_{32}\rho_{1\ell}$ for $\ell \in \{4, 5\}$; and similar expressions hold for $j = 4, 5$.

The truncated vine model (2.3) hence leads to a parameterization $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\varphi_{21}, \ldots, \varphi_{d1}, \varphi_{32}, \ldots, \varphi_{d2}) = \boldsymbol{\Sigma}(\rho_{12}, \{\rho_{jk_1(j)}, \rho_{jk_2(j);k_1(j)} : j = 3, \ldots, d\})$ of the correlation matrix. The number of parameters in the model is $2d - 3$, which corresponds to the number of edges in T_1 and T_2. In general, an ℓ-truncated vine model with $\ell \leq d - 1$ has $\sum_{i=1}^{\ell}(d - i) = \ell(d - (\ell + 1)/2)$ parameters which are fewer than in the unstructured model with $d(d - 1)/2$ parameters if $d > \ell + 1$.

2.2. Factor models

Classical factor models (see, e.g., Harman (1967), Lawley and Maxwell (1971), Johnson and Wichern (2002)) are a sub-class of the model formulation (2.1) when setting $\boldsymbol{\varphi}_j = \mathbf{0}$ for all $j = 1, \ldots, d$. That is, we have in matrix-vector notation

$$\boldsymbol{Z} = \Delta \boldsymbol{V} + \Psi \boldsymbol{\varepsilon}, \tag{2.6}$$

where $\Delta = (\delta_{ji})_{j=1,\ldots,d,\ i=1,\ldots,p}$ is called the loading matrix, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_d)'$ and Ψ is diagonal with entries $\Psi_{jj} = \psi_j = \sqrt{1 - \boldsymbol{\delta}_j'\boldsymbol{\delta}_j}$ for $j = 1, \ldots, d$. Then the correlation matrix of $\boldsymbol{Z}$ is given by

$$\boldsymbol{\Sigma}(\Delta) = \Delta\Delta' + \Psi^2. \tag{2.7}$$

This representation however does not uniquely identify the loading matrix Δ, since it is not invariant with respect to orthogonal transformations. Let $U \in \mathbb{R}^{p \times p}$ such that $UU' = I$. Then the loading matrix $\widetilde{\Delta} := \Delta U$ implies the same parameterization as in (2.7): $\widetilde{\Delta}\widetilde{\Delta}' = (\Delta U)(\Delta U)' = \Delta UU'\Delta' = \Delta\Delta'$, that is, $\boldsymbol{\Sigma}(\Delta) = \boldsymbol{\Sigma}(\widetilde{\Delta})$.

One possibility to ensure uniqueness (up to sign) of the loadings is to choose the orthogonal matrix U such that $\widetilde{\Delta}$ has zeros in the upper right triangle, $\delta_{ji} = 0$ for $1 \leq j < i \leq p$, and strictly positive diagonal entries, $\delta_{ii} > 0$ for all $i = 1, \ldots, p$ (Geweke and Zhou (1996)). These and alternative constraints imply that the factor model (2.6) has $\nu = \frac{1}{2}((d - p)^2 - (d + p))$ degrees of freedom. Requiring $\nu \geq 0$ leads to a bound on the maximum number of factors for given d. For instance, in three and four dimensions there can be at most one factor, in five dimensions at most two factors, while three factors are only possible if $d \geq 6$.

The above constraint then implies that the total number of parameters of the p-factor model (2.6) with d sufficiently large is $\sum_{i=1}^{p}(d-i+1) = p(d-(p-1)/2)$. A one-factor model therefore has d parameters, while a 2-factor model has $2d-1$.

2.3. Combined model

While the model classes in the preceding two subsections assume either direct between-variable dependence or latent variables, one may also combine both models. That is, we assume $\boldsymbol{\varphi}_j \neq \mathbf{0}$ and $\boldsymbol{\delta}_j \neq \mathbf{0}$ in Model (2.1). In the simplest case, we have one factor V and one neighbor $Z_{k(j)}$ of each variable $Z_j,\ j = 1, \ldots, d$, in a Markov tree, which corresponds to a 1-truncated vine model. For $d \geq 5$, Model (2.1) can therefore be stated as

$$\begin{aligned} Z_1 &= \delta_1 V + \psi_1 \varepsilon_1, \\ Z_2 &= \varphi_2 Z_1 + \delta_2 V + \psi_2 \varepsilon_2, \\ Z_j &= \varphi_j Z_{k(j)} + \delta_j V + \psi_j \varepsilon_j, \quad j = 3, \ldots, d, \end{aligned} \tag{2.8}$$

where, without loss of generality, $\{1, 2\}$ is an edge of the Markov tree and $\varphi_1 = 0$. Further, the correlations between Z_i and V are denoted by $\rho_{i,V}$, and $\psi_j = \sqrt{1 - \varphi_j^2 - \delta_j^2 - 2\varphi_j\delta_j\rho_{k(j),V}}$. Model (2.8) is similar to a 2-truncated vine, where $Z_{k_1(j)} = V$ and $Z_{k_2(j)} = Z_{k(j)}$.

This is illustrated in Figure 3, which shows two graphical representations of the model with one latent variable and $d = 5$ observed variables. In a psychological context, V is a general latent variable for an instrument and there is some conditional dependence in the items after conditioning on V; items 2, 4 and 5 are conditionally linked to item 1, and item 3 has stronger conditional dependence with item 2. See the applications in Section 5 for more concrete meanings of the individual items. In a context of financial assets, V is a general economic variable, and there is some conditional dependence in the asset prices after conditioning on V; some pairs of assets may have stronger conditional dependence if they are in the same sector. Section 5 also has a financial application.

From an interpretation point of view, this is a good model to consider when the 1-factor model has moderate to large absolute loadings and there are some significant deviations of the observed correlation matrix and the fitted correlation matrix based on the 1-factor structure. The 'residual' deviations might be explained with some partial correlations (conditioned on

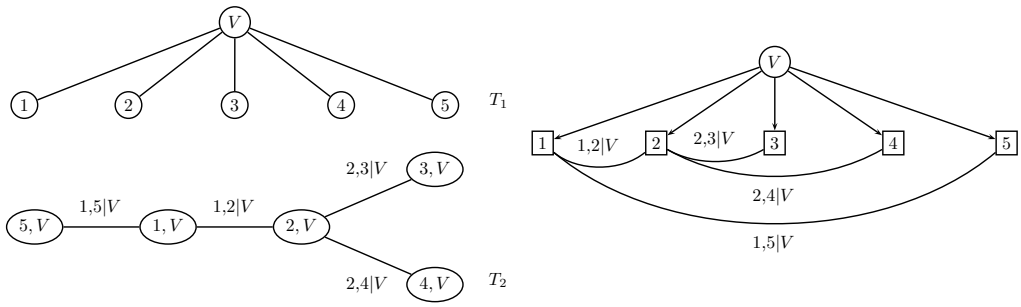

Figure 3: A five-dimensional combined 1-factor and 1-truncated vine model represented as a 2-truncated vine (left) and as a structural equation model with residual dependence (right; see Section 2.4).

the latent variable) that are smaller than the correlations with the latent variable. Then the dependence in the model is explained as being based on a latent variable with some additional residual dependence.

Using short-hand notation k for $k(j)$, and $\rho_{jk;V}$ as the partial correlation of variables j, k given V, we get similarly for the 2-truncated vine model (2.3) that

$$\begin{aligned}
&\rho_{j,V} = \mathrm{Cov}(Z_j, V) = \varphi_j \rho_{k,V} + \delta_j, \\
&1 = \mathrm{Var}(Z_j) = \varphi_j^2 + \delta_j^2 + 2\varphi_j \delta_j \rho_{k,V} + \psi_j^2, \\
&\rho_{jk;V} \sqrt{(1-\rho_{j,V}^2)(1-\rho_{k,V}^2)} = \mathrm{Cov}(Z_j, Z_k|V) = \varphi_j \mathrm{Var}(Z_k|V) = \varphi_j (1-\rho_{j,V}^2);
\end{aligned}$$

this has the loading parameters δ_j as well as the between-variable parameters φ_j in terms of correlations and partial correlations as

$$\begin{aligned}
\varphi_j &= \rho_{jk;V} \sqrt{(1-\rho_{j,V}^2)/(1-\rho_{k,V}^2)}\,, \\
\delta_j &= \rho_{j,V} - \varphi_j \rho_{k,V}, \\
\psi_j^2 &= 1 - \varphi_j^2 - \delta_j^2 - 2\varphi_j \delta_j \rho_{k,V} = (1-\rho_{j,V}^2)(1-\rho_{jk;V}^2).
\end{aligned}$$

Assuming that correlations $\rho_{k,V}$ are known, the full correlation matrix may then be obtained as well by interpreting Model (2.8) as a 2-truncated vine.

We thus have a correlation matrix parameterization $\mathbf{\Sigma} = \mathbf{\Sigma}(\delta_1, \ldots, \delta_d, \varphi_2, \ldots, \varphi_d) = \mathbf{\Sigma}(\rho_{1,V}, \ldots, \rho_{d,V}, \{\rho_{jk(j);V}, j = 2, \ldots, d\})$ in terms of $2d-1$ parameters, where $d \geq 5$. Extensions to additional factors and more vine levels are straightforward using the models of Sections 2.1 and 2.2.

Somewhat related to factor models, there are near non-identifiability issues in that different versions of the linear representation (2.8) can lead to similar correlation matrices. This is because the edges in the tree for residual dependence can vary for the smaller residual correlations and lead to roughly the same correlation matrix; larger residual correlations are less sensitive to tree structure.

2.4. Relationship to structural equation models

The general model formulation (2.1) can be considered as a structural equation model (SEM). These models are commonly used to specify and assess causal relationships between observed and unobserved quantities of data. They are particularly popular in psychometrics, where they are utilized to build meaningful models around and including factor models. A standard reference for SEMs is Bollen (1989), see also Steiger (2001),

However, for observational data, we are using these models to get parsimonious structures and not trying to infer causation. Latent variable models are reasonable in many applications as an explanation of the dependence in the observed variables, and we want some flexibility in the dependence conditioned on the latent variables. Without experimental data, we do not have a specific structure in mind, but a class of plausible parsimonious structures.

As noted in the introduction, our main interest in the representation given in (2.1) is to derive models that lead to correlation matrices that are based on $\mathcal{O}(d)$ parameters. The parameters can be in the form of regression coefficients or partial correlations. SEMs are more general and not all can be converted to the form of a truncated vine. While truncated vines guarantee that the model is well-defined, in general one has to be careful to come up with identifiable SEMs.

3. Estimation and model selection

In this section we discuss how the models presented in the previous section can be estimated for data $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{id})'$, $i = 1, \ldots, n$, which are the realizations of a multivariate normal random vector with means of 0 and variances of 1. We assume that the observations of each variable have been standardized to have mean 0 and variance 1; that is, $\frac{1}{n}\sum_{i=1}^n z_{ij} = 0$ and $\frac{1}{n}\sum_{i=1}^n z_{ij}^2 = 1$ for $j = 1, \ldots, d$. Further, let $\boldsymbol{R} = (r_{jk})_{j,k=1,\ldots,d}$ be the sample correlation matrix with entries $r_{jk} = \frac{1}{n}\sum_{i=1}^n z_{ij} z_{ik}$. These assumptions will

make the log-likelihoods at the maximum likelihood estimate (MLE) have a simpler (interpretable) form.

The major challenges of the model estimation are twofold.

(i) The latent factors in Models (2.6) and (2.8) are unobserved.
(ii) The Markov and vine tree structures in Models (2.2), (2.3) and (2.8) are unknown.

If the vine for residual dependence is completely specified, existing methods for SEMs can be used. In the following, it is shown how to deal with both challenges. Also the goodness-of-fit of the fitted correlation matrix is discussed.

3.1. Markov trees and truncated vines

Although the Markov tree model (2.2) corresponds to a 1-truncated vine model, we discuss estimation of both models separately, since the case of more than one vine tree is considerably more complicated, while the 1-level case is quite illustrative. For both models we use the parameterization in terms of (partial) correlations, since they do not impose any restrictions on the parameters to ensure positive definiteness of the correlation matrix Kurowicka and Cooke (2003).

Markov trees

Based on a product of conditional regression-type densities $f_{Z_j|Z_{k(j)}}$, $j \geq 2$,

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = f_{Z_1}(z_1) \prod_{j=2}^{d} f_{Z_j|Z_{k(j)}}(z_j|z_{k(j)}), \quad \boldsymbol{z} = (z_1, \ldots, z_d)',$$

the log-likelihood of the Markov tree model (2.2) with correlation matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\rho_{j,k(j)}, j = 2, \ldots, d)$ has the form

$$\begin{aligned} L(\rho_{j,k(j)}, j = 2, \ldots, d) &= -\frac{n}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^{n} \boldsymbol{z}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{z}_i \\ &= -\frac{n}{2} \sum_{j=2}^{d} \log(1 - \rho_{j,k(j)}^2) - \frac{1}{2} \Big\{ \sum_{i=1}^{n} z_{i1}^2 + \sum_{j=2}^{d} \sum_{i=1}^{n} \frac{(z_{ij} - \rho_{j,k(j)} z_{i,k(j)})^2}{(1 - \rho_{j,k(j)}^2)} \Big\}, \end{aligned} \tag{3.1}$$

where $|\cdot|$ denotes the matrix determinant and constant terms with 2π have been omitted.

From the above, one could get $\mathbf{\Sigma}^{-1}$ in closed form, but that is not needed to get the MLE of the $\rho_{j,k(j)}$s. Maximizing the above reduces to $d-1$ separate one-parameter optimizations of the form

$$L_j(\rho) := -\frac{n}{2}\log(1-\rho^2) - \frac{1}{2}\sum_{i=1}^{n}(z_{ij} - \rho z_{i,k(j)})^2/(1-\rho^2), \quad \rho = \rho_{j,k(j)}.$$

Using that $1 = \frac{1}{n}\sum_{i=1}^{n} z_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n} z_{i,k(j)}^2$ and the form of the sample correlation $r_{j,k(j)}$, this simplifies to

$$L_j(\rho) = -\frac{n}{2}\log(1-\rho^2) - \frac{n}{2}(1+\rho^2 - 2\rho r_{j,k(j)})/(1-\rho^2)$$

with derivative

$$L_j'(\rho) = \frac{n\rho}{1-\rho^2} - \frac{n(\rho - r_{j,k(j)})}{(1-\rho^2)} - \frac{\rho(1+\rho^2-2\rho r_{j,k(j)})}{(1-\rho^2)^2}.$$

Solving $L_j'(\widehat{\rho}) = 0$ for $\widehat{\rho}$ leads to $\widehat{\rho} = r_{j,k(j)}$. Hence, the MLE of (3.1) is $\widehat{\rho}_{j,k(j)} = r_{j,k(j)}$ for $j = 2, \ldots, d$. This however still leaves us with the problem of selecting the optimal Markov tree with edges $\{j, k(j)\}$. This can be solved as follows.

Substitution of $\widehat{\rho}_{j,k(j)} = r_{j,k(j)}$ back into the log-likelihood (3.1) leads to

$$\begin{aligned} &L(\widehat{\rho}_{j,k(j)}, j = 2, \ldots, d) \\ &= -\frac{n}{2}\sum_{j=2}^{d}\log(1-r_{j,k(j)}^2) - \frac{n}{2}\Big\{1 + \sum_{j=2}^{d}(1 + r_{j,k(j)}^2 - 2r_{j,k(j)}^2)/(1-r_{j,k(j)}^2)\Big\} \\ &= -\frac{n}{2}\sum_{j=2}^{d}\log(1-r_{j,k(j)}^2) - \frac{n}{2}d. \end{aligned} \tag{3.2}$$

Optimizing over all possible Markov trees therefore reduces to finding the tree T with $d-1$ edges $\{e = \{j,k\}\}$ that minimizes

$$\sum_{e=\{j,k\}\in T}\log(1-r_{jk}^2). \tag{3.3}$$

The optimal solution can easily be found using minimum spanning tree algorithms such as the one by Prim (1957), which is guaranteed to find the optimal solution when edge weights between nodes $1 \le k < j \le d$ are given

by $\log(1-r_{jk}^2)$. In other words, estimating Markov tree models (2.2) boils down to a simple model selection problem. Note that (3.3) is written in a way that does not depend on which variable is indexed as variable 1; the previous assumption of edges $\{1,2\}$ and $\{j,k(j)\}$ for $j=3,\ldots,d$ was made for notational convenience.

Truncated vines

As in the model definition we concentrate on the case of a 2-truncated vine model (2.3). The generalization to ℓ-truncated vines is briefly discussed subsequently.

For briefer notation we write $k_{1j} = k_1(j)$, $k_{2j} = k_2(j)$ and set $\alpha_j := \rho_{j,k_2(j);k_1(j)}$. Similar to above, the product of conditional regression-type densities $f_{Z_j|Z_{k_{1j}},Z_{k_{2j}}}$,

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = f_{Z_1}(z_1) f_{Z_2|Z_1}(z_1|z_2) \prod_{j=3}^{d} f_{Z_j|Z_{k_{1j}},Z_{k_{2j}}}(z_j|z_{k_{1j}}, z_{k_{2j}}), \quad \boldsymbol{z} = (z_1,\ldots,z_d)',$$

then leads to the log-likelihood of the 2-truncated vine model (2.3) with correlation matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\rho_{12}, \rho_{jk_{1j}}, \alpha_j, j=3,\ldots,d)$ as

$$\begin{aligned} L(\rho_{12}, \rho_{jk_{1j}}, \alpha_j, j=3,\ldots,d) &= -\frac{n}{2}\Big[\log(1-\rho_{12}^2) + \sum_{j=3}^{d} \log \psi_j^2\Big] \\ &- \frac{1}{2}\Big\{\sum_{i=1}^{n} z_{i1}^2 + \sum_{i=1}^{n} \frac{(z_{i2}-\rho_{12}z_{i1})^2}{(1-\rho_{12}^2)} + \sum_{j=3}^{d}\sum_{i=1}^{n} \frac{(z_{ij}-\varphi_{j1}z_{i,k_{1j}}-\varphi_{j2}z_{i,k_{2j}})^2}{\psi_j^2}\Big\}. \end{aligned} \tag{3.4}$$

After estimating $\widehat{\rho}_{12} = r_{12}$ as in the Markov model (2.2), this can be split into sequential log-likelihoods for $\varphi_{j1}, \varphi_{j2}$, $j=3,\ldots,d$. For the conditional density $f_{Z_j|Z_{k_{1j}},Z_{k_{2j}}}$, $\rho_{k_{1j}k_{2j}}$ is estimated from a previous stage, and then estimates of φ_{j1} and φ_{j2}, or equivalently estimates of $\rho_{jk_{1j}}$ and α_j, are obtained.

That is, consider a log-likelihood of the form

$$L_j(\varphi_1,\varphi_2) := -\frac{n}{2}\log\psi_j^2 - \frac{1}{2}\sum_{i=1}^{n}(z_{ij}-\varphi_1 z_{ik_{1j}} - \varphi_2 z_{ik_{2j}})^2/\psi_j^2 \tag{3.5}$$

where $\psi_j = 1-\varphi_{j1}^2-\varphi_{j2}^2-2\varphi_{j1}\varphi_{j2}\rho$, $\varphi_1=\varphi_{j1}$ and $\varphi_2=\varphi_{j2}$, and $\widehat{\rho} = r_{k_{1j}k_{2j}}$ is estimated from earlier in the sequence. The MLE of $\rho_{jk_{1j}}$ and α_j can then

be obtained after some calculus (see Appendix A) as $\widehat{\rho}_{jk_{1j}} = r_{jk_{1j}}$ and

$$\widehat{\alpha}_j = \frac{r_{jk_{2j}} - r_{jk_{1j}} r_{k_{1j}k_{2j}}}{\sqrt{(1 - r^2_{jk_{1j}})(1 - r^2_{k_{1j}k_{2j}})}} =: r_{jk_{2j};k_{1j}}, \tag{3.6}$$

which is the sample partial correlation. This also implies that $\widehat{\rho}_{jk_{2j}} = r_{jk_{2j}}$.

In the next step, an optimal 2-truncated vine tree structure needs to be determined. For this, substitute $\widehat{\rho}_{jk_{1j}} = r_{jk_{1j}}$ and $\widehat{\alpha}_j = r_{jk_{2j};k_{1j}}$ into (3.4) to obtain

$$\begin{aligned} L(\widehat{\rho}_{12}, \widehat{\rho}_{jk_{1j}}, \widehat{\alpha}_j, j = 3, \ldots, d) &= -\frac{n}{2}\Big[\log(1 - \widehat{\rho}^2_{12}) + \sum_{j=3}^{d} \log \widehat{\psi}^2_j\Big] - \frac{n}{2} d \\ &= -\frac{n}{2} d - \frac{n}{2} \log(1 - r^2_{12}) - \frac{n}{2} \sum_{j=3}^{d} \log(1 - r^2_{jk_{1j}}) - \frac{n}{2} \sum_{j=3}^{d} \log(1 - r^2_{jk_{2j};k_{1j}}), \end{aligned} \tag{3.7}$$

where the relationships (2.4) are exploited to express the MLE $\widehat{\psi}_j$ of ψ_j in terms of $\widehat{\rho}_{jk_{1j}}$ and $\widehat{\alpha}_j$.

Hence, optimizing over all possible 2-truncated vine tree structure amounts to finding trees T_1 and T_2 which minimize

$$\sum_{e=\{j,k\} \in T_1} \log(1 - r^2_{j,k}) + \sum_{e=\{j,m|k\} \in T_2} \log(1 - r^2_{jm;k}),$$

where sample partial correlations are obtained from the sample correlation matrix $\boldsymbol{R}$ as in (3.6).

The pattern of the 2-level case extends to ℓ-truncated vines by working with the regression formulation $Z_j | Z_{k_1(j)}, \ldots, Z_{k_\ell(j)}$. The optimal ℓ-truncated vine tree structure is then given by trees $T_1, \ldots, T_\ell$ that minimize

$$\sum_{e \in T_1, \ldots, T_\ell} \log(1 - r^2_e),$$

where r_e is a sample partial correlation for edge e in trees $T_2, \ldots, T_{d-1}$. Note that the log determinant of the correlation matrix $\boldsymbol{\Sigma}$ can be expressed using a partial correlation vine without any truncation as $\sum_{e \in T_1, \ldots, T_{d-1}} \log(1 - \rho^2_e)$, where ρ_e is a partial correlation for edge e in trees $T_2, \ldots, T_{d-1}$ (Kurowicka

and Cooke (2006)). With truncation at level ℓ, the higher order partial correlations are set to 0, and the log determinant is $\sum_{e \in T_1, \dots, T_\ell} \log(1 - \rho_e^2)$.

As for Markov tree models, estimation of truncated vine models corresponds to the model selection problem of choosing the right trees. While for Markov tree models the optimal tree structure can simply be selected using a minimum spanning tree algorithm, the sequential selection of trees in a truncated vine model with more than one level is more complicated, since the set of edges that can be used to construct a tree depends on the previous tree according to the construction principles of regular vines. For instance, in the case of a 2-truncated vine, the selection of minimum spanning trees in terms of edge weights $\log(1 - r_e^2)$ for both trees may not lead to the optimal solution. While the minimum spanning tree is the optimal choice in the second tree given the first tree, the minimum spanning tree may not be the overall best choice for the first tree. The question therefore is how to identify the overall best—or, at least, a near best—solution.

By Cayley's theorem, the search space of 2-truncated vine structures on d variables consists of the d^{d-2} possible first trees, since the second tree is optimally determined as a minimum spanning tree given the first tree. Other than for small d, this search space is too large to evaluate all possible models and then select the best one. We therefore propose two heuristic procedures to better exploit this search space; these two procedures with neighbors of trees and best spanning trees were not considered in Kurowicka (2011) and Brechmann et al. (2012).

Neighbors of trees. Although it is only a locally optimal solution, a minimum spanning tree may still be seen as a reasonable starting point for the search of a better solution. One approach therefore is to look at neighbors of the minimum spanning tree. We define a 1-neighbor of a tree T as follows:

(i) Choose an edge $e \notin T$ and set $T' = T \cup \{e\}$. Then the graph T' is no longer a tree but has a cycle C of edges including e.

(ii) Remove an edge $f \in C \setminus \{e\}$ from T' to obtain $T'' = T' \setminus \{f\}$. T'' is a tree and called a 1-neighbor of T.

In a similar way, we can define m-neighbors, where $m \in \mathbb{N}$ edges are added to a tree.

As an example for a 1-neighbor, consider T_1 in Figure 2. If one adds the edge $\{3, 4\}$, there is a cycle $C = \{\{2, 3\}, \{2, 4\}, \{3, 4\}\}$. By removing for example the edge $\{2, 3\}$ one obtains a valid tree again.

For the selection of 2-truncated vines starting from the minimum spanning tree T, we propose to try out a number of edges e such that they are the edges with the smallest weights $\log(1 - r_e^2)$ that are not included in T. To obtain a tree with large overall weight, we then select f as the edge with the largest weight in $C \setminus \{e\}$. There are $d(d-1)/2 - (d-1) = (d-1)(d-2)/2$ 1-neighbors which may be constructed in this way. A similar approach has been used by Gruber and Czado (2013) in the context of Bayesian model selection of vines.

Best spanning trees. A straightforward extension to go beyond the minimum spanning tree is to take into account a range of K best spanning trees, that is, the K spanning trees with smallest weight. The problem of identifying these K best spanning trees is however known to be NP-hard. Several algorithms have been described in the literature to optimize the running time and we implement the one by Gabow (1977). Note that there may of course be 1-neighbors of the minimum spanning tree among the K best spanning trees but this is not necessarily so.

If an ℓ-truncated vine structure rather than a 2-level one is built, then again only the ℓth tree is selected best as a minimum spanning tree given the first $\ell - 1$ trees and the sequential one-tree-at-a-time selection of minimum spanning trees is very likely to not lead to the best fit. The search space now consist of all $(\ell - 1)$-truncated regular vines and can clearly not be explored entirely. Combinations of the above two approaches may be used to find a reasonably good solution.

3.2. Factor models

There is a rich literature on factor models, which describes how to estimate the factor loadings. We will therefore not go into details here. Common approaches to estimation are direct MLE or the use of the iterative EM algorithm by Dempster et al. (1977). An EM algorithm for factor analysis can for example be found in Lange (2010).

3.3. Combined factor-vine model

Unlike the factor model (2.6), the factor-vine model (2.8) cannot be estimated by direct MLE, since not only the factors are unobserved but the tree structure of the truncated vine is unknown, too. We therefore propose

a modified EM algorithm which can deal with both issues for the case of one factor and a 1-truncated vine.

Clearly, $\boldsymbol{Z} = (Z_1, \ldots, Z_d)'$ and V are jointly multivariate normal with correlation matrix given by

$$\widetilde{\boldsymbol{\Sigma}} = \begin{pmatrix} 1 & \boldsymbol{\gamma}' \\ \boldsymbol{\gamma} & \boldsymbol{\Sigma} \end{pmatrix},$$

where $\boldsymbol{\gamma} := \boldsymbol{\rho}_V = (\rho_{1,V}, \ldots, \rho_{d,V})'$ is unknown, since V is unobserved. The observed correlation matrix of the data $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{id})'$, $i = 1, \ldots, n$, is denoted by $\boldsymbol{R}$ as before.

The modified EM algorithm will then proceed as follows: First, we obtain the expected value of V given the observations $\boldsymbol{z}_i$ of $\boldsymbol{Z}$ (E-step). Then we first maximize this expected value with respect to $\boldsymbol{\gamma}$ to obtain a new estimate of the factor loadings (first part of the M-step). In the second part of the M-step the 1-truncated vine structure is estimated given the new estimate of the factor loadings.

Let $\boldsymbol{\gamma}_0$ be the current estimate of the factor loadings and $\boldsymbol{\Sigma}_0$ that of the correlation matrix in terms of the 1-truncated vine. Then

$$[V|\boldsymbol{Z} = \boldsymbol{z}] \sim N(\boldsymbol{\gamma}_0'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{z}, 1 - \boldsymbol{\gamma}_0'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\gamma}_0).$$

Assume that we observe "complete" data $(V_i, \boldsymbol{z}_i')'$, $i = 1, \ldots, n$. Then n^{-1} times the log-likelihood is

$$-\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{R}) - \frac{1}{2}\log(1 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}) - \frac{1}{2n}\sum_{i=1}^{n}\frac{(V_i - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_i)^2}{(1 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma})}, \quad (3.8)$$

where constant terms with 2π have been ignored and $\mathrm{tr}(\cdot)$ is the matrix trace, that is, the sum of diagonal elements. Taking the expectation of (3.8) for V_i

random given $\boldsymbol{z}_i$ fixed for $i = 1, \ldots, n$ then leads to

$$\begin{aligned} Q(\boldsymbol{\gamma}, \boldsymbol{\Sigma} | \boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_0) = & -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{R}) - \frac{1}{2}\log(1 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}) \\ & - \frac{1}{2n}\sum_{i=1}^{n}(\boldsymbol{\gamma}_0'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{z}_i - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_i)^2/(1 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}) \\ & - \frac{1}{2}(1 - \boldsymbol{\gamma}_0'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\gamma}_0)/(1 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}) \\ = & -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{R}) - \frac{1}{2}\log(1 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}) \\ & - \frac{1}{2}(\boldsymbol{\gamma}_0'\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1})\boldsymbol{R}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\gamma}_0 - \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma})/(1 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}) \\ & - \frac{1}{2}(1 - \boldsymbol{\gamma}_0'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\gamma}_0)/(1 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}). \end{aligned} \tag{3.9}$$

If one formally applies the M-step to (3.9), then (3.9) is maximized with respect to the parameters. As $\boldsymbol{\Sigma}$ depends on $\boldsymbol{\gamma}$ we split the M-step into two parts. First we obtain a new estimate $\boldsymbol{\gamma}_1$ of $\boldsymbol{\gamma}$ pretending $\boldsymbol{\Sigma}$ does not depend on $\boldsymbol{\gamma}$ and then we update the correlation matrix of $\boldsymbol{Z}$ by $\boldsymbol{\Sigma}_1$. This strategy is inspired by the Expectation Conditional Maximization (ECM) algorithm proposed by Meng and Rubin (1993), which is a generalized version of the standard EM algorithm and guarantees an increase in the model likelihood.

If we pretend $\boldsymbol{\Sigma}$ to be constant and independent of $\boldsymbol{\gamma}$, then taking the derivative of (3.9) with respect to $\boldsymbol{\gamma}$ and setting to $\mathbf{0}$ to get the root $\boldsymbol{\gamma}_1$ leads to

$$\begin{aligned} \mathbf{0} = & \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1}{1 - \boldsymbol{\gamma}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1} - \frac{(1 - \boldsymbol{\gamma}_0'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\gamma}_0)\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1}{(1 - \boldsymbol{\gamma}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1)^2} \\ & - \frac{(\boldsymbol{\gamma}_0'\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\gamma}_1'\boldsymbol{\Sigma}^{-1})\boldsymbol{R}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\gamma}_0 - \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1)\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1}{(1 - \boldsymbol{\gamma}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1)^2} \\ & + \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{R}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\gamma}_0 - \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1)}{1 - \boldsymbol{\gamma}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1}. \end{aligned} \tag{3.10}$$

This can be solved by the Newton-Raphson method, with $\boldsymbol{\Sigma}$ replaced by $\boldsymbol{R}$ as it is unknown.

An alternative estimate $\boldsymbol{\gamma}_1$ is given by the regression-type moment estimator

$$\boldsymbol{\gamma}_1 = \boldsymbol{R}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\gamma}_0/\sigma_0^2, \quad \sigma_0^2 = \boldsymbol{\gamma}_0'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{R}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\gamma}_0 + 1 - \boldsymbol{\gamma}_0'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\gamma}_0, \tag{3.11}$$

where the numerator is an estimate of the covariance of $\boldsymbol{Z}$ and V and the denominator estimates the variance of V. In numerical examples, the use of this moment estimator resulted in a slightly slower convergence of the EM algorithm. This is however balanced in running time by a faster computation of the M-step as no iterative Newton-Raphson method is required.

Using the moment estimator (3.11) it can be verified that

$$\begin{aligned} & Q(\boldsymbol{\gamma}_1, \boldsymbol{\Sigma} | \boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_0) \\ &= -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\log\left(1 - \boldsymbol{\gamma}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1\right) - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{R}\right) \\ &\quad - \frac{1}{2}\frac{\boldsymbol{\gamma}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{R}\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1}{1 - \boldsymbol{\gamma}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1} - \frac{1}{2}\frac{\sigma_0^2(1 - 2\boldsymbol{\gamma}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1)}{1 - \boldsymbol{\gamma}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1} \\ &= -\frac{1}{2}\log|\boldsymbol{\Sigma} - \boldsymbol{\gamma}_1\boldsymbol{\gamma}_1'| - \frac{1}{2}\mathrm{tr}\left((\boldsymbol{\Sigma} - \boldsymbol{\gamma}_1\boldsymbol{\gamma}_1')^{-1}\boldsymbol{R}\right) - \frac{1}{2}\frac{\sigma_0^2(1 - 2\boldsymbol{\gamma}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1)}{1 - \boldsymbol{\gamma}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_1}, \end{aligned} \tag{3.12}$$

where we used that $|\boldsymbol{\Sigma}||1 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}| = |\widetilde{\boldsymbol{\Sigma}}| = |\boldsymbol{\Sigma} - \boldsymbol{\gamma}\boldsymbol{\gamma}'|$ and $(\boldsymbol{\Sigma} - \boldsymbol{\gamma}\boldsymbol{\gamma}')^{-1} = \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}/(1 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma})$.

The dominant term of $Q(\boldsymbol{\gamma}_1, \boldsymbol{\Sigma}|\boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_0)$ comes from the log determinant term because the other terms approach a constant near $-(d+1)/2$, with $-\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{R}\right)/2$ becoming close to $-d/2$, while the two terms in (3.12) approach $-1/2$. For a 2-truncated vine (Section 2.1),

$$|\widetilde{\boldsymbol{\Sigma}}| = |\boldsymbol{\Sigma} - \boldsymbol{\gamma}\boldsymbol{\gamma}'| = \prod_{j=1}^{d}(1 - \gamma_j^2)\prod_{j=2}^{d}(1 - \rho_{j,k(j);V}^2).$$

So, with $\boldsymbol{\gamma}$ fixed at the current estimate $\boldsymbol{\gamma}_1$, we can therefore approximately maximize $Q(\boldsymbol{\gamma}_1, \boldsymbol{\Sigma}|\boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_0)$ in terms of $\boldsymbol{\Sigma}$, which is given as a 1-truncated vine (or Markov tree), by minimizing

$$\sum_{e=\{j,k;V\}\in T} \log(1 - r_{jk;V}^2),$$

where $r_{jk;V} = (r_{jk} - \gamma_{j1}\gamma_{k1})/\sqrt{(1 - \gamma_{j1}^2)(1 - \gamma_{k1}^2)}$ is the current estimate of the partial correlation of Z_j and Z_k given V using $\boldsymbol{\gamma}_1 = (\gamma_{11}, \ldots, \gamma_{d1})'$. The solution is given again as a minimum spanning tree in terms of weights $\log(1 - r_{jk;V}^2)$, which then yields a new estimate $\boldsymbol{\Sigma}_1$ of $\boldsymbol{\Sigma}$ by computing the correlation matrix implied by a 2-truncated vine as described in Section 2.1.

By iterating these steps, a sequence of estimates for $\boldsymbol{\gamma}$ and, most importantly, for $\boldsymbol{\Sigma}$ is obtained. Convergence of the modified EM algorithm can be assumed for example when the change in estimates between two steps falls below a pre-specified threshold or after a certain number of iterations. The problem at hand—parameterization of correlation matrices—however provides a more effective way of convergence control, since the optimal solution is known. For the unstructured correlation matrix, the MLE of $\boldsymbol{\Sigma}$ is $\boldsymbol{R}$ and ignoring the term with 2π, the log-likelihood is

$$-\frac{n}{2}\log|\boldsymbol{R}| - \frac{n}{2}d. \tag{3.13}$$

The algorithm can therefore be stopped, when no more significant improvement with respect to this upper bound is made; this is more reliable than the decision based on the relative change in parameters. Monitoring of the log-likelihood is especially important, since we use a modified EM algorithm with approximate M-step. Convergence of the ignored part in $Q(\boldsymbol{\gamma}_1, \boldsymbol{\Sigma}|\boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_0)$ to the constant $-(d+1)/2$ was very fast in all numerical examples that we considered (typically within the first 10–50 iterations).

In general, the above algorithm will converge to a local minimum of the negative log-likelihood. Depending on the relative magnitudes of the correlations with the latent variable and partial correlations given the latent variable, the starting point for the modified EM algorithm has an effect. If the partial correlations tend to be smaller in magnitude than the correlations or loadings with the latent variable, then the model is a small variation of the 1-factor model, and loadings from the 1-factor correlation structure provide a good starting point. If some of the partial correlations are larger in magnitude, then several different starting points should be used.

For simulated correlation matrices from the combined factor-vine model, the modified EM algorithm works well, when the partial correlations are not too large, in that it tends to converge to a local minimum with loadings that are not far from the "true" loadings, with a Markov tree that can differ from the original. As d increases past 10, the modified EM algorithm can more often find the original tree. The reason for this is that for $d < 10$, it is easier to find perturbations of the loadings and the tree, and get approximately the same correlation matrix. The larger partial correlations in the residual dependence (as represented by the vine) are stable, but the vine structure for the smaller (more negligible) residual dependence can change with resampling, as it is mentioned at the end of Section 2.3. For example,

with $d = 5$ (the smallest dimension to consider the combined 1-factor and 1-truncated vine model), there are ten correlations and nine parameters, and the algorithm can more easily reach slightly perturbed correlation matrices from different starting points.

These ideas can be extended to more general combined factor-vine models. For a 1-factor model with residual dependence specified in terms of an ℓ-truncated vine, the general truncated vine algorithm described in Section 3.1 could be used in the modified EM iterations. For a 2-factor model (common factors, or factors with structured zeros) with residual dependence in one or more vine levels, the E-step is similar but details are more cumbersome. The extended models could be considered for larger d when $3d < d(d-1)/2$ (number of structured parameters is less than the total number of correlations).

3.4. Goodness-of-fit and model selection

Before moving on to discuss some examples, we like to note that the goodness-of-fit of a fitted structured correlation matrix can easily be assessed in terms of the following distance to the observed correlation matrix:

$$D(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})|\boldsymbol{R}) := \log|\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})| - \log|\boldsymbol{R}| + \mathrm{tr}(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\boldsymbol{R}) - \dim(\boldsymbol{R}) \geq 0, \quad (3.14)$$

where $\widehat{\boldsymbol{\theta}}$ is an estimate of the model parameters $\boldsymbol{\theta}$. (3.14) is $1/n$ times the likelihood ratio test statistic of two multivariate normal distributions with correlation matrices $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$ and $\boldsymbol{R}$, respectively (see, e.g., Kullback (1967)).

This distance D is common in practice and can conveniently be used for model selection, since it does not depend on an unknown population correlation matrix. It essentially assesses whether a structured correlation matrix with fewer parameters, say $\mathcal{O}(d)$, is an adequate approximation to the saturated sample correlation matrix $\boldsymbol{R}$ with $d(d-1)/2$ parameters. In our setting this means that, given a desired degree of closeness to the sample correlation matrix, models can be extended by adding additional vine trees and/or factors until this degree of closeness is attained.

To take into account the model complexity, often the Bayesian information criterion (BIC) is used for model selection. For a multivariate normal model with fitted structured correlation matrix, it is given by

$$\mathrm{BIC}(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})) := nd\log(2\pi) + n\log|\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})| + n\mathrm{tr}(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\boldsymbol{R}) + \log(n)\dim(\widehat{\boldsymbol{\theta}}), \quad (3.15)$$

where $\dim(\widehat{\boldsymbol{\theta}})$ denotes the number of model parameters in $\boldsymbol{\theta}$. When comparing different models with the BIC, the model with the smallest BIC value is considered to provide the best fit.

From (3.2) and (3.7) it follows that $\operatorname{tr}(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})^{-1}\boldsymbol{R}) = d = \dim(\boldsymbol{R})$ in the truncated vine model. Therefore D in (3.14) reduces to $D(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})|\boldsymbol{R}) = \log|\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})| - \log|\boldsymbol{R}|$, and the quantity $(\log|\boldsymbol{R}| - \log|\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})|)/\log|\boldsymbol{R}|$ may be used as a measure for the relative distance to the sample correlation matrix. Similarly, the BIC in (3.15) reduces to $\mathrm{BIC}(\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})) := nd[\log(2\pi)+1] + n\log|\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})| + \log(n)\dim(\widehat{\boldsymbol{\theta}})$.

4. Simulation study

Since direct MLE is unrealistic for truncated vines and combined factor-vine models due to the tree selection problem, we described heuristics to identify a good truncated vine structure in Section 3.1 as well as a modified EM algorithm for the combined case in Section 3.3. In order to validate the approaches, we conducted an extensive Monte Carlo study. For this, we randomly constructed correlation matrices in dimensions $d \in \{5, 7, 10, 15, 30\}$ according to either a 2-truncated vine structure, a 2-factor model or a combined 1-factor and 1-truncated vine model. To each correlation matrix, we then fitted 2-truncated vines with sequential one-tree-at-a-time selection of minimum spanning trees (2T/MST), with first tree selected from all possible 1-neighbors of the minimum spanning tree (2T/N), and with first tree selected from the 1000 best spanning trees (2T/bMST), as well as the combined 1-factor and 1-truncated vine model (1F/1T) and the 2-factor model (2F). The 2-truncated models have 7, 11, 17, 27 or 57 parameters, respectively; the combined and the factor models have 9, 13, 19, 29 or 59, respectively. The modified EM algorithm is run for 1000 iterations with Newton-Raphson method to update the loadings $\boldsymbol{\gamma}$ of the combined factor-vine model, where the loadings of the best fitting 1-factor model are used as the starting point. For comparison, the model is also fitted using different starting values: The loadings of the best fitting 1-factor model as starting values are followed by the loadings with random noise added, all values equal to 0.2 and three times completely random starting values. The results according to the best starting values are denoted by 1F/1T*.

Table 1 reports results for 1000 repetitions in each row. As evaluation criteria we consider the mean distance D (3.14) to the sample correlation matrix over all 1000 repetitions as well as the count of the distance D being

True	d	2T/ MST	2T/ N	2T/ bMST	1F/ 1T	1F/ 1T*	2F	2T/ MST	2T/ N	2T/ bMST	1F/ 1T	1F/ 1T*	2F
1F/1T	5	0.190	0.076	0.049	0.026	0.003	0.099	0	0	0	50	214	1
	7	0.508	0.273	0.166	0.107	0.020	0.479	0	0	0	48	202	0
	10	1.112	0.695	0.613	0.236	0.037	1.406	0	0	0	148	326	0
	15	2.145	1.550	1.638	0.214	0.022	3.558	0	0	0	403	507	0
	30	5.634	4.598	5.153	0.121	0.007	11.034	0	0	0	604	686	0
2T	5	0.306	0.059	0.000	0.028	0.001	0.207	182	472	1000	49	94	1
	7	0.705	0.211	0.015	0.169	0.072	0.859	37	144	745	0	0	0
	10	1.167	0.465	0.266	0.663	0.392	2.521	2	21	102	0	0	0
	15	2.033	1.024	1.097	1.692	1.255	6.540	0	0	5	0	0	0
	30	4.653	3.074	3.749	5.987	4.860	20.585	0	0	0	0	0	0
2F	5	0.092	0.025	0.015	0.005	0.001	0.002	0	0	0	6	17	904
	7	0.264	0.120	0.062	0.029	0.012	0.014	0	0	0	0	1	861
	10	0.628	0.331	0.335	0.086	0.037	0.037	0	0	0	2	5	828
	15	1.317	0.866	1.124	0.188	0.089	0.091	0	0	0	0	2	691
	30	3.271	2.592	3.189	0.458	0.208	0.320	0	0	0	2	3	517

Table 1: Results of the simulation study. Rows 2–6 shows results according to random factor-vine models, those for random 2-truncated vines are shown in rows 7–11 and those for random 2-factor models in rows 12–16. Columns 3–8: Mean distance D, as given in (3.14), to sample correlation matrices over all 1000 repetitions. Columns 9–14: Count of distance D being smaller than 10^{-8}.

smaller than 10^{-8}. If the correct model is identified, D should be equal to 0 but may be different from that due to numeric precision. In particular, in the combined factor-vine model the value of D for the correctly identified model depends on the estimation accuracy of the loadings $\boldsymbol{\gamma}$. Therefore we choose the value 10^{-8} as threshold. As mentioned in Section 3.3, when D is close to 0, and the dimension d is not much larger than 5, the solution found from the modified EM algorithm might be a correlation matrix very similar to the original but with some perturbations in the loadings $\boldsymbol{\gamma}$ and in the tree for residual dependence.

The results can be summarized as follows. The combined factor-vine model is well fitted using the modified EM algorithm, it seems to do better as the dimensions gets larger than 10 because near non-identifiability (different parameters leading to approximately the same correlation matrix) is less of an issue. On the other hand, the 2-truncated vine obviously does not well approximate the simulated correlation matrices. When the true model however is a 2-truncated vine, it becomes clear that sequential one-tree-

at-a-time selection of minimum spanning trees is not sufficient, especially in higher dimensions. The heuristics, which consider 1-neighbors and best spanning trees, strongly improve this sequential selection. The case $d = 30$ also illustrates that a large number of best spanning trees (at least larger than 1000) may be required to be taken into account in higher dimensions, since 1-neighbors, which explore the search space differently, here do a better job on average. Interestingly, the combined factor-vine model also does a reasonably good job to approximate the correlation matrices simulated according to 2-truncated vines. Moreover, when different starting values are considered for the combined model, the fit can be significantly improved. However, none of the tested choices of starting values uniformly produced better fits than the others. This means that several different starting values should be taken into account to verify the fit. Finally, the results indicate that the 2-factor model is incapable of approximating the correlation matrices simulated according to the truncated vine and combined factor-vine models. As a result, a factor model with a small number of factors cannot appropriately account for the presence of residual dependence; more factors and hence a less parsimonious model would be needed for this purpose. Conversely, correlation matrices generated according to a 2-factor model can rather well be approximated by combined 1-factor and 1-truncated vine models. This is not the case for 2-truncated vines. To summarize, the combined factor-vine model can better approximate either a factor model or a truncated vine model than vice versa.

5. Applications

To illustrate the different parameterizations of correlation matrices presented in the previous sections, we analyze three examples from psychology and finance. In both areas, factor models are common approaches to explain dependence among variables: Factor modeling in psychology is used to identify unobserved character traits of individuals. Relationships between observed and unobserved variables are then often expressed in terms of structural equation models (see, e.g., Bollen (1989)). In finance, asset prices and returns are driven by factors such as the general state of the economy or sectorial dependence, but it may be more common to model dependence without any latent structure. Our model framework allows for both approaches as well as for a novel combination of the two. We will compare the following 2- and 3-level models in each example:

(i) 2-truncated vine with sequential one-tree-at-a-time selection of minimum spanning trees (2T/MST);
(ii) 2-truncated vine with first tree selected from all possible 1-neighbors of the minimum spanning tree (2T/N);
(iii) 2-truncated vine with first tree selected from the 1000 best spanning trees (2T/bMST);
(iv) 3-truncated vine with sequential one-tree-at-a-time selection of minimum spanning trees (3T/MST);
(v) 3-truncated vine with first tree selected from all possible 1-neighbors of the minimum spanning tree and second tree as minimum spanning tree (3T/N);
(vi) 3-truncated vine with first tree selected from all possible 1-neighbors of the minimum spanning tree and second tree as first 1-neighbor of the minimum spanning tree (3T/N1);
(vii) combined 1-factor and 1-truncated vine model (1F/1T), where the modified EM algorithm of Section 3.3 is run for 10 000 iterations to ensure convergence, which is checked using convergence plots (see the second example for an illustration; the Newton-Raphson method is used to obtain the update of $\boldsymbol{\gamma}$);
(viii) 2-factor (2F) and 3-factor (3F) models estimated by direct maximum likelihood estimation.

We analyze two examples from psychology, each having nine variables. Then we investigate a parsimonious parameterization of the correlation matrix of the stock returns of all 30 Dow Jones constituents. Finally, a note on the computing time of the different methods is made.

For all three examples, it is interpretable from the context that the dependence comes predominantly from a latent variable. There is some deviation of the observed correlation matrix and the fitted correlation matrix based on the 1-factor structure. With the 2-factor structure, the deviation is somewhat reduced in the first example (Thurstone data) but not in the second example (Harman data) or third example (Dow Jones data). The 1F/1T correlation structure is a better fit than the 2-factor structure in all three examples—marginally better in the first and much better in the other two. In all three examples, we also check on the interpretability of the parameter estimates. The loadings for the 1F/1T model are similar to that of the 1-factor model and the partial correlations given the latent variables are somewhat smaller in magnitude, so that the interpretation of residual dependence is reasonable.

	1	2	3	4	5	6	7	8	9
1	-	0.43	0.60	0.53	0.63	0.38	0.61	0.56	0.59
2	0.48	-	0.41	0.41	0.43	0.59	0.43	0.50	0.40
3	0.62	0.40	-	0.50	0.63	0.36	0.58	0.52	0.53
4	0.52	0.40	0.47	-	0.51	0.44	0.51	0.46	0.45
5	0.62	0.41	0.63	0.52	-	0.37	0.59	0.60	0.76
6	0.42	0.59	0.36	0.44	0.32	-	0.38	0.39	0.34
7	0.58	0.40	0.59	0.54	0.59	0.36	-	0.59	0.53
8	0.56	0.50	0.50	0.48	0.59	0.43	0.59	-	0.63
9	0.59	0.44	0.56	0.40	0.76	0.33	0.53	0.63	-

Table 2: Thurstone data: The lower triangle shows sample correlations, the upper fitted correlations according to the combined factor and truncated vine model.

5.1. Thurstone data

The first data set we analyze are measurements of nine cognitive variables such as ability of classification and number series completion which were collected from 4175 students Thurstone (1933). The correlation matrix of the data is available as the data set `Thurstone.33` in the R-package `psych` Revelle (2012) and shown in the lower left triangle of Table 2.

Table 3 shows the number of parameters, distances D (3.14) to the sample correlation matrix and BIC values (3.15) for each of the nine different models taken into account here. For comparison, the number of parameters of an unstructured parametric correlation matrix estimate is 36 in this case. For this reason, 2-level models are preferable in terms of model complexity, since they can reduce the number of parameters by more than a factor of 2. From the results shown in Table 3, it is obvious that there is an underlying factor structure, which cannot be accounted for using only truncated vines. The combined factor-vine model is however clearly superior to a 2-factor model and does almost as good as the 3-factor model, which has however more parameters, so that the BIC values are actually quite close. The upper triangle of Table 2 shows the fitted correlations of the combined model. In most cases, they are very close to the sample correlations. Furthermore, Figure 4 illustrates the selected 1-truncated vine tree of the combined model. Most conditional dependencies are rather small after removal of factor dependence, but there is some stronger dependence left for mathematical and reading ability, respectively.

Finally, it should be noted that the fit of truncated vines can strongly

	2T/ MST	2T/ N	2T/ bMST	3T/ MST	3T/ N	3T/ N1	1F/ 1T	2F	3F
#par.	15	15	15	21	21	21	17	17	24
D	0.473	0.422	0.340	0.284	0.257	0.227	0.079	0.233	0.062
BIC/n	21.19	21.14	21.06	21.02	20.99	20.96	20.80	20.96	20.80

Table 3: Thurstone data: Numbers of parameters, distances D (3.14) to the sample correlation matrix and scaled BIC values (3.15) for the fitted correlation matrices.

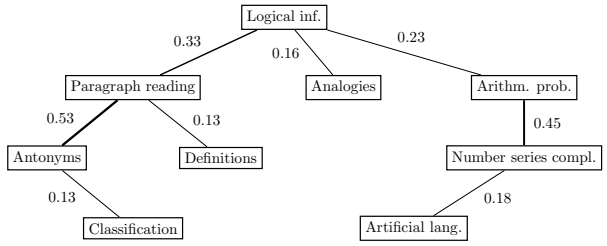

Figure 4: Thurstone data: Selected tree structure in the combined factor-vine model with partial correlations $\widehat{\rho}_{jk|V}$ as edge labels.

be improved by considering 1-neighbors and best spanning trees. This overcomes the only locally optimal fit of minimum spanning trees in each step and corresponds to the results of the simulation study in Section 4. Of course, the inclusion of an additional tree always increases the goodness-of-fit, i.e., decreases the distance D to the sample correlation matrix.

5.2. Harman data

In the second psychology example, we investigate measurements from nine ability tests of 696 participants Harman (1967). The data set can again be found in the R-package `psych` by Revelle (2012), where it is called `Harman.Holzinger`. The sample correlations are shown in the lower triangle of Table 4.

Results from the model fits are reported in Table 5. Conclusions are similar to the previous subsection. In this case however, the 2- and 3-truncated vine models are also superior to the 2-factor model, which indicates that there may be some unobserved heterogeneity in the data that cannot be appropriately captured using two factors. This is in line with Harman (1967) who identifies three distinct factors.

	1	2	3	4	5	6	7	8	9
1	-	0.75	0.78	0.43	0.45	0.49	0.24	0.32	0.30
2	0.75	-	0.70	0.52	0.54	0.58	0.23	0.35	0.33
3	0.78	0.72	-	0.47	0.48	0.55	0.32	0.37	0.36
4	0.44	0.52	0.47	-	0.82	0.82	0.32	0.34	0.34
5	0.45	0.53	0.48	0.82	-	0.74	0.34	0.36	0.36
6	0.51	0.58	0.54	0.82	0.74	-	0.35	0.37	0.37
7	0.21	0.23	0.28	0.33	0.37	0.35	-	0.39	0.52
8	0.30	0.32	0.37	0.33	0.36	0.38	0.45	-	0.67
9	0.31	0.30	0.37	0.31	0.36	0.38	0.52	0.67	-

Table 4: Harman data: The lower triangle shows sample correlations, the upper fitted correlations according to the combined factor and truncated vine model.

	2T/ MST	2T/ N	2T/ bMST	3T/ MST	3T/ N	3T/ N1	1F/ 1T	2F	3F
#par.	15	15	15	21	21	21	17	17	24
D	0.172	0.144	0.106	0.083	0.048	0.067	0.044	0.727	0.016
BIC/n	19.95	19.92	19.88	19.91	19.88	19.90	19.84	20.52	19.88

Table 5: Harman data: Numbers of parameters, distances D (3.14) to the sample correlation matrix and scaled BIC values (3.15) for the fitted correlation matrices.

The combined 1-factor and 1-truncated vine model takes into account both factor as well as between-variable dependence. As a result, it is clearly superior to a 2-factor one and again performs similarly as good as the 3-factor model: Although the distance to the sample correlation matrix is larger, the BIC value of the combined model is even smaller, since it is more parsimonious. It also improves over 2-truncated vine models and even slightly over 3-truncated ones. Once again, the truncated vine models are clearly improved through 1-neighbors and best spanning trees. Fitted correlations of the combined factor-vine model are shown in the upper triangle of Table 4, and the fitted tree after removal of factor dependence on the unobserved factor is illustrated in the left panel of Figure 5. The tree shows three groups of three variables each, which are also identified by the 3-factor model.

The convergence of the EM algorithm presented in Section 3.3 is illustrated in the right panel of Figure 5, where the log-likelihood of the first 200 EM iterations is shown. Convergence is very fast and the 10 000 itera-

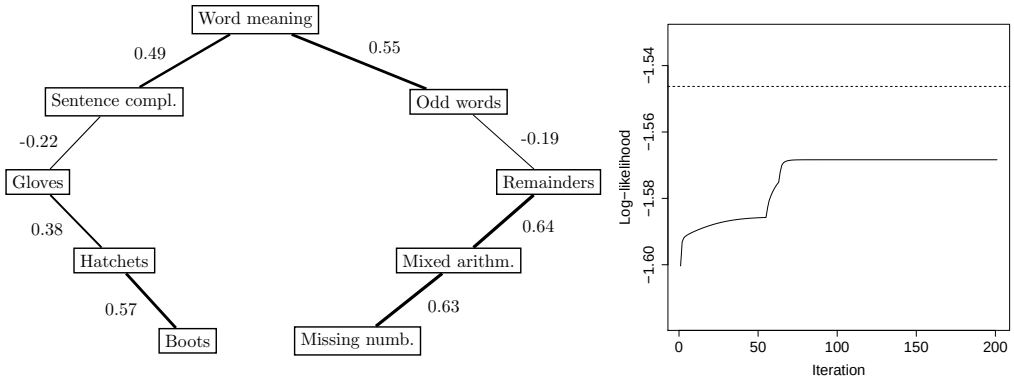

Figure 5: Harman data: Left panel: Selected tree structure in the combined factor-vine model with fitted partial correlations $\widehat{\rho}_{jk|V}$ as edge labels. Right panel: Fitted log-likelihoods of the first 200 EM iterations. The dashed line gives the best obtainable log-likelihood (3.13).

tions would not have been needed to ensure convergence. The jumps in the log-likelihoods in the first 70 iterations correspond to changes in the selected edges of the Markov tree.

5.3. Dow Jones returns

We also give a relevant example from finance and analyze the correlations of $n = 251$ stock returns of all $d = 30$ Dow Jones constituents in the year 2011. Financial log returns of a single asset are defined as $y_t = \log(P_t/P_{t-1})$ for day t, where $\{P_t\}$ is the time series of prices. It is known (e.g., Jondeau et al. (2007)) that returns are heavier tailed than normal and commonly used distributions for fitting are Student's t or skew-t. Over time, $\{y_t\}$ is nearly serially uncorrelated but the absolute returns have some serial autocorrelations so that generalized autoregressive conditional heteroscedastic (GARCH) models are often used. For illustration here, we are mainly interested in the dependence structure, so for each asset we just empirically transformed to normal. Suppose we have observations $(y_{i1}, \dots, y_{id})'$ of returns on d assets on day i for $i = 1, \dots, n$. The jth variable is converted to ranks $R_{1j}, \dots, R_{nj}$ with the rank of 1 for the smallest and n for the largest value. Then we convert to normal scores with $z_{ij} = \Phi^{-1}([R_{ij}+a]/[n+2a+1])$ for appropriately chosen a near -0.5 so that $\sum_{i=1}^n z_{ij} = 0$ and $\frac{1}{n}\sum_{i=1}^n z_{ij} = 1$ for each j.

	2T/ MST	2T/ N	2T/ bMST	3T/ MST	3T/ N	3T/ N1	1F/ 1T	2F	3F
#par.	57	57	57	84	84	84	59	59	87
D	4.771	4.603	4.622	3.839	3.598	3.596	2.147	4.048	3.118
BIC/n	57.73	57.56	57.58	57.39	57.15	57.15	55.15	57.05	56.73

Table 6: Dow Jones data: Numbers of parameters, distances D (3.14) to the sample correlation matrix and scaled BIC values (3.15) for the fitted correlation matrices.

The numbers of parameters, distances D (3.14) to the sample correlation matrix and BIC values (3.15) are shown in Table 6 for all model fits. The reported number of parameters should be compared to the 435 parameters of an unstructured parametric correlation matrix estimate. Therefore, also 3-level models constitute a strong improvement in the number of parameters.

For this data set, the combined 1-factor and 1-truncated vine model is overall best among all models under consideration, in particular it is even clearly superior to the 3-factor model, which was not the case in the two psychology examples. This shows that stock returns are driven by an overall factor, which may be interpreted as the general state of the economy, while remaining dependence can be explained best through between-stock dependence. In particular, the selected tree after removal of factor dependence (not shown here) identifies groups of stocks from common sectors such as IT or financials. Both the factor and the truncated vine model are not able to take these characteristics into account appropriately.

Note that the distance D reported for the best 2-truncated vine based on 1-neighbors is smaller than for the one based on the 1000 best spanning trees. This illustrates that the best 1-neighbor of the minimum spanning tree need not necessarily be among the very best spanning trees and shows that a possibly large number of best spanning trees is required to find a model improvement, as it was also observed in the simulation study in Section 4.

5.4. Computing time

While the estimation of loadings for factor models and one-tree-at-a-time selection of minimum spanning trees for truncated vines is very fast in terms of computing time (less than one second in nine dimensions, less than 10 seconds in 30 dimensions), the heuristics for finding good truncated vine models (see Section 3.1) and the modified EM algorithm to fit the combined

1-factor and 1-truncated vine model (see Section 3.3) are computationally more demanding.

In the two nine-dimensional psychology examples the neighbor-based fit of the Models (ii), (v) and (vi) took only 7–8 seconds using our implementation in R on a 2.6Ghz AMD Opteron. For the 30-dimensional Dow Jones data set, this however increased to 42–46 minutes, since also the number of possible 1-neighbors increased from 28 to 406. This is similar to the running time for the 2-truncated vine based on the 1000 best spanning trees: in the 30-dimensional example it was 46 minutes. This is unlike the nine-dimensional examples, where the running time was much slower (3.5 minutes) than for 1-neighbors. This is because there are only 28 1-neighbors to be investigated in nine dimensions, while the 1000 best spanning trees only represent 0.02% of the total number of trees on nine nodes.

Our implementation of the modified EM algorithm for the combined factor-vine model also proved to be reasonably fast. Although we ran it for 10 000 iterations it only took 2.5 minutes in the nine-dimensional case and less than 8 minutes for the 30-dimensional Dow Jones data set. Choosing fewer iterations such as 1000, for which we also could have assumed convergence, therefore even gets the running time under one minute.

6. Conclusion and outlook

In this paper, we have initiated the study of combined factor-vine models as another parsimonious dependence structure that is an alternative to the factor and truncated vine structures. The structure has conditional or residual dependence given latent variables, and this is interpretable in various contexts in psychometrics and finance. Because the exact form of the residual dependence is not specified a priori, computational methods of SEMs do not apply and we develop an efficient modified EM algorithm.

This paper is the first to have data examples that compare the dependence models of factor, truncated vine and combined factor-vine models. The examples show the better fit of the combined factor-vine model, so that this direction of research should be pursued further. Here, we have focused on Gaussian dependence but the motivation and planned future research are the copula versions with this dependence structure but with the capability of handling tail dependence and tail asymmetry.

The theory and applications of vine copulas or the pair-copula construction has been well developed; see Kurowicka and Joe (2011), Brechmann et al.

(2012), Dißmann et al. (2013). The theory in this paper can provide alternative algorithms in the sequential fitting of a combined factor-vine structure with a bivariate copula for each edge of the vine. After finding some well-fitting models based on an assumption of a multivariate normal copula, we can convert to a parametrization with correlations in the first tree and partial correlations in subsequent trees, and then extend to a vine copula model by replacing each correlation by a bivariate copula and each partial correlation by a bivariate copula applied to conditional distributions. For example, in finance applications, where it is important to account for tail dependence (more dependence in the joint tails than would be obtained with the multivariate normal copula), the bivariate copulas could be chosen to have upper and/or lower tail dependence.

Future research therefore includes the following. We will compare our vine tree selection approaches to previously proposed approaches; in particular, to those by Dißmann et al. (2013) who maximize the sum of absolute Kendall's τ, and by Kurowicka (2011) based on partial correlations. In light of the work on truncation by Brechmann et al. (2012), the criterion $(\log|\boldsymbol{R}|-\log|\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})|)/\log|\boldsymbol{R}|$ (see Section 3.4) can be used for truncation level selection: include vine trees, using model selection techniques as discussed in Section 3.1, until a pre-specified degree of closeness to the empirical correlation matrix is attained. In addition, the copula version of the combined factor-vine model proposed in this paper will be investigated in comparison to truncated vine copulas as well as to copula-based factor models as proposed by Krupskii and Joe (2013).

The factor model with residual dependence is intuitively plausible for the applications in this paper as well as others. The model fits so well in comparison with 2- and 3-level truncated vines and 3-factor models that also the general combined model, with arbitrary number of factors and vine levels, is definitely worthy of further study. For applications leading to a parsimonious dependence structure, we expect that two to four combined factors/vine levels will often be adequate. For models with two or more factors, we might also consider 'confirmatory' or 'structured' factors, where non-zero loadings are decided based on the context of the application. For example, for financial portfolios that include stocks and bonds, a 2-factor model with residual dependence might have a factor for stocks only and another factor for bonds only, and dependent 'residuals' satisfying a truncated vine structure.

With these factor models with residual dependence, this paper is the first

to make connections between SEMs and truncated vines. This connection can be developed further for linear or nonlinear SEMs where the path diagram has some specified edges for latent variables to observed variables plus additional unspecified edges for the residual dependence.

Acknowledgment

The authors gratefully acknowledge the helpful comments of the referees, which further improved the manuscript. The first author is thankful for support of the TUM Graduate School's International School of Applied Mathematics as well as of Allianz Deutschland AG. The second author is supported by an NSERC Canada Discovery Grant.

Appendix A. Maximum likelihood estimation of 2-truncated vine

We consider the log-likelihood $L_j(\varphi_1, \varphi_2)$ given in (3.5), where $\psi_j = 1 - \varphi_{j1}^2 - \varphi_{j2}^2 - 2\varphi_{j1}\varphi_{j2}\rho$, and $\widehat{\rho} = r_{k_{1j}k_{2j}}$ is known from earlier in the sequence. Let the MLEs be $\widehat{\rho}_{jk_{1j}}$, $\widehat{\rho}_{jk_{2j}}$, $\widehat{\alpha}_j$, $\widehat{\varphi}_{j1}$, $\widehat{\varphi}_{j2}$ and $\widehat{\psi}_j$, and let $e_{ij} = z_{ij} - \varphi_{j1} z_{i,k_{1j}} - \varphi_{j2} z_{i,k_{2j}}$ and $\widehat{e}_{ij} = z_{ij} - \widehat{\varphi}_{j1} z_{i,k_{1j}} - \widehat{\varphi}_{j2} z_{i,k_{2j}}$.

Differentiation of (3.5) with respect to φ_{jm} $(m = 1, 2)$ followed by multiplication by ψ_j^4 leads to:

$$-\frac{1}{2}\psi_j^2 \frac{\partial \psi_j^2}{\partial \varphi_{jm}} + \psi_j^2 (r_{jk_{mj}} - \varphi_{jm} - \varphi_{j,3-m} r_{k_{1j}k_{2j}}) + \frac{1}{2n}\sum_i e_i^2 \frac{\partial \psi_j^2}{\partial \varphi_{jm}}, \quad m = 1, 2, \tag{A.1}$$

where

$$\frac{\partial \psi_j^2}{\partial \varphi_{jm}} = -2\varphi_{jm} - 2\varphi_{j,3-m}\rho_{k_{1j}k_{2j}},$$

and the middle terms come from

$$\frac{1}{n}\sum_{i=1}^{n} e_{ij} z_{ik_{mj}} = \frac{1}{n}\sum_{i=1}^{n} (z_{ij} - \varphi_{j1} z_{ik_{1j}} - \varphi_{j2} z_{ik_{2j}}) z_{ik_{mj}}.$$

Assume by induction that the MLE of $\rho_{k_{1j}k_{2j}}$ is $r_{k_{1j}k_{2j}}$ based on the previous step, so that

$$\widehat{\alpha}_j = \frac{(\widehat{\rho}_{jk_{2j}} - \widehat{\rho}_{jk_{1j}} r_{k_{1j}k_{2j}})}{\sqrt{(1 - \widehat{\rho}_{jk_{1j}}^2)(1 - r_{k_{1j}k_{2j}}^2)}}.$$

Setting the two equations in (A.1) to 0 then leads to

$$\widehat{\psi}_j^2(\widehat{\varphi}_{j1} + \widehat{\varphi}_{j2} r_{k_{1j}k_{2j}}) + \widehat{\psi}_j^2(r_{jk_{1j}} - \widehat{\varphi}_{j1} - \widehat{\varphi}_{j2} r_{k_{1j}k_{2j}}) - (\widehat{\varphi}_{j1} + \widehat{\varphi}_{j2} r_{k_{1j}k_{2j}}) \frac{1}{n} \sum_{i=1}^{n} \widehat{e}_{ij}^2 = 0,$$

$$\widehat{\psi}_j^2(\widehat{\varphi}_{j2} + \widehat{\varphi}_{j1} r_{k_{1j},k_{2j}}) + \widehat{\psi}_j^2(r_{jk_{2j}} - \widehat{\varphi}_2 - \widehat{\varphi}_{j1} r_{k_{1j}k_{2j}}) - (\widehat{\varphi}_{j2} + \widehat{\varphi}_{j1} r_{k_{1j}k_{2j}}) \frac{1}{n} \sum_{i=1}^{n} \widehat{e}_{ij}^2 = 0,$$

which, after canceling some terms, is equivalent to

$$\widehat{\psi}_j^2 r_{jk_{1j}} = (\widehat{\varphi}_{j1} + \widehat{\varphi}_{j2} r_{k_{1j}k_{2j}}) \frac{1}{n} \sum_{i=1}^{n} \widehat{e}_{ij}^2, \tag{A.2}$$

$$\widehat{\psi}_j^2 r_{jk_{2j}} = (\widehat{\varphi}_{j2} + \widehat{\varphi}_{j1} r_{k_{1j}k_{2j}}) \frac{1}{n} \sum_{i=1}^{n} \widehat{e}_{ij}^2. \tag{A.3}$$

Standard results for regression yield $\widehat{\psi}_j^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{e}_{ij}^2$, since ψ_j^2 is the variance of $Z_{ij} - \varphi_{j1} Z_{ik_{1j}} - \varphi_{j2} Z_{ik_{2j}}$. From the equation for φ_{j1} (see (2.4)), we get $\widehat{\rho}_{jk_{1j}} = \widehat{\varphi}_{j1} + \widehat{\varphi}_{j2} r_{k_{1j}k_{2j}}$, so that from (A.2), $r_{jk_{1j}} = \widehat{\rho}_{jk_{1j}}$. Next, from (A.3) and the equations for φ_{j1} and φ_{j2} (see (2.4)), we get $r_{jk_{2j}} = \widehat{\varphi}_{j2} + \widehat{\varphi}_{j1} r_{k_{1j}k_{2j}} = \widehat{\varphi}_{j2} + (r_{jk_{1j}} - \widehat{\varphi}_{j2} r_{k_{1j}k_{2j}}) r_{k_{1j}k_{2j}}$ or

$$(r_{jk_{2j}} - r_{jk_{1j}} r_{k_{1j}k_{2j}})/(1 - r_{k_{1j}k_{2j}}^2) = \widehat{\varphi}_{j2} = \widehat{\alpha}_j \sqrt{(1 - r_{jk_{1j}}^2)/(1 - r_{k_{1j}k_{2j}}^2)}\,.$$

Hence

$$\widehat{\alpha}_j = \frac{r_{jk_{2j}} - r_{jk_{1j}} r_{k_{1j}k_{2j}}}{\sqrt{(1 - r_{jk_{1j}}^2)(1 - r_{k_{1j}k_{2j}}^2)}} = r_{jk_{2j};k_{1j}},$$

the sample partial correlation, and $\widehat{\rho}_{jk_{2j}} = r_{jk_{2j}}$.

References

Bedford, T. and R. M. Cooke (2002). Vines — a new graphical model for dependent random variables. *Annals of Statistics 30*, 1031–1068.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Bollen, K. A. and J. S. Long (1992). Tests for structural equation models: Introduction. *Sociological Methods & Research 21*, 123–131.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (2008). *Time Series: Forecasting and Control* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Braeken, J., F. Tuerlinckx, and P. De Boeck (2007). Copulas for residual dependency. *Psychometrika 72*, 393–411.

Brechmann, E. C., C. Czado, and K. Aas (2012). Truncated regular vines in high dimensions with applications to financial data. *Canadian Journal of Statistics 40*(1), 68–85.

Daniels, M. J. and R. E. Kass (2001). Shrinkage estimators for covariance matrices. *Biometrics 57*, 1173–1184.

Dempster, A. P. (1972). Covariance selection. *Biometrics 28*, 157–175.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B 39*(1), 1–38.

Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis 59*(1), 52–69.

Gabow, H. N. (1977). Two algorithms for generating weighted spanning trees in order. *SIAM Journal on Computing 6*(1), 139–150.

Geweke, J. and G. Zhou (1996). Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies 9*, 557–587.

Gruber, L. and C. Czado (2013). Sequential Bayesian model selection of regular vine copulas. Preprint.

Harman, H. H. (1967). *Modern Factor Analysis*. Chicago: University of Chicago Press.

Johnson, R. and D. W. Wichern (2002). *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NJ: Prentice Hall.

Jondeau, E., S.-H. Poon, and M. Rockinger (2007). *Financial Modeling Under Non-Gaussian Distributions*. London: Springer.

Krupskii, P. and H. Joe (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis 120*, 85–101.

Kullback, S. (1967). On testing correlation matrices. *Journal of the Royal Statistical Society. Series C 16*(1), 80–85.

Kurowicka, D. (2011). Optimal truncation of vines. In D. Kurowicka and H. Joe (Eds.), *Dependence Modeling: Vine Copula Handbook*. Singapore: World Scientific Publishing Co.

Kurowicka, D. and R. M. Cooke (2003). A parametrization of positive definite matrices in terms of partial correlation vines. *Linear Algebra and its Applications 372*, 225–251.

Kurowicka, D. and R. M. Cooke (2006). Completion problem with partial correlation vines. *Linear Algebra and its Applications 418*, 188–200.

Kurowicka, D. and H. Joe (2011). *Dependence Modeling: Vine Copula Handbook*. Singapore: World Scientific Publishing Co.

Lange, K. (2010). *Numerical Analysis for Statisticians* (2nd ed.). New York: Springer.

Lawley, D. N. and A. E. Maxwell (1971). *Factor Analysis as a Statistical Method* (2nd ed.). London: Butterworth.

Lewandowski, D., D. Kurowicka, and H. Joe (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis 100*(9), 1989–2001.

Meng, X.-L. and D. B. Rubin (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika 80*(2), 267–278.

Prim, R. C. (1957). Shortest connection networks and some generalizations. *Bell System Technical Journal 36*, 1389–1401.

Revelle, W. (2012). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois: Northwestern University. R package version 1.2.8.

Steiger, J. H. (2001). Driving fast in reverse: The relationship between software development, theory, and education in structural equation modeling. *Journal of the American Statistical Association 96*, 331–338.

Thurstone, L. L. (1933). *The Theory of Multiple Factors*. Ann Arbor: Edwards Brothers, Inc.

Whittaker, J. (2008). *Graphical Models in Applied Multivariate Statistics*. Wiley.