Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Parameter estimation for pair-copula constructions

Note

Note no SAMBA/36/10
Author Ingrid Hobæk Haff
Date 6th October 2010

Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modelling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Parameter estimation for pair-copula constructions**
Author	**Ingrid Hobæk Haff**
Date	6th October 2010
Publication number	SAMBA/36/10

Abstract

We explore various estimators for the parameters of a pair-copula construction (PCC), among those the stepwise semiparametric (SSP) estimator, designed for this dependence structure. We present its asymptotic properties, as well as the estimation algorithm for the two most common types of PCCs. Compared to the considered alternatives, i.e. maximum likelihood, inference functions for margins and semiparametric estimation, SSP is in general asymptotically less efficient. As we show in an example, this loss of efficiency may however be rather low. Furthermore, SSP is semiparametrically efficient for the Gaussian copula. More importantly, it is computationally tractable even in high dimensions, as opposed to its competitors. In any case, SSP may provide start values, required by the other estimators. It is also well suited for selecting the pair-copulae of a PCC for a given data set.

Keywords	copulae; efficiency; empirical distribution functions; hierarchical construction; stepwise estimation; vines
Target group	
Availability	
Project	
Project number	
Research field	
Number of pages	35

1 Introduction

The last decades' technological revolution have considerably increased the relevance of multivariate modelling. Copulae are now regularly used within fields such as finance, survival analysis and actuarial sciences. Although the list of parametric bivariate copulae is long and varied, the choice is rather limited in higher dimensions (Genest et al., 2009). Accordingly, a number of hierarchical, copula-based structures have been proposed, among those the pair-copula construction (PCC) of Joe (1996), further studied and considered by Bedford and Cooke (2001, 2002), Kurowicka and Cooke (2006) and Aas et al. (2009).

A PCC is a treelike construction, built from pair-copulae with conditional distributions as their two arguments (see Figure **??**). The number of conditioning variables is zero at the ground, and increases by one for each level, to ensure coherence of the structure. Despite its simple structure, the PCC is highly flexible and covers a wide range of complex dependencies (Hobæk Haff et al., 2010; Joe et al., 2010). After Aas et al. (2009) set it in an inferential context, it has made several appearances in the literature (Chollete et al., 2009; Czado and Min, 2010; Czado et al., 2009; Fischer et al., 2007; Heinen and Valdesogo, 2009; Kolbjørnsen and Stien, 2008; Schirmacher and Schirmacher, 2008), exhibiting its adequacy for various applications.

Regardless of its recent popularity, estimation of PCC parameters has so far been addressed mostly in an applied setting. The aim of this work is to explore the properties of alternative estimators. As the PCC is a member of the multivariate copula family, one may exploit the large collection of estimators proposed for that model class, among which moments type procedures, based on, for instance, the matrix of pairwise Kendall's tau coefficients (Clayton, 1978; Genest, 1987; Genest and Rivest, 1993; Oakes, 1982). Such methods may be well-suited for particular copula families. We are however interested in more general procedures, allowing for broader model classes. Moreover, we wish to exploit the specific structure of the PCC.

More specifically, the number of parameters of a PCC grows quickly with the dimension, even if all pair-copulae constituting the structure are from one-parameter families. In medium to high dimension, the existing copula estimators may simply become too demanding computationally, and will at least require good start values in the optimisation procedure. Furthermore, due to the PCC's tree structure, selection of appropriate pair-copulae for a given data set must be done level by level. Procedures that estimate all parameters simultaneously are therefore unfit for this task.

In all, we contemplate four estimators. The first is the classical maximum likelihood (ML), followed by the inference functions for margins (IFM) and semiparametric estimators, that have been developed specifically for multivariate copulae. These three estimators are treated in Section 2, and are included mostly for comparison. Section 3 is devoted to the fourth one, the stepwise semiparametric estimator (SSP). Unlike the others, it is designed for the PCC structure. Although it has been suggested and used earlier (Aas et al., 2009), it has never been formally presented, nor have its asymptotic properties been explored. In Section 4, we compare the four estimators in a few examples. Finally, Section 5 presents some concluding remarks.

The setting is as follows. Consider the observations $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ of n independent d-variate stochastic vectors $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$, originating from the same pair-copula construction. Assume further that the joint distribution is absolutely continuous, with strictly increasing margins. The corresponding copula is then unique (Sklar, 1959). Letting $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ denote the parameters of the margins and copula, respectively, the joint probability density function (pdf) may then be expressed as (McNeil et al. (2006), pp. 197)

$$f_{1..d}(x_1, \dots, x_d; \boldsymbol{\alpha}, \boldsymbol{\theta}) = c_{1..d}(F_1(x_1; \boldsymbol{\alpha}_1), \dots, F_d(x_d; \boldsymbol{\alpha}_d); \boldsymbol{\theta}) \prod_{l=1}^{d} f_l(x_l; \boldsymbol{\alpha}_l). \qquad (1.1)$$

Here, F_l and $f_l,\ l = 1, \dots, d$, are the marginal cumulative distribution functions and probability density functions, respectively, and $c_{1..d}$ is the corresponding copula density. Since this is a PCC, the copula density is, in turn, a product of pair-copulae.

Define the index sets $v_{ij} = \{i{+}1, \dots, i{+}j{-}1\}$, with $v_{i1} = \emptyset$, and $w_{ij} = \{i, v_{ij}, i{+}j\}$, for $1 \le i \le d - j,\ 1 \le j \le d - 1$. Thus, for a d-dimensional vector $\boldsymbol{a} = (a_1, \dots, a_d)$, we write $\boldsymbol{a}_{v_{ij}} = (a_{i+1}, \dots, a_{i+j-1})$ and $\boldsymbol{a}_{w_{ij}} = (a_i, \dots, a_{i+j})$. Further, let $F_{k|l}$ be the conditional cdf of X_k given $\boldsymbol{X}_l = \boldsymbol{x}_l$, and $c_{i,i+j|v_{ij}}$ the copula density corresponding to the conditional distribution $F_{i,i+j|v_{ij}}$ of (X_i, X_{i+j}) given $\boldsymbol{X}_{v_{ij}} = \boldsymbol{x}_{v_{ij}}$. Finally, let $\boldsymbol{\theta}_{i,i+j|v_{ij}}$ be the parameters of the copula density $c_{i,i+j|v_{ij}}$, and define
$\boldsymbol{\theta}_{i\to i+j} = \{\boldsymbol{\theta}_{s,s+t|v_{st}} : (s, s+t) \in w_{ij}\}$, with $\boldsymbol{\theta}_{i\to i} = \emptyset$, and
$\boldsymbol{\theta}_i = \{\boldsymbol{\theta}_{s,s+t|v_{st}} : |v_{st}| = i - 1\}$, where $|\cdot|$ denotes the cardinality, (i.e. $\boldsymbol{\theta}_i$ gathers all parameters at level i of the structure). For a so-called D-vine (Bedford and Cooke,

2001, 2002), the joint pdf (1.1) can now be re-expressed as (Aas et al., 2009)

$$
\begin{aligned}
f_{1..d}(x_1,\dots,x_d;\boldsymbol{\alpha},\boldsymbol{\theta}) = \\
&\prod_{l=1}^{d} f_l(x_l;\boldsymbol{\alpha}_l) \\
&\cdot \prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{i,i+j|v_{ij}}\Big(F_{i|v_{ij}}(x_i|\boldsymbol{x}_{v_{ij}};\boldsymbol{\alpha}_{w_{i,j-1}},\boldsymbol{\theta}_{i\to i+j-1})\,, \\
&\qquad F_{i+j|v_{ij}}(x_{i+j}|\boldsymbol{x}_{v_{ij}};\boldsymbol{\alpha}_{w_{i+1,j-1}},\boldsymbol{\theta}_{i+1\to i+j});\boldsymbol{\theta}_{i,i+j|v_{ij}}\Big)\,.
\end{aligned}
\tag{1.2}
$$

In four dimensions, this becomes

$$
\begin{aligned}
f_{1234}(x_1,x_2,x_3,x_4;\boldsymbol{\alpha},\boldsymbol{\theta}) = \\
&f_1(x_1;\boldsymbol{\alpha}_1)\cdot f_2(x_2;\boldsymbol{\alpha}_2)\cdot f_3(x_3;\boldsymbol{\alpha}_3)\cdot f_4(x_4;\boldsymbol{\alpha}_4) \\
&\cdot c_{12}(F_1(x_1;\boldsymbol{\alpha}_1),F_2(x_2;\boldsymbol{\alpha}_2);\boldsymbol{\theta}_{12})\cdot c_{23}(F_2(x_2;\boldsymbol{\alpha}_2),F_3(x_3;\boldsymbol{\alpha}_3);\boldsymbol{\theta}_{23}) \\
&\cdot c_{34}(F_3(x_3;\boldsymbol{\alpha}_3),F_4(x_4;\boldsymbol{\alpha}_4);\boldsymbol{\theta}_{34}) \\
&\cdot c_{13|2}(F_{1|2}(x_1|x_2;\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\theta}_{12}),F_{3|2}(x_3|x_2;\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\theta}_{23});\boldsymbol{\theta}_{13|2}) \\
&\cdot c_{24|3}(F_{2|3}(x_2|x_3;\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\theta}_{23}),F_{4|3}(x_4|x_3;\boldsymbol{\alpha}_3,\boldsymbol{\alpha}_4,\boldsymbol{\theta}_{34});\boldsymbol{\theta}_{24|3}) \\
&\cdot c_{14|23}(F_{1|23}(x_1|x_2,x_3;\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\theta}_{12},\boldsymbol{\theta}_{23},\boldsymbol{\theta}_{13|2}), \\
&\qquad F_{4|23}(x_4|x_2,x_3;\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\alpha}_4,\boldsymbol{\theta}_{23},\boldsymbol{\theta}_{34},\boldsymbol{\theta}_{24|3});\boldsymbol{\theta}_{14|23}).
\end{aligned}
\tag{1.3}
$$

For simplicity, we will start by assuming that the distribution in question is a D-vine, represented to the left in Figure **??** for $d = 5$. Similar results can be obtained for C-vines (Section 3.3) and other regular vines (Bedford and Cooke, 2001, 2002). We also assume that the PCC is of a simplified form (Hobæk Haff et al., 2010), i.e. that the parameters $\boldsymbol{\theta}_{i,i+j|v_{ij}}$ of the copulae $C_{i,i+j|v_{ij}}$, combining conditional distributions, are not functions of the conditioning variables $\boldsymbol{x}_{v_{ij}}$. Without this assumption, inference on these models is not doable in practice.

2 Multivariate copula estimators

Supposing the model is true, the maximum likelihood (ML) estimator is a natural choice, due to its asymptotic efficiency and other advantageous characteristics. According to (1.2), the log-likelihood function of a D-vine is given by

$$\begin{aligned}
l(\boldsymbol{\alpha},\boldsymbol{\theta};\boldsymbol{x}) \\
&= \sum_{k=1}^{n} \log(f_{1..d}(x_{1k},\ldots,x_{dk};\boldsymbol{\alpha},\boldsymbol{\theta})) \\
&= \sum_{k=1}^{n}\sum_{l=1}^{d} \log(f_l(x_{lk};\boldsymbol{\alpha}_l))+ \\
&\quad \sum_{k=1}^{n}\sum_{j=1}^{d-1}\sum_{i=1}^{d-j} \log\Big(c_{i,i+j|v_{ij}}\Big(F_{i|v_{ij}}(x_i|\boldsymbol{x}_{v_{ij}};\boldsymbol{\alpha}_{w_{i,j-1}},\boldsymbol{\theta}_{i\to i+j-1}), \\
&\qquad F_{i+j|v_{ij}}(x_{i+j}|\boldsymbol{x}_{v_{ij}};\boldsymbol{\alpha}_{w_{i+1,j-1}},\boldsymbol{\theta}_{i+1\to i+j});\boldsymbol{\theta}_{i,i+j|v_{ij}}\Big)\Big) \\
&= l_M(\boldsymbol{\alpha};\boldsymbol{x}) + l_C(\boldsymbol{\alpha},\boldsymbol{\theta};\boldsymbol{x}).
\end{aligned} \tag{2.1}$$

The ML estimator $\hat{\boldsymbol{\theta}}^{ML}$ is obtained by maximising the above log-likelihood function over all parameters, $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$, simultaneously. Under the additional assumptions $(M1)-(M8)$ of Lehmann (2004) (pp. 499–501), this corresponds to solving the set $\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\phi}_{ML}\left(X_{1k},\ldots,X_{dk};\hat{\boldsymbol{\alpha}}^{ML},\hat{\boldsymbol{\theta}}^{ML}\right) = \mathbf{0}$ of estimating equations (one equation per parameter), which is a vector of functions, with elements

$$\begin{aligned}
\boldsymbol{\phi}_{ML,l}(x_1,\ldots,x_d;\boldsymbol{\alpha},\boldsymbol{\theta}) &= \frac{\partial \log\left(f_{1\ldots d}(x_1,\ldots,x_d;\boldsymbol{\alpha},\boldsymbol{\theta})\right)}{\partial \boldsymbol{\alpha}_l}, \\
\boldsymbol{\phi}_{ML,d+(j-1)\left(d-\frac{j}{2}\right)+i}(x_1,\ldots,x_d;\boldsymbol{\alpha},\boldsymbol{\theta}) &= \frac{\partial \log\left(f_{1\ldots d}(x_1,\ldots,x_d;\boldsymbol{\alpha},\boldsymbol{\theta})\right)}{\partial \boldsymbol{\theta}_{i,i+j|v_{ij}}}, \\
& l=1,\ldots,d,\, i=1,\ldots,d-j,\ j=1,\ldots,d-1.
\end{aligned} \tag{2.2}$$

Define $\boldsymbol{\mathcal{I}}$ as the corresponding Fischer information matrix

$$\begin{aligned}
\boldsymbol{\mathcal{I}} &= \mathrm{E}\left(\left(\frac{\partial \log\left(f_{1..d}(\boldsymbol{X};\boldsymbol{\alpha},\boldsymbol{\theta})\right)}{\partial(\boldsymbol{\alpha},\boldsymbol{\theta})}\right)\left(\frac{\partial \log\left(f_{1..d}(\boldsymbol{X};\boldsymbol{\alpha},\boldsymbol{\theta})\right)}{\partial(\boldsymbol{\alpha},\boldsymbol{\theta})}\right)^T\right) \\
&= \mathrm{E}\left(-\frac{\partial^2 \log\left(f_{1..d}(\boldsymbol{X};\boldsymbol{\alpha},\boldsymbol{\theta})\right)}{\partial(\boldsymbol{\alpha},\boldsymbol{\theta})\partial(\boldsymbol{\alpha},\boldsymbol{\theta})^T}\right) = \begin{pmatrix} \boldsymbol{\mathcal{I}}_\alpha & \boldsymbol{\mathcal{I}}_{\alpha,\theta} \\ \boldsymbol{\mathcal{I}}_{\alpha,\theta}^T & \boldsymbol{\mathcal{I}}_\theta \end{pmatrix}.
\end{aligned}$$

In the last expression, it is partitioned according to marginal and dependence parameters. The corresponding inverse is

$$\mathcal{I}^{-1} = \begin{pmatrix} \mathcal{I}^{(\alpha)} & \mathcal{I}^{(\alpha,\theta)} \\ \left(\mathcal{I}^{(\alpha,\theta)}\right)^T & \mathcal{I}^{(\theta)} \end{pmatrix}, \quad \begin{aligned} \mathcal{I}^{(\alpha)} &= \left(\mathcal{I}_\alpha - \mathcal{I}_{\alpha,\theta}\mathcal{I}_\theta^{-1}\mathcal{I}_{\alpha,\theta}^T\right)^{-1} \\ \mathcal{I}^{(\alpha,\theta)} &= -\mathcal{I}^{(\alpha)}\mathcal{I}_{\alpha,\theta}\mathcal{I}_\theta^{-1} \\ \mathcal{I}^{(\theta)} &= \mathcal{I}_\theta^{-1} + \mathcal{I}_\theta^{-1}\mathcal{I}_{\alpha,\theta}^T\mathcal{I}^{(\alpha)}\mathcal{I}_{\alpha,\theta}\mathcal{I}_\theta^{-1}. \end{aligned} \tag{2.3}$$

It is well-known that the estimator $\hat{\boldsymbol{\theta}}^{ML}$ is consistent for $\boldsymbol{\theta}$ and asymptotically normal, i.e.

$$\begin{aligned} &\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{ML} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{V}^{ML}\right), \\ &\boldsymbol{V}^{ML} = \mathcal{I}^{(\theta)}. \end{aligned}$$

In general, ML estimation of PCC parameters will require numerical optimisation. Even in rather low dimensions, such as four or five, the number of parameters is high, especially if several of the model components have more than one parameter. For instance, a five-dimensional PCC, consisting of Student's t-copulae, has 20 parameters, before one has taken the margins into account. Thus, the optimisation becomes numerically challenging and highly time consuming. In fact, the ML estimator may not be an option in practice. Therefore, one needs faster and computationally easier estimation procedures.

Moreover, the above results require the chosen model to be the true model, i.e. the one that produced the data. If the specified model is close to the truth in the Kullback-Leibler (KL) sense, the ML estimator may behave very well (Claeskens and Hjort, 2008). However, it is in general non-robust to larger KL-divergences from the true model.

2.1 Two-step estimators

The next two estimators are not particularly designed for pair-copula constructions, but for multivariate copula models in general. Both consist of two steps, the first being estimation of the marginal parameters.

2.1.1 Inference function for margins estimator

The inference function for margins (IFM) estimator, introduced by Joe (Joe, 1997, 2005), addresses the computational inefficiency of the ML estimator by performing the estimation in two steps. First, one estimates the marginal parameters by maximising the marginal log-likelihood function l_M from (2.1). The resulting estimates $\hat{\boldsymbol{\alpha}}^{IFM}$ are plugged into the PCC log-likelihood function l_C to obtain $\hat{\boldsymbol{\theta}}^{IFM}$.

Under conditions $(M1) - (M8)$ (see above), this corresponds to solving the estimating equations $\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\phi}_{IFM}\left(X_{1k},\ldots,X_{dk};\hat{\boldsymbol{\alpha}}^{IFM},\hat{\boldsymbol{\theta}}^{IFM}\right)=\mathbf{0}$, with elements

$$\begin{aligned}\boldsymbol{\phi}_{IFM,l}(x_l;\boldsymbol{\alpha}_l) &= \frac{\partial \log\left(f_l(x_l;\boldsymbol{\alpha}_l)\right)}{\partial \boldsymbol{\alpha}_l} \\ \boldsymbol{\phi}_{IFM,d+(j-1)\left(d-\frac{j}{2}\right)+i}(x_1,\ldots,x_d;\boldsymbol{\alpha},\boldsymbol{\theta}) &= \frac{\partial \log\left(f_{1\ldots d}(x_1,\ldots,x_d;\boldsymbol{\alpha},\boldsymbol{\theta})\right)}{\partial \boldsymbol{\theta}_{i,i+j|v_{ij}}} \\ l=1,\ldots,d,\ i&=1,\ldots,d-j,\ j=1,\ldots,d-1.\end{aligned} \tag{2.4}$$

Compared to the ML equations (2.2), the full log-pdf, $\log f_{1\ldots d}$, is replaced with the marginal log-pdfs, $\log f_j$, for the estimation of $\boldsymbol{\alpha}$.

Consider a four-dimensional D-vine (1.3), consisting of Student's t-copulae, each having their own correlation and degrees of freedom parameter, combined with Student's t-margins. The parameter vectors are then $\boldsymbol{\alpha}=(\nu_1,\nu_2,\nu_3,\nu_4)$ and $\boldsymbol{\theta}=(\rho_{12},\rho_{23},\rho_{34},\rho_{13|2},\rho_{24|3},\rho_{14|23},\nu_{12},\nu_{23},\nu_{34},\nu_{13|2},\nu_{24|3},\ \nu_{14|23})$. IFM estimation of this model starts with a separate estimation of ν_i, $i=1,2,3,4$, margin by margin. The next step is to optimise $l_C\left(\hat{\nu}_1,\ldots,\hat{\nu}_4,\boldsymbol{\theta};\boldsymbol{x}\right)$ over $\boldsymbol{\theta}$, where l_C is given in (2.1), i.e. the sum of the log-copulae in line 3 to 8 of (1.3), over all observations.

For the margins, define $\boldsymbol{\mathcal{K}}_\alpha$ as the matrix with blocks $\boldsymbol{\mathcal{K}}_{\alpha,i,j}=$ $\mathrm{E}\left(\boldsymbol{\phi}_{IFM,i}(X_i;\boldsymbol{\alpha}_i)\boldsymbol{\phi}_{IFM,j}(X_j;\boldsymbol{\alpha}_j)\right)=\mathrm{E}\left(\left(\frac{\partial\log(f_i(X_i;\boldsymbol{\alpha}_i))}{\partial\boldsymbol{\alpha}_i}\right)\left(\frac{\partial\log(f_j(X_j;\boldsymbol{\alpha}_j))}{\partial\boldsymbol{\alpha}_j}\right)^T\right)$, $i,j=1,\ldots,d$, and $\boldsymbol{\mathcal{J}}_\alpha$ the block diagonal matrix with the diagonal blocks $\boldsymbol{\mathcal{J}}_{\alpha,i,i}=\mathrm{E}\left(-\frac{\partial}{\partial\boldsymbol{\alpha}_i^T}\boldsymbol{\phi}_{IFM,i}(X_i;\boldsymbol{\alpha}_i)\right)$ $=\mathrm{E}\left(-\frac{\partial^2\log(f_i(X_i;\boldsymbol{\alpha}_i))}{\partial\boldsymbol{\alpha}_i\partial\boldsymbol{\alpha}_i^T}\right)=\boldsymbol{\mathcal{K}}_{\alpha,i,i}$ of the $\boldsymbol{\mathcal{K}}_\alpha$, each block corresponding to one of the margins. If all margins are one-parameter families, $\boldsymbol{\mathcal{K}}_\alpha$ and $\boldsymbol{\mathcal{J}}_\alpha$ are $d\times d$ matrices. More generally, their dimension depends on the number of parameters of each margin.

Joe (2005) has shown that under the mentioned conditions, the estimator $\hat{\boldsymbol{\theta}}^{IFM}$ is consistent for $\boldsymbol{\theta}$, as well as asymptotically normal:

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{IFM}-\boldsymbol{\theta}\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\boldsymbol{V}^{IFM}\right),$$

with

$$\boldsymbol{V}^{IFM}=\boldsymbol{\mathcal{I}}_\theta^{-1}+\boldsymbol{\mathcal{I}}_\theta^{-1}\boldsymbol{\mathcal{I}}_{\alpha,\theta}^T\boldsymbol{\mathcal{J}}_\alpha^{-1}\boldsymbol{\mathcal{K}}_\alpha\boldsymbol{\mathcal{J}}_\alpha^{-1}\boldsymbol{\mathcal{I}}_{\alpha,\theta}\boldsymbol{\mathcal{I}}_\theta^{-1}. \tag{2.5}$$

We see that the above covariance matrix is obtained by replacing the marginal block $\boldsymbol{\mathcal{I}}^{(\alpha)}$ in (2.3) with $\boldsymbol{\mathcal{J}}_\alpha^{-1}\boldsymbol{\mathcal{K}}_\alpha\boldsymbol{\mathcal{J}}_\alpha^{-1}$, which is the asymptotic covariance matrix of $\hat{\boldsymbol{\alpha}}^{IFM}$. This shows how one loses asymptotic efficiency by discarding the information the dependence structure might have on the margins. Several studies, including Joe (2005) and Kim et al. (2007), have demonstrated that unless the dependence between the variables is extreme, this efficiency loss is likely to be rather small. Moreover, the IFM method is faster than the ML estimator, and can at least be used to set the start values in ML optimisation. Of course, for high-dimensional $\boldsymbol{\theta}$, the IFM estimator is still too slow to be used for PCCs.

2.1.2 Semiparametric estimator for copula parameters

The semiparametric (SP) estimator was introduced by Genest et al. (1995) and generalised by Tsukahara (2005). The censored case was treated by Shih and Louis (1995). Aas et al. (2009) suggest this method for PCCs. Just like IFM, it is a two-step estimator, treating the margins separately.

As seen from (1.2), the pair-copula arguments at the ground level of a pair-copula construction (T1 in Figure **??**) are marginal distributions $F_i(x_i)$. From the second level, they are conditional distributions, whose number of conditioning variables increases by one with each level. As a matter of fact, these conditional distributions may be written as functions of the margins. Let i, j be distinct indices, i.e. $i \neq j$, and v a non-empty set of indices, all from $\{1, \ldots, d\}$, such that $i, j \notin v$. Then, in a simplified pair-copula construction (Joe, 1997)

$$F_{i|v\cup j}(x_i|\boldsymbol{x}_{v\cup j}) = \left.\frac{\partial C_{ij|v}(u_i, u_j)}{\partial u_j}\right|_{u_i=F_{i|v}(x_i|\boldsymbol{x}_v),u_j=F_{j|v}(x_j|\boldsymbol{x}_v)}. \tag{2.6}$$

Thus, one can express $F_{i|v\cup j}$ as a function of the two conditional distributions $F_{i|v}$ and $F_{j|v}$ with one conditioning variable less, by extracting one of the variables j from the conditioning set $v \cup j$. Likewise, $F_{i|v}$ and $F_{j|v}$ may be written as bivariate functions of conditional distributions with a conditioning set reduced by one. Proceeding in this way, one finally obtains recursive functions of the margins.

The type of pair-copula construction determines the conditional distributions that are needed. At level $j \geq 2$ of a D-vine, these are the pairs $(F_{i|v_{ij}}(x_i|\boldsymbol{x}_{v_{ij}}), F_{i+j|v_{ij}}(x_{i+j}|\boldsymbol{x}_{v_{ij}}))$, $i = 1, \ldots, d-j$. Now, define the functions $h_{i,i+j|v_{ij}}$ and $h_{i+j,i|v_{ij}}$ as

$$\begin{aligned} h_{i,i+j|v_{ij}}(u_i, u_{i+j}) &\equiv \frac{\partial C_{i,i+j|v_{ij}}(u_i, u_{i+j})}{\partial u_{i+j}} \\ h_{i+j,i|v_{ij}}(u_{i+j}, u_i) &\equiv \frac{\partial C_{i,i+j|v_{ij}}(u_i, u_{i+j})}{\partial u_i}, \end{aligned} \tag{2.7}$$

for $i = 1, \ldots, d-j,\ j = 1, \ldots, d-1$. Using (2.6), one obtains

$$\begin{aligned} &F_{i|v_{ij}}(x_i|\boldsymbol{x}_{v_{ij}}) = \\ &\quad h_{i,i+j-1|v_{i,j-1}}(F_{i|v_{i,j-1}}(x_i|\boldsymbol{x}_{v_{i,j-1}}), F_{i+j-1|v_{i,j-1}}(x_{i+j-1}|\boldsymbol{x}_{v_{i,j-1}})) \\ &F_{i+j|v_{ij}}(x_{i+j}|\boldsymbol{x}_{v_{i,j-1}}) = \\ &\quad h_{i+j,i+1|v_{i+1,j-1}}(F_{i+j|v_{i+1,j-1}}(x_{i+j}|\boldsymbol{x}_{v_{i+1,j-1}}), F_{i+1|v_{i+1,j-1}}(x_{i+1}|\boldsymbol{x}_{v_{i+1,j-1}})), \end{aligned}$$

which are bivariate functions of the conditional distributions constituting the arguments of the copulae at the previous level, $j-1$. As one continues this recursion, one achieves, as earlier mentioned, functions of the margins $F_i(x_i), \ldots, F_{i+j}(x_{i+j})$.

Since these are needed in the asymptotics, we denote them $g^1_{i,i+i}$ and $g^2_{i,i+i}$, and explicitly define them below. Note however, that for all practical purposes, such as in the estimation algorithm (Algorithm 1), one will use the nested h-functions from (2.7). Define

$$\begin{aligned} g^1_{i,i+j}(u_i,\ldots,u_{i+j-1}) &\equiv F_{i|v_{ij}}(F_i^{-1}(u_i)|F_{i+1}^{-1}(u_{i+1}),\ldots,F_{i+j-1}^{-1}(u_{i+j-1})) \\ g^2_{i,i+j}(u_{i+1},\ldots,u_{i+j}) &\equiv F_{i+j|v_{ij}}(F_{i+j}^{-1}(u_{i+j})|F_{i+1}^{-1}(u_{i+1}),\ldots,F_{i+j-1}^{-1}(u_{i+j-1})) \end{aligned}, \tag{2.8}$$

for $i = 1,\ldots,d-j,\ j = 1,\ldots,d-1$. Now, one may rewrite (1.2) in terms of the these g-functions:

$$\begin{aligned} &f_{1..d}(x_1,\ldots,x_d;\boldsymbol{\alpha},\boldsymbol{\theta}) = \\ &\prod_{l=1}^{d} f_l(x_l;\boldsymbol{\alpha}_l) \\ &\cdot\prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{i,i+j|v_{ij}}\left(g^1_{i,i+j}(F_i(x_i;\boldsymbol{\alpha}_i),\ldots,F_{i+j-1}(x_{i+j};\boldsymbol{\alpha}_{i+j-1});\boldsymbol{\theta}_{i\to i+j-1}),\right. \\ &\qquad g^2_{i,i+j}(F_{i+1}(x_{i+1};\boldsymbol{\alpha}_{i+1}),\ldots,F_{i+j}(x_{i+j};\boldsymbol{\alpha}_{i+j});\boldsymbol{\theta}_{i+1\to i+j}); \\ &\qquad \left.\boldsymbol{\theta}_{i,i+j|v_{ij}}\right). \end{aligned} \tag{2.9}$$

Recall that IFM estimates are obtained by plugging the estimated marginal parameters $\hat{\boldsymbol{\alpha}}$ into the PCC log-likelihood function l_C, when one estimates $\boldsymbol{\theta}$. Semiparametric estimation consists in replacing the parametric marginal cdfs $u_j = F_j(x_j;\boldsymbol{\alpha}_j)$ in l_C with the corresponding empirical ones

$$u_{jn} = F_{jn}(x_j) = \frac{1}{n+1}\sum_{k=1}^{n} I(x_{jk}\le x_j),\ I(A) = \begin{cases} 1, & \text{if A is true} \\ 0, & \text{otherwise} \end{cases}.$$

The resulting pseudo log-likelihood function $l_{C,P}(\boldsymbol{\theta};\boldsymbol{x})$, given by

$$\begin{aligned} &l_{C,P}(\boldsymbol{\theta};\boldsymbol{x}) = \sum_{k=1}^{n}\log\left(c_{1\ldots d}\left(F_{1n}(x_{1k}),\ldots,F_{dn}(x_{dk});\boldsymbol{\theta}\right)\right) = \\ &\sum_{k=1}^{n}\sum_{j=1}^{d-1}\sum_{i=1}^{d-j}\log\left(c_{i,i+j|v_{ij}}\left(g^1_{i,i+j}(F_{i,n}(x_{ik}),\ldots,F_{i+j-1,n}(x_{i+j-1,k});\boldsymbol{\theta}_{i\to i+j-1}),\right.\right. \\ &\qquad g^2_{i,i+j}(F_{i+1,n}(x_{i+1,k}),\ldots,F_{i+j,n}(x_{i+j,k});\boldsymbol{\theta}_{i+1\to i+j}); \\ &\qquad \left.\left.\boldsymbol{\theta}_{i,i+j|v_{ij}}\right)\right), \end{aligned}$$

is just a function of $\boldsymbol{\theta}$. To obtain the semiparametric estimator $\hat{\boldsymbol{\theta}}^{SP}$, one simply maximises $l_{C,P}(\boldsymbol{\theta};\boldsymbol{X})$ with respect to $\boldsymbol{\theta}$.

Consider again the four-dimensional Student's t-vine of Section 2.1.1. Using the SP estimator for this model, one starts with a separate estimation of the marginal parameters ν_i, just as with IFM. However, SP estimation also requires a preliminary computation of the so-called pseudo-observations $u_{ik,n} = F_{in}(x_{ik}),\ i =$

$1, 2, 3, 4$, $k = 1, \ldots, n$. The estimate $\hat{\boldsymbol{\theta}}^{SP}$ is obtained by maximising $l_{C,P}(\boldsymbol{x}; \boldsymbol{\theta})$, in this case

$$\sum_{k=1}^{n} \left(\log\left(c_{12}(u_{1k,n}, u_{2k,n}; \boldsymbol{\theta}_{12})\right) + \ldots + \log\left(c_{14|23}(g_{14}^{1}(u_{1k,n}, \ldots, u_{3k,n}; \boldsymbol{\theta}_{1\to 3}),\right.\right.$$
$$\left.\left. g_{14}^{2}(u_{2k,n}, \ldots, u_{4k,n}; \boldsymbol{\theta}_{2\to 4}); \boldsymbol{\theta}_{14|23})\right)\right),$$

with $\boldsymbol{\theta}_{1\to 3} = (\boldsymbol{\theta}_{12}, \boldsymbol{\theta}_{23}, \boldsymbol{\theta}_{13|2})$ and $\boldsymbol{\theta}_{2\to 4} = (\boldsymbol{\theta}_{23}, \boldsymbol{\theta}_{34}, \boldsymbol{\theta}_{24|3})$, over $\boldsymbol{\theta}$.

In addition to the assumptions made for the ML estimator (see above), assume that the copula density $c_{1\ldots d}$ fulfills condition $(A.1)$ from Tsukahara (2005). Then, the procedure corresponds to solving the estimating equations $\frac{1}{n}\sum_{k=1}^{n} \boldsymbol{\phi}_{SP}\left(F_{1n}(X_{1k}), \ldots, F_{1n}(X_{dk}\right.$ $\mathbf{0}$, with elements

$$\boldsymbol{\phi}_{SP,(j-1)\left(d-\frac{j}{2}\right)+i}(u_1, \ldots, u_d; \boldsymbol{\theta}) = \frac{\partial \log\left(c_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta})\right)}{\partial \boldsymbol{\theta}_{i,i+j|v_{ij}}}, \quad i = 1, \ldots, d-j,\ j = 1, \ldots, d-1. \tag{2.10}$$

The full log-pdf, $\log f_{1\ldots d}$, is now substituted with the copula log-density, $\log c_{1\ldots d}$, in the estimating equations.

Let $\boldsymbol{U}$ be a d-variate stochastic vector distributed according to the copula $C_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta})$, and define

$$\boldsymbol{W}_j^{SP}(\boldsymbol{U}; \boldsymbol{\theta}) = \int \frac{\partial^2 \log c_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial u_j} I(U_j \leq u_j) dC_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta}).$$

Further, define the matrix

$$\begin{aligned}\boldsymbol{B}_\theta^{SP} &= \mathrm{Var}\left(\sum_{j=1}^{d} \boldsymbol{W}_j^{SP}(\boldsymbol{U}; \boldsymbol{\theta})\right) + \sum_{j=1}^{d} \mathrm{Cov}\left(\boldsymbol{\phi}_{SP}(\boldsymbol{U}; \boldsymbol{\theta}), \boldsymbol{W}_j^{SP}(\boldsymbol{U}; \boldsymbol{\theta})\right) \\ &= \mathrm{Var}\left(\sum_{j=1}^{d} \boldsymbol{W}_j^{SP}(\boldsymbol{U}; \boldsymbol{\theta})\right) + \sum_{j=1}^{d} \mathrm{Cov}\left(\frac{\partial \log c_{1\ldots d}(\boldsymbol{U}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \boldsymbol{W}_j^{SP}(\boldsymbol{U}; \boldsymbol{\theta})\right),\end{aligned}$$

where $\boldsymbol{A} = \mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{Z})$ for two stochastic vectors $\boldsymbol{Y}$ and $\boldsymbol{Z}$ is the matrix with elements

$$A_{ij} = \mathrm{Cov}(Y_i, Z_j) + \mathrm{Cov}(Y_j, Z_i).$$

The matrix $\boldsymbol{B}_\theta^{SP}$ quantifies the effect of replacing the parametric marginal distributions with empirical ones. In two dimensions, the covariance terms of $\boldsymbol{B}_\theta^{SP}$ are 0 (Genest and Werker, 2002). The semiparametric estimator $\hat{\boldsymbol{\theta}}^{SP}$ has been shown to be consistent and asymptotically normal (Genest et al., 1995):

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{SP} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{V}^{SP}\right),$$

with

$$\boldsymbol{V}^{SP} = \boldsymbol{\mathcal{I}}_\theta^{-1} + \boldsymbol{\mathcal{I}}_\theta^{-1}\boldsymbol{B}_\theta^{SP}\boldsymbol{\mathcal{I}}_\theta^{-1}. \tag{2.11}$$

Due to the completely separate and independent estimation of marginal and dependence parameters, the semiparametric estimator is more robust to misspecification of the margins than ML and IFM (Kim et al., 2007), but not to misspecification of the pair-copulae. Computationally, it is comparable to IFM. Hence, for high-dimensional $\boldsymbol{\theta}$, although better than ML, this procedure may be too demanding for PCCs, and will at least require good start values in the optimising routine.

3 PCC parameter estimators

If the number of PCC parameters $\boldsymbol{\theta}$ is high enough, the estimators considered so far will be computationally too heavy. In any case, they necessitate appropriate start values. The next estimator, designed especially for pair-copula constructions, addresses this particular issue.

3.1 Stepwise semiparametric estimator (SSP)

As in semiparametric estimation, the marginal parameters are handled separately, and the parametric margins in the PCC log-likelihood function l_C are replaced with the non-parametric ones. The idea is to estimate the PCC parameters level by level, conditioning on the parameters from preceding levels of the structure. Define the functions

$$\begin{aligned}\psi_j(u_1,\dots,u_d;\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_j) = \\ \sum_{i=1}^{d-j} \log\big(c_{i,i+j|v_{ij}}(g^1_{i,i+j}(u_i,\dots,u_{i+j-1};\boldsymbol{\theta}_{i\to i+j-1}), \\ g^2_{i,i+j}(u_{i+1},\dots,u_{i+j};\boldsymbol{\theta}_{i+1\to i+j});\boldsymbol{\theta}_{i,i+j|v_{ij}})\big)\end{aligned} \tag{3.1}$$

and the level pseudo log-likelihood functions

$$l_{C,P,j}(\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_j;\boldsymbol{x}) = \sum_{k=1}^{n}\sum_{l=1}^{j}\psi_l(F_{1n}(x_{1k}),\dots,F_{dn}(x_{dk});\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_l),$$

for $j = 1,\dots,d-1$. Hence, $l_{C,P,j}$ is the sum over all log pair-copulae up to, and including, level j. To obtain the parameter estimates $\hat{\boldsymbol{\theta}}_j^{SSP}$ for a particular level j, one plugs the estimates $\hat{\boldsymbol{\theta}}_1^{SSP},\dots,\hat{\boldsymbol{\theta}}_{j-1}^{SSP}$ from preceding levels into $l_{C,P,j}$ and maximises it with respect to $\boldsymbol{\theta}_j$. Assuming the standard conditions for the ML estimator are fulfilled (see Section 2), this corresponds to solving the estimating equations $\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\phi}_{SSP}\left(F_{1n}(X_{1k}),\dots,F_{dn}(X_{dk});\hat{\boldsymbol{\theta}}^{SSP}\right) = \mathbf{0}$, with

$$\begin{aligned}&\boldsymbol{\phi}_{SSP,(j-1)\left(d-\frac{j}{2}\right)+i}(u_1,\dots,u_d;\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_j) \\ &\qquad = \frac{\partial}{\partial\boldsymbol{\theta}_{i,i+j|v_{ij}}}\sum_{l=1}^{j}\psi_l(u_1,\dots,u_d;\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_l) \\ &\qquad = \frac{\partial}{\partial\boldsymbol{\theta}_{i,i+j|v_{ij}}}\psi_j(u_1,\dots,u_d;\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_j), \\ &i = 1,\dots,d-j,\ j = 1,\dots,d-1.\end{aligned} \tag{3.2}$$

Compared to the SP equations (2.10), the full log copula density $\log c_{1\ldots d}$ is now replaced by the sum of log copula densities up to, and including, the level the parameter belongs to. The corresponding estimation procedure is presented in Algorithm 1, where $a \mathrel{+}= b$ is short for $a = a + b$. If none of the pair-copulae constituting the structure share parameters, which will usually be the case, the estimating equations are reduced to $\frac{\partial}{\partial \boldsymbol{\theta}_{i,i+j|v_{ij}}} \log(c_{i,i+j|v_{ij}})$. This means that the optimisation is performed for each copula, individually.

Let us return to the four-dimensional D-vine considered in Section 2.1.2. As in the SP procedure, one estimates the marginal parameters $\boldsymbol{\alpha}$ in a separate step, and computes the pseudo-observations $u_{ik,n} = F_{in}(x_{ik})$, $i = 1, 2, 3, 4$, $k = 1, \ldots, n$. One proceeds to the level 1 parameters, estimating each of the pairs $(\rho_{i,i+1}, \nu_{i,i+1})$ by optimising $\sum_{k=1}^{n} \log(c_{i,i+1}(u_{ik,n}, u_{i+1,k,n}; \rho_{i,i+1}, \nu_{i,i+1}))$, for $i = 1, 2, 3$. One subsequently computes the copula arguments $u_{i|i+1,k,n} = h_{i,i+1}(u_{ik,n}, u_{i+1,k,n}; \hat{\rho}_{i,i+1}, \hat{\nu}_{i,i+1}))$ and $u_{i+2|i+1,k,n} =$
$h_{i+2,i+1}(u_{i+2,k,n}, u_{i+1,k,n}; \hat{\rho}_{i+1,i+2}, \hat{\nu}_{i+1,i+2}))$, $i = 1, 2, 3$, $k = 1, \ldots, n$, for level 2, by plugging the resulting estimates into the adequate h-functions (2.7). At level 2, one estimates each of the pairs $(\rho_{i,i+2|i+1}, \nu_{i,i+2|i+1})$, for $i = 1, 2$, by maximising $\sum_{k=1}^{n} \log(c_{i,i+2|i+1}(u_{i|i+1,k,n}, u_{i+2|i+1,k,n}; \rho_{i,i+2|i+1}, \nu_{i,i+2|i+1}))$. Next, one computes the copula arguments $u_{1|23,k,n}$ and $u_{4|23,k,n}$ for level 3 by plugging the estimates from level 2 into $h_{13|2}$ and $h_{24|3}$. Finally, one optimises
$\sum_{k=1}^{n} \log(c_{14|23}(u_{1|23,k,n}, u_{4|23,k,n}; \rho_{14|23}, \nu_{14|23}))$ to obtain the estimates
$(\hat{\rho}_{14|23}, \hat{\nu}_{14|23})$.

When some of the copulae share parameters, the procedure is a little different. Let us for instance replace the previously described D-vine with a four-dimensional Student's t-copula with correlations $(\rho_{12}, \rho_{23}, \rho_{34}, \rho_{13}, \rho_{24}, \rho_{14})$ and ν degrees of freedom. This is also a D-vine consisting of Student's t-copulae. The correlation parameters of these copulae are now the corresponding partial correlations $(\rho_{12}, \rho_{23}, \rho_{34}, \rho_{13|2}, \rho_{24|3}, \rho_{14}$
However, the degrees of freedom parameter is shared. More specifically, it is ν for the three copulae at the ground level, $\nu + 1$ for the two at level 2 and $\nu + 2$ for the top level copula. The SSP estimation procedure is now as follows. After having computed the pseudo-observations, one maximises the function

$$\sum_{k=1}^{n} \psi_1(u_{1k,n}, \ldots, u_{4k,n}; \rho_{12}, \rho_{23}, \rho_{34}, \nu) = \sum_{k=1}^{n} \sum_{i=1}^{3} (\log(c_{i,i+1}(u_{i,k,n}, u_{i+1,k,n}; \rho_{i,i+1}, \nu))$$

over $(\rho_{12}, \rho_{23}, \rho_{34}, \nu)$. Then, one computes the copula arguments for level 2 by plugging the resulting estimates into the adequate h-functions, as described above. At the second level, one estimates the two correlations ρ_{13} and ρ_{24}, which are not shared by $c_{13|2}$ and $c_{24|3}$. The optimisation can therefore be done separately. More specifically, one optimises each of
$\sum_{k=1}^{n} \log(c_{i,i+2|i+1}(u_{i|i+1,k,n}, u_{i+2|i+1,k,n}; \rho_{i,i+2}, \hat{\rho}_{i,i+1}, \hat{\rho}_{i+1,i+2}, \hat{\nu}))$, over $\rho_{i,i+2}$,

$i = 1,2$ (note that $\hat{\rho}_{i,i+1}, \hat{\rho}_{i+1,i+2}$ are needed to compute the partial correlations $\rho_{i,i+2|i+1}$). Next, one computes the copula arguments for the top level copula, and finally, one maximises $\sum_{k=1}^{n} \log(c_{14|23}(u_{1|23,k,n}, u_{4|23\},k,n}; \rho_{14}, \hat{\rho}_{12}, \ldots, \hat{\rho}_{24|3}, \hat{\nu}))$ over ρ_{14}. Note however that although it is possible to estimate the parameters of a multivariate Student's t-copula as described above, it is unnecessarily complex. In practice, one would typically estimate the correlation parameters via the corresponding Kendall's τ coefficients, and subsequently optimise the pseudo log-likelihood function $l_{C,P}$ over ν, plugging in the estimated correlations, as described in for instance McNeil et al. (2006). The main purpose of the PCC is to model pairs that behave differently. If one does not really need that flexibility, then using a PCC is like using a sledgehammer to crack a nut.

Let us now consider conditions $(A.1) - (A.5)$ from Tsukahara (2005). The last four of these are covered by the standard conditions for the ML estimator. Further, define

$$\boldsymbol{\phi}_{(j-1)\left(d-\frac{j}{2}\right)+i}(\boldsymbol{u}; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_j) = \frac{\partial}{\partial \boldsymbol{\theta}_{i,i+j|v_{ij}}} \psi_j(\boldsymbol{u}; \boldsymbol{\theta}_j) \equiv \boldsymbol{\psi}_{j,\theta}(\boldsymbol{u}; \boldsymbol{\theta}_j)$$
$$\frac{\partial}{\partial u_k} \boldsymbol{\phi}_{(j-1)\left(d-\frac{j}{2}\right)+i}(\boldsymbol{u}; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_j) = \frac{\partial}{\partial u_k} \boldsymbol{\psi}_{j,\theta}(\boldsymbol{u}; \boldsymbol{\theta}_j) \equiv \boldsymbol{\psi}_{j,\theta,u_k}(\boldsymbol{u}; \boldsymbol{\theta}_j).$$

Let $\mathcal{Q}$ and $\mathcal{R}$ be the sets of positive, symmetric, inverse square integrable functions on $[0,1]$ and reproducing u-shaped functions on $[0,1]$, respectively, as defined in Tsukahara (2005). For the SSP estimator, Condition $(A.1)$ may then be phrased in the following way.

Condition 1. For each $\boldsymbol{\theta}$, $\boldsymbol{\psi}_{j,\theta} = (\psi_{j,\theta,1}, \ldots, \psi_{j,\theta,l_{ij}})$ and $\boldsymbol{\psi}_{j,\theta,u_k} = (\psi_{j,\theta,u_k,1}, \ldots, \psi_{j,\theta,u_k,l_{ij}}),\ j = 1, \ldots, d-1,\ i = 1, \ldots, d-j$ are continuous, and there exist functions $r_{ij,k}, \tilde{r}_{ij,k} \in \mathcal{R}$ and $q_{ij,k} \in \mathcal{Q}$, such that

$$\begin{array}{ll} |\psi_{j,\theta,m}(\boldsymbol{u}; \boldsymbol{\theta}_j)| & \leq \prod_{l=1}^{d} r_{ij,l}(u_l) \\ |\psi_{j,\theta,u_k,m}(\boldsymbol{u}; \boldsymbol{\theta}_j)| & \leq \tilde{r}_{ij,k}(u_k) \prod_{l \neq k} r_{ij,l}(u_l) \end{array}, \quad \begin{array}{l} k,l = 1, \ldots, d,\ j = 1, \ldots, d-1, \\ i = 1, \ldots, d-j,\ m = 1, \ldots, l_{ij}, \end{array}$$

with

$$\begin{array}{rl} \int \left(\prod_{l=1}^{d} r_{ij,l}(u_l) \right)^2 dC_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta}) & < \infty, \\ \int \left(q_{ij,k}(u_k) \tilde{r}_{ij,k}(u_k) \prod_{l \neq k} r_{ij,l}(u_l) \right)^2 dC_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta}) & < \infty, \end{array}$$

where $|\boldsymbol{\theta}_{ij|v_{ij}}| = l_{ij}$ is the number of parameters of the pair-copula $C_{i,i+j|v_{ij}}$. When none of the pair-copulae share parameters, Condition 1 becomes a condition on each of them, individually.

Once more let $\boldsymbol{U}$ be a d-variate stochastic vector distributed according to $C_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta})$, and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{d-1})$. Define

$$\begin{aligned} \boldsymbol{W}_j^{SSP}(\boldsymbol{U}; \boldsymbol{\theta}) &= \int \frac{\partial}{\partial u_j} \boldsymbol{\phi}_{SSP}(u_1, \ldots, u_d; \boldsymbol{\theta}) I(U_j \leq u_j) dC_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta}) \\ &= \int \frac{\partial^2}{\partial \boldsymbol{\theta} \partial u_j} \boldsymbol{\psi}(u_1, \ldots, u_d; \boldsymbol{\theta}) I(U_j \leq u_j) dC_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta}) \end{aligned}$$

$\psi_1(\boldsymbol{\theta}_1) = 0$
for $j = 1, \ldots, d$ **do**
for $k = 1, \ldots, n$ **do**
$u_{jk} = x_{jk}$.
end for
end for
for $i = 1, \ldots, d-1$ **do**
for $k = 1, \ldots, n$ **do**
$\psi_1(\boldsymbol{\theta}_1) \mathrel{+}= \log c_{i,i+1}(u_{i,k}, u_{i+1,k}; \boldsymbol{\theta}_{i,i+1})$
end for
end for
$\hat{\boldsymbol{\theta}}_1 = \underset{\boldsymbol{\theta}_1}{\text{argmax}}(\psi_1(\boldsymbol{\theta}_1))$
for $i = 1, \ldots, d-2$ **do**
for $k = 1, \ldots, n$ **do**
$u_{i|i+1,k} = h_{i,i+1}(u_{ik}, u_{i+1,k}, \hat{\boldsymbol{\theta}}_{i,i+1})$
$u_{i+2|i+1,k} = h_{i+2,i+1}(u_{i+2,k}, u_{i+1,k}, \hat{\boldsymbol{\theta}}_{i+1,i+2})$
end for
end for
for $j = 2, \ldots, d-1$ **do**
$\psi_j(\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_{j-1}, \boldsymbol{\theta}_j) = 0$
for $i = 1, \ldots, d-j$ **do**
for $k = 1, \ldots, n$ **do**
$\psi_j(\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_{j-1}, \boldsymbol{\theta}_j) \mathrel{+}= \log c_{i,i+j|v_{ij}}(u_{i|v_{ij},k}, u_{i+j|v_{ij},k}; \boldsymbol{\theta}_{i,i+j|v_{ij}})$
end for
end for
$\hat{\boldsymbol{\theta}}_j = \underset{\boldsymbol{\theta}_j}{\text{argmax}}(\psi_j(\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_{j-1}, \boldsymbol{\theta}_j))$
if $j == d-1$ **then**
Stop
end if
for $i = 1, \ldots, d-j-1$ **do**
for $k = 1, \ldots, n$ **do**
$u_{i|v_{i,j+1},k} = h_{i,i+j|v_{ij}}(u_{i|v_{ij},k}, u_{i+j|v_{ij},k}, \hat{\boldsymbol{\theta}}_{i,i+j|v_{ij}})$
$u_{i+j+1|v_{i,j+1},k} = h_{i+j+1,i+1|v_{i+1,j}}(u_{i+j+1|v_{i+1,j},k}, u_{i+1|v_{i+1,j},k}, \hat{\boldsymbol{\theta}}_{i+1,i+j+1|v_{i+1,j}})$
end for
end for
end for

Algorithm 1. SSP for a D-vine

and the matrix

$$\boldsymbol{B}_\theta^{SSP} = \text{Var}\left(\sum_{j=1}^{d} \boldsymbol{W}_j^{SSP}(\boldsymbol{U};\boldsymbol{\theta})\right) + \sum_{j=1}^{d} \text{Cov}\left(\boldsymbol{\phi}_{SSP}(\boldsymbol{U};\boldsymbol{\theta}), \boldsymbol{W}_j^{SSP}(\boldsymbol{U};\boldsymbol{\theta})\right)$$

$$= \text{Var}\left(\sum_{j=1}^{d} \boldsymbol{W}_j^{SSP}(\boldsymbol{U};\boldsymbol{\theta})\right) + \sum_{j=1}^{d} \text{Cov}\left(\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\psi}(\boldsymbol{U};\boldsymbol{\theta}), \boldsymbol{W}_j^{SSP}(\boldsymbol{U};\boldsymbol{\theta})\right).$$

Moreover, define the two matrices

$$\boldsymbol{\mathcal{K}}_\theta = \text{E}\left(\boldsymbol{\phi}_{SSP}\boldsymbol{\phi}_{SSP}^T\right) = \begin{pmatrix} \boldsymbol{\mathcal{K}}_{\theta,1,1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0}^T & \boldsymbol{\mathcal{K}}_{\theta,d-2,d-2} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0}^T & \boldsymbol{\mathcal{I}}_{\theta,d-1,d-1} \end{pmatrix},$$

and

$$\boldsymbol{\mathcal{J}}_\theta = \text{E}\left(-\frac{\partial \boldsymbol{\phi}_{SSP}}{\partial \boldsymbol{\theta}^T}\right) = \begin{pmatrix} \boldsymbol{\mathcal{J}}_{\theta,1,1} & & & \mathbf{0} \\ \vdots & & & \\ \vdots & \ddots & & \\ \boldsymbol{\mathcal{J}}_{\theta,d-2,1} & \cdots & \boldsymbol{\mathcal{J}}_{\theta,d-2,d-2} & \mathbf{0} \\ \boldsymbol{\mathcal{I}}_{\theta,d-1,1} & \cdots & \boldsymbol{\mathcal{I}}_{\theta,d-1,d-2} & \boldsymbol{\mathcal{I}}_{\theta,d-1,d-1} \end{pmatrix},$$

where the blocks $\boldsymbol{\mathcal{K}}_{\theta,i,j} = \text{E}\left(\left(\frac{\partial \psi_i}{\partial \boldsymbol{\theta}_i}\right)\left(\frac{\partial \psi_j}{\partial \boldsymbol{\theta}_j}\right)^T\right)$ and $\boldsymbol{\mathcal{J}}_{\theta,i,j} = -\text{E}\left(\frac{\partial^2 \psi_i}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j^T}\right)$, $i, j = 1, \ldots, d-1$, correspond to each of the construction's levels. The block diagonal and block lower triangular forms of $\boldsymbol{\mathcal{K}}_\theta$ and $\boldsymbol{\mathcal{J}}_\theta$, respectively, follow from the structure of the estimating equations (see Appendix A.1). More specifically, the ψ functions depend on all the parameters from previous levels but not from following levels. Further, the estimating equations for the top level copula parameters are based on the full copula, as for the SP estimator. This accounts for the appearance of blocks from the Fischer matrix $\boldsymbol{\mathcal{I}}_\theta$ in the last rows of $\boldsymbol{\mathcal{K}}_\theta$ and $\boldsymbol{\mathcal{J}}_\theta$. If all pair-copulae are from one-parameter families, then $\boldsymbol{\mathcal{K}}_\theta$ and $\boldsymbol{\mathcal{J}}_\theta$ are $d(d-1)/2 \times d(d-1)/2$ matrices.

We now have all the necessary components to establish the asymptotic properties of the stepwise semiparametric estimator.

Theorem 1. *Under Condition 1, as well as Conditions $(M1)-(M8)$ of Lehmann (2004), the SSP estimator $\hat{\boldsymbol{\theta}}^{SSP}$ is consistent for $\boldsymbol{\theta}$ and asymptotically normal:*

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{SSP} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{V}^{SSP}\right),$$

with

$$\boldsymbol{V}^{SSP} = \boldsymbol{\mathcal{J}}_\theta^{-1}\boldsymbol{\mathcal{K}}_\theta\left(\boldsymbol{\mathcal{J}}_\theta^{-1}\right)^T + \boldsymbol{\mathcal{J}}_\theta^{-1}\boldsymbol{B}_\theta^{SSP}\left(\boldsymbol{\mathcal{J}}_\theta^{-1}\right)^T. \tag{3.3}$$

Proof. Theorem 1 follows directly from Joe (1997) and Theorem 1 of Tsukahara (2005), with the estimating equations (3.2). □

Theorem 2. *Under the conditions of Theorem 1, the SSP estimator $\hat{\boldsymbol{\theta}}^{SSP}$ is asymptotically semiparametrically efficient for the parameters $\boldsymbol{\theta}$ of the Gaussian copula.*

The proof is given in Appendix A.2.

In general, the stepwise semiparametric estimator $\hat{\boldsymbol{\theta}}^{SSP}$ has a lower asymptotic efficiency than $\hat{\theta}^{SP}$, since it at a given level discards all information from following levels. Nonetheless, the levelwise estimation significantly improves the computational efficiency. The SSP estimator is therefore adequate for medium to high-dimensional models, and as a start value for the SP estimator. Moreover, it is inherently suited for determining an appropriate PCC for a data set, which consists in choosing an ordering of the variables and a set of parametric pair-copulae in a stepwise manner. Once the ordering is fixed, one finds suitable copulae for the ground level, based on the pseudo-observations. At the second level, the necessary pair-copula arguments are obtained by transforming the pseudo-observations with the adequate h-functions, which depend on the chosen ground level copulae. This requires ground level parameter estimates, which can be provided by the SSP estimator. After one has selected copulae for the second level, one proceeds in the same manner for the remaining levels. Of course, one could construct a similar, stepwise estimator with a different transformation to uniform margins. Instead of the empirical margins, one might for instance use the parametric margins as in IFM estimation. That particular estimator was in fact proposed by Joe and Xu (1996).

3.2 Robustness

The SSP estimator is a substantial improvement over the three former in terms of computational speed. However, it presupposes that the specified model is the true one. If the amount of data available is high enough, it should, in most cases, be possible to find adequate marginal distributions. For the pair-copulae, the task is more complex. Using the pseudo-observations, one may obtain a reasonable model for the ground level. Subsequently, however, one must condition on choices from previous levels, as described above. One would therefore expect the quality of the model to decrease with the construction level.

SSP estimation consists in replacing the parametric margins in the PCC log-likelihood function l_C with the non-parametric ones, while keeping the parametric forms of the conditional distributions, i.e. the g-functions (2.8). The resulting estimator is robust to misspecification of the margins, but not of the pair-copulae. By replacing also the conditional distributions with non-parametric versions, one would reduce this sensitivity to chosen pair-copulae preceding in the structure. One possibility is the empirical conditional distribution proposed by Stute (1986):

$$F_{i|v,n}(x_i|\boldsymbol{x}_v) = \frac{n}{n+1}\frac{\sum_{k=1}^{n} I(x_{ik} \leq x_i) K_h\left(\frac{\boldsymbol{x}_v - \boldsymbol{x}_{v,k}}{h^l}\right)}{\sum_{k=1}^{n} K_h\left(\frac{\boldsymbol{x}_v - \boldsymbol{x}_{v,k}}{h^l}\right)}, \tag{3.4}$$

where K_h is a kernel on $\mathbb{R}^l$ with bandwidth parameter h, and l is the dimension of $\boldsymbol{x}_v$. The definition (3.4) is slightly modified here to avoid boundary problems in 0 and 1. It converges almost surely to the true conditional distribution, though at a rather slow pace. Provided $h \to 0$ and $nh^l \to \infty$ as $n \to \infty$, the rate of convergence is of order $(nh^l)^{1/2}$. The quality of the estimates will therefore significantly decrease with the level number. Alternative definitions of the empirical conditional distribution function, such as the one proposed by Hall and Yao (2005), share this unfortunate property.

Recall that the conditional distributions of interest can be written as a recursion of h-functions (2.7), i.e. bivariate functions of transformed variables from the preceding level. The h-functions are in fact conditional distributions of their first variable, given the second. They can therefore be estimated non-parametrically by (3.4) with $l = 1$. Seemingly, one can exploit this to avoid the curse of dimensionality. However, the transformed variables at a given level are obtained from the transformed variables at the preceding level. Hence, the error propagates from level to level, and as expected, the resulting rate of convergence is of the same order as for the original variables, that is $(nh^l)^{1/2}$.

Accordingly, the estimator suggested above becomes unreliable already at the fourth or fifth level of the structure, depending on the amount of data. Since the intention is to improve the quality of estimates at higher levels, it is in practice useless, unless the rate of convergence is increased by additional assumptions on the conditional distributions.

3.3 C-vines

For simplicity, we have only considered D-vines so far. The same results are however easily obtained for other regular vines, and C-vines, in particular (see Figure **??**).

Define the index set $z_{ji} = \{1, \ldots, j-1, j+i\}$, with $z_{0i} = \varnothing$ and $z_{1i} = j+i$, as well as $\boldsymbol{\theta}_{j\_i} = \{\boldsymbol{\theta}_{s,s+t|z_{s-1,0}} : (s, s+t) \in z_{ji}\}$, with $\boldsymbol{\theta}_{1\_i} = \varnothing$, for $0 \leq i \leq d-j$,,

$0 \leq j \leq d-1$. The joint density of a C-vine is given then by (Aas et al., 2009)

$$
\begin{aligned}
&f_{1..d}(x_1,\dots,x_d;\boldsymbol{\alpha},\boldsymbol{\theta}) = \\
&\quad \prod_{l=1}^{d} f_l(x_l;\boldsymbol{\alpha}_l) \\
&\quad \cdot \prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{j,j+i|z_{j-1,0}} \Big(F_{j|z_{j-1,0}}(x_j|\boldsymbol{x}_{z_{j-1,0}};\boldsymbol{\alpha}_{z_{j0}},\boldsymbol{\theta}_{j\_0}), \\
&\qquad\qquad F_{j+i|z_{j-1,0}}(x_{j+i}|\boldsymbol{x}_{z_{j-1,0}};\boldsymbol{\alpha}_{z_{ji}},\boldsymbol{\theta}_{j\_i});\boldsymbol{\theta}_{j,j+i|z_{j-1,0}} \Big).
\end{aligned}
\tag{3.5}
$$

Hence, the log-likelihood function of n independent observations from a C-vine is

$$
\begin{aligned}
&l(\boldsymbol{\alpha},\boldsymbol{\theta};\boldsymbol{x}) \\
&= \sum_{k=1}^{n} \log(f_{1..d}(x_{1k},\dots,x_{dk};\boldsymbol{\alpha},\boldsymbol{\theta})) \\
&= \sum_{k=1}^{n}\sum_{l=1}^{d} \log(f_l(x_{lk};\boldsymbol{\alpha}_l)) + \\
&\quad \sum_{k=1}^{n}\sum_{j=1}^{d-1}\sum_{i=1}^{d-j} \log\big(c_{j,j+i|z_{j-1,0}} \big(F_{j|z_{j-1,0}}(x_j|\boldsymbol{x}_{z_{j-1,0}};\boldsymbol{\alpha}_{z_{j0}},\boldsymbol{\theta}_{j\_0}), \\
&\qquad\qquad F_{j+i|z_{j-1,0}}(x_{j+i}|\boldsymbol{x}_{z_{j-1,0}};\boldsymbol{\alpha}_{z_{ji}},\boldsymbol{\theta}_{j\_i});\boldsymbol{\theta}_{j,j+i|z_{j-1,0}} \big)\big) \\
&= l_M(\boldsymbol{\alpha};\boldsymbol{x}) + l_C(\boldsymbol{\alpha},\boldsymbol{\theta};\boldsymbol{x}).
\end{aligned}
\tag{3.6}
$$

Replacing l_C for D-vines (from (2.1)) with l_C from (3.6), one retrieves the results from Section 2 for C-vines. To achieve the SSP estimator, one must simply replace the psi-function (3.1) in the estimating equations (3.2) with

$$
\begin{aligned}
&\psi_j(u_1,\dots,u_d;\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_j) = \\
&\quad \sum_{i=1}^{d-j} \log\big(c_{j,j+i|z_{j-1,0}} \big(F_{j|z_{j-1,0}}(x_j|\boldsymbol{x}_{z_{j-1,0}};\boldsymbol{\alpha}_{z_{j0}},\boldsymbol{\theta}_{j\_0}), \\
&\qquad\qquad F_{j+i|z_{j-1,0}}(x_{j+i}|\boldsymbol{x}_{z_{j-1,0}};\boldsymbol{\alpha}_{z_{ji}},\boldsymbol{\theta}_{j\_i});\boldsymbol{\theta}_{j,j+i|z_{j-1,0}} \big)\big).
\end{aligned}
\tag{3.7}
$$

Also, the h-functions (2.7) are redefined as

$$
h_{j+i,j|z_{j-1,0}}(u_{j+i},u_j) \equiv \frac{\partial C_{j,j+i|z_{j-1,0}}(u_j,u_{j+i})}{\partial u_j}, \tag{3.8}
$$

for $i = 1,\dots,d-j$, $j = 1,\dots,d-1$. The estimation procedure for a C-vine is described in Algorithm 2 below.

$\psi_1(\boldsymbol{\theta}_1) = 0$
for $j = 1, \ldots, d$ **do**
  for $k = 1, \ldots, n$ **do**
    $u_{jk} = x_{jk}$.
  end for
end for
for $i = 1, \ldots, d-1$ **do**
  for $k = 1, \ldots, n$ **do**
    $\psi_1(\boldsymbol{\theta}_1) \mathrel{+}= \log c_{1,i+1}(u_{1k}, u_{i+1,k}; \boldsymbol{\theta}_{1,i+1})$
  end for
end for
$\hat{\boldsymbol{\theta}}_1 = \underset{\boldsymbol{\theta}_1}{\text{argmax}}(\psi_1(\boldsymbol{\theta}_1))$
for $j = 1, \ldots, d-1$ **do**
  $\psi_j(\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_{j-1}, \boldsymbol{\theta}_j) = 0$
  for $i = 1, \ldots, d-j$ **do**
    for $k = 1, \ldots, n$ **do**
      $\psi_j(\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_{j-1}, \boldsymbol{\theta}_j) \mathrel{+}= \log c_{j,j+i|z_{j-1,0}}(u_{j|z_{j-1,0},k}, u_{j+i|z_{j-1,0},k}; \boldsymbol{\theta}_{j,j+i|z_{j-1,0}})$
    end for
  end for
  $\hat{\boldsymbol{\theta}}_j = \underset{\boldsymbol{\theta}_j}{\text{argmax}}(\psi_j(\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_{j-1}, \boldsymbol{\theta}_j))$
  if $j == d-1$ **then**
    Stop
  end if
  for $i = 1, \ldots, d-j$ **do**
    for $k = 1, \ldots, n$ **do**
      $u_{j+i|z_{j,0},k} = h_{j+i,j|z_{j-1,0}}(u_{j+i|z_{j-1,0},k}, u_{j|z_{j-1,0},k}, \hat{\boldsymbol{\theta}}_{j,j+i|z_{j-1,0}})$
    end for
  end for
end for

Algorithm 2. SSP for a C-vine

4 Examples

To compare the various estimators' performance, we have carried out asymptotic computations on a couple of simple three-dimensional examples.

Example 4.0.1. Consider the three-dimensional Gaussian distribution

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim \mathcal{N}_3\left(\mathbf{0}, \boldsymbol{SRS}\right), \boldsymbol{S} = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix}, \boldsymbol{R} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}.$$

This distribution can be represented by a D-vine consisting of Gaussian pair-copulae and margins. Among the three possible decompositions, we choose

$$(1) - (2) - (3).$$

In practice, there are scarcely any other models for which it is feasible to do all computations analytically. It is also one of the few distributions the IFM and SP estimators are asymptotically efficient for, as explained below.

The maximum likelihood estimators $\boldsymbol{\alpha}^{ML}$ and $\boldsymbol{\theta}^{ML}$ are of course the empirical standard deviations and correlations, respectively. It follows immediately that the IFM estimator $\boldsymbol{\alpha}^{IFM}$ is the same as the ML estimator. Thus, both $\boldsymbol{\alpha}^{IFM}$ and $\boldsymbol{\theta}^{IFM}$ are asymptotically efficient. Moreover, the SP estimator $\boldsymbol{\theta}^{SP}$ is semiparametrically efficient for the copula parameters $\boldsymbol{\theta}$, as mentioned earlier.

For SSP, we must compute the matrices $\boldsymbol{\mathcal{K}}_C$, $\boldsymbol{\mathcal{J}}_C$ and $\boldsymbol{B}_\theta^{SSP}$, as defined in Section 3.1. The chosen D-vine representation consists of the copulae C_{12}, C_{23} and $C_{13|2}$. These are Gaussian copulae with parameters ρ_{12}, ρ_{23} and the partial correlation $\rho_{13|2} = (\rho_{13} - \rho_{12}\rho_{23})/\sqrt{(1-\rho_{12}^2)(1-\rho_{23}^2)}$, respectively. The covariance matrix $\boldsymbol{V}^{SSP}$ of $\rho_{12},\ \rho_{23}$, and ρ_{13}, in that order (corresponding to the PCC levels), is shown in Appendix A.3, along with $\boldsymbol{V}^{ML}$. We see that $\boldsymbol{V}^{SSP} = \boldsymbol{V}^{ML}$. Hence, $\boldsymbol{V}^{SSP} = \boldsymbol{V}^{SP}$. As the SP estimator is asymptotically efficient for $\boldsymbol{\theta}$, so must the SSP estimator be.

Example 4.0.2. Consider the three-dimensional PCC with exponential margins and Gumbel pair-copulae:

$$f_{123}(x_1, x_2, x_3; \boldsymbol{\lambda}, \boldsymbol{\delta}) = \prod_{j=1}^{3} f_j(x_j;\lambda_j) c_{12}(u_1, u_2; \delta_{12}) c_{23}(u_2, u_3; \delta_{12}) c_{13|2}(u_{1|2}, u_{3|2}; \delta_{13|2}),$$

with

$$
\begin{aligned}
f_j(x_j;\lambda_j) =& \lambda_j \exp\{-\lambda_j x_j\},\\
c_{ij}(u_i,u_j;\delta) =& \exp\{-(\tilde{u}_i^\delta+\tilde{u}_j^\delta)^{1/\delta}\}(u_iu_j)^{-1}(\tilde{u}_i\tilde{u}_j)^{\delta-1}(\tilde{u}_i^\delta+\tilde{u}_j^\delta)^{-2+1/\delta}\\
&\cdot((\tilde{u}_i^\delta+\tilde{u}_j^\delta)^{1/\delta}+\delta-1),\\
u_{i|j} =& \frac{\partial C_{ij}(u_i,u_j;\delta)}{\partial u_j} = \exp\{-(\tilde{u}_i^\delta+\tilde{u}_j^\delta)^{1/\delta}\}u_j^{-1}(\tilde{u}_j)^{\delta-1}(\tilde{u}_i^\delta+\tilde{u}_j^\delta)^{1/\delta-1},
\end{aligned}
$$

where $u_j = 1-\exp\{-\lambda_j x_j\}$ and $\tilde{u}_j = -\log(u_j),\ i,j = 1,2,3$. For various parameter sets, we have computed the covariance matrices by numerical derivation and integration. Since the dependence parameters $\boldsymbol{\delta}$ are our primary interest, we let $\lambda_1 = \lambda_3 = \lambda_3 = 1$ in all sets. Moreover, we let $\delta_{12} = \delta_{23}$. Table 4.1 shows the resulting asymptotic relative efficiencies of the ground and top level parameter estimators, $(\hat{\delta}_{12}, \hat{\delta}_{23})$ and $\delta_{13|2}$, respectively, i.e. the ratios between the variances of ML and the estimators in question. In a Gumbel copula, the dependence increases with the parameter δ. Kendall's τ is 0 when $\delta = 1$ and tends to 1 as $\delta \to \infty$. We see that all three estimators (IFM, SP and SSP) are rather efficient in general. They all lose asymptotic efficiency with increasing dependence at the ground level, i.e. for δ_{12} and δ_{23}, whereas the two latter gain efficiency at the top level. The asymptotic variances of all three estimators actually decrease with increasing dependence at both levels, though not as fast as for ML. Moreover, SSP is overall less efficient than IFM and SP, as expected. However, the difference is quite small at the top level.

	IFM		SP		SSP	
	$\hat{\delta}_{12}$	$\hat{\delta}_{13\|2}$	$\hat{\delta}_{12}$	$\hat{\delta}_{13\|2}$	$\hat{\delta}_{12}$	$\hat{\delta}_{13\|2}$
			$\delta_{12} = \delta_{23} = 1.2$			
$\delta_{13\|2} = 1.2$	0.997	0.997	0.921	0.955	0.904	0.953
$\delta_{13\|2} = 2$	0.985	0.996	0.902	0.984	0.891	0.981
$\delta_{13\|2} = 3$	0.971	0.994	0.846	0.990	0.837	0.987
			$\delta_{12} = \delta_{23} = 2$			
$\delta_{13\|2} = 1.2$	0.995	0.985	0.913	0.851	0.879	0.843
$\delta_{13\|2} = 2$	0.981	0.983	0.896	0.950	0.850	0.936
$\delta_{13\|2} = 3$	0.956	0.969	0.832	0.976	0.815	0.962
			$\delta_{12} = \delta_{23} = 3$			
$\delta_{13\|2} = 1.2$	0.995	0.974	0.912	0.814	0.861	0.808
$\delta_{13\|2} = 2$	0.973	0.954	0.871	0.921	0.843	0.887
$\delta_{13\|2} = 3$	0.944	0.932	0.825	0.951	0.777	0.931

Table 4.1. Asymptotic relative efficiencies of $\hat{\delta}_{12}$ and $\hat{\delta}_{13|2}$ for various parameter sets.

5 Concluding remarks

There are various estimators for the parameters of a pair-copula construction, among those the stepwise semiparametric estimator, which is designed for this particular dependence structure. Although previously suggested, it has never been formally introduced. In this paper, we have presented its asymptotic properties, as well as the estimation algorithm for the two most common types of PCCs, namely D- and C-vines.

Compared to alternatives such as maximum likelihood, inference functions for margins and semiparametric estimation, SSP is in general asymptotically less efficient. However, it is semiparametrically efficient for the Gaussian copula. A toy example involving a three-dimensional D-vine consisting of Gumbel copulae, exhibits the SSP estimator's higher variance relative to the alternatives. Nonetheless, the loss of efficiency is rather low, and decreases with the construction level. To truly compare the alternative estimators' performance, we plan to perform a large simulation study.

One of the main advantages of the SSP estimator, is that it is computationally tractable even in high dimensions, as opposed to its competitors. Moreover, it provides start values required by the other estimators. Finally, determining the pair-copulae of a PCC is a stepwise procedure, that involves parameter estimates from previous levels. The SSP estimator lends itself perfectly to this task.

For simplicity, we have only considered C- and D-vines. Equivalent results are, however, easily obtained for the more general class of regular vines. Moreover, we have assumed the observations to be independent, identically distributed. In practice, the estimation often includes a preliminary step to deal with deviations from these assumptions, for instance GARCH filtration of time series data. The effect of such an additional step on the SSP estimator is a subject for future work.

6 Acknowledgements

This work is funded by Statistics for Innovation, (sfi)^2. I thank my supervisors Arnoldo Frigessi and Kjersti Aas for helpful discussions and comments.

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2).

Bedford, T. and Cooke, R. (2001). Probabilistic density decomposition for conditionally dependent random variables modeled by vines. *Annals of mathematics and Artificial Intelligence*, 32:245–268.

Bedford, T. and Cooke, R. (2002). Vines – a new graphical model for dependent random variables. *Annals of Statistics*, 30(4):1031–1068.

Chollete, L., Heinen, A., and Valdesogo, A. (2009). Modeling international financial returns with a multivariate regime switching copula. *Journal of Financial Econometrics*, 7:437–480.

Claeskens, G. and Hjort, N. L. (2008). *Model Selection and Model Averaging*. Cambridge series in statistical and probibalistic mathematics. Cambridge University Press.

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141–151.

Czado, C. and Min, A. (2010). Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics*, 8:511–546.

Czado, C., Min, A., Baumann, T., and Dakovic, R. (2009). Pair-copula constructions for modeling exchange rate dependence. Submitted for publication.

Fischer, M., Köck, C., Schlüter, S., and Weigert, F. (2007). Multivariate copula models at work: Outperforming the "desert island copula"? Technical Report 79, Universität Erlangen-Nürnberg, Lehrstuhl für Statistik und Ökonometrie.

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika*, 74:549–555.

Genest, C., Gerber, H. U., Goovaerts, M. J., and Laeven, R. J. A. (2009). Editorial to the special issue on modeling and measurement of multivariate risk in insurance and finance. *Insurance: Mathematics and Economics*, 44(2).

Genest, C., Ghoudi, K., and Rivest, L. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82:543–552.

Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, 88:1034–1043.

Genest, C. and Werker, B. (2002). Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. In *Distributions with Given Marginals and Statistical Modelling*, pages 103–112. Kluwer Academic Publishers, Dordrecht.

Hall, P. and Yao, Q. (2005). Approximating conditional distribution functions using dimension reduction. *The Annals of Statistics*, 33:1404–1421.

Heinen, A. and Valdesogo, A. (2009). Asymmetric capm dependence for large dimensions: the canonical vine autoregressive model. CORE Discussion Paper (2009/69).

Hobæk Haff, I., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction – simply useful or too simplistic? *Journal of Multivariate Analysis*, 101:1296–1310.

Joe, H. (1996). *Distributions with Fixed Marginals and Related Topics*, chapter Families of m-variate distributions with given margins and m(m-1)/2 dependence parameters. IMS, Hayward, CA.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94:401–419.

Joe, H., Li, H., and Nikoloulopoulos, A., K. (2010). Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, 101:252–270.

Joe, H. and Xu, J. (1996). The estimation method of inference functions for margins for multivariate models. Technical Report 166, University of British Columbia, Department of Statistics.

Kim, G., Silvapulle, M., and Silvapulle, P. (2007). Comparison of semiparametric and parametric models for estimating copulas. *Computational Statistics and Data Analysis*, 51:2836–2850.

Kolbjørnsen, O. and Stien, M. (2008). D-vine creation of non-gaussian random field. In *Procedings of the Eight International Geostatistics Congress*, pages 399–408. GECAMIN Ltd.

Kurowicka, D. and Cooke, R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley, New York.

Lehmann, E. (2004). *Elements of Large-Sample Theory*. Springer Verlag, New York.

McNeil, A., Frey, R., and Embrechts, P. (2006). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Priceton University Press.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B*, 44:414–422.

Schirmacher, D. and Schirmacher, E. (2008). Multivariate dependence modeling using pair-copulas. Presented at The 2008 ERM Symposium, Chicago.

Shih, J. and Louis, T. (1995). Inferences on the association parameter in copula models for survival data. *Biometrics*, 51:1384–1399.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris*, 8.

Stute, W. (1986). On almost sure convergence of conditional emprircal distribution functions. *The Annals of Probability*, 11:891–901.

Tsukahara, H. (2005). Semiparametric estimation of copula models. *The Canadian Journal of Statistics*, 33:357–375.

A Appendix

A.1 Matrices $\mathcal{K}_\theta$ and $\mathcal{J}_\theta$

As stated in Section 3.1, the matrices $\boldsymbol{\mathcal{K}}_\theta = \mathrm{E}\left(\left(\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{\theta}}\right)^T\right)$ and $\boldsymbol{\mathcal{J}}_\theta = \mathrm{E}\left(-\frac{\partial^2 \boldsymbol{\psi}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right)$ are block diagonal and block lower triangular, respectively, i.e. $\boldsymbol{\mathcal{K}}_{\theta,i,j} = \mathbf{0},\ i \neq j$ and $\boldsymbol{\mathcal{J}}_{\theta,i,j} = \mathbf{0},\ i < j$. This follows from the structure of the ψ-functions, as shown below.

We start with $\boldsymbol{\mathcal{J}}_{\theta,i,j}$, where $i < j$. Then,

$$\boldsymbol{\mathcal{J}}_{\theta,i,j} = \mathrm{E}\left(-\frac{\partial^2 \psi_i(u_1,\ldots,u_d;\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j^T}\right)$$

with

$$\begin{aligned}
&\psi_i(u_1,\ldots,u_d;\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_i) = \\
&\quad \sum_{k=1}^{d-i} \log\left(c_{k,k+i|v_{ki}}(g^1_{k,k+i}(u_k,\ldots,u_{k+i-1};\boldsymbol{\theta}_{k\to k+i-1}),\right. \\
&\qquad\qquad \left. g^2_{k,k+i}(u_{k+1},\ldots,u_{k+i};\boldsymbol{\theta}_{k+1\to k+i});\boldsymbol{\theta}_{k,k+i|v_{ki}})\right).
\end{aligned}$$

Since none of the copulae at level i are functions of the parameters at a following level j,

$$\frac{\partial \psi_i(u_1,\ldots,u_d;\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_j} = \mathbf{0}.$$

Hence, $\boldsymbol{\mathcal{J}}_{\theta,i,j} = \mathbf{0},\ i < j$.

Assume now that $i < j$. Moreover, let $\boldsymbol{u} = (u_1,\ldots,u_d) = (\boldsymbol{u}_{w_{ki}}, \boldsymbol{u}_{-w_{ki}})$. Then,

$$\begin{aligned}
\boldsymbol{\mathcal{K}}_{\theta,i,j} &= \mathrm{E}\left(\left(\frac{\partial \psi_i(u_1,\ldots,u_d;\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i}\right)\left(\frac{\partial \psi_j(u_1,\ldots,u_d;\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}\right)^T\right) \\
&= \int_{\boldsymbol{u}} \frac{\partial}{\partial \boldsymbol{\theta}_i} \sum_{k=1}^{d-i} \log c_{k,k+i|v_{ki}} \frac{\partial}{\partial \boldsymbol{\theta}_j^T} \sum_{l=1}^{d-j} \log c_{l,l+j|v_{lj}} c_{1\ldots d} d\boldsymbol{u} \\
&= \sum_{k=1}^{d-i}\sum_{l=1}^{d-j} \int_{\boldsymbol{u}} \frac{\partial}{\partial \boldsymbol{\theta}_i} \log c_{k,k+i|v_{ki}} \frac{1}{c_{l,l+j|v_{lj}}} \frac{\partial}{\partial \boldsymbol{\theta}_j^T} c_{l,l+j|v_{lj}} c_{1\ldots d} d\boldsymbol{u} \\
&= \sum_{k=1}^{d-i}\sum_{l=1}^{d-j} \int_{\boldsymbol{u}_{w_{ki}}} \frac{\partial}{\partial \boldsymbol{\theta}_i} \log c_{k,k+i|v_{ki}} \int_{\boldsymbol{u}_{-w_{ki}}} \frac{1}{c_{l,l+j|v_{lj}}} \frac{\partial}{\partial \boldsymbol{\theta}_j^T} c_{l,l+j|v_{lj}} c_{1\ldots d} d\boldsymbol{u}_{-w_{ki}} d\boldsymbol{u}_{w_{ki}}.
\end{aligned}$$

Under the conditions of Theorem 1, we may exchange the integration and differentiation in the inner integral. Thus,

$$
\begin{aligned}
&\mathcal{K}_{\theta,i,j} \\
&= \sum_{k=1}^{d-i}\sum_{l=1}^{d-j}\int_{\boldsymbol{u}_{w_{ki}}} \frac{\partial}{\partial \boldsymbol{\theta}_i}\log c_{k,k+i|v_{ki}} \frac{\partial}{\partial \boldsymbol{\theta}_j^T}\left(\int_{\boldsymbol{u}_{-w_{ki}}} \frac{1}{c_{l,l+j|v_{lj}}} c_{l,l+j|v_{lj}} c_{1\ldots d} d\boldsymbol{u}_{-w_{ki}}\right) d\boldsymbol{u}_{w_{ki}} \\
&= \sum_{k=1}^{d-i}\sum_{l=1}^{d-j}\int_{\boldsymbol{u}_{w_{ki}}} \frac{\partial}{\partial \boldsymbol{\theta}_i}\log c_{k,k+i|v_{ki}} \frac{\partial}{\partial \boldsymbol{\theta}_j^T}\left(\int_{\boldsymbol{u}_{-w_{ki}}} c_{1\ldots d} d\boldsymbol{u}_{-w_{ki}}\right) d\boldsymbol{u}_{w_{ki}} \\
&= \sum_{k=1}^{d-i}\sum_{l=1}^{d-j}\int_{\boldsymbol{u}_{w_{ki}}} \frac{\partial}{\partial \boldsymbol{\theta}_i}\log c_{k,k+i|v_{ki}} \frac{\partial}{\partial \boldsymbol{\theta}_j^T} c_{w_{ki}} d\boldsymbol{u}_{w_{ki}}.
\end{aligned}
$$

The pair-copulae composing $c_{w_{ki}}$, situated in levels $1, \ldots, i$, are not functions of parameters from a following level j. Thus, $\frac{\partial}{\partial \boldsymbol{\theta}_j} c_{w_{ki}} = \mathbf{0}$. Consequently, $\mathcal{K}_{\theta,i,j} = \mathbf{0},\ i < j$. The exact same argument can be repeated for $i > j$. Hence, $\mathcal{K}_{\theta,i,j} = \mathbf{0},\ i \neq j$.

A.2 Proof of Theorem 2

Proof. In two dimensions, the SSP estimator is the same as the SP estimator, which was shown to be semiparametrically efficient by **?**. In three dimensions, we have computed the asymptotic covariance matrices for comparison. As shown in Example 4.0.1, the covariance matrices of the SP and SSP estimators, $\boldsymbol{V}^{SP}$ and $\boldsymbol{V}^{SSP}$, respectively, are equal. Thus, the SSP estimator is semiparametrically efficient also for the three-dimensional Gaussian copula.

Assume now that it is true for the $(d-1)$-dimensional Gaussian copula. As the SP estimator is semiparametrically efficient, the asymptotic covariance matrix of the ML estimator $\hat{\boldsymbol{\theta}}^{ML}$ for the Gaussian copula must be the same, regardless of the margins. Moreover, when all margins are normal, the maximum likelihood estimator of the copula parameters is simply the empirical correlation matrix. Adding an extra dimension leaves the remaining estimators unchanged. Hence, the blocks of the covariance matrix corresponding to the $(d-1)$ dimensional sub-model will be the same as for the $(d-1)$ dimensional Gaussian copula. The same argument can repeated for all $(d-1)$ dimensional sub-models, covering all levels but the top. Due to its levelwise structure, the SSP estimator for a given sub-model is unaffected when adding an extra dimension, and so must the corresponding block of its asymptotic covariance matrix be. Accordingly, we must have $\boldsymbol{V}^{SSP}_{1\ldots d-2,1\ldots d-2} = \boldsymbol{V}^{SP}_{1\ldots d-2,1\ldots d-2} = \boldsymbol{V}_{1\ldots d-2,1\ldots d-2}$. Hence, it remains to show that $\boldsymbol{V}^{SSP}_{1\ldots d-2,d-1} = \boldsymbol{V}^{SP}_{1\ldots d-2,d-1}$ and $V^{SSP}_{d-1,d-1} = V^{SP}_{d-1,d-1}$, related to the estimators $\hat{\theta}^{SP}_{1d|v_{1d}}$ and $\hat{\theta}^{SSP}_{1d|v_{1d}}$ for the top level copula. According to Theorem 1 of Tsukahara (2005) and Theorem 1, respectively,

$$\sqrt{n}\left(\hat{\theta}^{SP}_{1d|v_{1d}} - \theta_{1d|v_{1d}}\right) \xrightarrow{d} Z_{SP} \sim \mathcal{N}\left(0, V^{SP}_{d-1,d-1}\right)$$

and

$$\sqrt{n}\left(\hat{\theta}^{SSP}_{1d|v_{1d}} - \theta_{1d|v_{1d}}\right) \xrightarrow{d} Z_{SSP} \sim \mathcal{N}\left(0, V^{SSP}_{d-1,d-1}\right),$$

as $n \to \infty$. Now, define $\boldsymbol{U}_n = (\boldsymbol{U}_{n1}, \ldots, \boldsymbol{U}_{nn})$, with $\boldsymbol{U}_{nj} = (F_{1n}(X_{1j}), \ldots, F_{dn}(X_{dj})),\ j = 1, \ldots, d$, and let

$$\boldsymbol{\Psi}^{SP}(\boldsymbol{U}_n; \hat{\boldsymbol{\theta}}^{SP}) = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{\phi}_{SP}\left(\boldsymbol{U}_n; \hat{\boldsymbol{\theta}}^{SP}\right) = \boldsymbol{0}$$
$$\boldsymbol{\Psi}^{SSP}(\boldsymbol{U}_n; \hat{\boldsymbol{\theta}}^{SSP}) = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{\phi}_{SSP}\left(\boldsymbol{U}_n; \hat{\boldsymbol{\theta}}^{SSP}\right) = \boldsymbol{0},$$

be the estimating equations of the SP and SSP estimators, respectively. Further, let

$$\Psi(\boldsymbol{U}_n; \boldsymbol{\theta}) = \Psi^{SP}_{\frac{d(d-1)}{2}}(\boldsymbol{U}_n; \boldsymbol{\theta}) = \Psi^{SSP}_{\frac{d(d-1)}{2}}(\boldsymbol{U}_n; \boldsymbol{\theta}) = \frac{1}{n}\sum_{k=1}^{n} \frac{\partial}{\partial \theta_{1d|v_{1d}}} \log\left(c_{1\ldots d}(\boldsymbol{U}_n; \boldsymbol{\theta})\right).$$

According to Theorem 1 of Tsukahara (2005),

$$\begin{aligned}\Psi\left(\boldsymbol{U}_n; \hat{\boldsymbol{\theta}}^{SP}\right) =& \Psi(\boldsymbol{U}_n; \boldsymbol{\theta}) + \frac{\partial \Psi(\boldsymbol{U}_n; \boldsymbol{\theta})}{\partial \theta_{1d|v_{1d}}}\left(\hat{\theta}^{SP}_{1d|v_{1d}} - \theta_{1d|v_{1d}}\right) \\ &+ \frac{\partial \Psi(\boldsymbol{U}_n; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T_{1\to d-2}}\left(\hat{\boldsymbol{\theta}}^{SP}_{1\to d-2} - \boldsymbol{\theta}_{1\to d-2}\right) + o_P\left(\frac{1}{n}\right) = 0.\end{aligned}$$

Likewise, using Theorem 1, one obtains

$$\begin{aligned}\Psi\left(\boldsymbol{U}_n; \hat{\boldsymbol{\theta}}^{SSP}\right) =& \Psi(\boldsymbol{U}_n; \boldsymbol{\theta}) + \frac{\partial \Psi(\boldsymbol{U}_n; \boldsymbol{\theta})}{\partial \theta_{1d|v_{1d}}}\left(\hat{\theta}^{SSP}_{1d|v_{1d}} - \theta_{1d|v_{1d}}\right) \\ &+ \frac{\partial \Psi(\boldsymbol{U}_n; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T_{1\to d-2}}\left(\hat{\boldsymbol{\theta}}^{SSP}_{1\to d-2} - \boldsymbol{\theta}_{1\to d-2}\right) + o_P\left(\frac{1}{n}\right) = 0.\end{aligned}$$

Hence,

$$\begin{aligned}&\sqrt{n}\left(\hat{\theta}^{SSP}_{1d|v_{1d}} - \hat{\theta}^{SP}_{1d|v_{1d}}\right) \\ &= \frac{\boldsymbol{A}_1}{A_2}\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{SP}_{1\to d-2} - \hat{\boldsymbol{\theta}}^{SSP}_{1\to d-2}\right) + o_P\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

with $\boldsymbol{A}_1 = \frac{\partial \Psi(\boldsymbol{U}_n;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T_{1\to d-2}} = \frac{1}{n}\sum_{k=1}^{n} \frac{\partial^2}{\partial \theta_{1d|v_{1d}} \partial \boldsymbol{\theta}^T_{1\to d-2}} \log\left(c_{1\ldots d}(\boldsymbol{U}_n; \boldsymbol{\theta})\right)$ and $A_2 = \frac{\partial \Psi(\boldsymbol{U}_n;\boldsymbol{\theta})}{\partial \theta_{1d|v_{1d}}} = \frac{1}{n}\sum_{k=1}^{n} \frac{\partial^2}{\partial \theta^2_{1d|v_{1d}}} \log\left(c_{1\ldots d}(\boldsymbol{U}_n; \boldsymbol{\theta})\right)$. Now, according to the assumption,

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{SSP}_{1\to d-2} - \boldsymbol{\theta}_{1\to d-2}\right) \xrightarrow{d} \boldsymbol{Y} \sim \mathcal{N}_{\frac{d(d-1)}{2}-1}\left(\boldsymbol{0}, \boldsymbol{V}_{1\ldots d-2, 1\ldots d-2}\right),\ n \to \infty.$$

Thus,

$$\begin{aligned}\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{SP}_{1\to d-2} - \hat{\boldsymbol{\theta}}^{SSP}_{1\to d-2}\right) &= \sqrt{n}\left(\hat{\boldsymbol{\theta}}^{SP}_{1\to d-2} - \boldsymbol{\theta}_{1\to d-2}\right) - \sqrt{n}\left(\hat{\boldsymbol{\theta}}^{SSP}_{1\to d-2} - \boldsymbol{\theta}_{1\to d-2}\right) \\ &\xrightarrow{p} \boldsymbol{0},\ n \to \infty.\end{aligned}$$

Moreover, under the assumed conditions,

$$\begin{aligned}
\boldsymbol{A}_1 &\xrightarrow{p} \mathrm{E}\left(\frac{\partial^2}{\partial \theta_{1d|v_{1d}} \partial \boldsymbol{\theta}_{1\to d-2}^T} \log\left(c_{1\dots d}(\boldsymbol{U}_n; \boldsymbol{\theta})\right)\right) = -\boldsymbol{\mathcal{I}}^T_{\theta,1\dots d-2,d-1} \\
A_2 &\xrightarrow{p} \mathrm{E}\left(\frac{\partial^2}{\partial \theta^2_{1d|v_{1d}}} \log\left(c_{1\dots d}(\boldsymbol{U}_n; \boldsymbol{\theta})\right)\right) = -\mathcal{I}_{\theta,d-1,d-1},
\end{aligned}$$

as $n \to \infty$. Hence,

$$\sqrt{n}\left(\hat{\theta}^{SSP}_{1d|v_{1d}} - \theta^{SP}_{1d|v_{1d}}\right) \xrightarrow{p} 0,$$

which means that $Z_{SP} \overset{d}{=} Z_{SSP}$. In other words, $V^{SSP}_{d-1,d-1} = V^{SP}_{d-1,d-1}$. Moreover,

$$\begin{aligned}
\boldsymbol{V}^{SSP}_{1\dots d-2,d-1} =& -\frac{1}{\mathcal{I}_{\theta,d}} \boldsymbol{V}_{1\dots d-2,1\dots d-2} \boldsymbol{\mathcal{I}}_{\theta,1\dots d-2,d-1} + \frac{1}{\mathcal{I}_{\theta,d}} \boldsymbol{\mathcal{J}}^{-1}_{\theta,1\dots d-2,d-1} \boldsymbol{B}^{SSP}_{1\dots d-2,d-1} \\
V^{SSP}_{d-1,d-1} =& \frac{1}{\mathcal{I}_{\theta,d}} + \frac{B^{SSP}_{d-1,d-1}}{\mathcal{I}^2_{\theta,d}} + \frac{1}{\mathcal{I}_{\theta,d}} \boldsymbol{\mathcal{I}}^T_{\theta,1\dots d-2,d-1} \boldsymbol{V}_{1\dots d-2,1\dots d-2} \boldsymbol{\mathcal{I}}_{\theta,1\dots d-2,d-1} \\
& - \frac{2}{\mathcal{I}^2_{\theta,d}} \boldsymbol{\mathcal{I}}^T_{\theta,1\dots d-2,d-1} \boldsymbol{\mathcal{J}}^{-1}_{\theta,1\dots d-2,d-1} \boldsymbol{B}^{SSP}_{1\dots d-2,d-1} \\
=& \frac{1}{\mathcal{I}_{\theta,d}} + \frac{B^{SSP}_{d-1,d-1}}{\mathcal{I}^2_{\theta,d}} - \frac{1}{\mathcal{I}_{\theta,d}} \boldsymbol{\mathcal{I}}^T_{\theta,1\dots d-2,d-1} \boldsymbol{V}_{1\dots d-2,1\dots d-2} \boldsymbol{\mathcal{I}}_{\theta,1\dots d-2,d-1} \\
& - 2\boldsymbol{\mathcal{I}}^T_{\theta,1\dots d-2,d-1} \boldsymbol{V}^{SSP}_{1\dots d-2,d-1}.
\end{aligned}$$

Correspondingly for SP,

$$\begin{aligned}
\boldsymbol{V}^{SP}_{1\dots d-2,d-1} =& -\frac{1}{\mathcal{I}_{\theta,d}} \boldsymbol{V}_{1\dots d-2,1\dots d-2} \boldsymbol{\mathcal{I}}_{\theta,1\dots d-2,d-1} \\
& + \frac{1}{\mathcal{I}_{\theta,d}} \boldsymbol{\mathcal{I}}^{(\theta)}_{1\dots d-2,1\dots d-2} \left(\boldsymbol{B}^{SP}_{1\dots d-2,d-1} - \frac{B^{SP}_{d-1,d-1}}{\mathcal{I}_{\theta,d}} \boldsymbol{\mathcal{I}}_{\theta,1\dots d-2,d-1}\right) \\
V^{SP}_{d-1,d-1} =& \frac{1}{\mathcal{I}_{\theta,d}} + \frac{B^{SP}_{d-1,d-1}}{\mathcal{I}^2_{\theta,d}} + \frac{1}{\mathcal{I}_{\theta,d}} \boldsymbol{\mathcal{I}}^T_{\theta,1\dots d-2,d-1} \boldsymbol{V}_{1\dots d-2,1\dots d-2} \boldsymbol{\mathcal{I}}_{\theta,1\dots d-2,d-1} \\
& - \frac{2}{\mathcal{I}^2_{\theta,d}} \boldsymbol{\mathcal{I}}^T_{\theta,1\dots d-2,d-1} \boldsymbol{\mathcal{I}}^{(\theta)}_{1\dots d-2,1\dots d-2} \left(\boldsymbol{B}^{SP}_{1\dots d-2,d-1}\right. \\
& \left. - \frac{B^{SP}_{d-1,d-1}}{\mathcal{I}_{\theta,d}} \boldsymbol{\mathcal{I}}_{\theta,1\dots d-2,d-1}\right) \\
=& \frac{1}{\mathcal{I}_{\theta,d}} + \frac{B^{SP}_{d-1,d-1}}{\mathcal{I}^2_{\theta,d}} - \frac{1}{\mathcal{I}_{\theta,d}} \boldsymbol{\mathcal{I}}^T_{\theta,1\dots d-2,d-1} \boldsymbol{V}_{1\dots d-2,1\dots d-2} \boldsymbol{\mathcal{I}}_{\theta,1\dots d-2,d-1} \\
& - 2\boldsymbol{\mathcal{I}}^T_{\theta,1\dots d-2,d-1} \boldsymbol{V}^{SP}_{1\dots d-2,d-1}.
\end{aligned}$$

Since the estimating equation for the top copula parameter is the same for SP and SSP, $B^{SSP}_{d-1,d-1} = B^{SP}_{d-1,d-1}$. Moreover, $V^{SSP}_{d-1,d-1} = V^{SP}_{d-1,d-1}$. Consequently, $\boldsymbol{V}^{SSP}_{1\dots d-2,d-1} = \boldsymbol{V}^{SP}_{1\dots d-2,d-1}$. □

A.3 Covariance matrices from Example 4.0.1

The asymptotic covariance matrix of the ML estimator is given by

$$\boldsymbol{V}^{ML} = \frac{1}{2}\begin{pmatrix} 2(1-\rho_{12}^2)^2 & v_1 & v_2 \\ v_1 & 2(1-\rho_{23}^2)^2 & v_3 \\ v_2 & v_3 & 2(1-\rho_{13}^2)^2 \end{pmatrix},$$

with

$$\begin{aligned} v_1 &= 2\rho_{13}(1-\rho_{12}^2)(1-\rho_{23}^2) - \rho_{12}\rho_{23}|\boldsymbol{R}|, \\ v_2 &= 2\rho_{23}(1-\rho_{12}^2)(1-\rho_{13}^2) - \rho_{12}\rho_{13}|\boldsymbol{R}|, \\ v_3 &= 2\rho_{12}(1-\rho_{13}^2)(1-\rho_{23}^2) - \rho_{13}\rho_{23}|\boldsymbol{R}|. \end{aligned}$$

For the SSP estimator, we have

$$\boldsymbol{V}^{SSP} = \boldsymbol{\mathcal{J}}_\theta^{-1}\boldsymbol{\mathcal{K}}_\theta\left(\boldsymbol{\mathcal{J}}_\theta^{-1}\right)^T + \boldsymbol{\mathcal{J}}_\theta^{-1}\boldsymbol{B}_\theta^{SSP}\left(\boldsymbol{\mathcal{J}}_\theta^{-1}\right)^T,$$

with

$$\boldsymbol{\mathcal{K}}_\theta = \begin{pmatrix} \frac{1+\rho_{12}^2}{(1-\rho_{12}^2)^2} & \frac{k_1}{(1-\rho_{12}^2)(1-\rho_{23}^2)} & 0 \\ \frac{k_1}{(1-\rho_{12}^2)(1-\rho_{23}^2)} & \frac{1+\rho_{23}^2}{(1-\rho_{23}^2)^2} & 0 \\ 0 & 0 & \frac{|\boldsymbol{R}|+2(\rho_{13}-\rho_{12}\rho_{23})^2}{|\boldsymbol{R}|^2} \end{pmatrix},$$

$$\boldsymbol{\mathcal{J}}_\theta = \begin{pmatrix} \frac{1+\rho_{12}^2}{(1-\rho_{12}^2)^2} & 0 & 0 \\ 0 & \frac{1+\rho_{23}^2}{(1-\rho_{23}^2)^2} & 0 \\ \frac{j_1}{|\boldsymbol{R}|^2} & \frac{j_2}{|\boldsymbol{R}|^2} & \frac{|\boldsymbol{R}|+2(\rho_{13}-\rho_{12}\rho_{23})^2}{|\boldsymbol{R}|^2} \end{pmatrix},$$

where

$$\boldsymbol{B}_\theta^{SSP} = \begin{pmatrix} \frac{\rho_{12}^2(1+\rho_{12}^2)}{(1-\rho_{12}^2)^2} & \frac{b_1}{2(1-\rho_{12}^2)(1-\rho_{23}^2)} & \frac{(\rho_{13}-\rho_{12}\rho_{23})b_2}{2(1-\rho_{12}^2)|\boldsymbol{R}|} \\ \frac{b_1}{2(1-\rho_{12}^2)(1-\rho_{23}^2)} & \frac{\rho_{23}^2(1+\rho_{23}^2)}{(1-\rho_{23}^2)^2} & \frac{(\rho_{13}-\rho_{12}\rho_{23})b_3}{2(1-\rho_{23}^2)|\boldsymbol{R}|} \\ \frac{(\rho_{13}-\rho_{12}\rho_{23})b_2}{2(1-\rho_{12}^2)|\boldsymbol{R}|} & \frac{(\rho_{13}-\rho_{12}\rho_{23})b_3}{2(1-\rho_{23}^2)|\boldsymbol{R}|} & \frac{(1+\rho_{13}^2)(\rho_{13}-\rho_{12}\rho_{23})^2}{|\boldsymbol{R}|^2} \end{pmatrix},$$

$$\begin{aligned} k_1 &= (\rho_{13}-\rho_{12}\rho_{23})(|\boldsymbol{R}| + \rho_{13}^2 - \rho_{12}^2\rho_{23}^2), \\ j_1 &= -\rho_{23}|\boldsymbol{R}| + 2(\rho_{12}-\rho_{13}\rho_{23})(\rho_{13}-\rho_{12}\rho_{23}), \\ j_2 &= -\rho_{12}|\boldsymbol{R}| + 2(\rho_{13}-\rho_{12}\rho_{23})(\rho_{23}-\rho_{12}\rho_{13}), \end{aligned}$$

$$
\begin{aligned}
b_1 =& \rho_{12}\rho_{23}(1+\rho_{12}^2+\rho_{13}^2+\rho_{23}^2)\left(1-\frac{2}{1-\rho_{12}^2}-\frac{2}{1-\rho_{23}^2}\right) \\
&+2\rho_{23}(1+\rho_{12}^2)\frac{\rho_{12}+\rho_{13}\rho_{23}}{1-\rho_{12}^2}+2\rho_{12}(1+\rho_{23}^2)\frac{\rho_{23}+\rho_{12}\rho_{13}}{1-\rho_{23}^2}, \\
b_2 =& \rho_{12}(1+\rho_{12}^2+\rho_{13}^2+\rho_{23}^2)\left(1-\frac{2}{1-\rho_{12}^2}\right)+2(1+\rho_{12}^2)\frac{\rho_{12}+\rho_{13}\rho_{23}}{1-\rho_{12}^2}, \\
b_3 =& \rho_{23}(1+\rho_{12}^2+\rho_{13}^2+\rho_{23}^2)\left(1-\frac{2}{1-\rho_{23}^2}\right)+2(1+\rho_{23}^2)\frac{\rho_{23}+\rho_{12}\rho_{13}}{1-\rho_{23}^2}.
\end{aligned}
$$

The resulting covariance matrix is

$$
\boldsymbol{V}^{SSP} = \frac{1}{2}\begin{pmatrix} 2(1-\rho_{12}^2)^2 & v_1 & v_2 \\ v_1 & 2(1-\rho_{23}^2)^2 & v_3 \\ v_2 & v_3 & 2(1-\rho_{13}^2)^2 \end{pmatrix} = \boldsymbol{V}^{ML}.
$$