Journal of Multivariate Analysis 110 (2012) 74–90

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Beyond simplified pair-copula constructions

Elif F. Acar *, Christian Genest, Johanna Nešlehová

Department of Mathematics and Statistics, McGill University, Montréal, QC, Canada H3A 0B9

A R T I C L E I N F O

Article history:
Available online 17 February 2012

AMS subject classifications:
62F12
62H12
62H20

Keywords:
Conditional copulas
Kendall's tau
Local likelihood
Pair-copula constructions
Ranks
Vines

A B S T R A C T

Pair-copula constructions (PCCs) offer great flexibility in modeling multivariate dependence. For inference purposes, however, conditional pair-copulas are often assumed to depend on the conditioning variables only indirectly through the conditional margins. The authors show here that this assumption can be misleading. To assess its validity in trivariate PCCs, they propose a visual tool based on a local likelihood estimator of the conditional copula parameter which does not rely on the simplifying assumption. They establish the consistency of the estimator and assess its performance in finite samples via Monte Carlo simulations. They also provide a real data application.

1. Introduction

Over the past two decades, dependence modeling via copulas has evolved considerably and has found applications in areas as diverse as actuarial science, biostatistics, finance, hydrology, and machine learning. In the bivariate case, many parametric copula families have been proposed that can represent a broad range of dependence patterns. In higher dimensions, however, parametric copula families are harder to construct and their tractability often comes at the cost of flexibility. For example, meta-elliptical copulas are somewhat of a straightjacket, if only because all lower-dimensional margins belong to the same class. In many applications, this property is too restrictive as pairs of variables may exhibit very different dependence patterns.

A more flexible way to model multivariate dependences is offered by pair-copula constructions (PCCs), also known as *vine copulas* [7,15,16]. Vines are graphical models that provide a systematic way to decompose a multivariate copula into a cascade of bivariate copulas, some of which are conditional. A simple example of a PCC in the trivariate case consists of writing the joint density c of a random vector (U_1, U_2, U_3) with uniform margins on $(0, 1)$ in the form

$$c(u_1, u_2, u_3) = c_{12}(u_1, u_2)c_{23}(u_2, u_3)c_{13|2}(u_{1|2}, u_{3|2}; u_2). \tag{1}$$

Here, c_{12} and c_{23} are the copula densities of the pairs (U_1, U_2) and (U_2, U_3), respectively. Furthermore, $c_{13|2}$ is the conditional copula density of the pair (U_1, U_3) given $U_2 = u_2$, evaluated at $u_{k|2} = \Pr(U_k \le u_k | U_2 = u_2)$ for $k = 1, 3$. Any choice of c_{12}, c_{13} and $c_{13|2}$ leads to a valid trivariate copula density. More generally, using different bivariate copulas as building blocks in a d-variate PCC, one can construct highly flexible multivariate copula models.

Inference for a given PCC is typically carried out by specifying a parametric copula for each building block. Copula parameters are then estimated sequentially starting with the unconditional pair-copulas and moving up the hierarchy [1].

* Corresponding author.
E-mail address: eacar@math.mcgill.ca (E.F. Acar).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2012.02.001

In the above example, this amounts to estimating the parameters of c_{12} and c_{23} first, and those of $c_{13|2}$ in the second step. A standard assumption is that the conditional pair-copulas of the PCC depend on the conditioning variable(s) only through the conditional margins. In (1), this is equivalent to assuming that the conditional copula $c_{13|2}$ of the pair (U_1, U_3) given $U_2 = u_2$ is *the same* for all values of $u_2 \in (0, 1)$.

This simplifying assumption seems to have been made mainly for convenience at a time when inference tools for conditional copulas were still under development [20]. Through examples, it is shown in [14] that simplified PCCs can provide a good approximation in some cases. This paper revisits this issue and introduces a nonparametric smoothing methodology that relaxes this simplifying assumption for trivariate PCCs.

After a brief summary of vine copula constructions in the trivariate case in Section 2, estimation for simplified three-dimensional PCCs is described in Section 3. Through simulations, it is then shown in Section 4 that inference based on simplified PCCs can be misleading and may even conduce the belief that some pairs of variables are conditionally independent when in fact they are not. The new methodology, which derives from recent work [4], is described in Section 5. This approach is seen to perform well in simulations and in a data application, as detailed in Sections 6 and 8, respectively. The consistency of the proposed method is presented in Section 7 and Section 9 concludes with a short discussion.

The following notation is used throughout the paper. Vectors in $\mathbb{R}^3$ are denoted by bold letters, e.g., $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$. If $A \subset \{1, 2, 3\}$ is non-empty, $\mathbf{x}_A$ stands for an $|A|$-dimensional vector with components x_k, $k \in A$. If $\mathbf{X}$ is a random vector with distribution function F and density f, then for arbitrary disjoint index sets A and B, the symbols $F_{A|B}$ and $f_{A|B}$ denote the conditional distribution function and density of $\mathbf{X}_A$ given $\mathbf{X}_B = \mathbf{x}_B$, respectively.

2. Trivariate PCCs

Let X_1, X_2, X_3 be random variables with joint distribution function F and continuous margins F_1, F_2, F_3, respectively. Sklar's Representation Theorem [22] states that, for all $x_1, x_2, x_3 \in \mathbb{R}$,

$$F(x_1, x_2, x_3) = C\{F_1(x_1), F_2(x_2), F_3(x_3)\},$$

where C is a copula, i.e., a distribution function with margins that are uniform on $(0, 1)$. If F is absolutely continuous, its density can be written in terms of the density c of C as

$$f(x_1, x_2, x_3) = c\{F_1(x_1), F_2(x_2), F_3(x_3)\} \prod_{k=1}^{3} f_k(x_k),$$

where, for each $k \in \{1, 2, 3\}$, f_k is the density of F_k.

A PCC is based on the fact that f can be decomposed as

$$f(x_1, x_2, x_3) = f_3(x_3) \times f_{2|3}(x_2|x_3) \times f_{1|23}(x_1|x_2, x_3). \tag{2}$$

Note that this factorization is unique up to relabeling. For any index set $A \subset \{1, 2, 3\}$ and $k \in A$, let $A - k = A \setminus \{k\}$. Using Sklar's Representation Theorem, one can then write, for arbitrary $j \notin A$,

$$f_{j|A} = c_{jk|A-k}(F_{j|A-k}, F_{k|A-k}) f_{j|A-k}. \tag{3}$$

Repeated applications of relation (3) in (2) make it possible to express f as

$$\begin{aligned} f(x_1, x_2, x_3) = {} & f_1(x_1) f_2(x_2) f_3(x_3) \times c_{12}\{F_1(x_1), F_2(x_2)\} \times c_{23}\{F_2(x_2), F_3(x_3)\} \\ & \times c_{13|2}\{F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2\}, \end{aligned} \tag{4}$$

which reduces to (1) if the margins of F are uniform. The univariate conditional distributions featuring in (4) are given by

$$F_{j|k}(x_j|x_k) = h_{jk}\{F_j(x_j), F_k(x_k)\},$$

where, for all $u, v \in (0, 1)$,

$$h_{jk}(u, v) = \frac{\partial}{\partial v} C_{jk}(u, v). \tag{5}$$

3. Inference for simplified PCCs

Now suppose the density f of (X_1, X_2, X_3) follows a simplified PCC model, i.e., f is of the form (4), where the conditional copula density $c_{13|2}$ does not depend on the conditioning variable. The last term in (4) thus reduces to

$$c_{13|2}\{F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)\}.$$

To ease the presentation, assume that all copulas appearing in (4) are parametrized by scalar parameters $\theta_{12}, \theta_{23}, \theta_{13|2}$ that are indexed in the same way as the corresponding copula.

Suppose that $\{(X_{11}, X_{21}, X_{31}), \ldots, (X_{1n}, X_{2n}, X_{3n})\}$ is a random sample from (X_1, X_2, X_3). If the margins F_1, F_2, F_3 of the latter random vector are known, then a random sample from the underlying copula is given by $\mathcal{U} = \{(U_{11}, U_{21}, U_{31}), \ldots, (U_{1n}, U_{2n}, U_{3n})\}$, where, for each $i \in \{1, \ldots, n\}$,

$$(U_{1i}, U_{2i}, U_{3i}) = (F_1(X_{1i}), F_2(X_{2i}), F_3(X_{3i})).$$

Inference for the unknown parameter $\boldsymbol{\theta} = (\theta_{12}, \theta_{23}, \theta_{13|2})$ can thus be based on the log-likelihood function given by

$$\begin{aligned} L(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \left\{ \ln c_{12}(U_{1i}, U_{2i}; \theta_{12}) + \ln c_{23}(U_{2i}, U_{3i}; \theta_{23}) + \ln c_{13|2}(U_{(1|2)i}, U_{(3|2)i}; \theta_{13|2}) \right\} \\ &\equiv L_{12}(\theta_{12}) + L_{23}(\theta_{23}) + L_{13|2}(\theta_{12}, \theta_{23}, \theta_{13|2}), \end{aligned} \tag{6}$$

where

$$U_{(1|2)i} = h_{12}(U_{1i}, U_{2i}; \theta_{12}) \quad \text{and} \quad U_{(3|2)i} = h_{32}(U_{3i}, U_{2i}; \theta_{23}).$$

Given that θ_{12} and θ_{23} appear in $U_{(k|2)i}$ for $k = 1, 3$ and $i \in \{1, \ldots, n\}$, the last term $L_{13|2}$ of the log-likelihood function depends on all parameters. Therefore, (6) has to be maximized jointly. As this can be computationally demanding, a sequential estimation approach suggested in [1] is often used.

In this approach, the stepwise estimates $\check{\theta}_{12}$ and $\check{\theta}_{23}$ are obtained first by maximizing $L_{12}(\theta_{12})$ and $L_{23}(\theta_{23})$, respectively. Next, pseudo observations are constructed by letting, for every $i \in \{1, \ldots, n\}$,

$$U^*_{(1|2)i} = h_{12}(U_{1i}, U_{2i}; \check{\theta}_{12}), \qquad U^*_{(3|2)i} = h_{32}(U_{3i}, U_{2i}; \check{\theta}_{23}). \tag{7}$$

An estimator $\check{\theta}_{13|2}$ can then be obtained through the maximization of

$$L^*_{13|2}(\theta_{13|2}) = \sum_{i=1}^{n} \ln c_{13|2}(U^*_{(1|2)i}, U^*_{(3|2)i}; \theta_{13|2}).$$

Note that the stepwise estimates $\check{\theta}_{12}$, $\check{\theta}_{23}$, $\check{\theta}_{13|2}$ do not maximize (6), but nonetheless provide a good approximation of the joint estimate of $\boldsymbol{\theta}$.

When the margins are unknown, the sample $\mathcal{U}$ is no longer available. However, the margins can be estimated nonparametrically by the corresponding empirical distribution functions defined here, for all $k \in \{1, 2, 3\}$ and $x \in \mathbb{R}$, by

$$F_{kn}(x) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbf{1}(X_{ki} \le x),$$

where division by $n + 1$ instead of n is chosen to avoid boundary problems. Setting, for each $i \in \{1, \ldots, n\}$,

$$(\widehat{U}_{1i}, \widehat{U}_{2i}, \widehat{U}_{3i}) = (F_{1n}(X_{1i}), F_{2n}(X_{2i}), F_{3n}(X_{3i})),$$

one obtains the pseudo sample $\widehat{\mathcal{U}} = \{(\widehat{U}_{11}, \widehat{U}_{21}, \widehat{U}_{31}), \ldots, (\widehat{U}_{1n}, \widehat{U}_{2n}, \widehat{U}_{3n})\}$ that can be used to make inference on the dependence parameters. Specifically, to estimate $\boldsymbol{\theta}$, one can replace $\mathcal{U}$ by $\widehat{\mathcal{U}}$ in (6) and maximize jointly the resulting pseudo log-likelihood function $\widehat{L}(\boldsymbol{\theta})$. This estimator, denoted by $\tilde{\boldsymbol{\theta}}$, is consistent and asymptotically Normal under regularity conditions given in [11,21,23].

As an alternative, the sequential estimation procedure proposed in [1] can be adapted. In the first step, parameter values $\hat{\theta}_{12}$ and $\hat{\theta}_{23}$ are found that maximize $\widehat{L}_{12}(\theta_{12})$ and $\widehat{L}_{23}(\theta_{23})$, respectively. In the second step, the estimate $\hat{\theta}_{13|2}$ is obtained by maximizing

$$\widehat{L}^*_{13|2}(\theta_{13|2}) = \sum_{i=1}^{n} \ln c_{13|2}(\widehat{U}^*_{(1|2)i}, \widehat{U}^*_{(3|2)i}; \theta_{13|2}),$$

where, for each $i \in \{1, \ldots, n\}$,

$$\widehat{U}^*_{(1|2)i} = h_{12}(\widehat{U}_{1i}, \widehat{U}_{2i}; \hat{\theta}_{12}), \qquad \widehat{U}^*_{(3|2)i} = h_{32}(\widehat{U}_{3i}, \widehat{U}_{2i}; \hat{\theta}_{23}). \tag{8}$$

Consistency and asymptotic normality of the stepwise semiparametric estimator $\hat{\boldsymbol{\theta}} = (\hat{\theta}_{12}, \hat{\theta}_{23}, \hat{\theta}_{13|2})$ are established in [12]. Simulation studies in [12,13] further reveal that $\hat{\boldsymbol{\theta}}$ is asymptotically slightly less efficient than the estimator $\tilde{\boldsymbol{\theta}}$ obtained through joint maximization. In general, $\tilde{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$ are in close agreement. However, $\hat{\boldsymbol{\theta}}$ is often the only feasible solution, especially in high dimensions where a joint maximization is too computationally intensive.

4. A critical look at the simplifying assumption

The estimation techniques presented in Section 3 rely critically on the assumption that the conditional copula $C_{13|2}$ in model (4) does not depend on the conditioning variable. It is argued in [14] that this simplification is not only required for fast, flexible, and robust inference, but that it provides "a rather good approximation, even when the simplifying assumption is far from being fulfilled by the actual model". It will be shown below that this view is too optimistic.

Table 1
Parameter estimates for the data displayed in Fig. 1.

Model	Known margins	Unknown margins
$\gamma = 0$	$\check{\theta}_{12} = 1.300$ $\check{\theta}_{23} = 2.964$ $\check{\theta}_{13\|2} = 0.047$	$\hat{\theta}_{12} = 1.252$ $\hat{\theta}_{23} = 2.973$ $\hat{\theta}_{13\|2} = 0.037$
$\gamma = 1$	$\check{\theta}_{12} = 1.249$ $\check{\theta}_{23} = 2.737$ $\check{\theta}_{13\|2} = 0.215$	$\hat{\theta}_{12} = 1.203$ $\hat{\theta}_{23} = 2.699$ $\hat{\theta}_{13\|2} = 0.262$

To this end, assume for simplicity that $(X_1, X_2, X_3) = (U_1, U_2, U_3)$ is a random vector with standard uniform margins. Further suppose that

(a) C_{12} is a Clayton copula with parameter $\theta_{12} = 1.2$;
(b) C_{23} is a Gumbel–Hougaard copula with parameter $\theta_{23} = 3$;
(c) given $U_2 = u_2$, $C_{13|2}$ is a Frank copula with parameter

$$\theta_{13|2}(u_2) = \gamma(4u_2 - 2)^3,$$

where $\gamma \in \{0, 1\}$. When $\gamma = 0$, the variables U_1 and U_3 are conditionally independent given U_2, and hence the simplifying assumption is satisfied. When $\gamma = 1$, however, the conditional copula $C_{13|2}$ depends on the value of U_2 and the resulting model is not a simplified PCC.

The algorithm below describes how to simulate from the above model.

Algorithm 1. To generate the random triple (U_1, U_2, U_3), proceed as follows.

1. Simulate independent standard uniform variates W_1, W_2, W_3.
2. Set $U_1 = W_1$.
3. Set $U_2 = h_{21}^{-1}(W_2, U_1; \theta_{12})$.
4. (a) If $\gamma = 0$, set $U_3 = h_{32}^{-1}(W_3, U_2; \theta_{23})$.
   (b) If $\gamma = 1$, set
   $$U_3 = h_{32}^{-1}\left[h_{31|2}^{-1}\{W_3, h_{12}(U_1, U_2; \theta_{12}); \theta_{13|2}(U_2)\}, U_2; \theta_{23}\right].$$

Here, h_{12}, h_{21}, and h_{32} are defined as in (5) but their dependence on parameters is made explicit for additional clarity. Similarly,

$$h_{31|2}(u, v) = \frac{\partial}{\partial v} C_{31|2}(u, v)$$

but its dependence on $\theta_{13|2}(u_2)$ is emphasized. Finally, $h^{-1}(u, v)$ generally refers to the inverse of the map $u \mapsto h(u, v)$ with fixed $v \in (0, 1)$.

Fig. 1 shows pairwise scatter plots derived from random samples of size $n = 500$ from (U_1, U_2, U_3) generated using Algorithm 1 when $\gamma = 0$ (top panel) and $\gamma = 1$ (bottom panel), respectively. As can be seen, these samples look fairly similar. For pairs (U_1, U_2) and (U_2, U_3), this similarity is entirely expected as by construction, each of these two pairs has exactly the same distribution whether $\gamma = 0$ or 1. For this reason, the estimates of θ_{12} and θ_{23} are equal within sampling variation, as can be seen in Table 1.

While it is not possible to tell whether $\gamma = 0$ or 1 from Fig. 1, one could hope to distinguish between these two models by looking at scatter plots of the pairs of pseudo observations $(U^*_{1|2}, U^*_{3|2})$ or $(\widehat{U}^*_{1|2}, \widehat{U}^*_{3|2})$, depending on whether the margins are known or not. Assuming that copulas C_{12} and C_{23} are known, one can construct these pairs using the estimates of θ_{12} and θ_{23} given in Table 1; the relevant equation is (7) when margins are known and (8) when they are not. The resulting scatter plots are displayed in Fig. 2. The four graphs are very similar and suggest that under both models, the variables U_1 and U_2 are conditionally independent, given U_2.

Now suppose that a simplified PCC model is fitted in which the conditional copula $C_{13|2}$ belongs to the Frank family. One is then led to the estimates $\check{\theta}_{13|2}$ and $\hat{\theta}_{13|2}$ reported in Table 1. All of them seem close to zero, in line with Fig. 2. To test whether this conclusion is statistically significant, the experiment was repeated 1000 times. Displayed in Fig. 3 are boxplots showing the dispersion of the estimates $\check{\theta}_{13|2}$ and $\hat{\theta}_{13|2}$. These pictures confirm that whether the margins are known or not, one cannot reject the hypothesis of conditional independence, which corresponds to $\theta_{13|2} = 0$.

While this conclusion is valid when $\gamma = 0$, it is clearly mistaken when $\gamma = 1$. The problem is that when the simplifying assumption is unwarranted, as in the case $\gamma = 1$, the pairs $(U^*_{1|2}, U^*_{3|2})$ or $(\widehat{U}^*_{1|2}, \widehat{U}^*_{3|2})$ are misinterpreted as a pseudo sample from a single copula $C_{13|2}$ that does not actually exist. It will be shown in the following section how this issue can be resolved using local likelihood techniques.

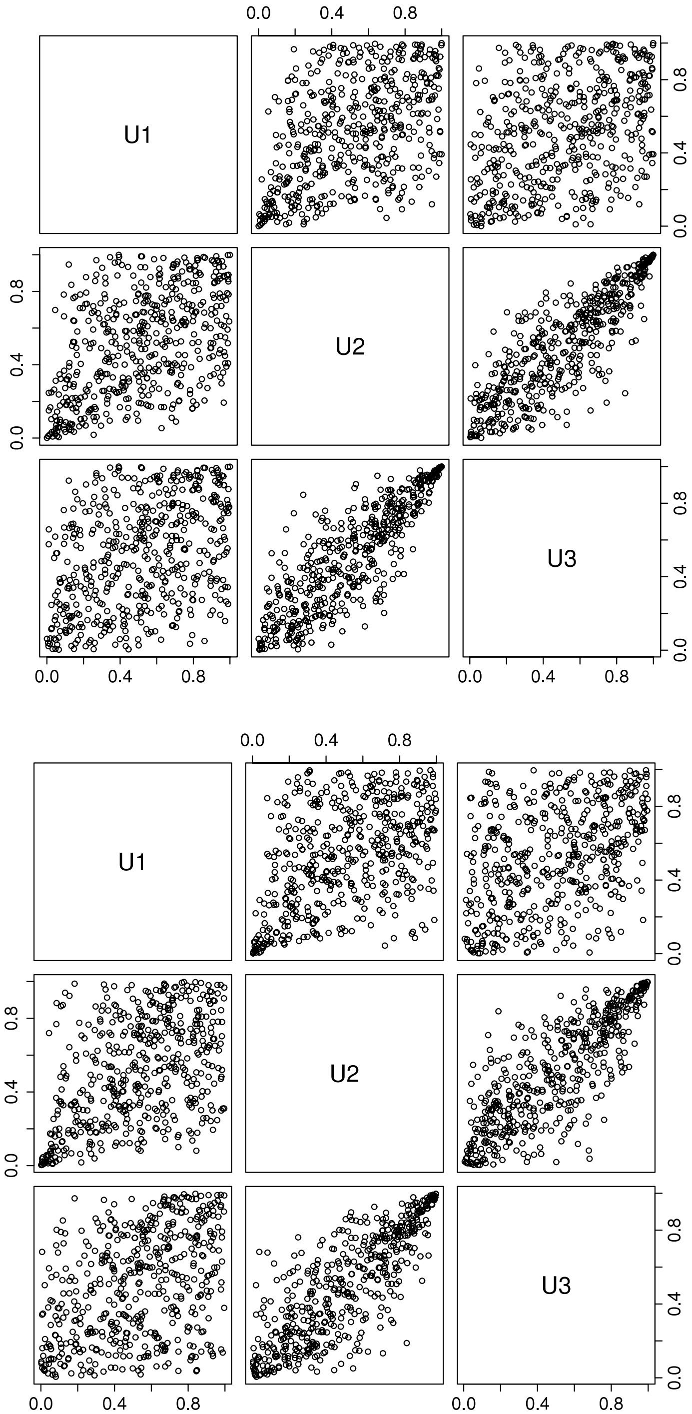

Fig. 1. Pairwise scatter plots of variables derived from a random sample of size $n = 500$ from (U_1, U_2, U_3) generated by Algorithm 1 when $\gamma = 0$ (top) and $\gamma = 1$ (bottom).

5. Inference for general PCCs

As illustrated in Section 4, the assumption that conditional copulas do not depend on the conditioning variables should not be made blindly. Because it may have undesirable consequences, its validity should be assessed, at least graphically. One such technique is presented below, based on the assumption that the conditional copula $C_{13|2}$ has the same parametric form for all values of the corresponding conditioning variable X_2.

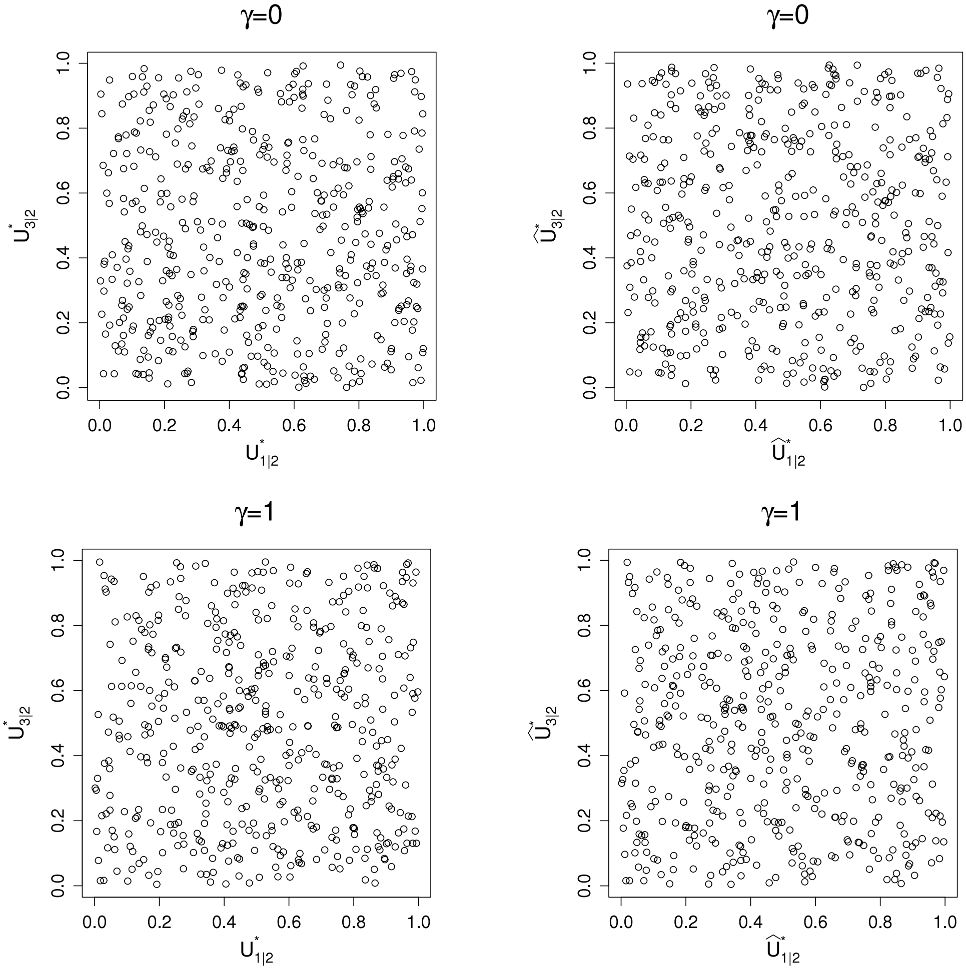

Fig. 2. Scatter plots of the pairs of pseudo conditional marginal distributions obtained using known margins (left panel) and rank-based margins (right panel).

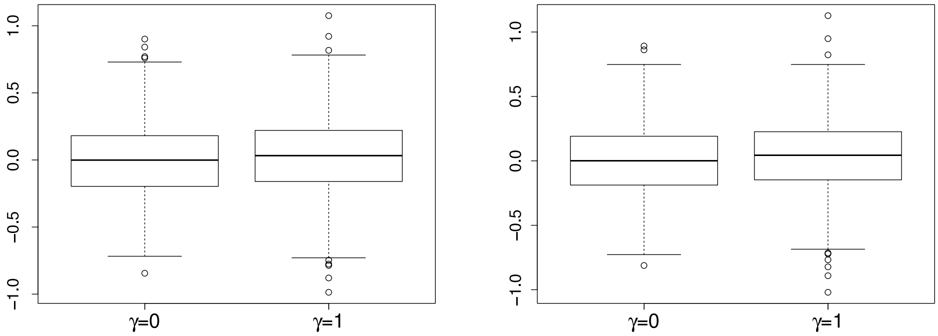

Fig. 3. Boxplots of estimates $\check{\theta}_{13|2}$ (left) and $\hat{\theta}_{13|2}$ (right) obtained from 1000 Monte Carlo samples of size $n = 500$.

In such a case, nonparametric local likelihood methods can be used to estimate $\theta_{13|2}$ as a function of x_2. To avoid complications that occur when the set of values taken by $\theta_{13|2}$ is restricted, introduce a reparametrization

$$\eta_{13|2}(x_2) = g\{\theta_{13|2}(x_2)\},$$

where g is any convenient link function ensuring that $\eta_{13|2}(x_2)$ can take potentially any value in $\mathbb{R}$ as x_2 varies over its domain. The choice of g is entirely arbitrary and does not affect inference. When $C_{13|2}$ belongs to Frank's family, g can be the identity; when $C_{13|2}$ is a Clayton copula with unspecified positive association, $g(x) = \ln(x)$ is a convenient choice.

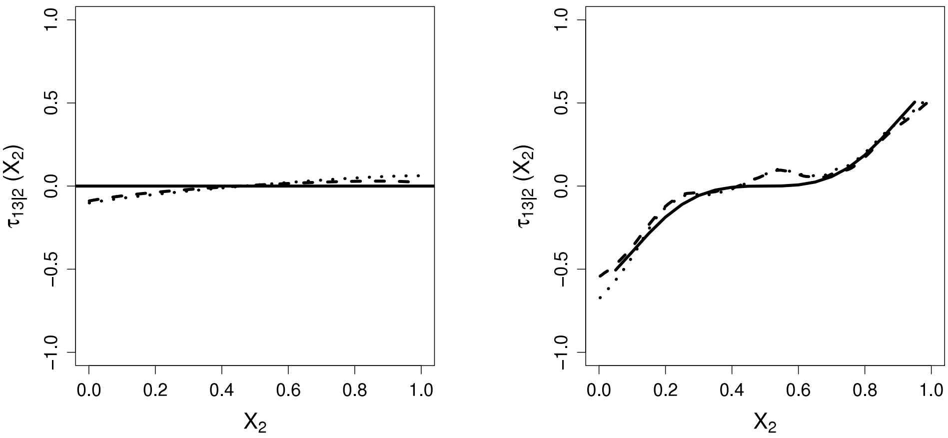

Fig. 4. Plots of $\tau(X_1, X_3|X_2 = x_2)$ as a function of x_2 assuming a Frank copula for $C_{13|2}$, as derived from $\hat{\theta}_{13|2}$ (dashed) and $\check{\theta}_{13|2}$ (dotted) for the data of Fig. 1 when $\gamma = 0$ (left) and $\gamma = 1$ (right). In both graphs, the true function is shown as a solid curve.

Following [4], suppose that $\eta_{13|2}$ is twice continuously differentiable at any interior point x in the support of F_2. For any observation X_{2i} in a neighborhood of such an x, one can then write

$$\eta_{13|2}(X_{2i}) \approx \eta_{13|2}(x) + \eta'_{13|2}(x)(X_{2i} - x) \equiv \beta_{0x} + \beta_{1x}(X_{2i} - x).$$

In principle, a higher order polynomial could also be used in the local approximation, though at the cost of estimating more parameters. As a local linear fit often suffices to represent the underlying function [10], this is the approach taken here.

An estimate of $\boldsymbol{\beta}_x = (\beta_{0x}, \beta_{1x})$, and hence of $\eta_{13|2}(x) = \beta_{0x}$, can be obtained by maximizing a local version of the log-likelihood function in which the contribution of the ith observation is weighted by its proximity to x. To be specific, let K be a smooth kernel function and for arbitrary $t \in \mathbb{R}$, define $K_{\lambda_n}(t) = K(t/\lambda_n)/\lambda_n$, where $\lambda_n > 0$ is a bandwidth parameter controlling the size of the neighborhood around x. When the marginal distributions are known, a kernel-weighted local log-likelihood function is then given by

$$\mathcal{L}^*(\boldsymbol{\beta}_x, x) = \sum_{i=1}^{n} K_{\lambda_n}(X_{2i} - x) \ln c_{13|2}\left[U^*_{(1|2)i}, U^*_{(3|2)i}; g^{-1}\{\beta_{0x} + \beta_{1x}(X_{2i} - x)\}\right].$$

When the margins are unknown, a rank-based equivalent is given by

$$\widehat{\mathcal{L}}^*(\boldsymbol{\beta}_x, x) = \sum_{i=1}^{n} K_{\lambda_n}(X_{2i} - x) \ln c_{13|2}\left[\widehat{U}^*_{(1|2)i}, \widehat{U}^*_{(3|2)i}; g^{-1}\{\beta_{0x} + \beta_{1x}(X_{2i} - x)\}\right].$$

Let $\check{\boldsymbol{\beta}}_x$ and $\hat{\boldsymbol{\beta}}_x$ be parameter values maximizing $\mathcal{L}^*(\boldsymbol{\beta}_x, x)$ and $\widehat{\mathcal{L}}^*(\boldsymbol{\beta}_x, x)$, respectively. Typically, the choice of kernel has little influence on these estimates; in the current study, computations were based on the Epanechnikov kernel defined, for all $t \in \mathbb{R}$, by $K(t) = 0.75 \max(0, 1 - t^2)$. However, the local linear estimates $\check{\boldsymbol{\beta}}_x$ and $\hat{\boldsymbol{\beta}}_x$ are sensitive to the choice of bandwidth λ_n. The data-driven procedure of [4] can be used to make a suitable selection.

Once $\check{\boldsymbol{\beta}}_x$ or $\hat{\boldsymbol{\beta}}_x$ has been obtained, an estimate of $\theta_{13|2}(x)$ is given by

$$\check{\theta}_{13|2}(x) = g^{-1}(\check{\beta}_{0x}) \quad \text{or} \quad \hat{\theta}_{13|2}(x) = g^{-1}(\hat{\beta}_{0x}),$$

respectively. Repeating this procedure for a large number of values of x over the range of X_2, an estimate of $\theta_{13|2}$ is obtained as a function of x_2. By plotting this function, one can develop a sense of whether $\theta_{13|2}$ is functionally dependent on X_2, i.e., whether the simplifying assumption is reasonable.

This is illustrated in Fig. 4 for the data from Section 4. Displayed there are the plots of $\tau(X_1, X_3|X_2 = x_2)$ as a function of x_2, where τ denotes the value of Kendall's tau, obtained from either $\check{\theta}_{13|2}$ or $\hat{\theta}_{13|2}$, according as the margins are known or not. The left and right panels correspond to $\gamma = 0$ and $\gamma = 1$, respectively. In both cases, the estimate reproduces quite closely the true curve (solid), regardless of whether the margins are known or not. The flat curves in the left panel are in line with the simplifying assumption for the case $\gamma = 0$, while the nonlinear pattern in the right panel indicates that the simplifying assumption may not be reasonable for the case $\gamma = 1$.

To obtain Fig. 4, a local linear estimation was performed at each observed value of the conditioning variable X_2 under the (correct) assumption that the conditional copula $C_{13|2}$ belongs to Frank's family. The bandwidths used were $\lambda_n = 1$ and 0.17 when $\gamma = 0$ and 1, respectively. These bandwidths were selected among six pilot bandwidth values ranging from 0.05 to 1, equally spaced on a logarithmic scale.

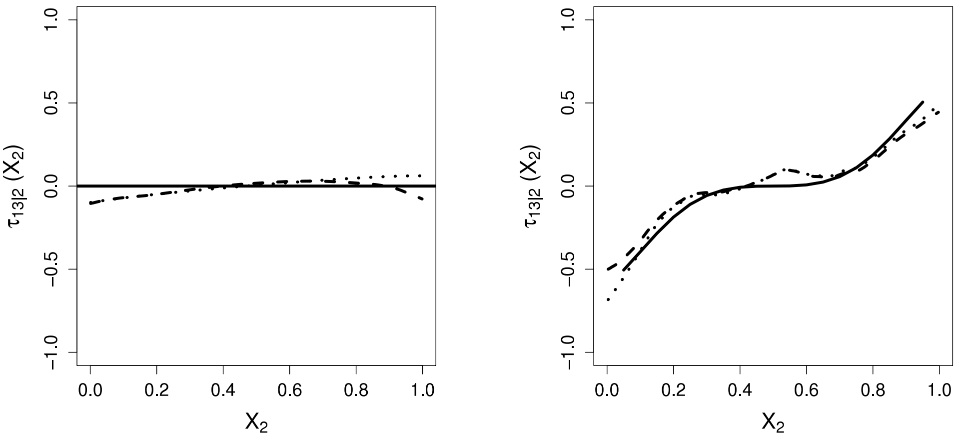

Fig. 5. Plots of $\tau(X_1, X_3|X_2 = x_2)$ as a function of x_2 assuming a Plackett copula for $C_{13|2}$, as derived from $\hat{\theta}_{13|2}$ (dashed) and $\check{\theta}_{13|2}$ (dotted) for the data of Fig. 1 when $\gamma = 0$ (left) and 1 (right). In both graphs, the true function is shown as a solid curve.

The bandwidth selection was based on the cross-validated likelihood criterion described in [4]. As it happens, this data-driven procedure led to the same choice whether the margins are known or estimated through standardized ranks. When $\gamma = 0$, the bandwidth value $\lambda_n = 1$ corresponds to a global fit, as might have been expected from the fact that $C_{13|2}$ does not depend on X_2. When $\gamma = 1$, however, the selected bandwidth turned out to be much smaller to reflect the local features of the underlying dependence function.

The estimates in Fig. 4 are obtained under the true Frank copula. One may wonder whether misspecifying the conditional copula would lead to a different conclusion. To investigate the performance of the local linear estimator under copula misspecification, the exercise was repeated assuming Plackett copulas, which share some – but not all – properties of Frank copulas; for example, Plackett and Frank copulas are both symmetric and exhibit no tail dependence. However, Frank copulas are Archimedean whereas Plackett copulas are characterized by a constant odds ratio relationship; see, e.g., [19]. As Fig. 5 shows, the copula misspecification had a negligible effect on the estimates, at least in this case.

6. Simulation results

In order to assess the sampling performance of the local linear estimators $\hat{\theta}_{13|2}$ and $\check{\theta}_{13|2}$, a Monte Carlo experiment was conducted using the model from Section 4. More specifically, Algorithm 1 was used to generate 100 samples of size $n = 500$, both when $\gamma = 0$ and 1. For each Monte Carlo sample, the conditional copula parameter $\theta_{13|2}$ was estimated both under the assumption of a simplified PCC and under the more general hypothesis that the parameter $\theta_{13|2}$ depends on x_2, as described in Section 5.

Table 2 presents the results in terms of Integrated Square Bias (ISB), Variance (IVAR), and Mean Square Error (IMSE), as measured upon conversion of the parameter estimates to Kendall's tau scale. When working with ranks, for instance, these are defined by

$$\mathrm{ISB}(\hat{\tau}_{13|2}) = \int_{\mathcal{X}} [\mathrm{E}\{\hat{\tau}_{13|2}(x)\} - \tau_{13|2}(x)]^2 \, \mathrm{d}x,$$

$$\mathrm{IVAR}(\hat{\tau}_{13|2}) = \int_{\mathcal{X}} \mathrm{E}([\hat{\tau}_{13|2}(x) - \mathrm{E}\{\hat{\tau}_{13|2}(x)\}]^2) \, \mathrm{d}x,$$

and

$$\mathrm{IMSE}(\hat{\tau}_{13|2}) = \mathrm{ISB}(\hat{\tau}_{13|2}) + \mathrm{IVAR}(\hat{\tau}_{13|2}).$$

In these expressions, the expectations were replaced by averages over the 100 Monte Carlo samples and the integrals were approximated by taking a sum over 19 equally spaced grid points from 0.05 to 0.95, inclusively.

Based on Table 2, the following observations can be made:

(a) Whether in the simplified or general PCC framework, the rank-based estimator $\hat{\tau}_{13|2}$ seems to perform at least as well as the estimator $\check{\tau}_{13|2}$ designed for the case when the margins are known.
(b) When the simplifying assumption holds ($\gamma = 0$), the local linear estimator performs worse than the likelihood estimator, especially in terms of variance.
(c) When the simplifying assumption is violated ($\gamma = 1$), the likelihood estimator is severely biased while the local linear estimator performs well, despite its slightly higher variance.

In support of the above conclusions, a graphical summary of the local linear estimates on Kendall's tau scale is given in Fig. 6. Regardless of whether $\gamma = 0$ or 1, the local linear estimator is close to $\tau_{13|2}$ both when the margins are known and estimated. To assess the variability of local linear estimates across the range of X_2, 90% pointwise Monte Carlo confidence

Table 2
Integrated square bias, variance, and mean square error ($\times 100$) of two estimators of $\tau_{13|2}$ in the model of Section 4 with $\gamma = 0$ or 1. The last column shows the average $\bar{\lambda}_n$ of the data-driven bandwidths.

		Simplified PCC			General PCC			
γ		ISB	IVAR	IMSE	ISB	IVAR	IMSE	$\bar{\lambda}_n$
0	$\check{\tau}_{13\|2}$	0.000	0.083	0.084	0.006	0.394	0.400	0.768
	$\hat{\tau}_{13\|2}$	0.000	0.086	0.086	0.005	0.365	0.370	0.796
1	$\check{\tau}_{13\|2}$	5.718	0.102	5.819	0.029	0.511	0.540	0.206
	$\hat{\tau}_{13\|2}$	5.718	0.098	5.816	0.035	0.482	0.517	0.224

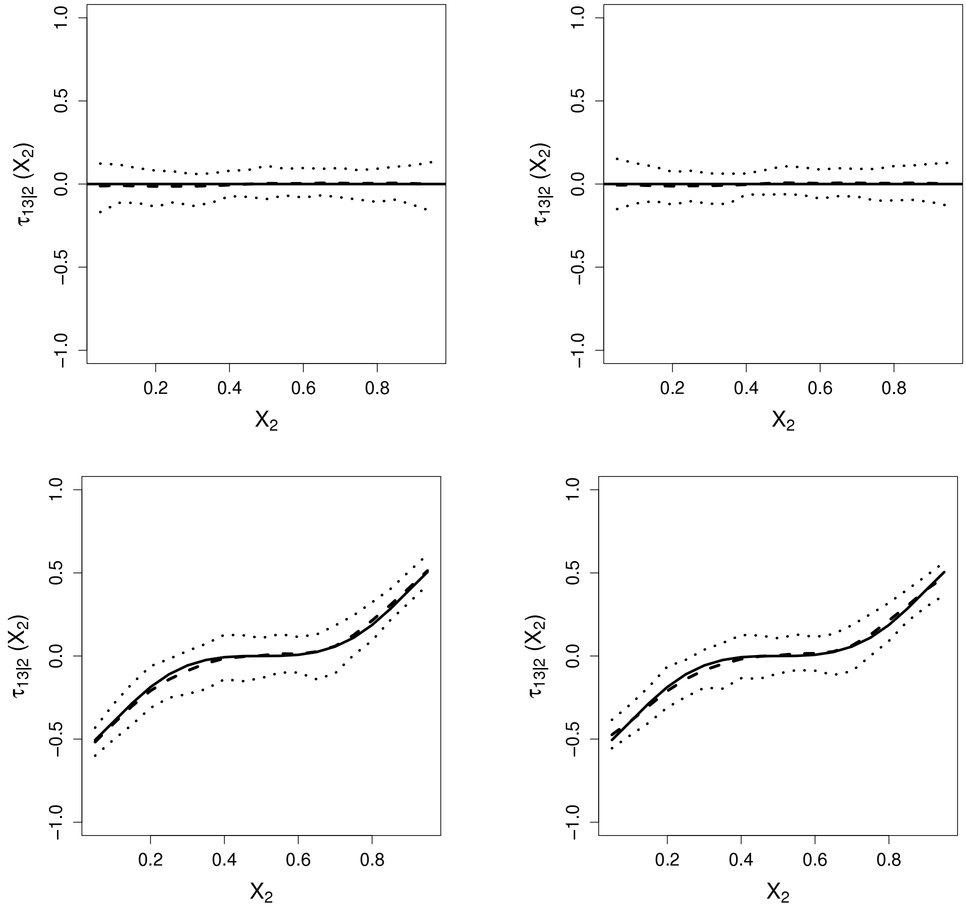

Fig. 6. Plots of $\tau_{13|2}$ and its estimates $\check{\tau}_{13|2}$ (left) and $\hat{\tau}_{13|2}$ (right) as a function of x_2 when $\gamma = 0$ (top) and 1 (bottom). In each panel, the true function is shown as a solid curve, the average of the estimates taken over 100 Monte Carlo samples is displayed by the dashed curve and the 90% Monte Carlo confidence intervals are given by dotted curves.

intervals are provided. As can be seen in Fig. 6, the true functional form of $\tau_{13|2}$ falls within the confidence intervals both when $\gamma = 0$ and 1. Overall, these results suggest that $\check{\theta}_{12|3}$ and $\hat{\theta}_{12|3}$ are consistent estimates of the conditional copula parameter $\theta_{13|2}$.

7. Consistency of the estimators

The purpose of this section is to establish the consistency of the local linear estimator of $\theta_{13|2}$ defined in Section 5. When the margins F_1, F_2, F_3 and the parameters θ_{12}, θ_{23} are known, the proposed procedure reduces to the estimator considered in [3,4]. It is shown in [3] that the latter estimator is asymptotically Normal. Furthermore, expressions for its limiting bias and variance are given in [4].

When the parameters θ_{12} and θ_{23} are unknown, as in the present context, the results in [3,4] are not directly applicable. It is shown below that the local linear estimator remains consistent, both when the margins F_1, F_2, F_3 are known and unknown. The proof is detailed here in the most realistic case, i.e., when the margins are unknown. The argument is similar to the proof given elsewhere in this Special Issue [2] for the case where nonparametric estimates of $F_{1|2}$ and $F_{3|2}$ are used.

Recall that to obtain the local linear estimator $\hat{\theta}_{13|2}$, the distribution functions of X_1 and X_3 given $X_2 = x_2$ are first estimated by

$$\widehat{F}_{1|2}(x_1|x_2) = h_{12}\{F_{1n}(x_1), F_{2n}(x_2); \hat{\theta}_{12}\},$$
$$\widehat{F}_{3|2}(x_3|x_2) = h_{32}\{F_{3n}(x_1), F_{2n}(x_2); \hat{\theta}_{23}\},$$

respectively. Here, h_{k2} is defined as in (5) for $k = 1, 3$. For $i \in \{1, \ldots, n\}$, denote the linear approximation of $\eta_{13|2}(X_{2i})$ by

$$\bar{\eta}_{13|2}(x, X_{2i}) = \beta_{0x} + \beta_{1x}(X_{2i} - x) = \beta_{0x} + \beta_{1xn}\left(\frac{X_{2i} - x}{\lambda_n}\right),$$

where $\beta_{1xn} = \lambda_n \beta_{1x}$. Further write

$$\ell(\eta, u, v) = \ln c_{13|2}\{u, v; g^{-1}(\eta)\}$$

and for arbitrary integers r, s, t, let

$$\ell_{rst}(\eta, u, v) = \frac{\partial^{r+s+t}}{\partial \eta^r\, \partial u^s\, \partial v^t}\, \ell(\eta, u, v).$$

The reparametrization $\boldsymbol{\beta}_{xn} = (\beta_{0xn}, \beta_{1xn}) = (\beta_{0x}, \beta_{1xn})$ then leads to the local log-likelihood function given by

$$\widehat{\mathcal{L}}^*_{\lambda_n}(\boldsymbol{\beta}_{xn}, x) = \sum_{i=1}^{n} K_{\lambda_n}(X_{2i} - x)\ell\left\{\bar{\eta}_{13|2}(x, X_{2i}), \widehat{F}_{1|2}(X_{1i}|X_{2i}), \widehat{F}_{3|2}(X_{3i}|X_{2i})\right\}.$$

Now assume that ℓ is sufficiently smooth that it can be expanded in Taylor series with respect to its first argument. Precise conditions are spelled out in Appendix A. If $\boldsymbol{b} = (b_0, b_1)$ is sufficiently close to $\boldsymbol{\beta}_{xn}$, one can then write

$$\frac{1}{n}\{\widehat{\mathcal{L}}^*_{\lambda_n}(\boldsymbol{b}, x) - \widehat{\mathcal{L}}^*_{\lambda_n}(\boldsymbol{\beta}_{xn}, x)\} = \hat{S}_{1n}(x) + \hat{S}_{2n}(x) + \hat{S}_{3n}(x)$$

in terms of

$$\hat{S}_{1n}(x) = \sum_{r=0}^{1} \hat{A}_{rn}(x)(b_r - \beta_{rxn}),$$
$$\hat{S}_{2n}(x) = \frac{1}{2}\sum_{r=0}^{1}\sum_{s=0}^{1} \hat{B}_{rsn}(x)(b_r - \beta_{rxn})(b_s - \beta_{sxn}),$$
$$\hat{S}_{3n}(x) = \frac{1}{6}\sum_{r=0}^{1}\sum_{s=0}^{1}\sum_{t=0}^{1} \hat{C}_{rstn}(x)(b_r - \beta_{rxn})(b_s - \beta_{sxn})(b_t - \beta_{txn}),$$

where

$$\hat{A}_{rn}(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_{2i} - x}{\lambda_n}\right)^{r} K_{\lambda_n}(X_{2i} - x)\ell_{100}\{\bar{\eta}_{13|2}(x, X_{2i}), \widehat{F}_{1|2}(X_{1i}|X_{2i}), \widehat{F}_{3|2}(X_{3i}|X_{2i})\},$$
$$\hat{B}_{rsn}(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_{2i} - x}{\lambda_n}\right)^{r+s} K_{\lambda_n}(X_{2i} - x)\ell_{200}\{\bar{\eta}_{13|2}(x, X_{2i}), \widehat{F}_{1|2}(X_{1i}|X_{2i}), \widehat{F}_{3|2}(X_{3i}|X_{2i})\},$$
$$\hat{C}_{rstn}(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_{2i} - x}{\lambda_n}\right)^{r+s+t} K_{\lambda_n}(X_{2i} - x)\ell_{300}\{\eta^*(x, X_{2i}), \widehat{F}_{1|2}(X_{1i}|X_{2i}), \widehat{F}_{3|2}(X_{3i}|X_{2i})\},$$

and, for $i \in \{1, \ldots, n\}$, $\eta^*(x, X_{2i})$ lies between $\bar{\eta}_{13|2}(x, X_{2i})$ and $b_0 + b_1(X_{2i} - x)/\lambda_n$.

The following result constitutes the first step in establishing the consistency of $\hat{\theta}_{13|2}$. Its proof is detailed in Appendix B.

Lemma 1. *Assume that regularity conditions* (A1)–(A3), (C1)–(C3) *and* (D) *listed in Appendix* A *hold. If* x *is in the interior* $\mathcal{X}_2$ *of the support of* F_2, *then, as* $n \to \infty$,

$$|\hat{A}_{rn}(x) - A_{rn}(x)| \xrightarrow{p} 0, \qquad |\hat{B}_{rsn}(x) - B_{rsn}(x)| \xrightarrow{p} 0, \qquad |\hat{C}_{rstn}(x) - C_{rstn}(x)| \xrightarrow{p} 0$$

for all $r, s, t \in \{0, 1\}$, *where*

$$A_{rn}(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_{2i}-x}{\lambda_n}\right)^{r} K_{\lambda_n}(X_{2i}-x)\ell_{100}\{\eta_{13|2}(X_{2i}), F_{1|2}(X_{1i}|X_{2i}), F_{3|2}(X_{3i}|X_{2i})\},$$

$$B_{rsn}(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_{2i}-x}{\lambda_n}\right)^{r+s} K_{\lambda_n}(X_{2i}-x)\ell_{200}\{\eta_{13|2}(X_{2i}), F_{1|2}(X_{1i}|X_{2i}), F_{3|2}(X_{3i}|X_{2i})\},$$

$$C_{rstn}(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_{2i}-x}{\lambda_n}\right)^{r+s+t} K_{\lambda_n}(X_{2i}-x)\ell_{300}\{\eta^{*}(x, X_{2i}), F_{1|2}(X_{1i}|X_{2i}), F_{3|2}(X_{3i}|X_{2i})\}.$$

Thus for arbitrary $r, s, t \in \{0, 1\}$, $\hat{A}_{rn}, \hat{B}_{rsn}, \hat{C}_{rstn}$ behave asymptotically as $A_{rn}, B_{rsn}, C_{rstn}$, respectively. The limiting behavior of the latter quantities is stated below and proved in Appendix B. In what follows,

$$\begin{aligned}\mathcal{I}(x) &= \mathrm{E}\left[\ell_{100}^{2}\left[g\{\theta_{13|2}(x)\}, F_{1|2}(X_1|x), F_{3|2}(X_3|x)|X_2 = x\right]\right]\\ &= -\mathrm{E}\left[\ell_{200}\left[g\{\theta_{13|2}(x)\}, F_{1|2}(X_1|x), F_{3|2}(X_3|x)|X_2 = x\right]\right]\end{aligned}$$

denotes the Fisher Information for $g\{\theta_{13|2}(x)\}$ at any possible x.

Lemma 2. *Assume that regularity conditions* (A1)–(A3), (B1)–(B2), (C1)–(C3) *and* (D) *listed in* Appendix A *hold. If* x *is in the interior* $\mathcal{X}_2$ *of the support of* F_2*, then, as* $n \to \infty$,

(a) $|A_{0n}(x)| \xrightarrow{p} 0$ *and* $|A_{1n}(x)| \xrightarrow{p} 0$;
(b) $|B_{01n}(x)| \xrightarrow{p} 0$ *and* $|B_{10n}(x)| \xrightarrow{p} 0$, *while*

$$B_{00n}(x) \xrightarrow{p} -\mathcal{I}(x)f_2(x) \quad \text{and} \quad B_{11n}(x) \xrightarrow{p} -\mathcal{I}(x)f_2(x)\mu_{11},$$

where $\mu_{11} = \int t^2K(t)\,\mathrm{d}t$;
(c) $\sum_{r,s,t}|C_{rstn}(x)| < M_x$ *in probability for some constant* $M_x > 0$.

The following theorem establishes the consistency of the rank-based local linear likelihood estimator $\hat{\theta}_{13|2}$. A similar result holds for the estimator $\breve{\theta}_{13|2}$ in the known-margin case.

Theorem 1. *Assume that regularity conditions* (A1)–(A3), (B1)–(B2), (C1)–(C3) *and* (D) *listed in* Appendix A *hold. For arbitrary* $x \in \mathcal{X}_2$*, there exist solutions* $\hat{\boldsymbol{\beta}} = (\hat{\beta}_{0x}, \hat{\beta}_{1x})$ *to the local likelihood equations* $\partial\widehat{\mathcal{L}}^{*}(\boldsymbol{\beta}_x, x)/\partial\boldsymbol{\beta}_x = 0$ *such that, as* $n \to \infty$,

$$\hat{\beta}_{0x} \xrightarrow{p} \beta_{0x}, \qquad \lambda_n(\hat{\beta}_{1x} - \beta_{1x}) \xrightarrow{p} 0.$$

Proof. For a given $x \in \mathcal{X}_2$, let

$$Q_x = \frac{1}{2}f_2(x)\mathcal{I}(x)\mu_{11}$$

and fix $\varepsilon \in (0, 3Q_x/M_x)$, where μ_{11} and $M_x > 0$ are as in Lemma 2. Let also

$$D_\varepsilon = \{\boldsymbol{b} \in \mathbb{R}^2 : |b_0 - \beta_{0xn}|^2 + |b_1 - \beta_{1xn}|^2 = \varepsilon^2\}.$$

It turns out that

$$\lim_{n\to\infty} \Pr\{\widehat{\mathcal{L}}^{*}_{\lambda_n}(\boldsymbol{b}, x) < \widehat{\mathcal{L}}^{*}_{\lambda_n}(\boldsymbol{\beta}_{xn}, x) \text{ for all } \boldsymbol{b} \in D_\varepsilon\} = 1. \tag{9}$$

To prove this claim, first observe that for arbitrary $\boldsymbol{b} \in D_\varepsilon$, one has

$$\{\widehat{\mathcal{L}}^{*}_{\lambda_n}(\boldsymbol{b}, x) \geq \widehat{\mathcal{L}}^{*}_{\lambda_n}(\boldsymbol{\beta}_{xn}, x)\} = \{\hat{S}_{1n}(x) + \hat{S}_{2n}(x) + \hat{S}_{3n}(x) \geq 0\},$$

which is a subset of

$$\left\{\hat{S}_{1n}(x) + \hat{S}_{2n}(x) + \hat{S}_{3n}(x) + \frac{1}{2}f_2(x)\mathcal{I}(x)\{(b_0 - \beta_{0xn})^2 + (b_1 - \beta_{1xn})^2\mu_{11}\} \geq Q_x\varepsilon^2\right\}.$$

The latter event is further contained in

$$\begin{aligned}E_n = \Bigg\{&\varepsilon|\hat{A}_{0n}(x)| + \varepsilon|\hat{A}_{1n}(x)| + \frac{\varepsilon^2}{2}|\hat{B}_{10n}(x)| + \frac{\varepsilon^2}{2}|\hat{B}_{01n}(x)|\\ &+ \frac{\varepsilon^2}{2}|\hat{B}_{00n}(x) + f_2(x)\mathcal{I}(x)| + \frac{\varepsilon^2}{2}|\hat{B}_{11n}(x) + f_2(x)\mathcal{I}(x)\mu_{11}| + \frac{\varepsilon^3}{6}\sum_{r,s,t}|\hat{C}_{rstn}(x)| \geq Q_x\varepsilon^2\Bigg\}.\end{aligned}$$

Because E_n does not depend on a particular choice of $\boldsymbol{b} \in D_\varepsilon$, one has

$$\{\widehat{\mathcal{L}}^*_{\lambda_n}(\boldsymbol{b}, x) \geq \widehat{\mathcal{L}}^*_{\lambda_n}(\boldsymbol{\beta}_{xn}, x) \text{ for at least one } \boldsymbol{b} \in D_\varepsilon\} \subseteq E_n.$$

It remains to show that $\Pr(E_n) \to 0$ as $n \to \infty$. To see this, write

$$\Pr(E_n) \leq \Pr(E_{1n}) + \Pr(E_{2n}),$$

where

$$E_{1n} = \left\{\varepsilon|\hat{A}_{0n}(x)| + \varepsilon|\hat{A}_{1n}(x)| + \frac{\varepsilon^2}{2}|\hat{B}_{10n}(x)| + \frac{\varepsilon^2}{2}|\hat{B}_{01n}(x)| \right.$$
$$\left. + \frac{\varepsilon^2}{2}|\hat{B}_{00n}(x) + f_2(x)\mathcal{I}(x)| + \frac{\varepsilon^2}{2}|\hat{B}_{11n}(x) + f_2(x)\mathcal{I}(x)\mu_{11}| \geq \frac{\varepsilon^2 Q_x}{2}\right\}$$

and

$$E_{2n} = \left\{\sum_{r,s,t} |\hat{C}_{rstn}(x)| \geq M_x\right\}.$$

Calling on Lemmas 1–2, one can deduce that both $\Pr(E_{1n})$ and $\Pr(E_{2n})$ go to zero as $n \to \infty$. This establishes claim (9). The latter implies that the probability of $\widehat{\mathcal{L}}^*_{\lambda_n}$ having a local maximum in the ball $B_\varepsilon(\boldsymbol{\beta}_{nx})$ with radius ε and centered at $\boldsymbol{\beta}_{xn}$ tends to 1 as $n \to \infty$, and this for any ε sufficiently small. To complete the argument, one can then proceed as in [17]. For any suitable $\varepsilon > 0$, there exists a sequence $(\hat{\beta}_{0x}, \hat{\beta}_{1x})$ of solutions to the local likelihood equations $\partial\widehat{\mathcal{L}}^*(\boldsymbol{\beta}_x)/\partial\boldsymbol{\beta}_x = 0$ for which one has simultaneously

$$\Pr(|\hat{\beta}_{0x} - \beta_{0x}| > \varepsilon) \to 0, \qquad \Pr(\lambda_n|\hat{\beta}_{1x} - \beta_{1x}| > \varepsilon) \to 0.$$

By choosing the root of the local likelihood equations closest to (β_{0x}, β_{1x}), one can thus obtain a sequence $(\hat{\beta}_{0x}, \hat{\beta}_{1x})$ of roots independently of ε for which the statement of Theorem 1 holds. □

8. Data application

As a practical illustration of the local linear estimation in PCCs, the classical hydro-geochemical stream and sediment reconnaissance data from [9] were revisited. They consist of the observed log-concentrations of seven chemicals in 655 water samples collected near Grand Junction, Colorado.

In particular, consider the pairwise scatter plots of the rank-transformed data shown in Fig. 7, which illustrate the dependence between cobalt (Co), titanium (Ti) and scandium (Sc). These variables are positively associated, as confirmed by the pairwise empirical values of Kendall's tau, viz.

$$\tau_n(\text{Co}, \text{Ti}) = 0.365, \qquad \tau_n(\text{Ti}, \text{Sc}) = 0.436, \qquad \tau_n(\text{Co}, \text{Sc}) = 0.535.$$

As argued in [8,9], the triplet (Co, Ti, Sc) can be jointly modeled neither by a meta-elliptical nor by an extreme-value copula. Fig. 7 suggests that a Student-t copula may be suitable for the dependence between each pair.

Suppose that each pair $(X_1, X_2) = (\text{Co}, \text{Ti})$, $(X_2, X_3) = (\text{Ti}, \text{Sc})$ and $(X_1, X_3) = (\text{Co}, \text{Sc})$ is modeled by some Student-t copula parameterized by $\theta = (\rho, \nu)$ as in Example 5.3.3 of [18]; the maximum pseudo likelihood parameter estimates are then $(\hat{\rho}_{12}, \hat{\nu}_{12}) = (0.53, 7)$, $(\hat{\rho}_{23}, \hat{\nu}_{23}) = (0.62, 6)$ and $(\hat{\rho}_{13}, \hat{\nu}_{13}) = (0.74, 8)$, respectively.

One may wonder whether or not the conditional dependence of the pair (X_1, X_3) given $X_2 = x_2$ can be modeled by a copula that does not depend on x_2. In other words, is the simplified PCC hypothesis justified in this case?

To address this issue, the pseudo observations $\widehat{U}^*_{1|2}$ and $\widehat{U}^*_{3|2}$ were first constructed as in (8). Their joint behavior is illustrated in the left panel of Fig. 8. The clustering of observations in the upper right corner suggests that the Gumbel–Hougaard copula family may be an appropriate model for $C_{13|2}$. Under the simplifying assumption, the maximum pseudo likelihood parameter estimate of $C_{13|2}$ is $\hat{\theta}_{13|2} = 1.65$, which corresponds to $\hat{\tau}_{13|2} = 0.39$. The latter value is represented by a solid line in the right panel of Fig. 8.

Under the more general assumption that $C_{13|2}$ belongs to the Gumbel–Hougaard family whose parameter depends on x_2, $\theta_{13|2}$ can be estimated using the local likelihood technique presented in Section 5. Given that the parameter range of the Gumbel–Hougaard copula is the interval $[1, \infty)$, a convenient link function is $g(t) = \ln(t-1)$. To select an appropriate bandwidth for the local linear estimation, six pilot bandwidth values were considered. They were equally spaced on the logarithmic scale, ranging from 0.30 to 1.52. Ultimately, the cross-validated likelihood criterion led to $\lambda_n = 0.57$. The dashed curve in the right panel of Fig. 8 shows the estimate of $\tau_{13|2}$ corresponding to $\hat{\theta}_{13|2}$ as a function of x_2. To assess the variation in $\theta_{13|2}(X_2)$, 90% pointwise confidence intervals at 31 equally spaced grid points in the range of x_2 were obtained by nonparametric bootstrapping of the original data with 100 bootstrap replicates. As the constant parameter estimate is not entirely contained within the intervals, it appears that the simplifying assumption is not appropriate in this vine construction.

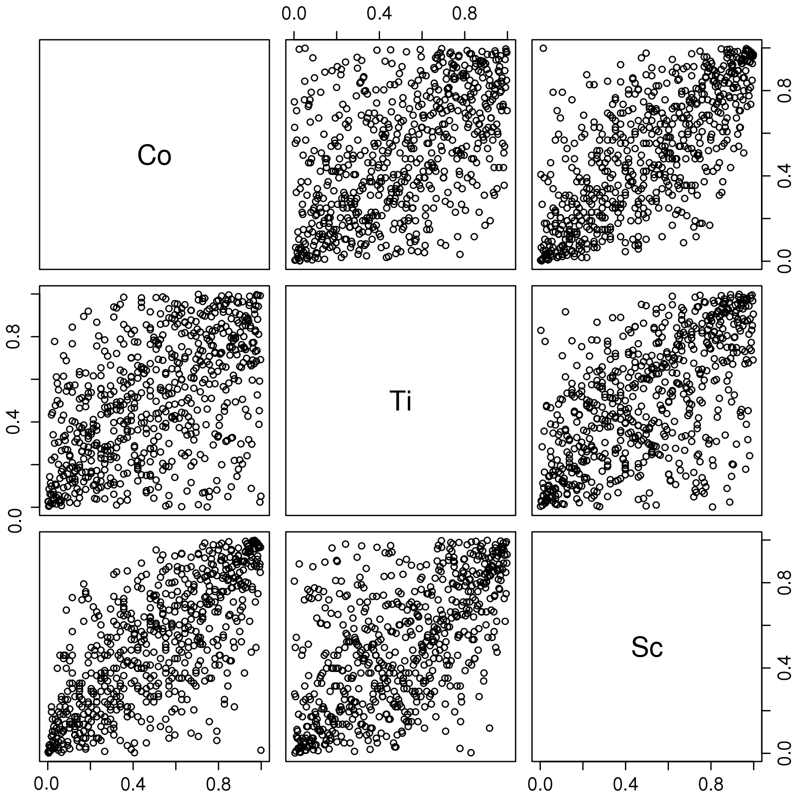

Fig. 7. Pairwise scatter plots of the empirical ranks of cobalt (Co), titanium (Ti) and scandium (Sc) in 655 water samples collected near Grand Junction, Colorado.

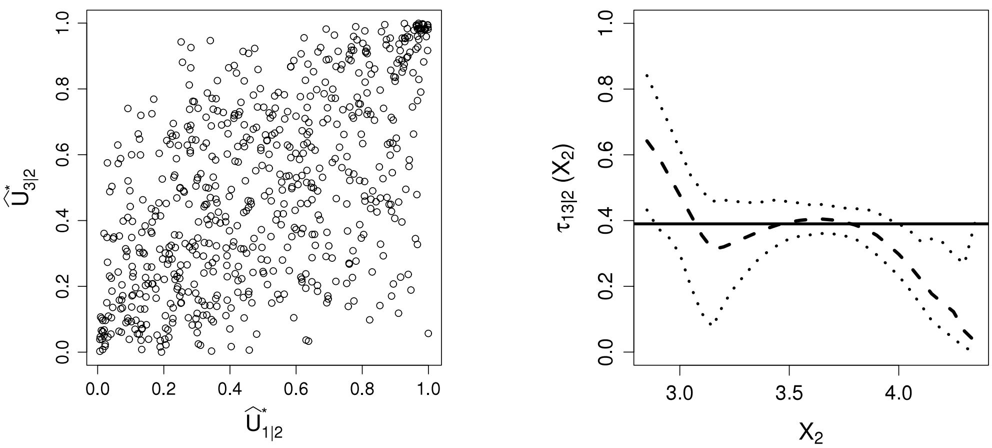

Fig. 8. Scatter plot of the pseudo observations $\widehat{U}^*_{1|2}$ and $\widehat{U}^*_{3|2}$ (left) and plot of estimates $\hat{\tau}_{13|2}$ of Kendall's tau (right) obtained under the simplifying assumption (solid) and using the local linear approach (dashed) as a function of X_2, along with the 90% bootstrap confidence intervals (dotted).

As a further illustration, the validity of the simplifying assumption was also investigated under an alternative vine construction, where the pairs that have the strongest dependence dictate the unconditional copulas of the PCC; this modeling strategy is suggested in [1]. Displayed in Fig. 9 are the pseudo observations $\widehat{U}^*_{1|3}$ and $\widehat{U}^*_{2|3}$ obtained from the fitted Student-t copulas C_{13} and C_{23} and the estimates of $\tau_{12|3}$ assuming a Frank copula for $C_{12|3}$, which can accommodate both positive and negative association. The nonparametric estimates and the corresponding bootstrap confidence intervals show considerable variation especially at small values of x_3, which is possibly due to computational instability close to the boundaries. Even excluding the latter, the constant parameter estimate $\hat{\theta}_{12|3} = 0.72$ (corresponding to $\hat{\tau}_{12|3} = 0.08$) does not fall in the bootstrap confidence intervals. Thus the simplifying assumption does not seem to hold under this construction either.

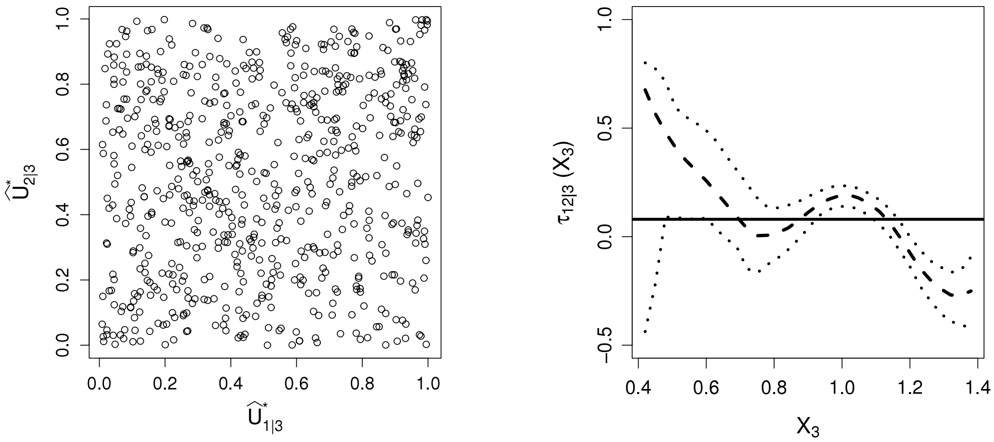

Fig. 9. Scatter plot of the pseudo observations $\widehat{U}^*_{1|3}$ and $\widehat{U}^*_{2|3}$ (left) and plot of estimates $\hat{\tau}_{13|2}$ of Kendall's tau (right) obtained under the simplifying assumption (solid) and using the local linear approach (dashed) as a function of X_3, along with the 90% bootstrap confidence intervals (dotted).

9. Conclusion and discussion

While simplified pair-copula constructions are quickly gaining popularity in multivariate data modeling, it was shown here that an uncritical use of the simplifying assumption may be misleading. As an alternative, this paper has presented a kernel-based nonparametric method for the estimation of the dependence parameter of a conditional copula in a trivariate PCC. This technique, which was proved to be consistent, makes it possible to assess the validity of the simplified PCC assumption according to which the conditional copula component of the model does not depend on the conditioning variable.

Although the proposed technique was introduced here as a visual tool, it could also be used to construct a formal test of the simplifying assumption, e.g., by adapting the generalized likelihood ratio approach of [3,5]. To accomplish this, it would be necessary to determine the asymptotic distribution of the local likelihood estimator within the PCC framework. Also of interest is the generalization of this methodology to multidimensional vines in which conditional copulas at higher levels of the hierarchy feature more than one conditioning variable. This will be the subject of future work.

Acknowledgments

Funding in support of this work was provided by the Canada Research Chairs Program, the Natural Sciences and Engineering Research Council of Canada, the Fonds québécois de la recherche sur la nature et les technologies, as well as by the Centre de recherches mathématiques de Montréal.

Appendix A

The following regularity conditions are required to establish the consistency of $\hat{\theta}_{13|2}$.

(A1) F_2 admits a differentiable density f_2 on the interior $\mathcal{X}_2$ of its support.

(A2) There exists an open subset Θ_0 of the parameter space Θ such that $c_{13|2}$ is strictly positive and admits all partial derivatives up to the order three on $(0, 1) \times (0, 1) \times \Theta_0$.

(A3) One has $g^{-1} \circ \eta_{13|2}(\mathcal{X}_2) \subset \Theta_0$ and, for $k \in \{1, 2, 3\}$, the partial derivatives ℓ_{k00} are Lipschitz on $(0, 1) \times (0, 1) \times \eta_{13|2}(\mathcal{X}_2)$.

(B1) There exists a constant $M_I > 0$ such that, for all $x \in \mathcal{X}_2$,

$$\mathcal{I}(x) = \mathrm{E}\left[\ell^2_{100}\left[g\{\theta_{13|2}(x)\}, F_{1|2}(X_1|x), F_{3|2}(X_3|x)\right]|X_2 = x\right] < M_I.$$

(B2) For any $k \in \{1, 2, 3\}$, there exists a function $J_k : (0, 1)^2 \to \mathbb{R}$ such that for all $u, v \in (0, 1)$ and $\theta \in \Theta_0$, $|\ell_{k00}\{g(\theta), u, v\}| \le J_k(u, v)$ and $\mathrm{E}_\theta\{J^2_k(U_{1|2}, U_{3|2})\} < \infty$ is uniformly bounded on Θ_0. Here $(U_{1|2}, U_{3|2})$ is a random pair distributed as $C_{13|2}(u, v; \theta)$.

(C1) The functions $\eta_{13|2}$ and g^{-1} are continuously differentiable up to the order two and three, respectively. Furthermore, the second order derivative $\eta^{(2)}_{13|2}$ is bounded.

(C2) The kernel K is a symmetric bounded probability density function with compact support, which is assumed to be $[-1, 1]$ without loss of generality.

(C3) The bandwidth λ_n depends on n in such a way that as $n \to \infty$, $\lambda_n \to 0$ and $n\lambda_n^{2+\delta} \to \infty$ for some $\delta > 0$.

(D) For $k \in \{1, 3\}$, the function $h_{k2}(u, v, \theta)$ is Lipschitz and C_{k2} satisfies the standard regularity conditions for maximum likelihood estimation; see, e.g., [17].

Appendix B

This Appendix contains proofs of Lemmas 1 and 2.

Proof of Lemma 1. First call on the triangle inequality to write

$$|\hat{A}_{rn}(x) - A_{rn}(x)| \le \frac{1}{n}\sum_{i=1}^{n}\left|\frac{X_{2i}-x}{\lambda_n}\right|^r K_{\lambda_n}(X_{2i}-x)\left|\ell_{100}\{\bar{\eta}_{13|2}(x, X_{2i}), \widehat{F}_{1|2}(X_{1i}|X_{2i}), \widehat{F}_{3|2}(X_{3i}|X_{2i})\}\right.$$
$$\left. - \ \ell_{100}\{\eta_{13|2}(X_{2i}), F_{1|2}(X_{1i}|X_{2i}), F_{3|2}(X_{3i}|X_{2i})\}\right|.$$

Given that ℓ_{100} is Lipschitz by condition (A3), a constant $M_1 > 0$ can be found that allows one to bound the right-hand side from above by

$$\frac{M_1}{n}\sum_{i=1}^{n}\left|\frac{X_{2i}-x}{\lambda_n}\right|^r K_{\lambda_n}(X_{2i}-x)\{|\bar{\eta}_{13|2}(x, X_{2i}) - \eta_{13|2}(X_{2i})|$$
$$+ |\widehat{F}_{1|2}(X_{1i}|X_{2i}) - F_{1|2}(X_{1i}|X_{2i})| + |\widehat{F}_{3|2}(X_{3i}|X_{2i}) - F_{3|2}(X_{3i}|X_{2i})|\}.$$

In particular, therefore,

$$|\hat{A}_{rn}(x) - A_{rn}(x)| \le \Delta_2\lambda_n^2\frac{M_1}{2n}\sum_{i=1}^{n}\left|\frac{X_{2i}-x}{\lambda_n}\right|^{r+2} K_{\lambda_n}(X_{2i}-x) + (\Delta_1+\Delta_3)\frac{M_1}{n}\sum_{i=1}^{n}\left|\frac{X_{2i}-x}{\lambda_n}\right|^{r} K_{\lambda_n}(X_{2i}-x),$$

where

$$\Delta_2 = \sup_{x\in\mathcal{X}_2}|\eta''_{13|2}(x)| < \infty$$

by condition (C1) and for $k \in \{1, 3\}$,

$$\Delta_k = \sup_{x_k, x_2\in\mathbb{R}}|\widehat{F}_{k|2}(x_k|x_2) - F_{k|2}(x_k|x_2)|.$$

Using the Lipschitz condition (A3) and analogous arguments, one can find constants $M_2 > 0$ and $M_3 > 0$ such that

$$|\hat{B}_{rsn}(x) - B_{rsn}(x)| \le \Delta_2\lambda_n^2\frac{M_2}{2n}\sum_{i=1}^{n}\left|\frac{X_{2i}-x}{\lambda_n}\right|^{r+s+2} K_{\lambda_n}(X_{2i}-x) + (\Delta_1+\Delta_3)\frac{M_2}{n}\sum_{i=1}^{n}\left|\frac{X_{2i}-x}{\lambda_n}\right|^{r+s} K_{\lambda_n}(X_{2i}-x)$$

and

$$|\hat{C}_{rstn}(x) - C_{rstn}(x)| \le (\Delta_1+\Delta_3)\frac{M_3}{n}\sum_{i=1}^{n}\left|\frac{X_{2i}-x}{\lambda_n}\right|^{r+s+t} K_{\lambda_n}(X_{2i}-x).$$

Now as $n \to \infty$, condition (C2) guarantees that, for any $w > 0$,

$$\frac{1}{n}\sum_{i=1}^{n}\left|\frac{X_{2i}-x}{\lambda_n}\right|^{w} K_{\lambda_n}(X_{2i}-x) = O_p(1).$$

Furthermore, let θ_k stand either for θ_{12} if $k = 1$ or for θ_{23} if $k = 3$. Then for $k \in \{1, 3\}$,

$$\Delta_k = \sup_{x_k, x_2\in\mathbb{R}}|h_{k2}\{F_{kn}(x_k), F_{2n}(x_2); \hat{\theta}_k\} - h_{k2}\{F_k(x_k), F_2(x_2); \theta_k\}|.$$

Using condition (B1), one can find a constant $M_4 > 0$ such that, for $k = 1, 3$,

$$\Delta_k \le M_4\left\{|\hat{\theta}_k - \theta| + \sup_{x_k\in\mathbb{R}}|F_{kn}(x_k) - F_k(x_k)| + \sup_{x_2\in\mathbb{R}}|F_{2n}(x_2) - F_2(x_2)|\right\}.$$

The consistency of the maximum pseudo likelihood estimator and the Glivenko–Cantelli Theorem together imply that for $k \in \{1, 3\}$, $\Delta_k = o_p(1)$. This shows that $|\hat{C}_{rstn}(x) - C_{rstn}(x)| = o_p(1)$. Similarly, it follows from condition (C3) that

$$|\hat{A}_{rn}(x) - A_{rn}(x)| = O_p(\lambda_n^2) + o_p(1) = o_p(1)$$

and

$$|\hat{B}_{rsn}(x) - B_{rsn}(x)| = O_p(\lambda_n^2) + o_p(1) = o_p(1). \quad \square$$

Proof of Lemma 2. To establish statement (a), fix $r \in \{0, 1\}$ and for each $i \in \{1, \dots, n\}$, let

$$Z_{in} = \left(\frac{X_{2i} - x}{\lambda_n}\right)^r K_{\lambda_n}(X_{2i} - x)\, \ell_{100}\{\eta_{13|2}(X_{2i}), F_{1|2}(X_{1i}|X_{2i}), F_{3|2}(X_{3i}|X_{2i})\},$$

so that $A_{rn} = (Z_{1n} + \cdots + Z_{nn})/n$. In view of condition (B1), it is immediate that, for all $i \in \{1, \dots, n\}$,

$$\mathrm{E}(Z_{in}) = \int \left(\frac{y - x}{\lambda_n}\right)^r K_{\lambda_n}(y - x) \left[\int \ell_{100}\{\eta_{13|2}(y), u, v\}\, \mathrm{d}C_{13|2}\{u, v; \theta_{13|2}(y)\}\right] \mathrm{d}F_2(y)$$

vanishes. The same condition further implies that

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n \mathrm{var}(Z_{in}) &= \frac{1}{n}\int \left(\frac{y - x}{\lambda_n}\right)^{2r} K^2_{\lambda_n}(y - x)\, \mathcal{I}(y)\, \mathrm{d}F_2(y)\\ &\le \frac{M_I}{n}\int \left(\frac{y - x}{\lambda_n}\right)^{2r} K^2_{\lambda_n}(y - x)\, \mathrm{d}F_2(y) = O\left(\frac{1}{n\lambda_n}\right).\end{aligned}$$

It follows from the Weak Law of Large Numbers (e.g., Theorem 10.2 in [6]) that, as $n \to \infty$, $A_{rn} \xrightarrow{p} 0$ for $r \in \{0, 1\}$.

Turning to claim (b), fix $r, s \in \{0, 1\}$ and, for each $i \in \{1, \dots, n\}$, redefine

$$Z_{in} = \left(\frac{X_{2i} - x}{\lambda_n}\right)^{r+s} K_{\lambda_n}(X_{2i} - x)\, \ell_{200}\{\eta_{13|2}(X_{2i}), F_{1|2}(X_{1i}|X_{2i}), F_{3|2}(X_{3i}|X_{2i})\},$$

so that $B_{rsn} = (Z_{1n} + \cdots + Z_{nn})/n$. In view of condition (B1), it is immediate that, for all $i \in \{1, \dots, n\}$,

$$\begin{aligned}\mathrm{E}(Z_{in}) &= \int \left(\frac{y - x}{\lambda_n}\right)^{r+s} K_{\lambda_n}(y - x) \left[\int \ell_{200}\{\eta_{13|2}(y), u, v\}\, \mathrm{d}C_{13|2}\{u, v; \theta_{13|2}(y)\}\right] \mathrm{d}F_2(y)\\ &= -\int \left(\frac{y - x}{\lambda_n}\right)^{r+s} K_{\lambda_n}(y - x)\mathcal{I}(y)\mathrm{d}F_2(y)\\ &= -\int z^{r+s} K(z)\mathcal{I}(\lambda_n z + x) f_2(\lambda_n z + x)\mathrm{d}z.\end{aligned}$$

Expanding the product function $\mathcal{I} \times f_2$ in Taylor series around x, one finds

$$\mathrm{E}(Z_{in}) = -\mathcal{I}(x) f_2(x)\mu_{rs} + O(\lambda_n),$$

where, for all $r, s \in \{0, 1\}$, $\mu_{rs} = \int t^{r+s} K(t)\, \mathrm{d}t$. Furthermore, it follows from condition (B2) that

$$\begin{aligned}\frac{1}{n^2}\sum_{i=1}^n \mathrm{var}(Z_{in}) &= \frac{1}{n}\int \left(\frac{y - x}{\lambda_n}\right)^{2r+2s} K^2_{\lambda_n}(y - x) \left[\int \ell^2_{200}\{\eta_{13|2}(y), u, v\}\, \mathrm{d}C_{13|2}\{u, v; \theta_{13|2}(y)\}\right] \mathrm{d}F_2(y)\\ &\le \frac{M_4}{n}\int \left(\frac{y - x}{\lambda_n}\right)^{2r+2s} K^2_{\lambda_n}(y - x)\mathrm{d}F_2(y) = O\left(\frac{1}{n\lambda_n}\right).\end{aligned}$$

In particular, therefore, the Weak Law of Large Numbers implies that, for all $r, s \in \{0, 1\}$, as $n \to \infty$,

$$B_{rsn} \xrightarrow{p} -\mathcal{I}(x) f_2(x)\mu_{rs}.$$

Claim (b) then follows because $\mu_{00} = 1$ and $\mu_{01} = \mu_{10} = 0$ by condition (C2).

Finally, to prove claim (c), use condition (B2) to see that, for each $r, s, t \in \{0, 1\}$, one has $|C_{rstn}(x)| \le W_{rstn}$, where

$$W_{rstn} = \frac{1}{n}\sum_{i=1}^n \left|\frac{X_{2i} - x}{\lambda_n}\right|^{r+s+t} K_{\lambda_n}(X_{2i} - x) J_3\{F_{1|2}(X_{1i}|X_{2i}), F_{3|2}(X_{3i}|X_{2i})\}.$$

Proceeding as above, one can invoke the Weak Law of Large Numbers and condition (B2) to show that there exists $M_5 \ge 0$ such that $\sum_{r,s,t} W_{rstn} \xrightarrow{p} M_5$. Thus, as $n \to \infty$,

$$\Pr\left(\sum_{r,s,t} |C_{rstn}| \le M_x\right) \to 1$$

whenever $M_x > M_5$. This completes the proof. □

References

[1] K. Aas, C. Czado, A. Frigessi, H. Bakken, Pair-copula constructions of multiple dependence, Insurance Math. Econom. 44 (2009) 182–198.
[2] F. Abegaz, I. Gijbels, N. Veraverbeke, Semiparametric estimation in copula models, J. Multivariate Anal. (2012) (in press).
[3] E.F. Acar, Nonparametric estimation and inference for the copula parameter in conditional copulas, Ph.D. Thesis, University of Toronto, Toronto, ON, 2010.
[4] E.F. Acar, R.V. Craiu, F. Yao, Dependence calibration in conditional copulas: a nonparametric approach, Biometrics 67 (2011) 445–453.
[5] E.F. Acar, R.V. Craiu, F. Yao, Statistical testing for conditional copulas, unpublished manuscript, 2011.
[6] H. Bauer, Wahrscheinlichkeitstheorie, fifth ed., Walter de Gruyter & Co., Berlin, 2002.
[7] T. Bedford, R.M. Cooke, Vines—a new graphical model for dependent random variables, Ann. Statist. 30 (2002) 1031–1068.
[8] N. Ben Ghorbal, C. Genest, J. Nešlehová, On the Ghoudi, Khoudraji, and Rivest test for extreme-value dependence, Canad. J. Statist. 37 (2009) 534–552.
[9] R.D. Cook, M.E. Johnson, A family of distributions for modelling nonelliptically symmetric multivariate data, J. Roy. Statist. Soc. Ser. B 43 (1981) 210–218.
[10] J. Fan, I. Gijbels, Local Polynomial Modelling and Its Applications, Chapman & Hall, London, 1996.
[11] C. Genest, K. Ghoudi, L.-P. Rivest, A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, Biometrika 82 (1995) 543–552.
[12] I. Hobæk Haff, Parameter estimation for pair-copula constructions, Bernoulli, 2012 (in press).
[13] I. Hobæk Haff, Comparison of estimators for pair-copula constructions, J. Multivariate Anal. 110 (2012) 91–105.
[14] I. Hobæk Haff, K. Aas, A. Frigessi, On the simplified pair-copula construction—simply useful or too simplistic? J. Multivariate Anal. 101 (2010) 1296–1310.
[15] H. Joe, Multivariate Models and Dependence Concepts, Chapman & Hall, London, 1997.
[16] D. Kurowicka, H. Joe, Dependence Modeling: Handbook on Vine Copulae, World Scientific Publishing, Singapore/SG, 2010.
[17] E.L. Lehmann, Theory of Point Estimation, Springer, New York, 1997.
[18] A.J. McNeil, R. Frey, P. Embrechts, Quantitative Risk Management: Concepts, Techniques and Tools, Princeton University Press, Princeton, NJ, 2005.
[19] R.B. Nelsen, An Introduction to Copulas, second ed., Springer, New York, 2006.
[20] A.J. Patton, Modelling asymmetric exchange rate dependence, Internat. Econom. Rev. 47 (2006) 527–556.
[21] J.H. Shih, T.A. Louis, Inferences on the association parameter in copula models for bivariate survival data, Biometrics 51 (1995) 1384–1399.
[22] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8 (1959) 229–231.
[23] H. Tsukahara, Semiparametric estimation in copula models, Canad. J. Statist. 33 (2005) 357–375.