Parameter estimation for pair-copula constructions

INGRID HOBÆK HAFF

Statistics for Innovation, Norwegian Computing Center, PB 114 Blindern, NO-0373 Oslo, Norway.
E-mail: ingrid@nr.no

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, that is, maximum likelihood, inference functions for margins and semiparametric estimation, SSP is in general asymptotically less efficient. As we show in a few examples, 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

1. Introduction

The last decades' technological revolution have considerably increased the relevance of multi-variate 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.* [15]). Accordingly, a number of hier-archical, copula-based structures have been proposed, among those the pair-copula construction (PCC) of Joe [22], further studied and considered by Bedford and Cooke [2,3], Kurowicka and Cooke [30] and Aas *et al.* [1].

A PCC is a treelike construction, built from pair-copulae with conditional distributions as their two arguments (see Figure 1). The number of conditioning variables is zero at the ground, and increases by one for each level, to ensure coherence of the construction. Despite its simple structure, the PCC is highly flexible and covers a wide range of complex dependencies (Joe *et al.* [25], Hobæk Haff *et al.* [21]). After Aas *et al.* [1] set it in an inferential context, it has made several appearances in the literature (Fischer *et al.* [13], Chollete *et al.* [6], Heinen and Valdesogo [19], Schirmacher and Schirmacher [35], Czado and Min [9], Kolbjørnsen and Stien [29], Czado *et al.* [10,11]), 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
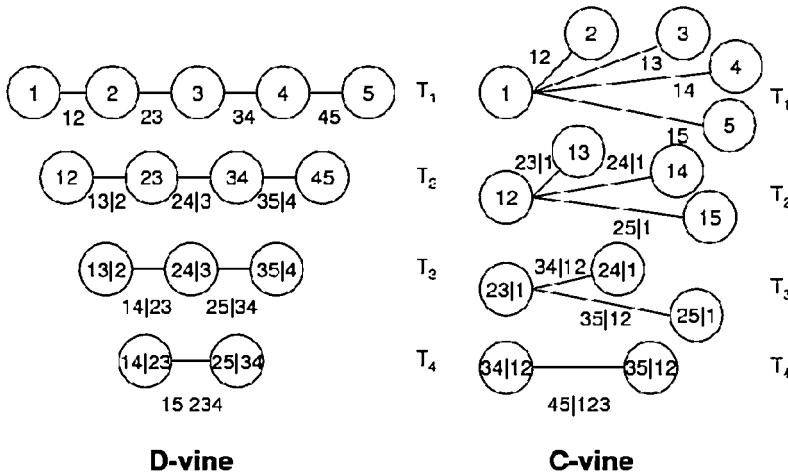
Figure 1. Five-dimensional D-vine (to the left) and C-vine (to the right).

large collection of estimators proposed for that model class, such as moments type procedures, based on, for instance, the matrix of pairwise Kendall's tau coefficients (Clayton [8], Oakes [34], Genest [14], Genest and Rivest [17]). 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.* [1]), 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 $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of n independent d-variate stochastic vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$, 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 [37]). Letting $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ denote the parameters of the margins and copula, respectively, the joint probability density function (p.d.f.) may then be ex-

Table 1. Notation overview

Symbol	$j = 0$	$j = 1$	$j = 2, \ldots, d-1$
v_{ij}^*	–	$\varnothing$	$\{i+1, \ldots, i+j-1\}$
w_{ij}^*	i	$\{i, i+j\}$	$\{i, v_{ij}, i+j\}$
$\boldsymbol{\theta}_j^*$	–	$\{\boldsymbol{\theta}_{i,i+1}: i = 1, \ldots, d-1\}$	$\{\boldsymbol{\theta}_{s,s+t\mid v_{st}}: \mid v_{st} \mid = j-1\}$
$\boldsymbol{\theta}_{i \to i+j}^*$	–	$\boldsymbol{\theta}_{i,i+1}$	$\{\boldsymbol{\theta}_{s,s+t\mid v_{st}}: (s, s+t) \in w_{ij}\}$
z_{ij}^{**}	$\varnothing$	$\{i+1\}$	$\{1, \ldots, j-1, j+i\}$
$\boldsymbol{\theta}_{j\_i}^{**}$	$\varnothing$	$\varnothing$	$\{\boldsymbol{\theta}_{s,s+t\mid z_{s-1,0}}: (s, s+t) \in z_{ji}\}$

*: $i = 1, \ldots, d-j$, **: $i = 0, \ldots, d-j$.

pressed as (McNeil *et al.* [32], page 197)

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

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

Define the index sets $v_{ij} = \{i+1, \ldots, i+j-1\}$, $w_{ij} = \{i, v_{ij}, i+j\}$, for $1 \leq i \leq d-j, 1 \leq j \leq d-1$, with $v_{i1} = \varnothing$, and $w_{i0} = i$. Thus, for a vector $\mathbf{a} = (a_1, \ldots, a_d)$, we write $\mathbf{a}_{v_{ij}} = (a_{i+1}, \ldots, a_{i+j-1})$ and $\mathbf{a}_{w_{ij}} = (a_i, \ldots, a_{i+j})$. Further, for an index k and a set of indices v, with $k \notin v$, let $F_{k\mid v}$ be the conditional c.d.f. of X_k given $\mathbf{X}_v = \mathbf{x}_v$, and $c_{i,i+j\mid v_{ij}}$ the copula density corresponding to the conditional distribution $F_{i,i+j\mid v_{ij}}$ of (X_i, X_{i+j}) given $\mathbf{X}_{v_{ij}} = \mathbf{x}_{v_{ij}}$. Finally, let $\boldsymbol{\theta}_{i,i+j\mid v_{ij}}$ be the parameters of the copula density $c_{i,i+j\mid v_{ij}}$, and define $\boldsymbol{\theta}_{i \to i+j} = \{\boldsymbol{\theta}_{s,s+t\mid v_{st}}: (s, s+t) \in w_{ij}\}$, with $\boldsymbol{\theta}_{i \to i} = \varnothing$, and $\boldsymbol{\theta}_j = \{\boldsymbol{\theta}_{s,s+t\mid v_{st}}: \mid v_{st} \mid = j-1\}$, where $\mid \cdot \mid$ denotes the cardinality (i.e., $\boldsymbol{\theta}_j$ gathers all parameters from level j of the structure). Table 1 gives an overview of the notation. For a so-called D-vine (Bedford and Cooke [2,3]), the joint p.d.f. (1) can now be written as (Aas *et al.* [1])

$$f_{1\ldots d}(x_1, \ldots, x_d; \boldsymbol{\alpha}, \boldsymbol{\theta})$$

$$= \prod_{l=1}^{d} f_l(x_l; \boldsymbol{\alpha}_l) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j\mid v_{ij}}(F_{i\mid v_{ij}}(x_i\mid \mathbf{x}_{v_{ij}}; \boldsymbol{\alpha}_{w_{i,j-1}}, \boldsymbol{\theta}_{i \to i+j-1}), \tag{2}$$

$$F_{i+j\mid v_{ij}}(x_{i+j}\mid \mathbf{x}_{v_{ij}}; \boldsymbol{\alpha}_{w_{i+1,j-1}}, \boldsymbol{\theta}_{i+1 \to i+j}); \boldsymbol{\theta}_{i,i+j\mid v_{ij}}).$$

In four dimensions, this becomes

$$f_{1234}(x_1, x_2, x_3, x_4; \boldsymbol{\alpha}, \boldsymbol{\theta}) = f_1(x_1; \boldsymbol{\alpha}_1) \cdot f_2(x_2; \boldsymbol{\alpha}_2) \cdot f_3(x_3; \boldsymbol{\alpha}_3) \cdot f_4(x_4; \boldsymbol{\alpha}_4)$$

$$\cdot c_{12}(F_1(x_1; \boldsymbol{\alpha}_1), F_2(x_2; \boldsymbol{\alpha}_2); \boldsymbol{\theta}_{12}) \cdot c_{23}(F_2(x_2; \boldsymbol{\alpha}_2), F_3(x_3; \boldsymbol{\alpha}_3); \boldsymbol{\theta}_{23})$$

$$\cdot c_{34}(F_3(x_3; \boldsymbol{\alpha}_3), F_4(x_4; \boldsymbol{\alpha}_4); \boldsymbol{\theta}_{34}) \tag{3}$$

$$\cdot c_{13|2}(F_{1|2}(x_1|x_2; \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\theta}_{12}), F_{3|2}(x_3|x_2; \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\theta}_{23}); \boldsymbol{\theta}_{13|2})$$

$$\cdot c_{24|3}(F_{2|3}(x_2|x_3; \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\theta}_{23}), F_{4|3}(x_4|x_3; \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4, \boldsymbol{\theta}_{34}); \boldsymbol{\theta}_{24|3})$$

$$\cdot c_{14|23}(F_{1|23}(x_1|x_2, x_3; \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\theta}_{12}, \boldsymbol{\theta}_{23}, \boldsymbol{\theta}_{13|2}),$$

$$F_{4|23}(x_4|x_2, x_3; \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4, \boldsymbol{\theta}_{23}, \boldsymbol{\theta}_{34}, \boldsymbol{\theta}_{24|3}); \boldsymbol{\theta}_{14|23}).$$

For simplicity, we will start by assuming that the distribution in question is a D-vine, represented to the left in Figure 1 for $d = 5$. Similar results can be obtained for C-vines (Section 3.3) and other regular vines (Bedford and Cooke [2,3]). We also assume that the PCC is of a simplified form (Hobæk Haff *et al.* [21]), that is, 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 $\mathbf{x}_{v_{ij}}$. Without this assumption, inference on these models is not doable in practice.

2. Multivariate copula estimators

As previously mentioned, a PCC is a multivariate copula. Hence, one may estimate its parameters with well-known methods, such as maximum likelihood or the two-step inference functions for margins and semiparametric estimators.

2.1. Maximum likelihood (ML) estimator

Supposing the model is true, the ML estimator is a natural choice, due to its asymptotic efficiency and other advantageous characteristics. According to (2), the log-likelihood function of a D-vine is given by

$$l(\boldsymbol{\alpha}, \boldsymbol{\theta}; \mathbf{x}) = \sum_{k=1}^{n} \log(f_{1\ldots d}(x_{1k}, \ldots, x_{dk}; \boldsymbol{\alpha}, \boldsymbol{\theta}))$$

$$= \sum_{k=1}^{n} \sum_{l=1}^{d} \log(f_l(x_{lk}; \boldsymbol{\alpha}_l))$$

$$+ \sum_{k=1}^{n} \sum_{j=1}^{d-1} \sum_{i=1}^{d-j} \log(c_{i,i+j|v_{ij}}(F_{i|v_{ij}}(x_i|\mathbf{x}_{v_{ij}}; \boldsymbol{\alpha}_{w_{i,j-1}}, \boldsymbol{\theta}_{i \to i+j-1}), \tag{4}$$

$$F_{i+j|v_{ij}}(x_{i+j}|\mathbf{x}_{v_{ij}}; \boldsymbol{\alpha}_{w_{i+1,j-1}}, \boldsymbol{\theta}_{i+1 \to i+j}); \boldsymbol{\theta}_{i,i+j|v_{ij}}))$$

$$= l_M(\boldsymbol{\alpha}; \mathbf{x}) + l_C(\boldsymbol{\alpha}, \boldsymbol{\theta}; \mathbf{x}),$$

where $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. The ML estimator $\hat{\boldsymbol{\theta}}^{\mathrm{ML}}$ is obtained by maximising the above log-likelihood function over all parameters, $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$, simultaneously. Under the additional assumptions (M1)–(M8) of Lehmann [31] (pages 499–501), this corresponds to solving the set of estimating equations (one equation per parameter), $\frac{1}{n} \sum_{k=1}^{n} \boldsymbol{\phi}^{\mathrm{ML}}(X_{1k}, \ldots, X_{dk}; \hat{\boldsymbol{\alpha}}^{\mathrm{ML}}, \hat{\boldsymbol{\theta}}^{\mathrm{ML}}) = \mathbf{0}$,

which is a vector of functions, with elements

$$\phi_l^{\mathrm{ML}}(x_1, \ldots, x_d; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \frac{\partial \log(f_{1\ldots d}(x_1, \ldots, x_d; \boldsymbol{\alpha}, \boldsymbol{\theta}))}{\partial \boldsymbol{\alpha}_l},$$

$$\phi_{d+(j-1)(d-j/2)+i}^{\mathrm{ML}}(x_1, \ldots, x_d; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \frac{\partial \log(f_{1\ldots d}(x_1, \ldots, x_d; \boldsymbol{\alpha}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_{i,i+j|v_{ij}}}$$

(5)

for $l = 1, \ldots, d$, $i = 1, \ldots, d-j$, $j = 1, \ldots, d-1$. Define $\mathcal{I}$ as the corresponding Fisher information matrix

$$
\mathcal{I} = \mathrm{E}\left(\left(\frac{\partial \log(f_{1\ldots d}(\mathbf{X}; \boldsymbol{\alpha}, \boldsymbol{\theta}))}{\partial(\boldsymbol{\alpha}, \boldsymbol{\theta})} \right) \left(\frac{\partial \log(f_{1\ldots d}(\mathbf{X}; \boldsymbol{\alpha}, \boldsymbol{\theta}))}{\partial(\boldsymbol{\alpha}, \boldsymbol{\theta})} \right)^T \right)
$$

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

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

$$
\mathcal{I}^{-1} = \begin{pmatrix} \mathcal{I}^{(\alpha)} & \mathcal{I}^{(\alpha,\theta)} \\ \left(\mathcal{I}^{(\alpha,\theta)}\right)^T & \mathcal{I}^{(\theta)} \end{pmatrix},
$$

$$
\mathcal{I}^{(\alpha)} = (\mathcal{I}_{\alpha} - \mathcal{I}_{\alpha,\theta}\mathcal{I}_{\theta}^{-1}\mathcal{I}_{\alpha,\theta}^T)^{-1},
$$

$$
\mathcal{I}^{(\alpha,\theta)} = -\mathcal{I}^{(\alpha)}\mathcal{I}_{\alpha,\theta}\mathcal{I}_{\theta}^{-1},
$$

$$
\mathcal{I}^{(\theta)} = \mathcal{I}_{\theta}^{-1} + \mathcal{I}_{\theta}^{-1}\mathcal{I}_{\alpha,\theta}^T\mathcal{I}^{(\alpha)}\mathcal{I}_{\alpha,\theta}\mathcal{I}_{\theta}^{-1}.
$$

(6)

It is well-known that under the mentioned conditions, the estimator $\hat{\boldsymbol{\theta}}^{\mathrm{ML}}$ is consistent for $\boldsymbol{\theta}$ and asymptotically normal, that is,

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

$$\mathbf{V}^{\mathrm{ML}} = \mathcal{I}^{(\theta)}.$$

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 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, to which one must add the ones of the margins. Finding the global maximum in such a high-dimensional space is numerically challenging, even with more elaborate optimisation schemes, such as the modified Newton–Raphson method with first and second order derivatives. It will in any case be highly time consuming, so 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, that is, 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 [7]). However, it is in general non-robust to larger KL-divergences from the true model.

2.2. 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.

2.2.1. *Inference function for margins (IFM) estimator*

The IFM estimator, introduced by Joe [23,24], addresses the computational inefficiency of the ML estimator by performing the estimation in two steps. First, one estimates $\boldsymbol{\alpha}$ by maximising the term l_M from (4). The resulting estimates $\hat{\boldsymbol{\alpha}}^{\text{IFM}}$ are plugged into the term l_C to obtain $\hat{\boldsymbol{\theta}}^{\text{IFM}}$.

Under conditions (M1)–(M8) (see Section 2.1), this corresponds to solving

$$\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\phi}^{\text{IFM}}(X_{1k},\ldots,X_{dk};\hat{\boldsymbol{\alpha}}^{\text{IFM}},\hat{\boldsymbol{\theta}}^{\text{IFM}})=\mathbf{0},$$

with elements

$$\boldsymbol{\phi}_l^{\text{IFM}}(x_l;\boldsymbol{\alpha}_l)=\frac{\partial\log(f_l(x_l;\boldsymbol{\alpha}_l))}{\partial\boldsymbol{\alpha}_l},$$

$$\boldsymbol{\phi}_{d+(j-1)(d-j/2)+i}^{\text{IFM}}(x_1,\ldots,x_d;\boldsymbol{\alpha},\boldsymbol{\theta})=\frac{\partial\log(f_{1\ldots d}(x_1,\ldots,x_d;\boldsymbol{\alpha},\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}_{i,i+j|v_{ij}}} \tag{7}$$

for $l=1,\ldots,d$, $i=1,\ldots,d-j$, $j=1,\ldots,d-1$. Compared to the ML equations (5), the full log-p.d.f., $\log f_{1\ldots d}$, is replaced with the marginal log-p.d.f.s, $\log f_j$, for the estimation of $\boldsymbol{\alpha}$.

Consider a four-dimensional D-vine (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 $\boldsymbol{\alpha}=(v_1,v_2,v_3,v_4)$ and $\boldsymbol{\theta}=(\rho_{12},\rho_{23},\rho_{34},\rho_{13|2},\rho_{24|3},\rho_{14|23},v_{12},v_{23},v_{34},v_{13|2},v_{24|3},v_{14|23})$. IFM estimation of this model starts with a separate estimation of v_i, $i=1,2,3,4$, margin by margin. The next step is to optimise $l_C(\hat{v}_1,\ldots,\hat{v}_4,\boldsymbol{\theta};\mathbf{x})$ over $\boldsymbol{\theta}$, l_C being the sum of the log-copula densities in line 3 to 8 of (3), over all observations.

Define the matrix $\mathcal{K}_\alpha$ with

$$\mathcal{K}_{\alpha,i,j}=\mathrm{E}(\boldsymbol{\phi}_i^{\text{IFM}}(X_i;\boldsymbol{\alpha}_i)\boldsymbol{\phi}_j^{\text{IFM}}(X_j;\boldsymbol{\alpha}_j))$$

$$=\mathrm{E}\left(\left(\frac{\partial\log(f_i(X_i;\boldsymbol{\alpha}_i))}{\partial\boldsymbol{\alpha}_i}\right)\left(\frac{\partial\log(f_j(X_j;\boldsymbol{\alpha}_j))}{\partial\boldsymbol{\alpha}_j}\right)^T\right),\qquad i,j=1,\ldots,d,$$

and the block diagonal matrix $\mathcal{J}_\alpha$ with

$$\mathcal{J}_{\alpha,i,i}=\mathrm{E}\left(-\frac{\partial}{\partial\boldsymbol{\alpha}_i^T}\boldsymbol{\phi}_i^{\text{IFM}}(X_i;\boldsymbol{\alpha}_i)\right)=\mathrm{E}\left(-\frac{\partial^2\log(f_i(X_i;\boldsymbol{\alpha}_i))}{\partial\boldsymbol{\alpha}_i\,\partial\boldsymbol{\alpha}_i^T}\right)=\mathcal{K}_{\alpha,i,i},$$

each block corresponding to one of the margins. If all margins are one-parameter families, $\mathcal{K}_\alpha$ and $\mathcal{J}_\alpha$ are $d\times d$ matrices. More generally, their dimension depends on the number of parameters of each margin.

Joe [24] showed that under the mentioned conditions, the estimator $\hat{\boldsymbol{\theta}}^{\text{IFM}}$ is consistent for $\boldsymbol{\theta}$, as well as asymptotically normal:

$$
\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\theta}}^{\text{IFM}} - \boldsymbol{\theta}) &\xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}^{\text{IFM}}), \\
\mathbf{V}^{\text{IFM}} &= \mathcal{I}_\theta^{-1} + \mathcal{I}_\theta^{-1} \mathcal{I}_{\alpha,\theta}^T \mathcal{J}_\alpha^{-1} \mathcal{K}_\alpha \mathcal{J}_\alpha^{-1} \mathcal{I}_{\alpha,\theta} \mathcal{I}_\theta^{-1}.
\end{aligned}
\tag{8}
$$

The above covariance matrix is obtained by replacing $\mathcal{I}^{(\alpha)}$ in (6) with the asymptotic covariance matrix $\mathcal{J}_\alpha^{-1} \mathcal{K}_\alpha \mathcal{J}_\alpha^{-1}$ of $\hat{\boldsymbol{\alpha}}^{\text{IFM}}$. This quantifies the loss of asymptotic efficiency from discarding information the dependence structure might have on the margins. Several studies, including Joe [24] and Kim *et al.* [27], have demonstrated that unless the dependence between the variables is extreme, this loss tends to be rather small. That is also the impression from Examples 4.2 and 4.3 (Section 4). Moreover, the IFM method is computationally 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.

2.2.2. *Semiparametric (SP) estimator for copula parameters*

Just like IFM, the SP estimator is a two-step estimator, treating the margins separately. It was introduced by Genest *et al.* [16], and for the censored case by Shih and Louis [36]. Later, it was generalised by Tsukahara [40]. Aas *et al.* [1] suggest this estimator for PCCs.

As seen from (2), the pair-copula arguments at the ground level of a pair-copula construction (T_1 in Figure 1) are marginal distributions F_i. From the second level, they are conditional distributions, whose conditioning set increases by one with each level. These conditional distributions may however be written as functions of the margins. Let i, j be distinct indices, that is, $i \neq j$, and v a nonempty set of indices, all from $\{1, \ldots, d\}$, such that $i, j \notin v$. Then, in a simplified pair-copula construction (Joe [22])

$$
F_{i|v \cup j}(x_i | \mathbf{x}_{v \cup j}) = \left. \frac{\partial C_{ij|v}(u_i, u_j)}{\partial u_j} \right|_{u_i = F_{i|v}(x_i | \mathbf{x}_v), u_j = F_{j|v}(x_j | \mathbf{x}_v)}.
\tag{9}
$$

Thus, by extracting one of the variables j from the conditioning set $v \cup j$, one can express $F_{i|v \cup j}$ as a function of two conditional distributions $F_{i|v}$ and $F_{j|v}$ with one conditioning variable less. 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 which conditional distributions are needed. At level $j \geq 2$ of a D-vine, these are the pairs $(F_{i|v_{ij}}(x_i | \mathbf{x}_{v_{ij}}), F_{i+j|v_{ij}}(x_{i+j} | \mathbf{x}_{v_{ij}}))$, $i = 1, \ldots, d - j$. Now, define the functions

$$
\begin{aligned}
h_{i,i+j|v_{ij}}(u_i, u_{i+j}) &\equiv \frac{\partial C_{i,i+j|v_{ij}}(u_i, u_{i+j})}{\partial u_{i+j}}, \\
h_{i+j,i|v_{ij}}(u_{i+j}, u_i) &\equiv \frac{\partial C_{i,i+j|v_{ij}}(u_i, u_{i+j})}{\partial u_i}
\end{aligned}
\tag{10}
$$

for $i = 1, \ldots, d - j$, $j = 1, \ldots, d - 1$. Using (9), one obtains

$$F_{i|v_{ij}}(x_i|\mathbf{x}_{v_{ij}})$$
$$= h_{i,i+j-1|v_{i,j-1}}(F_{i|v_{i,j-1}}(x_i|\mathbf{x}_{v_{i,j-1}}), F_{i+j-1|v_{i,j-1}}(x_{i+j-1}|\mathbf{x}_{v_{i,j-1}})),$$
$$F_{i+j|v_{ij}}(x_{i+j}|\mathbf{x}_{v_{i,j-1}})$$
$$= h_{i+j,i+1|v_{i+1,j-1}}(F_{i+j|v_{i+1,j-1}}(x_{i+j}|\mathbf{x}_{v_{i+1,j-1}}), F_{i+1|v_{i+1,j-1}}(x_{i+1}|\mathbf{x}_{v_{i+1,j-1}})),$$

which are functions of conditional distributions constituting the arguments of the copulae from the previous level, $j - 1$. As one continues this recursion, one achieves, as earlier mentioned, functions of the margins $F_i, \ldots, F_{i+j}$. Since these are needed in the asymptotics, we denote them $g_{i,i+j}^{(1)}$ and $g_{i,i+j}^{(2)}$, and explicitly define them below. Note however, that for all practical purposes, such as in the estimation algorithm (Algorithm 1 in Supplement A of Hobæk Haff [20]), one will use the nested h-functions from (10). Define

$$g_{i,i+j}^{(1)}(u_i, \ldots, u_{i+j-1})$$
$$\equiv F_{i|v_{ij}}(F_i^{-1}(u_i)|F_{i+1}^{-1}(u_{i+1}), \ldots, F_{i+j-1}^{-1}(u_{i+j-1})),$$
$$g_{i,i+j}^{(2)}(u_{i+1}, \ldots, u_{i+j})$$
$$\equiv F_{i+j|v_{ij}}(F_{i+j}^{-1}(u_{i+j})|F_{i+1}^{-1}(u_{i+1}), \ldots, F_{i+j-1}^{-1}(u_{i+j-1}))$$

(11)

for $i = 1, \ldots, d - j$, $j = 1, \ldots, d - 1$. Now, one may rewrite (2) as:

$$f_{1\ldots d}(x_1, \ldots, x_d; \boldsymbol{\alpha}, \boldsymbol{\theta})$$
$$= \prod_{l=1}^{d} f_l(x_l; \boldsymbol{\alpha}_l)$$
$$\cdot \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|v_{ij}}\big(g_{i,i+j}^{(1)}(F_i(x_i; \boldsymbol{\alpha}_i), \ldots, F_{i+j-1}(x_{i+j}; \boldsymbol{\alpha}_{i+j-1}); \boldsymbol{\theta}_{i \rightarrow i+j-1}),$$

(12)

$$g_{i,i+j}^{(2)}(F_{i+1}(x_{i+1}; \boldsymbol{\alpha}_{i+1}), \ldots, F_{i+j}(x_{i+j}; \boldsymbol{\alpha}_{i+j}); \boldsymbol{\theta}_{i+1 \rightarrow i+j});$$
$$\boldsymbol{\theta}_{i,i+j|v_{ij}}\big).$$

Recall that IFM estimates $\hat{\boldsymbol{\theta}}^{\text{IFM}}$ are obtained by plugging the estimated marginal parameters $\hat{\boldsymbol{\alpha}}^{\text{IFM}}$ into the function l_C. Semiparametric estimation consists in replacing the parametric marginal c.d.f.s $u_j = F_j(x_j; \boldsymbol{\alpha}_j)$ in l_C with the corresponding empirical ones

$$u_{jn} = F_{jn}(x_j) = \frac{1}{n+1} \sum_{k=1}^{n} I(x_{jk} \leq x_j),$$

$$I(A) = \begin{cases} 1, & \text{if } A \text{ is true,} \\ 0, & \text{otherwise.} \end{cases}$$

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

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

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

Returning to the four-dimensional Student's t-vine of Section 2.2.1, SP estimation of this model 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 $\hat{\boldsymbol{\theta}}^{\mathrm{SP}}$ is obtained by maximising $l_{C,P}(\mathbf{x}; \boldsymbol{\theta})$, in this case

$$
\begin{aligned}
\sum_{k=1}^{n}\big(&\log(c_{12}(u_{1k,n}, u_{2k,n}; \boldsymbol{\theta}_{12})) + \cdots \\
&+ \log\big(c_{14|23}\big(g_{14}^{(1)}(u_{1k,n}, \ldots, u_{3k,n}; \boldsymbol{\theta}_{1\to3}), g_{14}^{(2)}(u_{2k,n}, \ldots, u_{4k,n}; \boldsymbol{\theta}_{2\to4}); \boldsymbol{\theta}_{14|23})\big)\big),
\end{aligned}
$$

with $\boldsymbol{\theta}_{1\to3} = (\boldsymbol{\theta}_{12}, \boldsymbol{\theta}_{23}, \boldsymbol{\theta}_{13|2})$ and $\boldsymbol{\theta}_{2\to4} = (\boldsymbol{\theta}_{23}, \boldsymbol{\theta}_{34}, \boldsymbol{\theta}_{24|3})$, over $\boldsymbol{\theta}$.

In addition to the assumptions made for the ML estimator, assume that $c_{1\ldots d}$ fulfills condition (A.1) from Tsukahara [40]. Then, the procedure corresponds to solving

$$
\frac{1}{n} \sum_{k=1}^{n} \boldsymbol{\phi}^{\mathrm{SP}}(F_{1n}(X_{1k}), \ldots, F_{1n}(X_{dk}); \hat{\boldsymbol{\theta}}^{\mathrm{SP}}) = \mathbf{0},
$$

with

$$
\boldsymbol{\phi}_{(j-1)(d-j/2)+i}^{\mathrm{SP}}(u_1, \ldots, u_d; \boldsymbol{\theta}) = \frac{\partial \log(c_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_{i,i+j|v_{ij}}} \tag{13}
$$

for $i = 1, \ldots, d-j$, $j = 1, \ldots, d-1$.

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

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

Further, define

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

where $\mathbf{A} = \mathrm{Cov}(\mathbf{Y}, \mathbf{Z})$, for two stochastic vectors $\mathbf{Y}$ and $\mathbf{Z}$, is the matrix with elements $A_{ij} = \mathrm{Cov}(Y_i, Z_j) + \mathrm{Cov}(Y_j, Z_i)$. The matrix $\mathbf{B}_\theta^{SP}$ quantifies the effect of replacing the parametric marginal c.d.f.s with empirical ones. According to Genest *et al.* [16] and later shown by Tsukahara [40], $\hat{\boldsymbol{\theta}}^{SP}$ is, under the mentioned conditions, consistent and asymptotically normal:

$$
\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\theta}}^{SP} - \boldsymbol{\theta}) &\xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}^{SP}), \\
\mathbf{V}^{SP} &= \mathcal{I}_\theta^{-1} + \mathcal{I}_\theta^{-1} \mathbf{B}_\theta^{SP} \mathcal{I}_\theta^{-1}.
\end{aligned}
\tag{14}
$$

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.* [27]). If either of the latter two produce estimates that are rather different from the former, it indicates that the chosen margins or copulae are inadequate for the data.

Computationally, SP is comparable to IFM. Hence, for high-dimensional $\boldsymbol{\theta}$, although faster than ML, this procedure will require good start values, and may still be too demanding for PCCs.

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 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 nonparametric ones. The idea is to estimate the PCC parameters level by level, conditioning on the parameters from preceding levels of the structure. Define

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

with

$$\psi_j(u_1, \ldots, u_d; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_j)$$

$$= \sum_{i=1}^{d-j} \log\left(c_{i,i+j|v_{ij}}\left(g_{i,i+j}^{(1)}(u_i, \ldots, u_{i+j-1}; \boldsymbol{\theta}_{i \to i+j-1}),\right.\right. \tag{15}$$

$$\left.\left. g_{i,i+j}^{(2)}(u_{i+1}, \ldots, u_{i+j}; \boldsymbol{\theta}_{i+1 \to i+j}); \boldsymbol{\theta}_{i,i+j|v_{ij}}\right)\right)$$

for $j = 1, \ldots, d-1$. Hence, $l_{C,P,j}$ is the sum over all log pair-copula densities up to, and including, level j. To obtain the parameter estimates $\hat{\boldsymbol{\theta}}_j^{\text{SSP}}$ for level j, one plugs the estimates $\hat{\boldsymbol{\theta}}_1^{\text{SSP}}, \ldots, \hat{\boldsymbol{\theta}}_{j-1}^{\text{SSP}}$ 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 $\frac{1}{n} \sum_{k=1}^n \boldsymbol{\phi}^{\text{SSP}}(F_{1n}(X_{1k}), \ldots, F_{dn}(X_{dk}); \hat{\boldsymbol{\theta}}^{\text{SSP}}) = \mathbf{0}$, with

$$\boldsymbol{\phi}_{(j-1)(d-j/2)+i}^{\text{SSP}}(u_1, \ldots, u_d; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_j) = \frac{\partial}{\partial \boldsymbol{\theta}_{i,i+j|v_{ij}}} \sum_{l=1}^j \psi_l(u_1, \ldots, u_d; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_l)$$

$$= \frac{\partial}{\partial \boldsymbol{\theta}_{i,i+j|v_{ij}}} \psi_j(u_1, \ldots, u_d; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_j) \tag{16}$$

for $i = 1, \ldots, d-j$, $j = 1, \ldots, d-1$. Compared to the SP equations (13), 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 (Supplement A of Hobæk Haff [20]). 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 2.2.2. As in the SP procedure, one computes the pseudo-observations $u_{ik,n} = F_{in}(x_{ik})$, $i = 1, 2, 3, 4$, $k = 1, \ldots, n$. One starts with 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 for level 2, $u_{i|i+1,k,n} = h_{i,i+1}(u_{ik,n}, u_{i+1,k,n}; \hat{\rho}_{i,i+1}, \hat{\nu}_{i,i+1}))$ and $u_{i+2|i+1,k,n} = h_{i+2,i+1}(u_{i+2,k,n}, u_{i+1,k,n}; \hat{\rho}_{i+1,i+2}, \hat{\nu}_{i+1,i+2}))$, $i = 1, 2, 3$, $k = 1, \ldots, n$, by plugging the resulting estimates into the adequate h-functions (10). 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, to obtain $(\hat{\rho}_{14|23}, \hat{\nu}_{14|23})$, one optimises $\sum_{k=1}^n \log(c_{14|23}(u_{1|23,k,n}, u_{4|23,k,n}; \rho_{14|23}, \nu_{14|23}))$.

When some of the copulae share parameters, the procedure is a little different. Let us for instance consider 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 (see for instance Min and Czado [33]). The correlation parameters of these copulae are now the corresponding partial correlations $(\rho_{12}, \rho_{23}, \rho_{34}, \rho_{13|2}, \rho_{24|3}, \rho_{14|23})$. However, the degrees of freedom

parameter is shared. More specifically, it is ν for the three copulae at the ground level, $\nu + 1$ at level 2 and $\nu + 2$ for the top level copula. The SSP estimation procedure is now as follows. Having computed the pseudo-observations, one maximises the level 1 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 calculates the copula arguments for level 2 as described above. At the second level, one estimates ρ_{13} and ρ_{24}, which are not shared by $c_{13|2}$ and $c_{24|3}$. 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.* [32] (page 231). The main purpose of the PCC is to model pairs with different behaviour. 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 [40]. The last four of these are the standard conditions for the ML estimator, but on the score functions (16). Further, define

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

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

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 [40]. Further, let $|\boldsymbol{\theta}_{ij|v_{ij}}| = l_{ij}$ be the number of parameters of the pair-copula $C_{i,i+j|v_{ij}}$. For the SSP estimator, Condition (A.1) may then be phrased in the following way (note that a subscript 'j' on ϕ is missing in Tsukahara [40]).

Condition 1. *For each* $\boldsymbol{\theta}$, $\boldsymbol{\psi}_{ij,\theta} = (\psi_{ij,\theta,1}, \ldots, \psi_{ij,\theta,l_{ij}})$ *and* $\boldsymbol{\psi}_{ij,\theta,u_k} = (\psi_{ij,\theta,u_k,1}, \ldots, \psi_{ij,\theta,u_k,l_{ij}})$, $j = 1, \ldots, d-1, i = 1, \ldots, d-j$ *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*

$$|\psi_{ij,\theta,m}(\mathbf{u}; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_j)| \leq \prod_{l=1}^{d} r_{ij,l}(u_l),$$

$$|\psi_{ij,\theta,u_k,m}(\mathbf{u}; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_j)| \leq \tilde{r}_{ij,k}(u_k) \prod_{l \neq k} r_{ij,l}(u_l)$$

for $k, l = 1, \ldots, d$, $j = 1, \ldots, d-1$, $i = 1, \ldots, d-j$, $m = 1, \ldots, l_{ij}$, with

$$\int \left(\prod_{l=1}^{d} r_{ij,l}(u_l) \right)^2 \, dC_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta}) < \infty,$$

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

When none of the pair-copulae share parameters, Condition 1 becomes a condition on each of them, individually.

Once more let $\mathbf{U}$ be distributed according to $C_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta})$, as well as $\boldsymbol{\psi}_\theta = (\boldsymbol{\psi}_{11,\theta}, \ldots, \boldsymbol{\psi}_{1,d-1,\theta})$ and $\boldsymbol{\psi}_{\theta,u_j} = (\boldsymbol{\psi}_{11,\theta,u_j}, \ldots, \boldsymbol{\psi}_{1,d-1,\theta,u_j})$. Define

$$\mathbf{W}_j^{\mathrm{SSP}}(\mathbf{U}; \boldsymbol{\theta}) = \int \frac{\partial}{\partial u_j} \boldsymbol{\phi}^{\mathrm{SSP}}(u_1, \ldots, u_d; \boldsymbol{\theta}) I(U_j \leq u_j) \, dC_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta})$$

$$= \int \boldsymbol{\psi}_{\theta,u_j}(u_1, \ldots, u_d; \boldsymbol{\theta}) I(U_j \leq u_j) \, dC_{1\ldots d}(u_1, \ldots, u_d; \boldsymbol{\theta})$$

and the matrix

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

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

Moreover, define the two matrices

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

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

where the blocks $\mathcal{K}_{\theta,i,j} = \mathrm{E}((\frac{\partial \psi_i}{\partial \theta_i})(\frac{\partial \psi_j}{\partial \theta_j})^T)$ and $\mathcal{J}_{\theta,i,j} = -\mathrm{E}(\frac{\partial^2 \psi_i}{\partial \theta_i \, \partial \theta_j^T})$, $i, j = 1, \ldots, d-1$, correspond to each of the construction's levels. The block diagonal and block lower triangular forms of $\mathcal{K}_\theta$ and $\mathcal{J}_\theta$, 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 Fisher matrix $\mathcal{I}_\theta$ in the last rows of $\mathcal{K}_\theta$ and $\mathcal{J}_\theta$. If all pair-copulae are from one-parameter families, then $\mathcal{K}_\theta$ and $\mathcal{J}_\theta$ 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 step-wise semiparametric estimator.

Theorem 1. *Under Condition* 1, *as well as Conditions* (A.2)–(A.5) *of Tsukahara* [40], *the SSP estimator* $\hat{\boldsymbol{\theta}}^{\mathrm{SSP}}$ *is consistent for* $\boldsymbol{\theta}$ *and asymptotically normal*:

$$
\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\theta}}^{\mathrm{SSP}} - \boldsymbol{\theta}) &\xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}^{\mathrm{SSP}}), \\
\mathbf{V}^{\mathrm{SSP}} &= \mathcal{J}_\theta^{-1} \mathcal{K}_\theta (\mathcal{J}_\theta^{-1})^T + \mathcal{J}_\theta^{-1} \mathbf{B}_\theta^{\mathrm{SSP}} (\mathcal{J}_\theta^{-1})^T .
\end{aligned}
\tag{17}
$$

Proof. Theorem 1 follows directly Theorem 1 of Tsukahara [40], with the estimating equations (16). Note that Theorem 1 of Tsukahara [40] is valid for the multiparameter case $m > 1$, despite some misprints and imprecisions in the original paper. Specifically, Condition (A.1) is assumed valid for every element of the vector $\boldsymbol{\phi}$ of estimating equations, that is, a subscript '*j*' is needed. Proposition 3 holds for $m > 1$ thanks to the Cramer–Wold device. The rest of the argument works for $m > 1$, using $\|\cdot\|$ instead of $|\cdot|$ for the norm (Tsukahara [41]). $\qquad\square$

In order to construct confidence intervals for $\boldsymbol{\theta}$, one needs a consistent estimate of $\mathbf{V}^{\mathrm{SSP}}$. As noted in Tsukahara [40], one may estimate this covariance matrix consistently by replacing expectations and variances in (17) by sample equivalents, and plugging in the estimate $\hat{\boldsymbol{\theta}}^{\mathrm{SSP}}$. More specifically, letting $\boldsymbol{\psi}_{\theta,\theta} = \frac{\partial}{\partial \boldsymbol{\theta}^T} \boldsymbol{\psi}_\theta \mathbf{W}^{\mathrm{SSP}} = \sum_{j=1}^d \mathbf{W}_j^{\mathrm{SSP}}$, and u_{ik}, $i = 1, \ldots, d$, $k = 1, \ldots, n$, are the pseudo-observations,

$$
\hat{\mathbf{V}}^{\mathrm{SSP}} = \hat{\mathcal{J}}_\theta^{-1} \hat{\mathcal{K}}_\theta (\hat{\mathcal{J}}_\theta^{-1})^T + \hat{\mathcal{J}}_\theta^{-1} \hat{\mathbf{B}}_\theta^{\mathrm{SSP}} (\hat{\mathcal{J}}_\theta^{-1})^T ,
$$

with

$$
\hat{\mathcal{K}}_\theta = \frac{1}{n} \sum_{k=1}^n \boldsymbol{\psi}_\theta(u_{1k}, \ldots, u_{dk}; \hat{\boldsymbol{\theta}}^{\mathrm{SSP}})(\boldsymbol{\psi}_\theta(u_{1k}, \ldots, u_{dk}; \hat{\boldsymbol{\theta}}^{\mathrm{SSP}}))^T ,
$$

$$
\hat{\mathcal{J}}_\theta = -\frac{1}{n} \sum_{k=1}^n \boldsymbol{\psi}_{\theta,\theta}(u_{1k}, \ldots, u_{dk}; \hat{\boldsymbol{\theta}}^{\mathrm{SSP}}),
$$

$$
\hat{\mathbf{B}}_\theta^{\mathrm{SSP}} = \frac{1}{n} \sum_{k=1}^n \mathbf{W}^{\mathrm{SSP}}(u_{1k}, \ldots, u_{dk}; \hat{\boldsymbol{\theta}}^{\mathrm{SSP}})(\mathbf{W}^{\mathrm{SSP}}(u_{1k}, \ldots, u_{dk}; \hat{\boldsymbol{\theta}}^{\mathrm{SSP}}))^T
$$

$$
+ \frac{1}{n} \sum_{k=1}^n \sum_{j=1}^d \boldsymbol{\psi}_\theta(u_{1k}, \ldots, u_{dk}; \hat{\boldsymbol{\theta}}^{\mathrm{SSP}})(\mathbf{W}^{\mathrm{SSP}}(u_{1k}, \ldots, u_{dk}; \hat{\boldsymbol{\theta}}^{\mathrm{SSP}}))^T .
$$

In most cases, there is no analytic expression for the derivatives $\boldsymbol{\psi}_\theta$ and $\boldsymbol{\psi}_{\theta,\theta}$, but they can be approximated numerically. However, the computation of the $\mathbf{W}_j$-vectors involves d-dimensional integrals, which is more problematic. In practice, one will not be able to compute the above co-variance matrix estimate for $d > 3$. Instead, one will have to resort to some resampling technique, such as parametric bootstrap from $C_{1...d}(\cdot; \hat{\boldsymbol{\theta}}^{\mathrm{SSP}})$, as described in Example 4.4 of Section 4.

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

The proof is given in Appendix A.2.

In general, the stepwise semiparametric estimator $\hat{\boldsymbol{\theta}}^{\mathrm{SSP}}$ is asymptotically less efficient than $\hat{\boldsymbol{\theta}}^{\mathrm{SP}}$, 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 to produce start values for the SP estimator. Further, a substantial difference between SSP and SP estimates may be a sign that the copulae are unsuitable. Hence, one may use the SSP estimator to assess the sensitivity to the chosen copulae. 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, for instance using the parametric margins as in IFM estimation. That particular estimator was in fact proposed by Joe and Xu [26].

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 function l_C with the non-parametric ones, while keeping the parametric forms of the conditional distributions, that is, the g-functions (11). The resulting estimator is robust toward misspecification of the margins, but not of the pair-copulae. By replacing also the conditional distributions with nonparametric versions, one would reduce this sensitivity to chosen pair-copulae preceding in the structure. One

possibility is the empirical conditional distribution proposed by Stute [39]:

$$F_{i|v,n}(x_i|\mathbf{x}_v) = \frac{n}{n+1} \frac{\sum_{k=1}^{n} I(x_{ik} \leq x_i) K_s(\mathbf{x}_v - \mathbf{x}_{v,k})}{\sum_{k=1}^{n} K_s(\mathbf{x}_v - \mathbf{x}_{v,k})}, \qquad K_s(\mathbf{y}) = \frac{1}{s^l} K\left(\frac{\mathbf{y}}{s}\right), \qquad (18)$$

where l is the dimension of $\mathbf{x}_v$, K is a kernel function on $\mathbb{R}^l$ and s the bandwidth parameter. The definition (18) is slightly modified here to avoid boundary problems in 0 and 1. Provided $h \to 0$ and $ns^l \to \infty$ as $n \to \infty$, it converges almost surely to the true conditional distribution, though at a rather slow pace of order $(ns^l)^{1/2}$. The quality of the estimates will therefore significantly decrease with the level number. Alternative versions of the empirical conditional distribution function, such as the one proposed by Hall and Yao [18], share this unfortunate property.

Recall that the conditional distributions of interest are recursions of h-functions (10), which, in turn, are conditional distributions of uniform variables with a conditioning set of length one. These functions can therefore be estimated nonparametrically by (18) with $l = 1$. Seemingly, one can exploit this to avoid the curse of dimensionality. However, the two arguments of the h-functions are again h-functions from 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 $(ns^l)^{1/2}$.

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 C-vines (see Figure 1) and more general regular vines (though computation of the log-likelihood function is more complex (Dissmann *et al.* [12])).

The p.d.f. of a C-vine is given by (Aas *et al.* [1])

$$
\begin{aligned}
& f_{1\ldots d}(x_1, \ldots, x_d; \boldsymbol{\alpha}, \boldsymbol{\theta}) \\
& = \prod_{l=1}^{d} f_l(x_l; \boldsymbol{\alpha}_l) \\
& \quad \cdot \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i|z_{j-1,0}}(F_{j|z_{j-1,0}}(x_j|\mathbf{x}_{z_{j-1,0}}; \boldsymbol{\alpha}_{z_{j0}}, \boldsymbol{\theta}_{j\_0}), \\
& \qquad\qquad F_{j+i|z_{j-1,0}}(x_{j+i}|\mathbf{x}_{z_{j-1,0}}; \boldsymbol{\alpha}_{z_{ji}}, \boldsymbol{\theta}_{j\_i}); \boldsymbol{\theta}_{j,j+i|z_{j-1,0}}),
\end{aligned}
\tag{19}
$$

where $z_{ji} = \{1, \ldots, j-1, j+i\}$ and $\boldsymbol{\theta}_{j\_i} = \{\boldsymbol{\theta}_{s,s+t|z_{s-1,0}}: (s, s+t) \in z_{ji}\}$, for $0 \leq i \leq d - j$, $0 \leq j \leq d - 1$, with $z_{0i} = \varnothing$, $z_{1i} = \{i+1\}$ and $\boldsymbol{\theta}_{0\_i} = \boldsymbol{\theta}_{1\_i} = \varnothing$. Hence, the log-likelihood

function of n independent observations from a C-vine is

$$
\begin{aligned}
l(\boldsymbol{\alpha}, \boldsymbol{\theta}; \mathbf{x}) &= \sum_{k=1}^{n} \log(f_{1\ldots d}(x_{1k}, \ldots, x_{dk}; \boldsymbol{\alpha}, \boldsymbol{\theta})) \\
&= \sum_{k=1}^{n} \sum_{l=1}^{d} \log(f_l(x_{lk}; \boldsymbol{\alpha}_l)) \\
&\quad + \sum_{k=1}^{n} \sum_{j=1}^{d-1} \sum_{i=1}^{d-j} \log(c_{j,j+i|z_{j-1,0}}(F_{j|z_{j-1,0}}(x_j|\mathbf{x}_{z_{j-1,0}}; \boldsymbol{\alpha}_{z_{j0}}, \boldsymbol{\theta}_{j\_0}), \tag{20}
\end{aligned}
$$

$$
F_{j+i|z_{j-1,0}}(x_{j+i}|\mathbf{x}_{z_{j-1,0}}; \boldsymbol{\alpha}_{z_{ji}}, \boldsymbol{\theta}_{j\_i}); \boldsymbol{\theta}_{j,j+i|z_{j-1,0}}))
$$

$$
= l_M(\boldsymbol{\alpha}; \mathbf{x}) + l_C(\boldsymbol{\alpha}, \boldsymbol{\theta}; \mathbf{x}).
$$

Replacing l_C from (4) with l_C from (20), one retrieves the results from Section 2 for C-vines. To achieve the SSP estimator, one must simply replace the psi-function (15) in the estimating equations (16) with

$$
\begin{aligned}
&\psi_j(u_1, \ldots, u_d; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_j) \\
&= \sum_{i=1}^{d-j} \log(c_{j,j+i|z_{j-1,0}}(F_{j|z_{j-1,0}}(x_j|\mathbf{x}_{z_{j-1,0}}; \boldsymbol{\alpha}_{z_{j0}}, \boldsymbol{\theta}_{j\_0}), \tag{21}
\end{aligned}
$$

$$
F_{j+i|z_{j-1,0}}(x_{j+i}|\mathbf{x}_{z_{j-1,0}}; \boldsymbol{\alpha}_{z_{ji}}, \boldsymbol{\theta}_{j\_i}); \boldsymbol{\theta}_{j,j+i|z_{j-1,0}})).
$$

Also, the h-functions (10) are redefined as

$$
h_{j+i,j|z_{j-1,0}}(u_{j+i}, u_j) \equiv \frac{\partial C_{j,j+i|z_{j-1,0}}(u_j, u_{j+i})}{\partial u_j} \tag{22}
$$

for $i = 1, \ldots, d - j$, $j = 1, \ldots, d - 1$. The estimation procedure for a C-vine is described in Algorithm 2 (Supplement A of Hobæk Haff [20]).

4. Examples

To compare the four estimators' performance, we have carried out asymptotic computations on a few examples (Examples 4.1 to 4.3). We have also fitted a D-vine to a set of precipitation series, using each of the estimators (Example 4.4).

Example 4.1. Consider the three-dimensional Gaussian distribution

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

This distribution can be represented by a D-vine consisting of Gaussian pair-copulae and margins, more specifically

$$f_{123}(x_1, x_2, x_3; \mathbf{S}, \mathbf{R}) = c(u_1, u_2; \rho_{12})c(u_2, u_3; \rho_{23})c(u_{1|2}, u_{3|2}; \rho_{13|2})\prod_{i=1}^{3} f(x_i; \sigma_i),$$

where

$$c(u_i, u_j; \rho) = \frac{\exp\{-\rho/(2(1-\rho^2))(\rho\Phi^{-1}(u_i)^2 + \rho\Phi^{-1}(u_j)^2 - 2\Phi^{-1}(u_i)\Phi^{-1}(u_j))\}}{\sqrt{1-\rho^2}}$$

$$f(x; \sigma) = \frac{\exp\{-x^2/(2\sigma^2)\}}{\sqrt{2\pi}\sigma}, \qquad \rho_{13|2} = \frac{\rho_{13} - \rho_{12}\rho_{23}}{\sqrt{(1-\rho_{12}^2)(1-\rho_{23}^2)}},$$

$$u_{i|j} = h_{ij}(u_i, u_j; \rho) = \Phi\left(\frac{\Phi^{-1}(u_i) - \rho\Phi^{-1}(u_j)}{\sqrt{1-\rho^2}}\right),$$

with $u_i = \Phi(x_i)$, Φ being the c.d.f. of the standard Gaussian distribution. Note that this is one of the three possible decompositions of f_{123}.

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 ML estimators $\hat{\boldsymbol{\alpha}}^{\text{ML}}$ and $\hat{\boldsymbol{\theta}}^{\text{ML}}$ are of course the empirical standard deviations and correlations, respectively. It is easily verified that for this particular model, the IFM estimators $\hat{\boldsymbol{\alpha}}^{\text{IFM}}$ and $\hat{\boldsymbol{\theta}}^{\text{IFM}}$ are identical to $\hat{\boldsymbol{\alpha}}^{\text{ML}}$ and $\hat{\boldsymbol{\theta}}^{\text{ML}}$. Thus, they are asymptotically efficient. Moreover, the SP estimator $\hat{\boldsymbol{\theta}}^{\text{SP}}$ is semiparametrically efficient for $\boldsymbol{\theta}$, as shown by Klaassen and Wellner [28].

For SSP, we must compute the matrices $\mathcal{K}_\theta$, $\mathcal{J}_\theta$ and $\mathbf{B}_\theta^{\text{SSP}}$, defined in Section 3.1. The covariance matrix $\mathbf{V}^{\text{SSP}}$ of ρ_{12}, ρ_{23} and ρ_{13}, in that order (corresponding to the PCC levels), is shown in Appendix A.3, along with $\mathbf{V}^{\text{ML}}$. We see that $\mathbf{V}^{\text{SSP}} = \mathbf{V}^{\text{ML}} = \mathbf{V}^{\text{SP}}$. As the SP estimator is asymptotically semiparametrically efficient for $\boldsymbol{\theta}$, so must the SSP estimator be.

Example 4.2. Consider the three-dimensional PCC with exponential margins and Gumbel pair-copulae:

$$f_{123}(x_1, x_2, x_3; \boldsymbol{\lambda}, \boldsymbol{\delta}) = c(u_1, u_2; \delta_{12})c(u_2, u_3; \delta_{23})c(u_{1|2}, u_{3|2}; \delta_{13|2})\prod_{i=1}^{3} f(x_i; \lambda_i),$$

where

$$c(u_i, u_j; \delta) = \frac{\exp\{-(\tilde{u}_i^\delta + \tilde{u}_j^\delta)^{1/\delta}\}(\tilde{u}_i\tilde{u}_j)^{\delta-1}((\tilde{u}_i^\delta + \tilde{u}_j^\delta)^{1/\delta} + \delta - 1)}{u_i u_j(\tilde{u}_i^\delta + \tilde{u}_j^\delta)^{2-1/\delta}},$$

$$f(x; \lambda) = \lambda\exp\{-\lambda x\}, \qquad u_{i|j} = h_{ij}(u_i, u_j; \delta) = \frac{\exp\{-(\tilde{u}_i^\delta + \tilde{u}_j^\delta)^{1/\delta}(\tilde{u}_j)^{\delta-1}\}}{u_j(\tilde{u}_i^\delta + \tilde{u}_j^\delta)^{1-1/\delta}},$$

Table 2. Asymptotic relative efficiencies of $\hat{\delta}_{12}$ and $\hat{\delta}_{13|2}$ from Example 4.2, for various parameter sets

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

with $u_j = 1 - \exp\{-\lambda_j x_j\}$ 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 $\lambda_1 = \lambda_2 = \lambda_3 = 1$ in all sets. Moreover, we let $\delta_{12} = \delta_{23}$. Table 2 shows the resulting asymptotic relative efficiencies of the ground and top level parameter estimators, $(\hat{\delta}_{12}, \hat{\delta}_{23})$ and $\hat{\delta}_{13|2}$, respectively, that is, the ratios between the variances of the ML and alternative 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 $\delta \to \infty$. The three estimators are rather efficient in general, with IFM on top, followed by SP and finally SSP. As the true margins are known, this is not that surprising, and agrees with the results of Kim *et al.* [27]. All three estimators lose asymptotic efficiency with increasing dependence at the ground level, that is, for δ_{12} and δ_{23}, whereas SP and SSP 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. As expected, SSP is overall less efficient than IFM and SP, but the difference is quite small at the top level.

Example 4.3. Consider the five-dimensional D-vine with Student's t-margins and Student's t-copulae:

$$f_{12345}(x_1, x_2, x_3, x_4, x_5; \boldsymbol{v}_M, \boldsymbol{\rho}, \boldsymbol{v}_C) = \prod_{l=1}^{5} f(x_l; v_l) \prod_{j=1}^{4} \prod_{i=1}^{5-j} c(u_{i|v_{ij}}, u_{i+j|v_{ij}}; \rho_{i,i+j|v_{ij}}, v_{i,i+j|v_{ij}}),$$

with $\boldsymbol{v}_M = (v_1, \ldots, v_5)$, $\boldsymbol{\rho} = (\rho_{12}, \ldots, \rho_{15|234})$, $\boldsymbol{v}_C = (v_{12}, \ldots, v_{15|234})$, and

$$f(x; v) = \frac{\Gamma((v+1)/2)}{\sqrt{\pi v}\Gamma(v/2)}\left(1 + \frac{x^2}{v}\right)^{-(v+1)/2},$$

$$c(u_i, u_j; \rho, v) = \Gamma\left(\frac{v+2}{2}\right)\Gamma\left(\frac{v}{2}\right)\left(1 + \frac{t_v^{-1}(u_i)^2}{v}\right)^{(v+1)/2}\left(1 + \frac{t_v^{-1}(u_j)^2}{v}\right)^{(v+1)/2}$$

$$\Bigg/ \left(\Gamma\left(\frac{v+1}{2}\right)^2 \sqrt{1-\rho^2} \right.$$

$$\left. \times \left(1 + \frac{t_v^{-1}(u_i)^2 + t_v^{-1}(u_j)^2 - 2\rho t_v^{-1}(u_i) t_v^{-1}(u_j)}{v(1-\rho^2)} \right)^{(v+2)/2} \right),$$

$$u_{i|v_{ij}} = h(u_{i|v_{i,j-1}}, u_{i+j-1|v_{i,j-1}}; \rho_{i,i+j-1|v_{i,j-1}}, v_{i,i+j-1|v_{i,j-1}}),$$

$$u_{i+j|v_{ij}} = h(u_{i+j|v_{i+1,j-1}}, u_{i+1|v_{i+1,j-1}}; \rho_{i+1,i+j|v_{i+1,j-1}}, v_{i+1,i+j|v_{i+1,j-1}}),$$

$$h(u, v; \rho, v) = t_{v+1}\left(\sqrt{v+1}\left(t_v^{-1}(u) - \rho t_v^{-1}(v)\right)\left(\left(v + t_v^{-1}(v)^2\right)(1-\rho^2)\right)^{-1/2}\right),$$

with $u_i = t_v(x_i)$, t_v being the c.d.f. of the Student's t-distribution with v degrees of freedom. This is a five-dimensional extension of the example model considered in Sections 2.2.2 and 3.1, that is, none of the copulae share parameters, nor do the margins. The number of parameters is therefore 25.

In this case, it is infeasible to compute the asymptotic covariance matrices, both analytically and numerically. Therefore, we resort to simulation and Monte Carlo methods. More specifically, we have generated $N = 250$ samples of size $n = 10{,}000$ from the above distribution with four different parameter sets. For each sample, we have estimated the PCC parameters ρ and v_C using the four estimators. Finally, we have computed the sample covariance matrices of the resulting estimates. The four parameter sets we have considered are $(\rho, v) = (0.3, 6), (0.7, 6), (0.3, 20), (0.7, 20)$, where we let $\rho_{12} = \cdots = \rho_{15|234} = \rho$, $v_{12} = \cdots = v_{15|234} = v$, fixing the marginal parameters at $v_1 = \cdots = v_5 = 6$. Table 3 shows the resulting

Table 3. Asymptotic relative efficiencies of $\hat{\rho}$ and $\hat{v}_C$ from Example 4.3, averaged over each level, for various parameter sets

	Level 1		Level 2		Level 3		Level 4	
	$\hat{\rho}$	$\hat{v}$	$\hat{\rho}$	$\hat{v}$	$\hat{\rho}$	$\hat{v}$	$\hat{\rho}$	$\hat{v}$
				IFM				
$(\rho, v) = (0.3, 6)$	0.988	0.996	0.988	0.997	0.998	0.996	0.997	0.998
$(\rho, v) = (0.7, 6)$	0.935	0.913	0.961	0.988	0.984	0.979	0.968	0.996
$(\rho, v) = (0.3, 20)$	0.997	0.996	0.993	0.999	0.990	0.998	0.997	0.999
$(\rho, v) = (0.7, 20)$	0.952	0.992	0.962	0.993	0.969	0.988	0.991	0.989
				SP				
$(\rho, v) = (0.3, 6)$	0.952	0.985	0.973	0.987	0.991	0.992	0.998	0.997
$(\rho, v) = (0.7, 6)$	0.872	0.883	0.915	0.956	0.974	0.965	0.988	0.977
$(\rho, v) = (0.3, 20)$	0.965	0.963	0.994	0.970	0.989	0.991	0.983	0.992
$(\rho, v) = (0.7, 20)$	0.938	0.926	0.966	0.992	0.958	0.985	0.994	0.990
				SSP				
$(\rho, v) = (0.3, 6)$	0.890	0.907	0.932	0.937	0.938	0.985	0.946	0.996
$(\rho, v) = (0.7, 6)$	0.852	0.855	0.861	0.934	0.925	0.948	0.967	0.959
$(\rho, v) = (0.3, 20)$	0.941	0.955	0.992	0.964	0.981	0.975	0.973	0.974
$(\rho, v) = (0.7, 20)$	0.870	0.911	0.950	0.990	0.968	0.981	0.980	0.985

relative efficiencies averaged over each level. The three estimators behave rather similarly, although IFM once more appears to be the most efficient, SP the second and SSP the last. More specifically, their efficiency decreases with increasing dependence (either higher correlation or lower number of degrees of freedom) at all levels of the structure. Furthermore, they all become more efficient with increasing level number. In particular, the SSP estimator gains with respect to its competitors at the higher levels, just as for the Gumbel vine in Example 4.2. Note that an increased efficiency is not synonymous with a lower estimator variance, but only measures the behaviour relative to the ML estimator. Actually, the variances of all four estimators increase with the level of the structure, as one would expect.

Example 4.4. Finally, we have fitted a D-vine to a set of daily precipitation values recorded from 01.01.1990 to 31.12.2006 at five different meteorological stations in Norway; Vestby, Ski, Lørenskog, Nannestad and Hurdal, shown on the map (Figure 2, Supplement B of Hobæk Haff [20]). These data were provided by the Norwegian Meteorological Institute. Moreover, this is one of the data sets studied in Berg and Aas [4], extended with the series from Lørenskog. We have followed their example and modelled only the positive precipitation, that is, we have discarded all observations for which at least one of the stations has recorded zero precipitation, leaving 2013 out of the original 6209. The aim is to remove the temporal dependence between the observations, in accordance with our assumptions. Autocorrelation plots of the resulting data set indicate that this is reasonable.

Since rain showers tend to be very local, we expect the dependence between measurements from two proximate stations to be stronger than from stations that are farther apart. As the stations almost lie on a straight line (see Figure 2, Supplement B of Hobæk Haff [20]), a D-vine ordered according to geography is a very natural model. More specifically, the chosen dependence structure is the left-hand side of Figure 1, with Vestby, Ski, Lørenskog, Nannestad and Hurdal as variables 1, 2, 3, 4 and 5, respectively. To find adequate copulae for our structure, we computed the pseudo-observations, shown in Figure 3 in Supplement B of Hobæk Haff [20]. There are strong indications of upper, but not of lower tail dependence. We therefore chose Gumbel copulae at the ground level. An inspection of the data transformed with the estimated h-functions from the preceding level (as described in Section 3.1) indicated that Gaussian copulae would be reasonable for the three remaining levels. Finally, according to histograms of the data (shown in Figure 4, Supplement B of Hobæk Haff [20]), the generalised gamma distribution (Stacy [38]) with p.d.f.

$$f(x; \gamma, \beta, p) = \frac{p}{\beta^\gamma \Gamma(\gamma/p)} x^{\gamma-1} \exp\left\{-\left(\frac{x}{\beta}\right)^p\right\}$$

appears to be suitable for the margins. This distribution is gamma for $p = 1$ and exponential if in addition $\gamma = 1$. Both the ML and the IFM estimates of γ and p were rather different from 1, which confirms that the margins are neither exponential nor gamma distributions. The actual fitted marginal p.d.f.s are shown in the histograms of the data (Figure 4).

We have fitted the described model with each of the four estimators, using the R-routine `optim()`. The resulting estimates are shown in Table 4, along with 95% confidence intervals. These were computed by $\hat{\theta} \pm \Phi^{-1}(0.975)\hat{se}$, for each of the ten parameters, where $\hat{se}$ is an estimate of the parameter's asymptotic standard deviation. For the ML estimator, we computed the

Table 4. Estimated parameters with 95% confidence intervals for the precipitation data set of Example 4.4

Lev.	Par.	ML	IFM	SP	SSP
1	θ_{12}	4.56	4.37	4.32	4.32
		(4.44, 4.71)	(4.18, 4.56)	(4.14, 4.50)	(4.14, 4.50)
	θ_{23}	3.02	2.92	2.91	2.90
		(2.91, 3.13)	(2.80, 3.04)	(2.79, 3.03)	(2.79, 3.03)
	θ_{34}	2.53	2.47	2.47	2.47
		(2.44, 2.62)	(2.37, 2.57)	(2.37, 2.57)	(2.37, 2.56)
	θ_{45}	3.59	3.48	3.44	3.44
		(3.45, 3.73)	(3.34, 3.62)	(3.30, 3.58)	(3.30, 3.58)
2	$\theta_{13|2}$	-0.17	-0.17	-0.17	-0.17
		$(-0.21, -0.13)$	$(-0.21, -0.13)$	$(-0.21, -0.13)$	$(-0.21, -0.13)$
	$\theta_{24|3}$	0.21	0.20	0.21	0.21
		(0.15, 0.27)	(0.16, 0.24)	(0.17, 0.25)	(0.17, 0.25)
	$\theta_{35|4}$	0.066	0.067	0.061	0.061
		(0.022, 0.11)	(0.031, 0.10)	(0.023, 0.099)	(0.024, 0.098)
3	$\theta_{14|23}$	0.093	0.088	0.081	0.081
		(0.055, 0.13)	(0.053, 0.12)	(0.044, 0.12)	(0.044, 0.12)
	$\theta_{25|34}$	0.050	0.043	0.033	0.033
		(0.009, 0.091)	(0.008, 0.079)	$(-0.003, 0.070)$	$(-0.003, 0.070)$
4	$\theta_{15|234}$	0.040	0.045	0.046	0.046
		(0.006, 0.075)	(0.007, 0.083)	(0.012, 0.080)	(0.012, 0.080)

sample Fisher matrix

$$\hat{\mathcal{I}} = \frac{1}{n} \sum_{k=1}^{n} \frac{\partial}{\partial(\boldsymbol{\alpha}, \boldsymbol{\theta})} \log(f_{1\ldots5}(x_{1k}, \ldots, x_{5k}; \hat{\boldsymbol{\alpha}}^{\mathrm{ML}}, \hat{\boldsymbol{\theta}}^{\mathrm{ML}})),$$

where $\boldsymbol{\alpha} = (\gamma_1, \beta_1, p_1, \ldots, \gamma_5, \beta_5, p_5)$, the derivative being calculated numerically. The estimates $\hat{se}$ were then simply the square roots of the diagonal entries of $\hat{\mathcal{I}}^{(\theta)}$. For the three remaining estimators, we used parametric bootstrap to obtain $\hat{se}$. More specifically, we generated $B = 500$ bootstrap samples from $F_{1\ldots5}(x_1, \ldots, x_5; \hat{\boldsymbol{\alpha}}^{\mathrm{IFM}}, \hat{\boldsymbol{\theta}}^{\mathrm{IFM}})$, $C_{1\ldots5}(u_1, \ldots, u_5; \hat{\boldsymbol{\theta}}^{\mathrm{SP}})$ and $C_{1\ldots5}(u_1, \ldots, u_5; \hat{\boldsymbol{\theta}}^{\mathrm{SSP}})$, estimating the parameters $\hat{\boldsymbol{\alpha}}^{\mathrm{IFM},b}, \hat{\boldsymbol{\theta}}^{\mathrm{IFM},b}, \hat{\boldsymbol{\theta}}^{\mathrm{SP},b}, \hat{\boldsymbol{\theta}}^{\mathrm{SSP},b}, b = 1, \ldots, B$. Finally, we let the $\hat{se}$s be the sample standard deviations of the bootstrap estimates.

At the ground level, the parameter estimates are overall high. This indicates a strong positive dependence between large amounts of precipitation in stations that are close in distance, as anticipated. The IFM, SP and SSP estimators give similar values. However, the ML estimates are rather different, though the 95% confidence intervals overlap with the other estimators'. As noted earlier, this indicates that the chosen univariate margins or copulae are not quite adequate. Since the SP and SSP estimates are virtually the same, the problem appears to be the margins.

The second level models the conditional dependencies of two stations that are separated by one, given the one between them. All four estimators agree that this conditional dependence is

negative between Vestby and Lørenskog, positive between the pair Ski and Nannestad, whereas Lørenskog and Hurdal are almost conditionally independent. At the top two levels, the estimated copulae are close to the independence copula, as expected. Actually, the SP and SSP confidence intervals indicate that the copula $C_{25|34}$ is not significantly different from independence, which can be an important aspect for practical purposes.

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. The SSP estimator has a higher variance than the alternatives. Nonetheless, the loss of efficiency is rather low, and decreases with the construction level, as shown in a couple of examples. For the set of five precipitation series, the SSP estimates are actually almost indistinguishable from the SP ones. Moreover, the SSP estimator is semiparametrically so for the Gaussian copula. To compare the alternative estimators' performance more thoroughly, 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 preceding levels. The SSP estimator lends itself perfectly to that 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. Further, we have partitioned the parameter vector into marginal and dependence parameters. This excludes some distributions, such as the multivariate Student's *t*. However, if one does not need the flexibility to model the margins and dependence structure separately, as well different types of dependence between the various pairs of variables, a PCC is unnecessarily complex. Moreover, we have assumed the observations to be independent, identically distributed. In practice, the parameter estimation often includes a preliminary step to deal with deviations from these assumptions (Chen and Fan [5]), 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.

Appendix

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

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

We start with $\mathcal{J}_{\theta,i,j}$, where $i < j$. Then, $\mathcal{J}_{\theta,i,j} = \mathrm{E}(-\frac{\partial^2 \psi_i(u_1,\ldots,u_d;\theta_1,\ldots,\theta_i)}{\partial \theta_i \, \partial \theta_j^T})$, with ψ_i from (15). Since none of the copulae at level i are functions of the parameters at a following level j, $\frac{\partial \psi_i(u_1,\ldots,u_d;\theta_1,\ldots,\theta_i)}{\partial \theta_j} = \mathbf{0}$. Hence, $\mathcal{J}_{\theta,i,j} = \mathbf{0}, i < j$.

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

$$
\begin{aligned}
\mathcal{K}_{\theta,i,j} &= \mathrm{E}\left(\left(\frac{\partial \psi_i(u_1,\ldots,u_d;\theta_1,\ldots,\theta_i)}{\partial \theta_i}\right)\left(\frac{\partial \psi_j(u_1,\ldots,u_d;\theta_1,\ldots,\theta_j)}{\partial \theta_j}\right)^T\right) \\
&= \int_{\mathbf{u}} \frac{\partial}{\partial \theta_i} \sum_{k=1}^{d-i} \log c_{k,k+i|v_{ki}} \frac{\partial}{\partial \theta_j^T} \sum_{l=1}^{d-j} \log c_{l,l+j|v_{lj}} c_{1\ldots d} \, \mathrm{d}\mathbf{u} \\
&= \sum_{k=1}^{d-i}\sum_{l=1}^{d-j} \int_{\mathbf{u}_{w_{ki}}} \frac{\partial}{\partial \theta_i} \log c_{k,k+i|v_{ki}} \int_{\mathbf{u}_{-w_{ki}}} \frac{1}{c_{l,l+j|v_{lj}}} \frac{\partial}{\partial \theta_j^T} c_{l,l+j|v_{lj}} c_{1\ldots d} \, \mathrm{d}\mathbf{u}_{-w_{ki}} \, \mathrm{d}\mathbf{u}_{w_{ki}}.
\end{aligned}
$$

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

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

The pair-copulae composing $c_{w_{ki}}$, situated in levels $1,\ldots,i$, are not functions of parameters from a following level j. Thus, $\frac{\partial}{\partial \theta_j} c_{w_{ki}} = \mathbf{0}$. 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 Klaassen and Wellner [28]. In three dimensions, we have computed the asymptotic covariance matrices for comparison. As shown in Example 4.1, the covariance matrices of the SP and SSP estimators, $\mathbf{V}^{\mathrm{SP}}$ and $\mathbf{V}^{\mathrm{SSP}}$, 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. Further, for the d-dimensional model, partition the covariance matrix $\mathbf{V}^{\mathrm{SP}}$ as $\mathbf{V}_1^{\mathrm{SP}} = \mathbf{V}_{1\ldots d-2,1\ldots d-2}^{\mathrm{SSP}}$, $\mathbf{V}_{12}^{\mathrm{SP}} = \mathbf{V}_{1\ldots d-2,d-1}^{\mathrm{SSP}}$ and $V_2^{\mathrm{SP}} = V_{d-1,d-1}^{\mathrm{SSP}}$, and likewise for $\mathbf{V}^{\mathrm{SSP}}$, $\mathbf{V}^{\mathrm{ML}}$, $\mathbf{B}^{\mathrm{SP}}$, $\mathbf{B}^{\mathrm{SSP}}$, $\mathcal{I}_\theta$, $\mathcal{I}^{(\theta)}$ and $\mathcal{J}_\theta$.

As the SP estimator is semiparametrically efficient, $\mathbf{V}^{\mathrm{ML}}$ for the Gaussian copula must be the same, regardless of the margins. Moreover, when all margins are normal, $\hat{\boldsymbol{\theta}}^{\mathrm{ML}}$ is simply the empirical correlation matrix. Adding an extra dimension leaves the remaining estimators unchanged. Hence, $\mathbf{V}_1^{\mathrm{ML}}$, 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 $\mathbf{V}^{\mathrm{SSP}}$ be. Accordingly, we must have $\mathbf{V}_1^{\mathrm{SSP}} = \mathbf{V}_1^{\mathrm{SP}} = \mathbf{V}_1^{\mathrm{ML}}$. Hence, it remains to show that $\mathbf{V}_{12}^{\mathrm{SSP}} = \mathbf{V}_{12}^{\mathrm{SP}}$ and $V_2^{\mathrm{SSP}} = V_2^{\mathrm{SP}}$, related to the estimators $\hat{\theta}_{1d|v_{1d}}^{\mathrm{SP}}$ and $\hat{\theta}_{1d|v_{1d}}^{\mathrm{SSP}}$ for the top level copula. According to Theorem 1 from Tsukahara [40] and Theorem 1, respectively,

$$\sqrt{n}(\hat{\theta}_{1d|v_{1d}}^{\mathrm{SP}} - \theta_{1d|v_{1d}}) \xrightarrow{d} Z_{\mathrm{SP}} \sim \mathcal{N}(0, V_2^{\mathrm{SP}})$$

and

$$\sqrt{n}(\hat{\theta}_{1d|v_{1d}}^{\mathrm{SSP}} - \theta_{1d|v_{1d}}) \xrightarrow{d} Z_{\mathrm{SSP}} \sim \mathcal{N}(0, V_2^{\mathrm{SSP}}),$$

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

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

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

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

$$\Psi(\mathbf{U}_n; \boldsymbol{\theta}) = \Psi_{d(d-1)/2}^{\mathrm{SP}}(\mathbf{U}_n; \boldsymbol{\theta}) = \Psi_{d(d-1)/2}^{\mathrm{SSP}}(\mathbf{U}_n; \boldsymbol{\theta}) = \frac{1}{n} \sum_{k=1}^n \frac{\partial}{\partial \theta_{1d|v_{1d}}} \log(c_{1\dots d}(\mathbf{U}_n; \boldsymbol{\theta})).$$

According to Theorem 1 from Tsukahara [40],

$$\Psi(\mathbf{U}_n; \hat{\boldsymbol{\theta}}^{\mathrm{SP}}) = \Psi(\mathbf{U}_n; \boldsymbol{\theta}) + \frac{\partial \Psi(\mathbf{U}_n; \boldsymbol{\theta})}{\partial \theta_{1d|v_{1d}}}(\hat{\theta}_{1d|v_{1d}}^{\mathrm{SP}} - \theta_{1d|v_{1d}})$$

$$+ \frac{\partial \Psi(\mathbf{U}_n; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{1 \to d-2}^T}(\hat{\boldsymbol{\theta}}_{1 \to d-2}^{\mathrm{SP}} - \boldsymbol{\theta}_{1 \to d-2}) + \mathrm{o}_P\left(\frac{1}{n}\right) = 0.$$

Likewise, using Theorem 1, one obtains

$$\Psi(\mathbf{U}_n; \hat{\boldsymbol{\theta}}^{\mathrm{SSP}}) = \Psi(\mathbf{U}_n; \boldsymbol{\theta}) + \frac{\partial \Psi(\mathbf{U}_n; \boldsymbol{\theta})}{\partial \theta_{1d|v_{1d}}}(\hat{\theta}_{1d|v_{1d}}^{\mathrm{SSP}} - \theta_{1d|v_{1d}})$$

$$+ \frac{\partial \Psi(\mathbf{U}_n; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{1 \to d-2}^T}(\hat{\boldsymbol{\theta}}_{1 \to d-2}^{\mathrm{SSP}} - \boldsymbol{\theta}_{1 \to d-2}) + \mathrm{o}_P\left(\frac{1}{n}\right) = 0.$$

Hence,

$$\sqrt{n}(\hat{\theta}_{1d|v_{1d}}^{\text{SSP}} - \hat{\theta}_{1d|v_{1d}}^{\text{SP}}) = \frac{\mathbf{A}_1}{A_2}\sqrt{n}(\hat{\theta}_{1\to d-2}^{\text{SP}} - \hat{\theta}_{1\to d-2}^{\text{SSP}}) + \text{o}_P\left(\frac{1}{n}\right),$$

with

$$\mathbf{A}_1 = \frac{\partial\Psi(\mathbf{U}_n;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_{1\to d-2}^T} = \frac{1}{n}\sum_{k=1}^{n}\frac{\partial^2}{\partial\theta_{1d|v_{1d}}\,\partial\boldsymbol{\theta}_{1\to d-2}^T}\log(c_{1\ldots d}(\mathbf{U}_n;\boldsymbol{\theta}))$$

and

$$A_2 = \frac{\partial\Psi(\mathbf{U}_n;\boldsymbol{\theta})}{\partial\theta_{1d|v_{1d}}} = \frac{1}{n}\sum_{k=1}^{n}\frac{\partial^2}{\partial\theta_{1d|v_{1d}}^2}\log(c_{1\ldots d}(\mathbf{U}_n;\boldsymbol{\theta})).$$

According to the assumption,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{1\to d-2}^{\text{SSP}} - \boldsymbol{\theta}_{1\to d-2}) \xrightarrow{d} \mathbf{Y} \sim \mathcal{N}_{d(d-1)/2-1}(\mathbf{0}, \mathbf{V}_1^{\text{ML}}), \qquad n\to\infty.$$

Thus,

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

Moreover, under the assumed conditions, $\mathbf{A}_1 \xrightarrow{p} -\mathcal{I}_{\theta,12}^T$ and $A_2 \xrightarrow{p} -\mathcal{I}_{\theta,2}$, as $n\to\infty$. Hence,

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

which means that $Z_{\text{SP}} \stackrel{d}{=} Z_{\text{SSP}}$. In other words, $V_2^{\text{SSP}} = V_2^{\text{SP}}$. Moreover,

$$\mathbf{V}_{12}^{\text{SSP}} = \frac{1}{\mathcal{I}_{\theta,2}}(-\mathbf{V}_1^{\text{ML}}\mathcal{I}_{\theta,12} + \mathcal{J}_{\theta,12}^{-1}\mathbf{B}_{12}^{\text{SSP}}),$$

$$\begin{aligned}
V_2^{\text{SSP}} &= \frac{1}{\mathcal{I}_{\theta,2}}\left(1 + \frac{B_2^{\text{SSP}}}{\mathcal{I}_{\theta,2}} + \frac{1}{\mathcal{I}_{\theta,2}}\mathcal{I}_{\theta,12}^T\mathbf{V}_1^{\text{ML}}\mathcal{I}_{\theta,12} - \frac{2}{\mathcal{I}_{\theta,2}}\mathcal{I}_{\theta,12}^T\mathcal{J}_{\theta,12}^{-1}\mathbf{B}_{12}^{\text{SSP}}\right) \\
&= \frac{1}{\mathcal{I}_{\theta,2}^2}(\mathcal{I}_{\theta,2} + B_2^{\text{SSP}} - \mathcal{I}_{\theta,12}^T\mathbf{V}_1^{\text{ML}}\mathcal{I}_{\theta,12}) - 2\mathcal{I}_{\theta,12}^T\mathbf{V}_{12}^{\text{SSP}}.
\end{aligned}$$

Correspondingly for SP,

$$\mathbf{V}_{12}^{\text{SP}} = \frac{1}{\mathcal{I}_{\theta,2}}\left(-\mathbf{V}_1^{\text{ML}}\mathcal{I}_{\theta,12} + \mathcal{I}_1^{(\theta)}\left(\mathbf{B}_{12}^{\text{SP}} - \frac{B_2^{\text{SP}}}{\mathcal{I}_{\theta,2}}\mathcal{I}_{\theta,12}\right)\right),$$

$$V_2^{\mathrm{SP}} = \frac{1}{\mathcal{I}_{\theta,2}}\left(1 + \frac{B_2^{\mathrm{SP}}}{\mathcal{I}_{\theta,2}} + \frac{1}{\mathcal{I}_{\theta,2}}\mathcal{I}_{\theta,12}^T\mathbf{V}_1^{\mathrm{ML}}\mathcal{I}_{\theta,12} - \frac{2}{\mathcal{I}_{\theta,2}}\mathcal{I}_{\theta,12}^T\mathcal{I}_1^{(\theta)}\left(\mathbf{B}_{12}^{\mathrm{SP}} - \frac{B_2^{\mathrm{SP}}}{\mathcal{I}_{\theta,2}}\mathcal{I}_{\theta,12}\right)\right)$$

$$= \frac{1}{\mathcal{I}_{\theta,2}^2}(\mathcal{I}_{\theta,2} + B_2^{\mathrm{SP}} - \mathcal{I}_{\theta,12}^T\mathbf{V}_1^{\mathrm{ML}}\mathcal{I}_{\theta,12}) - 2\mathcal{I}_{\theta,12}^T\mathbf{V}_{12}^{\mathrm{SP}}.$$

Since the estimating equation for $\theta_{1d|v_{1d}}$ is the same for SP and SSP, $B_2^{\mathrm{SSP}} = B_2^{\mathrm{SP}}$. Moreover, $V_2^{\mathrm{SSP}} = V_2^{\mathrm{SP}}$. Consequently, $\mathbf{V}_{12}^{\mathrm{SSP}} = \mathbf{V}_{12}^{\mathrm{SP}}$. □

A.3. Covariance matrices from Example 4.1

The asymptotic covariance matrix of the ML estimator is given by

$$\mathbf{V}^{\mathrm{ML}} = \frac{1}{2}\begin{pmatrix} 2(1-\rho_{12}^2)^2 & v_{13} & v_{23} \\ v_{13} & 2(1-\rho_{23}^2)^2 & v_{12} \\ v_{23} & v_{12} & 2(1-\rho_{13}^2)^2 \end{pmatrix},$$

with $v_{ik} = 2\rho_{ik}(1-\rho_{il}^2)(1-\rho_{lk}^2) - \rho_{il}\rho_{lk}|\mathbf{R}|$. For the SSP estimator, we have

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

where

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

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

$$\mathbf{B}_\theta^{\mathrm{SSP}} = \begin{pmatrix} \dfrac{\rho_{12}^2(1+\rho_{12}^2)}{(1-\rho_{12}^2)^2} & \dfrac{\rho_{23}b_{12}+\rho_{12}b_{23}-\rho_{12}\rho_{23}a}{2(1-\rho_{12}^2)(1-\rho_{23}^2)} & \dfrac{(\rho_{13}-\rho_{12}\rho_{23})b_{12}}{2(1-\rho_{12}^2)|\mathbf{R}|} \\ \dfrac{\rho_{23}b_{12}+\rho_{12}b_{23}-\rho_{12}\rho_{23}a}{2(1-\rho_{12}^2)(1-\rho_{23}^2)} & \dfrac{\rho_{23}^2(1+\rho_{23}^2)}{(1-\rho_{23}^2)^2} & \dfrac{(\rho_{13}-\rho_{12}\rho_{23})b_{23}}{2(1-\rho_{23}^2)|\mathbf{R}|} \\ \dfrac{(\rho_{13}-\rho_{12}\rho_{23})b_{12}}{2(1-\rho_{12}^2)|\mathbf{R}|} & \dfrac{(\rho_{13}-\rho_{12}\rho_{23})b_{23}}{2(1-\rho_{23}^2)|\mathbf{R}|} & \dfrac{(1+\rho_{13}^2)(\rho_{13}-\rho_{12}\rho_{23})^2}{|\mathbf{R}|^2} \end{pmatrix},$$

with $a = 1 + \rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2$, $k_{12} = (\rho_{13} - \rho_{12}\rho_{23})(|\mathbf{R}| + \rho_{13}^2 - \rho_{12}^2\rho_{23}^2)$, $j_{ik} = -\rho_{ik}|\mathbf{R}| + 2(\rho_{il} - \rho_{lk}\rho_{ik})(\rho_{lk} - \rho_{il}\rho_{ik})$ and $b_{ik} = \rho_{ik}a(1 - \frac{2}{1-\rho_{ik}^2}) + 2(1 + \rho_{ik}^2)\frac{\rho_{ik}+\rho_{il}\rho_{lk}}{1-\rho_{ik}^2}$. The resulting covariance matrix is

$$\mathbf{V}^{\mathrm{SSP}} = \frac{1}{2} \begin{pmatrix} 2(1 - \rho_{12}^2)^2 & v_{13} & v_{23} \\ v_{13} & 2(1 - \rho_{23}^2)^2 & v_{12} \\ v_{23} & v_{12} & 2(1 - \rho_{13}^2)^2 \end{pmatrix} = \mathbf{V}^{\mathrm{ML}}.$$

Acknowledgements

Supplementary Material

Supplement A: SSP estimation algorithms for D- and C-vines (DOI: 10.3150/12-BEJ413SUPPA; .pdf). Estimation algorithms for the stepwise semiparametric estimator for D-and C-vines.

Supplement B: Figures and table from Example 4.4 (DOI: 10.3150/12-BEJ413SUPPB; .pdf). Figures 2, 3 and 4, as well as Table 4 from Example 4.4.

References

[1] Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance Math. Econom.* **44** 182–198. MR2517884

[2] Bedford, T. and Cooke, R.M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Ann. Math. Artif. Intell.* **32** 245–268. MR1859866

[3] Bedford, T. and Cooke, R.M. (2002). Vines – a new graphical model for dependent random variables. *Ann. Statist.* **30** 1031–1068. MR1926167

[4] Berg, D. and Aas, K. (2009). Models for construction of multivariate dependence. *European Journal of Finance* **15** 639–659.

[5] Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *J. Econometrics* **130** 307–335. MR2211797

[6] 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.

[7] Claeskens, G. and Hjort, N.L. (2008). *Model Selection and Model Averaging*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge Univ. Press. MR2431297

[8] Clayton, D.G. (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. MR0501698

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

[10] Czado, C., Min, A., Baumann, T. and Dakovic, R. (2009). Pair-copula constructions for modeling exchange rate dependence. Unpublished manuscript.

[11] Czado, C., Schepsmeier, U. and Min, A. (2010). Maximum likelihood estimation of mixed c-vines with application to exchange rates. Unpublished manuscript.

[12] Dissmann, J., Brechmann, E., Czado, C. and Kurowicka, K. (2011). Selecting and estimating regular vine copulae and application to financial returns. Unpublished manuscript.

[13] 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.

[14] Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika* **74** 549–555. MR0909358

[15] 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 Math. Econom.* **44** 143–145. MR2517879

[16] Genest, C., Ghoudi, K. and Rivest, L.P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82** 543–552. MR1366280

[17] Genest, C. and Rivest, L.P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *J. Amer. Statist. Assoc.* **88** 1034–1043. MR1242947

[18] Hall, P. and Yao, Q. (2005). Approximating conditional distribution functions using dimension reduction. *Ann. Statist.* **33** 1404–1421. MR2195640

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

[20] Hobæk Haff, I. (2012). Supplement to "Parameter estimation for pair-copula constructions." DOI:10.3150/12-BEJ413SUPPA, DOI:10.3150/12-BEJ413SUPPB.

[21] Hobæk Haff, I., Aas, K. and Frigessi, A. (2010). On the simplified pair-copula construction – simply useful or too simplistic? *J. Multivariate Anal.* **101** 1296–1310. MR2595309

[22] 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*. Hayward, CA: IMS.

[23] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. *Monographs on Statistics and Applied Probability* **73**. London: Chapman & Hall. MR1462613

[24] Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *J. Multivariate Anal.* **94** 401–419. MR2167922

[25] Joe, H., Li, H. and Nikoloulopoulos, A.K. (2010). Tail dependence functions and vine copulas. *J. Multivariate Anal.* **101** 252–270. MR2557632

[26] Joe, H. and Xu, J. (1996). The estimation method of inference functions for margins for multivariate models. Technical Report 166, Univ. British Columbia, Dept. Statistics.

[27] Kim, G., Silvapulle, M.J. and Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Comput. Statist. Data Anal.* **51** 2836–2850. MR2345609

[28] Klaassen, C.A.J. and Wellner, J.A. (1997). Efficient estimation in the bivariate normal copula model: Normal margins are least favourable. *Bernoulli* **3** 55–77. MR1466545

[29] Kolbjørnsen, O. and Stien, M. (2008). D-vine creation of non-Gaussian random field. In *Procedings of the Eight International Geostatistics Congress* 399–408. Santiago, Chile: GECAMIN Ltd.

[30] Kurowicka, D. and Cooke, R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR2216540

[31] Lehmann, E.L. (2004). *Elements of Large-Sample Theory*. *Springer Texts in Statistics*. New York: Springer.

[32] McNeil, A.J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management*: *Concepts*, *Techniques and Tools*. *Princeton Series in Finance*. Princeton, NJ: Princeton Univ. Press. MR2175089

[33] Min, A. and Czado, C. (2010). Scomdy models based on pair-copula constructions with application to exchange rates. Unpublished manuscript.

[34] Oakes, D. (1982). A model for association in bivariate survival data. *J. Roy. Statist. Soc. Ser. B* **44** 414–422. MR0693242

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

[36] Shih, J.H. and Louis, T.A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* **51** 1384–1399. MR1381050

[37] Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8** 229–231. MR0125600

[38] Stacy, E.W. (1962). A generalization of the gamma distribution. *Ann. Math. Statist.* **33** 1187–1192. MR0143277

[39] Stute, W. (1986). On almost sure convergence of conditional empirical distribution functions. *Ann. Probab.* **14** 891–901. MR0841591

[40] Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canad. J. Statist.* **33** 357–375. MR2193980

[41] Tsukahara, H. (2011). Personal communication.