SCORE TESTS FOR COVARIATE EFFECTS IN CONDITIONAL COPULAS

IRÈNE GIJBELS[1], MAREK OMELKA[2#], MICHAL PEŠTA[2] AND NOËL VERAVERBEKE[3,4]

[1] Department of Mathematics and Leuven Statistics Research Center (LStat), Katholieke Universiteit Leuven, Celestijnenlaan 200B, Box 2400, B-3001 Leuven (Heverlee), Belgium.
[2] Department of Probability and Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Praha 8, Czech Republic;
[3] Center for Statistics, Hasselt University, Agoralaan-building D, B-3590 Diepenbeek, Belgium;
[4] Unit for BMI, North-West University, Potchefstroom, South Africa

April 30, 2017

Abstract. We consider copula modeling of the dependence between two or more random variables in the presence of a multivariate covariate. The dependence parameter of the conditional copula possibly depends on the value of the covariate vector. In this paper we develop a new testing methodology for some important parametric specifications of this dependence parameter: constant, linear, quadratic, etc. in the covariate values, possibly after transformation with a link function. The margins are left unspecified. Our novel methodology opens plenty of new possibilities for testing how the conditional copula depends on the multivariate covariate and also for variable selection in copula model building. The suggested test is based on a Rao-type score statistic and regularity conditions are given under which the test has a limiting chi-square distribution under the null hypothesis. For small and moderate sample sizes, a permutation procedure is suggested to assess significance. In simulations it is shown that the test performs well (even under misspecification of the copula family and/or the dependence parameter structure) in comparison to available tests designed for testing for constancy of the dependence parameter. The test is illustrated on a real data set on concentrations of chemicals in water samples.

Keywords and phrases: Conditional copula, covariate effect, parametric dependence structure, Rao score test, specification test.

---

* This research was supported by the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy). The first author gratefully acknowledges support from the GOA/12/014 project of the Research Fund KU Leuven. The second and the third author gratefully acknowledge support from the grant GACR 15-04774Y. The fourth author is an extraordinary professor at the North-West University, Potchefstroom, South Africa.

Corresponding author. Email: omelka@karlin.mff.cuni.cz

AMS classification: 62H15, 62F05, 62G10

1. Introduction

Conditional copulas provide a convenient way to model the dependence between random variables whose dependence structure is possibly influenced by covariates. See Patton [18] for an early reference, Acar et al. [2] and Abegaz et al. [1] for semiparametric estimation, and Veraverbeke et al. [26] and Gijbels et al. [14] for nonparametric estimation of conditional copulas, among others.

A crucial fact in conditional copula modeling is that, in general, the covariates influence the conditional copula on two levels: the copula itself (the dependence structure) may change with the value of the covariate vector, and the margins may be influenced by the covariate vector. An important simplification occurs when the dependence structure remains unchanged whatever the realized value of the covariate vector. This simplification is often referred to as 'the simplifying assumption'; see, e.g., Hobæk Haff et al. [15], Acar et al. [4], Stöber et al. [23] and Gijbels et al. [11].

In this paper we contribute to testing for covariate effects on conditional copulas, and this within a parametric copula setting. The literature on such testing problems is quite limited. It includes a semiparametric likelihood ratio type of test—not assuming any structure on the functional dependence parameter in a given copula—that was proposed and studied in Acar et al. [3]; and nonparametric tests, leaving the copula dependence structure as well as the margins fully unspecified. In contrast to these semiparametric and nonparametric approaches, this paper presents a new approach, in which the starting point is to model parametrically the copula, but also its functional dependence parameter. The margins are left unspecified. Despite this parametric framework for both the copula and the dependence parameter, it turns out that the test proposed herein continues to have a very good performance under misspecification of one or both of these parts. This is an important first advantage. A second, more practical, advantage is that the proposed test does not require any choice of smoothing parameter (due to its major parametric setting). Thirdly, the developed test methodology can be applied for several testing problems: (i) testing for no covariate effect; (ii) testing for specific effects of a selection of covariates; (iii) testing for specific effects of all covariates (such as linear versus quadratic).

The paper is further organized as follows. In Section 2 we present the statistical framework, and briefly review semiparametric and nonparametric tests that are available in the literature. The new test methodology is exposed in Section 3, in which the essential elements of the derivation of the test and its asymptotic behavior are presented. Details about the theoretical results, including their proofs, are provided in the Appendix. The test methodology is applicable to several testing settings, as is explained in Section 3. In Sections 4 and 5 the test methodology is illustrated in various testing problems, in a univariate as well as a multivariate covariate setting. The use of the test in statistical analysis is demonstrated in a real data example in Section 6.

2. Statistical framework and state-of-the-art

In this section we first introduce the statistical framework, the main testing problem of interest, and briefly indicate major testing procedures available in the literature.

2.1. Statistical framework.

Suppose we have n independent and identically distributed observations $(Y_{11}, Y_{21}, \mathbf{X}_1), \ldots, (Y_{1n}, Y_{2n}, \mathbf{X}_n)$ from a random vector $(Y_1, Y_2, \mathbf{X})$, where Y_1 and Y_2 are univariate random variables and $\mathbf{X}$ is a d-dimensional random vector. Let $H(y_1, y_2, \mathbf{x})$ be the cumulative distribution function of $(Y_1, Y_2, \mathbf{X})$. Denote the joint and marginal distribution functions of (Y_1, Y_2), conditionally on $\mathbf{X} = \mathbf{x}$, as

$$H_{\mathbf{x}}(y_1, y_2) = \Pr(Y_1 \le y_1, Y_2 \le y_2 \,|\, \mathbf{X} = \mathbf{x}),$$

$$F_{1\mathbf{x}}(y_1) = \Pr(Y_1 \le y_1 \,|\, \mathbf{X} = \mathbf{x}), \quad F_{2\mathbf{x}}(y_2) = \Pr(Y_2 \le y_2 \,|\, \mathbf{X} = \mathbf{x}).$$

If $F_{1\mathbf{x}}$ and $F_{2\mathbf{x}}$ are continuous, then by Sklar's theorem (see, e.g., Sklar [22], Nelsen [17]) applied to the conditional probability distribution setting, there exists a unique copula $C_{\mathbf{x}}$ that links the conditional margins into the conditional joint distribution through the relation

$$H_{\mathbf{x}}(y_1, y_2) = C_{\mathbf{x}}\big\{F_{1\mathbf{x}}(y_1), F_{2\mathbf{x}}(y_2)\big\}.$$

The function $C_{\mathbf{x}}$ is called a *conditional copula.*

A first interest in this paper is to test whether the conditional copula $C_{\mathbf{x}}$ really depends on $\mathbf{x}$. More formally, we want to test the hypothesis

$$\mathcal{H}_0 : \forall_{\mathbf{x},\mathbf{x}' \in \mathbf{R}_{\mathbf{X}}} \; C_{\mathbf{x}} = C_{\mathbf{x}'} \tag{1}$$

versus the alternative

$$\mathcal{H}_A : \exists_{\mathbf{x},\mathbf{x}' \in \mathbf{R}_{\mathbf{X}}} \; C_{\mathbf{x}} \ne C_{\mathbf{x}'},$$

where $\mathbf{R}_{\mathbf{X}}$ denotes the domain of the covariate $\mathbf{X}$.

2.2. State-of-the-art.

In what follows, suppose for a moment that we can observe $U_{1i} = F_{1\mathbf{X}_i}(Y_{1i}), U_{2i} = F_{2\mathbf{X}_i}(Y_{1i})$ from the conditional copula $C_{\mathbf{X}_i}$. If these observations are not available (i.e., $F_{1\mathbf{X}}$ and $F_{2\mathbf{X}}$ are unknown), then one needs to estimate these, via estimation of the conditional margins, and work with pseudo-observations, denoted by $(\widehat{U}_{1i}, \widehat{U}_{2i})$.

2.2.1. *Semiparametric approach of Acar et al. [3].* In Acar et al. [3] the conditional copula function is modeled as

$$C_{\mathbf{x}}(u_1, u_2) = C(u_1, u_2; \theta(\mathbf{x})), \tag{2}$$

where $C(\cdot, \cdot; \theta)$ is a given parametric copula function, where for simplicity we assume here $\theta \in \mathbb{R}$. The dependence on $\mathbf{x}$ is brought in via the parameter θ that is allowed to depend on $\mathbf{x}$, i.e., by considering an unknown function $\theta(\mathbf{x})$ instead of an unknown parameter θ. Assume that the density associated with C exists, and denote it by $c(\cdot, \cdot; \theta(\mathbf{x}))$.

Roughly speaking the test statistic is based on comparing the log-likelihoods computed under the alternative and the null hypothesis. More formally, under the null hypothesis $\theta(\mathbf{x})$ does not depend on $\mathbf{x}$, thus the log-likelihood is given by

$$\ell_n(H_0) = \sum_{i=1}^{n} \ln c\big(U_{1i}, U_{2i}; \widehat{\theta}_n\big),$$

where $\widehat{\theta}_n$ is the maximizer of the log-likelihood above.

Under the alternative hypothesis, let $\widehat{\theta}_h(\mathbf{X}_i)$ be the estimate of the parameter $\theta(\mathbf{X}_i)$ using a local-likelihood method. Estimation of the unknown function $\theta(\cdot)$ requires smoothing/local techniques, and in the d-variate covariate setting this involves nonparametric estimation of the d-variate function $\theta(\mathbf{x})$. One thus has to face the usual curse of dimensionality issue. Acar et al. [3] focus on the univariate setting (i.e., $d = 1$) and hence, with h a smoothing parameter in estimating the univariate function $\theta(\cdot)$, consider the log-likelihood under the alternative hypothesis

$$\ell_n(H_A; h) = \sum_{i=1}^{n} \ln c\big(U_{1i}, U_{2i}; \widehat{\theta}_h(X_i)\big).$$

Finally the test statistic is given by

$$\lambda_n(h) = \ell_n(H_A; h) - \ell_n(H_0).$$

2.2.2. *Nonparametric approach of Gijbels et al. [13].* This test is based on the fact that when the conditional copula $C_{\mathbf{X}}$ does not depend on the realized covariate vector value $\mathbf{x}$, then the associated conditional Kendall's tau function $\tau(\mathbf{x})$ (see, e.g., Gijbels et al. [14]) does not depend on $\mathbf{x}$, i.e., is constant and equals $\tau^A = \mathrm{E}\{\tau(\mathbf{X})\}$, called the average conditional Kendall's tau. The test of Gijbels et al. [13] then consists of measuring the squared distance between a nonparametric estimator $\widehat{\tau}_n(\mathbf{x})$ of $\tau(\mathbf{x})$ and a nonparametric estimator $\bar{\tau}_n^A$ of τ^A. More precisely the test is based on the test statistic

$$V_{n1} = \frac{1}{n} \sum_{i=1}^{n} \{\widehat{\tau}_n(\mathbf{X}_i) - \bar{\tau}_n^A\}^2,$$

where $\bar{\tau}_n^A = \sum_{i=1}^{n} \widehat{\tau}_n(\mathbf{X}_i)/n$. See Veraverbeke et al. [26] and Gijbels et al. [14] for nonparametric conditional copula estimation (and association measures) in a univariate covariate setting, and Gijbels et al. [10] in a multivariate covariate setting.

Obviously, the test V_{n1} is based on estimating nonparametrically the conditional Kendall's tau $\tau(\mathbf{x})$ and the average conditional Kendall's tau τ^A. Hence, it is a nonparametric test involving also the choice of smoothing parameters (e.g., bandwidths).

3. New semiparametric approach

Similarly as in Acar et al. [3] the idea is based on the likelihood. Suppose that model (2) holds, where $\boldsymbol{\theta}(\mathbf{x}) = \big(\theta_1(\mathbf{x}), \dots, \theta_q(\mathbf{x})\big)^\top$ is an unknown q-dimensional parameter possibly depending on the value of the d-dimensional covariate $\mathbf{x} = (x_1, \dots, x_d)^\top$. Further suppose that $\boldsymbol{\theta}(\mathbf{x})$ is parametrized as $\boldsymbol{\theta}(\mathbf{x}) = \mathbf{a}(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\psi})$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^\top$ and $\boldsymbol{\psi} = (\psi_1, \dots, \psi_q)^\top$ are unknown parameters with values in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. Moreover, let the parametrization be done in such a way that if $\boldsymbol{\alpha}$ is the zero vector $\mathbf{0}_p = (0, \dots, 0)^\top$, then $\mathbf{a}(\mathbf{x}; \mathbf{0}_p; \boldsymbol{\psi})$ does not depend on $\mathbf{x}$ (and thus the simplifying assumption holds). Note that for notational simplicity, we used the same notation for the dimension of the dependence parameter $\boldsymbol{\theta}(\mathbf{x})$ and the parameter $\boldsymbol{\psi}$, which will be considered as a nuisance parameter. Since we will mainly deal with the case of a real-valued dependence parameter (i.e., $\boldsymbol{\theta}(\mathbf{x}) \in \mathbb{R}$), we chose to keep the notation simple. Throughout the paper, we assume that the parameters $(\boldsymbol{\alpha}, \boldsymbol{\psi})$ in the semiparametric model are identifiable.

Under the above described semiparametric framework, testing the null hypothesis (1) is equivalent to testing

$$\mathcal{H}_0 : \boldsymbol{\alpha} = \mathbf{0}_p, \quad \mathcal{H}_1 : \boldsymbol{\alpha} \neq \mathbf{0}_p. \tag{3}$$

Assume for a moment that the margins are known. Then, a suitable test statistic for testing (3) can be found within the framework of tests based on the likelihood in the presence of a nuisance parameter ($\boldsymbol{\psi}$). Although any of the three tests (likelihood ratio, Wald or Rao score test) can be used, in what follows we concentrate on the Rao score test. The reason is that the test statistic of this test does not require the estimation of $\boldsymbol{\alpha}$. This is convenient in particular when a permutation principle is used to improve the Type I error properties of the test in small or moderate samples. In Section 3.2, we describe the proposed test methodology.

3.1. **Examples of settings and testing problems.** Let us look more closely to the above framework in which the conditional copula and its functional dependence parameter are of parametric form, viz.

$$C_{\mathbf{x}}(u_1, u_2) = C(u_1, u_2; \boldsymbol{\theta}(\mathbf{x})) \quad \text{with} \quad \boldsymbol{\theta}(\mathbf{x}) = \mathbf{a}(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\psi})\ ,$$

where $\boldsymbol{\alpha}$ characterizes completely the dependence (or not) of $\boldsymbol{\theta}(\mathbf{x})$ (and hence of $C_{\mathbf{x}}(\cdot, \cdot)$) on $\mathbf{x}$, whereas the parameter $\boldsymbol{\psi}$ is a nuisance parameter.

For example, in a bivariate covariate setting (i.e., $d = 2$), one could model the conditional dependence between Y_1 and Y_2 by, say, a Frank copula with parameter $\theta(\mathbf{x})$, $\mathbf{x} = (x_1, x_2)^\top$, where $\theta(\mathbf{x})$ takes values in $\mathbb{R}$, since for a Frank copula, the parameter can take positive and negative values. There are many possible modeling choices for $\theta(\mathbf{x})$, such as

(i) $\theta(\mathbf{x}) = \psi + \alpha_1 x_1 + \alpha_2 x_2$, where thus $p = 2$ and $q = 1$;
(ii) $\theta(\mathbf{x}) = \psi + \alpha_1 x_1 + \alpha_2 x_1^2 + \alpha_3 x_2 + \alpha_4 x_1^2 + \alpha_5 x_1 x_2$, where $p = 5$ and $q = 1$;
(iii) $\theta(\mathbf{x}) = \psi_1 + \psi_2 x_1 + \alpha x_2$, where $p = 1$ and $q = 2$;
(iv) $\theta(\mathbf{x}) = \psi_1 + \psi_2 x_1 + \psi_3 x_1^2 + \alpha_1 x_2 + \alpha_2 x_2^2 + \alpha_3 x_1 x_2$, where thus $p = 3$ and $q = 3$.

Here q refers to the dimension of the nuisance parameter since $\theta(\mathbf{x}) \in \mathbb{R}$. Note that the considered testing problem (3) is of a distinct nature for the different cases:

a) For (i) and (ii) this means that we are testing for no covariate effect.
b) For (iii) this means that we are testing for no effect from the covariate x_2, whereas the covariate x_1 is assumed to have a linear effect.
c) For (iv) we are testing for no effect from the covariate x_2, whereas the covariate x_1 is assumed to have a quadratic effect and there is a possible first order interaction between x_1 and x_2.

An important further remark is that, in contrast with the Frank copula, other copula families have a restricted parameter space. For a Gaussian copula for example, $\theta \in (-1, 1)$, for a Clayton copula it is often assumed that $\theta \in (0, \infty)$, and for a Gumbel copula $\theta \in [1, \infty)$. In case of a restricted parameter space, one must work with a link function g, viz.

$$\boldsymbol{\theta}(\mathbf{x}) = g\{\boldsymbol{\eta}(\mathbf{x})\} = \mathbf{a}(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\psi})$$

with $\boldsymbol{\eta}(\mathbf{x})$ as, e.g., in (i)–(iv) above. For example, considering (i) above, in case of a Clayton or Gumbel copula, one could take $g(x) = e^x$ and $g(x) = e^x + 1$ as link functions, respectively, leading to

(iC) $\theta(\mathbf{x}) = \exp(\psi + \alpha_1 x_1 + \alpha_2 x_2)$;
(iG) $\theta(\mathbf{x}) = \exp(\psi + \alpha_1 x_1 + \alpha_2 x_2) + 1$.

Obviously there are many possible choices of link functions, and many parametrizations of the dependence parameter $\theta(\mathbf{x})$.

From the results in Sections 4 and 5 it can be seen that even when the parametric copula function and/or the parametric form of $\theta(\mathbf{x})$ (or the link function) are/is misspecified, the proposed tests are still performing quite well when it comes to testing for no covariate effect.

3.2. **Description of the suggested test.** A detailed derivation and discussion of the suggested test is provided for the case that the margins are not influenced by the covariate. The proposed test, however, is also applicable when the margins are influenced by the covariate, provided that appropriate pseudo-observations are used.

3.2.1. *Margins are not influenced by the covariate.* Assuming that the margins are known, Eq. (2) holds and the density of the covariate $\mathbf{X}$ does not depend on the parameters $\boldsymbol{\psi}$ and $\boldsymbol{\alpha}$, the log-likelihood is given by

$$\ell_n(\boldsymbol{\alpha}, \boldsymbol{\psi}) = \sum_{i=1}^{n} \ln c\big(U_{1i}, U_{2i}; \boldsymbol{\theta}(\mathbf{X}_i)\big) = \sum_{i=1}^{n} \ln c\big(U_{1i}, U_{2i}; \mathbf{a}(\mathbf{X}_i; \boldsymbol{\alpha}; \boldsymbol{\psi})\big),$$

where $c\big(u_1, u_2; \boldsymbol{\theta}(\mathbf{x})\big)$ is the assumed density of the copula of Y_1 and Y_2 when $\mathbf{X} = \mathbf{x}$. Note that if one could observe (U_{1i}, U_{2i}), then the Rao score test of the hypothesis (3) would be based on the score function

$$\frac{\partial}{\partial \boldsymbol{\alpha}}\, \ell_n(\boldsymbol{\alpha}, \boldsymbol{\psi})\Big|_{\boldsymbol{\alpha}=\mathbf{0}_p, \boldsymbol{\psi}=\widetilde{\boldsymbol{\psi}}_n},$$

where $\widetilde{\boldsymbol{\psi}}_n$ is the maximum likelihood estimator of $\boldsymbol{\psi}$ when assuming $\boldsymbol{\alpha} = \mathbf{0}_p$.

In practice the margins are usually unknown. If the covariate does not influence the marginal distributions, then one can replace the unobserved U_{1i}, U_{2i} with the pseudo-observations

$$\widetilde{U}_{1i} = \frac{n}{n+1}\, F_{1n}(Y_{1i}) \quad \text{and} \quad \widetilde{U}_{2i} = \frac{n}{n+1}\, F_{2n}(Y_{2i}), \tag{4}$$

where F_{1n} is the empirical distribution function of $Y_{11}, \ldots, Y_{1n}$ and analogously for F_{2n}.

In addition to the lack of knowledge about the margins, the second problem encountered in practice is that one can never be sure that the selected parametric form of the copula is correct. Thus our aim is to construct a test that performs well, even if the copula family is misspecified. That is why we denote

$$\rho(u_1, u_2, \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\psi}) = \ln c_R\big(u_1, u_2; \mathbf{a}(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\psi})\big),$$

where $c_R(u_1, u_2; \mathbf{a}(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\psi}))$ should be now understood as the copula density associated to a Reference copula C_R that is not necessarily the true copula C. The suggested test statistic will be based on the score

$$\mathbf{S}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{s}_{\boldsymbol{\alpha}}\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p; \widetilde{\boldsymbol{\psi}}_n\big),$$

where

$$\mathbf{s}_{\boldsymbol{\alpha}}(u_1, u_2, \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\psi}) = \frac{\partial}{\partial \boldsymbol{\alpha}}\, \rho(u_1, u_2, \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\psi}).$$

Finally for testing the null hypothesis (1), we suggest the following Rao-like test statistic

$$R_n = \mathbf{S}_n^\top \{\widehat{\mathrm{avar}(\mathbf{S}_n)}\}^{-1} \mathbf{S}_n, \tag{5}$$

where $\widehat{\mathrm{avar}(\mathbf{S}_n)}$ is an estimate of the asymptotic variance of $\mathbf{S}_n$, that is consistent under the null hypothesis.

Derivation of $\widehat{\operatorname{avar}(\mathbf{S}_n)}$. We now give an insight into the construction of $\widehat{\operatorname{avar}(\mathbf{S}_n)}$. Formal derivations and results can be found in the Appendix. Since here we suppose that the margins are not influenced by the covariate, we can work with the pseudo-observations $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$ defined in (4).

Denote $I(\boldsymbol{\alpha};\boldsymbol{\psi})$ the Fisher-like information matrix of a random vector $(U_{1i}, U_{2i}, \mathbf{X}_i)$ and note that this matrix can be written in block form as

$$I(\boldsymbol{\alpha},\boldsymbol{\psi}) = \begin{pmatrix} I_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\boldsymbol{\alpha},\boldsymbol{\psi}) & I_{\boldsymbol{\alpha}\boldsymbol{\psi}}(\boldsymbol{\alpha},\boldsymbol{\psi}) \\ I_{\boldsymbol{\psi}\boldsymbol{\alpha}}(\boldsymbol{\alpha},\boldsymbol{\psi}) & I_{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\alpha},\boldsymbol{\psi}) \end{pmatrix}, \tag{6}$$

where for instance $I_{\boldsymbol{\alpha}\boldsymbol{\psi}}(\boldsymbol{\alpha},\boldsymbol{\psi}) = -\mathrm{E}\left\{\partial^2\rho(U_{1i}, U_{2i}, \mathbf{X}_i;\boldsymbol{\alpha};\boldsymbol{\psi})/\left(\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\psi}^\top\right)\right\}$ is the upper right corner, which is a submatrix of dimension (p,q).

Let $\widetilde{\boldsymbol{\psi}}$ be the value of the parameter identified under the null hypothesis, i.e.,

$$\widetilde{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \mathrm{E}\,\rho(U_{1i}, U_{2i}, \mathbf{X}_i; \mathbf{0}_p; \boldsymbol{\psi}).$$

Under the null hypothesis one then gets the asymptotic representation (see Lemma 3 in the Appendix)

$$\sqrt{n}\left(\widetilde{\boldsymbol{\psi}}_n - \widetilde{\boldsymbol{\psi}}\right) = \{I_{\boldsymbol{\psi}\boldsymbol{\psi}}(\mathbf{0}_p,\widetilde{\boldsymbol{\psi}})\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{s}_{\boldsymbol{\psi}}\left(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p, \widetilde{\boldsymbol{\psi}}\right) + o_P(1), \tag{7}$$

where

$$\mathbf{s}_{\boldsymbol{\psi}}(u_1,u_2,\mathbf{x};\boldsymbol{\alpha},\boldsymbol{\psi}) = \frac{\partial}{\partial\boldsymbol{\psi}}\,\rho(u_1,u_2,\mathbf{x};\boldsymbol{\alpha},\boldsymbol{\psi}).$$

Provided that one can proceed as in standard regular models (see the Appendices for necessary conditions and technical details) one gets, with the help of (7),

$$\begin{aligned} \mathbf{S}_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{s}_{\boldsymbol{\alpha}}\left(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p; \widetilde{\boldsymbol{\psi}}\right) - I_{\boldsymbol{\alpha}\boldsymbol{\psi}}(\mathbf{0}_p,\widetilde{\boldsymbol{\psi}})\,\sqrt{n}\left(\widetilde{\boldsymbol{\psi}}_n - \widetilde{\boldsymbol{\psi}}\right) + o_P(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{s}_{\boldsymbol{\alpha}}\left(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p, \widetilde{\boldsymbol{\psi}}\right) \\ &\qquad - I_{\boldsymbol{\alpha}\boldsymbol{\psi}}(\mathbf{0}_p,\widetilde{\boldsymbol{\psi}})\,\{I_{\boldsymbol{\psi}\boldsymbol{\psi}}(\mathbf{0}_p,\widetilde{\boldsymbol{\psi}})\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{s}_{\boldsymbol{\psi}}\left(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p; \widetilde{\boldsymbol{\psi}}\right) + o_P(1). \end{aligned} \tag{8}$$

Now denote the joint score function as

$$\mathbf{s}(u_1,u_2,\mathbf{x};\boldsymbol{\psi}) = \left(\mathbf{s}_{\boldsymbol{\alpha}}^\top(u_1,u_2,\mathbf{x};\mathbf{0}_p;\boldsymbol{\psi}), \mathbf{s}_{\boldsymbol{\psi}}^\top(u_1,u_2,\mathbf{x};\mathbf{0}_p;\boldsymbol{\psi})\right)^\top,$$

where for notational simplicity we dropped the value of the parameter $\boldsymbol{\alpha}$ which in what follows is always set equal to $\mathbf{0}_p$. Then, under the null hypothesis by Theorem 1 of the Appendix,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{s}\left(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \widetilde{\boldsymbol{\psi}}\right) \rightsquigarrow \mathcal{N}_{p+q}\left(\mathbf{0}_{p+q}, \boldsymbol{\Sigma}\right) \tag{9}$$

as $n \to \infty$, where

$$\boldsymbol{\Sigma} = \operatorname{var}\Big\{\mathbf{s}\big(U_1, U_2, \mathbf{X}; \widetilde{\boldsymbol{\psi}}\big) + \int_{[0,1]^2}\int_{\mathbb{R}^d} \big[\mathbf{1}\{U_1 \leq v_1\} - v_1\big]\mathbf{s}^{(1)}\big(v_1, v_2, \mathbf{x}; \widetilde{\boldsymbol{\psi}}\big)\,\mathrm{d}C(v_1, v_2)\,\mathrm{d}F_{\mathbf{X}}(\mathbf{x})$$
$$+ \int_{[0,1]^2}\int_{\mathbb{R}^d} \big[\mathbf{1}\{U_2 \leq v_2\} - v_2\big]\mathbf{s}^{(2)}\big(v_1, v_2, \mathbf{x}; \widetilde{\boldsymbol{\psi}}\big)\,\mathrm{d}C(v_1, v_2)\,\mathrm{d}F_{\mathbf{X}}(\mathbf{x})\Big\},$$

with C being the copula of U_1 and U_2 (under the null hypothesis) and $\mathbf{s}^{(j)}(\cdot) = \partial\mathbf{s}(\cdot)/\partial u_j$.

Analogously as for the matrix in (6), let us write the matrix $\boldsymbol{\Sigma}$ in block form as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} & \boldsymbol{\Sigma}_{\boldsymbol{\alpha}\boldsymbol{\psi}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\psi}\boldsymbol{\alpha}} & \boldsymbol{\Sigma}_{\boldsymbol{\psi}\boldsymbol{\psi}} \end{pmatrix}.$$

Moreover for simplicity of notation put $I_{\boldsymbol{\alpha}\boldsymbol{\psi}} = I_{\boldsymbol{\alpha}\boldsymbol{\psi}}(\mathbf{0}_p, \widetilde{\boldsymbol{\psi}})$ and analogously for $I_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$, $I_{\boldsymbol{\psi}\boldsymbol{\psi}}$ and $I_{\boldsymbol{\psi}\boldsymbol{\alpha}}$. Now combining (8) and (9) yields

$$\mathbf{S}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{s}_{\boldsymbol{\alpha}}\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p, \widetilde{\boldsymbol{\psi}}_n\big) \rightsquigarrow \mathcal{N}_p\big(\mathbf{0}_p, \mathbb{V}\big), \tag{10}$$

where

$$\mathbb{V} = \big(\mathbb{I}_{p\times p}, -I_{\boldsymbol{\alpha}\boldsymbol{\psi}} I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\big)\,\boldsymbol{\Sigma}\,\big(\mathbb{I}_{p\times p}, -I_{\boldsymbol{\alpha}\boldsymbol{\psi}} I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\big)^{\top}$$
$$= \boldsymbol{\Sigma}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} - I_{\boldsymbol{\alpha}\boldsymbol{\psi}}\, I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\, \boldsymbol{\Sigma}_{\boldsymbol{\psi}\boldsymbol{\alpha}} - \boldsymbol{\Sigma}_{\boldsymbol{\alpha}\boldsymbol{\psi}}\, I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\, I_{\boldsymbol{\psi}\boldsymbol{\alpha}} + I_{\boldsymbol{\alpha}\boldsymbol{\psi}}\, I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\, \boldsymbol{\Sigma}_{\boldsymbol{\psi}\boldsymbol{\psi}}\, I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\, I_{\boldsymbol{\psi}\boldsymbol{\alpha}},$$

with $\mathbb{I}_{p\times p}$ being the identity matrix of dimension (p, p). This result is formally stated in Theorem 2 in the Appendix, where its proof is given.

The asymptotic variance of $\mathbf{S}_n$ can be estimated as

$$\widehat{\operatorname{avar}(\mathbf{S}_n)} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{Z}_i - \overline{\mathbf{Z}}_n)(\mathbf{Z}_i - \overline{\mathbf{Z}}_n)^{\top}, \tag{11}$$

where

$$\mathbf{Z}_i = \mathbf{s}_{\boldsymbol{\alpha}}\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p; \widetilde{\boldsymbol{\psi}}_n\big) + \widehat{\boldsymbol{\vartheta}}_{\boldsymbol{\alpha}}^{(1)}\big(\widetilde{U}_{1i}\big) + \widehat{\boldsymbol{\vartheta}}_{\boldsymbol{\alpha}}^{(2)}\big(\widetilde{U}_{2i}\big)$$
$$- \widehat{I}_{\boldsymbol{\alpha}\boldsymbol{\psi}}\, \widehat{I}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\Big\{\mathbf{s}_{\boldsymbol{\psi}}\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p; \widetilde{\boldsymbol{\psi}}_n\big) + \widehat{\boldsymbol{\vartheta}}_{\boldsymbol{\psi}}^{(1)}\big(\widetilde{U}_{1i}\big) + \widehat{\boldsymbol{\vartheta}}_{\boldsymbol{\psi}}^{(2)}\big(\widetilde{U}_{2i}\big)\Big\},$$

with

$$\widehat{I}_{\boldsymbol{\alpha}\boldsymbol{\psi}} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\,\mathbf{s}_{\boldsymbol{\alpha}}\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \boldsymbol{\alpha}; \boldsymbol{\psi}\big)}{\partial\boldsymbol{\psi}^{\top}}\Big|_{\boldsymbol{\alpha}=\mathbf{0}_p;\boldsymbol{\psi}=\widetilde{\boldsymbol{\psi}}_n},$$
$$\widehat{I}_{\boldsymbol{\psi}\boldsymbol{\psi}} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\,\mathbf{s}_{\boldsymbol{\psi}}\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \boldsymbol{\alpha}; \boldsymbol{\psi}\big)}{\partial\boldsymbol{\psi}^{\top}}\Big|_{\boldsymbol{\alpha}=\mathbf{0}_p;\boldsymbol{\psi}=\widetilde{\boldsymbol{\psi}}_n},$$

and for $k \in \{1, 2\}$,

$$\widehat{\boldsymbol{\vartheta}}^{(k)}_{\boldsymbol{\alpha}}(u) = \frac{1}{n}\sum_{j=1}^{n} \big[\mathbf{1}\{u \leq \widetilde{U}_{kj}\} - \widetilde{U}_{kj}\big]\, \mathbf{s}^{(k)}_{\boldsymbol{\alpha}}\big(\widetilde{U}_{1j}, \widetilde{U}_{2j}, \mathbf{X}_j; \mathbf{0}_p, \widetilde{\boldsymbol{\psi}}_n\big),$$

$$\widehat{\boldsymbol{\vartheta}}^{(k)}_{\boldsymbol{\psi}}(u) = \frac{1}{n}\sum_{j=1}^{n} \big[\mathbf{1}\{u \leq \widetilde{U}_{kj}\} - \widetilde{U}_{kj}\big]\, \mathbf{s}^{(k)}_{\boldsymbol{\psi}}\big(\widetilde{U}_{1j}, \widetilde{U}_{2j}, \mathbf{X}_j; \mathbf{0}_p, \widetilde{\boldsymbol{\psi}}_n\big).$$

3.2.2. *Margins are influenced by the covariate.* Often, the margins are influenced by the covariates. To remove this effect one can use various methods depending on what can be assumed about this effect; see, e.g., the beginning of Section 5 in Gijbels et al. [12]. Let $\widehat{Y}^a_{ji}$ be the values of Y_{ji} already adjusted for the effect of the covariate. Then mimicking (4) one can define the estimated 'pseudo-observations' as

$$\widehat{U}_{1i} = \frac{n}{n+1}\, \widehat{F}_{1n}\big(\widehat{Y}^a_{1i}\big) \quad \text{and} \quad \widehat{U}_{2i} = \frac{n}{n+1}\, \widehat{F}_{2n}\big(\widehat{Y}^a_{2i}\big),$$

where $\widehat{F}_{1n}(y) = \sum_{i=1}^n \mathbf{1}\{\widehat{Y}^a_{1i} \leq y\}/n$ is the empirical distribution function of the adjusted observations $\widehat{Y}^a_{11}, \ldots, \widehat{Y}^a_{1n}$ and analogously for $\widehat{F}_{2n}(y)$.

With properly adjusted pseudo-observations, the test statistic R_n, based now on the score

$$\mathbf{S}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{s}_{\boldsymbol{\alpha}}\big(\widehat{U}_{1i}, \widehat{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p; \widetilde{\boldsymbol{\psi}}_n\big),$$

will have the same asymptotic distribution under the null hypothesis as the test statistic, when margins are not influenced by the covariate. It should be mentioned that it is only in case of proper adjustments that the variance-covariance matrix $\mathbb{V}$ remains the same, and can be estimated similarly as before (replacing $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$ with $(\widehat{U}_{1i}, \widehat{U}_{2i})$ everywhere). See Gijbels et al. [12] for details about possible appropriate adjustments.

3.3. **Assessing significance.** Provided that under the null hypothesis (9) holds, the estimate of $\widehat{\text{avar}(\mathbf{S}_n)}$ given by (11) converges in probability to the regular matrix $\mathbb{V}$, then the Cramér–Slutsky theorem (under the null hypothesis) yields

$$R_n \rightsquigarrow \chi^2_p \tag{12}$$

as $n \to \infty$, where R_n is given by (5). Thus for large sample sizes one can assess the significance of the test statistic by using this asymptotic result.

In small and moderate samples we recommend to use a permutation procedure which reads as follows.

(1) Keep $(\widehat{U}_{11}, \widehat{U}_{21}), \ldots, (\widehat{U}_{1n}, \widehat{U}_{2n})$ fixed and permute $\mathbf{X}_1, \ldots, \mathbf{X}_n$ to obtain $\mathbf{X}^*_1, \ldots, \mathbf{X}^*_n$.
(2) Recalculate the test statistic based on triples $(\widehat{U}_{11}, \widehat{U}_{21}, \mathbf{X}^*_1)$, ..., $(\widehat{U}_{1n}, \widehat{U}_{2n}, \mathbf{X}^*_n)$.

These steps are carried out for a large number of permutations, say B, and the p-value of the test statistic is then calculated from the empirical distribution of these B 'observed' values of the test statistic. See for instance p. 158 of Davison and Hinkley [8]. The permutation tests are exact only in the situation that the marginal distributions are not affected by the covariate.

3.4. **Impact of the reference copula choice on size and power of the test.** Note that in constructing the test statistic R_n, we work with a reference copula C_R which may differ from the true (unknown) copula. Our theoretical results however (see, e.g., Theorem 2) are valid for any reference copula that satisfies the conditions of the theorems. This implies that the proposed test remains consistent even if the reference copula is misspecified (i.e., is different from the true copula). Consequently, one might wonder how restrictive are the assumptions under which the theoretical results hold. In the simulation study in Section 4, we illustrate the finite-sample impact of misspecifying the reference copula. As long as the parameter space of the reference copula allow to cover the type of dependence embodied by the true copula, such a choice is safe. For instance, if the true copula describes negative dependence, and the reference copula can only range over all non-negative dependencies, then this is not a good choice.

One might wonder further about the power properties of the test under local alternatives. It is outside the scope of the current paper to study this in detail. However, under an alternative hypothesis when there is a dependence on $\mathbf{x}$, the statistic S_n would still converge to a multivariate Gaussian distribution, but with no longer a zero mean, but with a mean that depends on how $\boldsymbol{\theta}(\mathbf{x})$ departs from a constant (say) under such a local alternative. As a result the statistic R_n would converge, under such local alternatives, to a non-central chi-square distribution. In case of a misspecified reference copula, this would influence the size of the non-centrality parameter, leading to a possible reduction of power.

3.5. **Extensions.** As was already pointed out via examples in Section 3.1, it is straightforward to apply our methodology for testing that only part of the covariate vector has an effect on the conditional dependence structure. To formalize this, let $\mathbf{X}_{i(S)}$ stand for the set of d_1 elements of $\mathbf{X}_i$ for which we want to test whether there is an effect. Denote by $\mathbf{X}_{i(NS)}$ the remaining $d - d_1$ covariates. Thus one can write

$$\mathbf{X}_i = \big(\mathbf{X}_{i(S)}^\top, \mathbf{X}_{i(NS)}^\top\big)^\top,$$

and analogously write $\mathbf{x} = \big(\mathbf{x}_{(S)}^\top, \mathbf{x}_{(NS)}^\top\big)$. The testing problem of interest is then

$$\mathcal{H}_0 : \forall_{\mathbf{x}\in\mathbf{R_X}}\, C_{\mathbf{x}} = C_{\mathbf{x}_{(S)}}, \quad \mathcal{H}_1 : \exists_{\mathbf{x}\in\mathbf{R_X}}\, C_{\mathbf{x}} \neq C_{\mathbf{x}_{(S)}}.$$

The only difference is that now the function $\mathbf{a}(\mathbf{x}; \mathbf{0}_p; \boldsymbol{\psi})$ can still depend on $\mathbf{x}_{(S)}$.

To assess the significance one can again use that the test statistic converges under the null hypothesis to a χ^2-distribution with $p = d - d_1$ degrees of freedom. Alternatively one can use the modified permutation procedure which is as follows

(1) Keep $(\widehat{U}_{11}, \widehat{U}_{21}, \mathbf{X}_{1(S)}), \dots, (\widehat{U}_{1n}, \widehat{U}_{2n}, \mathbf{X}_{n(S)})$ fixed and permute $\mathbf{X}_{1(NS)}, \dots, \mathbf{X}_{n(NS)}$ to obtain $\mathbf{X}^*_{1(NS)}, \dots, \mathbf{X}^*_{n(NS)}$.
(2) For each $i \in \{1, \dots, n\}$, put $\mathbf{X}^*_i = (\mathbf{X}_{i(S)}, \mathbf{X}^*_{i(NS)})$ and recalculate the test statistic based on triples $(\widehat{U}_{11}, \widehat{U}_{21}, \mathbf{X}^*_1)$, …, $(\widehat{U}_{1n}, \widehat{U}_{2n}, \mathbf{X}^*_n)$.

Note that the above tests can be used also for variable selection in model building.

4. Simulation study: Univariate covariate X

In this section we consider a univariate covariate X so that we can compare with the semiparametric test of Acar et al. [3] and the nonparametric test of Gijbels et al. [13]. First we consider the problem of testing any effect of the covariate. We consider situations in which the copula is correctly specified as well as settings where this is not the case. In a second subsection we want to test for the presence of other than a linear term of the covariate in the form of the conditional parameter, and in this case the parametric form of the copula is always considered to be correctly specified.

The R-computing environment [19] was used to perform the simulations.

4.1. **Testing for any effect of the covariate.**

Distribution of the covariate. In all simulation models the covariate X is distributed uniformly on the interval $[2, 5]$, as in Acar et al. [3].

Margins. The main purpose of this section is to illustrate that the proposed procedure can be used also in the same setting as in Acar et al. [3] and Gijbels et al. [13]. We therefore restrict ourselves mostly to the case that the margins are not influenced by the covariate. Only for one model we also present the results for the case that the margins are influenced by the covariate.

Copula models. We consider two sets of models. The first set of models is as in Acar et al. [3], and is listed in Table 1. These models constitute the semiparametric type of modeling discussed briefly in Section 2.2. Note that under Model 1 the null hypothesis (1) holds.

Table 1. *First set of models. The function $\theta(x) = g\{\eta(x)\}$ for the prespecified copula family.*

Family	Model 1	Model 2	Model 3	Link function
Frank	8	$25 - 4.2\,x$	$1 + 2.5(3 - x)^2$	$g(x) = x$
Clayton	e^1	$\exp(-1.2 + 0.8x)$	$\exp\{2 - 0.5(x - 3.8)^2\}$	$g(x) = e^x$
Gumbel	$e^{1/2} + 1$	$\exp(1.5 - 0.4x)+1$	$\exp\{-1 + 0.5(x - 4)^2\} + 1$	$g(x) = e^x + 1$

The second set of models consists of copulas that are mixtures of two copulas, viz.

$$C_x(u_1, u_2) = w(x)\, C_1(u_1, u_2) + \{1 - w(x)\}C_2(u_1, u_2),$$

with the weight function $w(x)$ and copulas C_1 and C_2 as given in Tables 2 and 3, respectively. Note that the null hypothesis (1) holds under Model 4 but that these copulas are not of the form $C_x(\cdot,\cdot) = C(\cdot,\cdot;\theta(x))$, and hence with Models 4–6 we are in misspecification settings.

Table 2. *Second set of models. The weight function $w(x)$.*

Model 4	Model 5	Model 6
$w(x) = 0.5$	$w(x) = (x-2)/3$	$w(x) = (x-3.5)^2/1.5^2$

Table 3. *Second set of models. The copulas C_1 and C_2.*

	C_1	C_2
Frank - mixt	independence copula	Frank copula with $\theta = 8$
Clayton - mixt	independence copula	Clayton copula with $\theta = 3$
Gumbel - mixt	independence copula	Gumbel copula with $\theta = 2.65$

For completeness we also consider Model 3, but now with margins that are influenced by the covariate, as follows

$$Y_{1i} = \sin\{2\,\pi\,(X_i - 1/2)\} + \varepsilon_{1i}, \quad Y_{2i} = \varepsilon_{2i}, \quad i \in \{1, \ldots, n\},$$

where the variables ε_{1i}, ε_{2i} and X_i are independent variables, and where each of the variables ε_{1i} and ε_{2i} has density $1 - |x|$ on the support $[-1, 1]$. We refer to this as Model 3X in the tables.

Sample sizes and number of samples. We consider the sample sizes $n = 50$, 100 and 200. We generate 5000 samples to estimate the levels of the tests. Further 1000 samples are generated to estimate the power of the tests.

Considered tests. For clarity of presentation we write the subscript R when denoting the reference copula, the reference dependence parameter and the reference link function for construction of our test.

The following tests are included in the simulation study:

- The test of Acar et al. [3] - $\lambda_n(h)$ in (2.2.1) with cross-validation choice of the bandwidth.
- The test of Gijbels et al. [13] based on the test statistic V_{n1} in (2.2.2).

Further tests based on the suggested test statistic (5) are included, with the parameter

$$\theta_R(x) = g_R\big(\psi + \alpha_1\, x + \alpha_2\, x^2\big), \tag{13}$$

where g_R is a reference link function and we are testing that $(\alpha_1, \alpha_2) = (0, 0)$. We used the following three reference copula families and reference link functions for constructing the test:

a) Frank family with $g_R(x) = x$ – R_n-*frank* and $R_n^{(perm)}$-*frank*
b) Clayton family with $g_R(x) = \exp\{x\}$ – R_n-*clayt.* and $R_n^{(perm)}$-*clayt.*
c) Gumbel family with $g_R(x) = \exp\{x\} + 1$ – R_n-*gumb.* and $R_n^{(perm)}$-*gumb.*

where the superscript $^{(perm)}$ indicates that the permutation test is used to assess significance. For the permutation procedure we use $B = 999$ throughout the paper. All tests are performed at the nominal 0.05 significance level.

sign. level = 0.05	Model 1			Model 2			Model 3			Model 3X		
	50	100	200	50	100	200	50	100	200	50	100	200
Frank − $\lambda_n(h)$	0.051	0.064	0.059	0.480	0.836	0.985	0.401	0.822	0.995	0.455	0.735	0.990
Frank − V_{n1}	0.053	0.056	0.044	0.424	0.781	0.995	0.308	0.635	0.940	0.280	0.596	0.950
Frank − R_n-*frank*	0.053	0.057	0.050	0.340	0.757	0.987	0.591	0.911	1.000	0.501	0.906	0.997
Frank − $R_n^{(perm)}$-*frank*	0.052	0.051	0.044	0.329	0.738	0.984	0.494	0.879	0.999	0.396	0.875	0.997
Clayton − $\lambda_n(h)$	0.073	0.071	0.069	0.837	0.994	1.000	0.490	0.832	0.985	0.340	0.670	0.940
Clayton − V_{n1}	0.061	0.064	0.064	0.574	0.933	1.000	0.190	0.345	0.645	0.130	0.274	0.570
Clayton − R_n-*clayt.*	0.064	0.059	0.060	0.747	0.992	1.000	0.380	0.735	0.979	0.153	0.606	0.963
Clayton − $R_n^{(perm)}$-*clayt.*	0.050	0.046	0.050	0.677	0.987	1.000	0.306	0.670	0.971	0.110	0.524	0.951
Gumbel − $\lambda_n(h)$	0.063	0.067	0.070	0.217	0.446	0.745	0.265	0.592	0.925	0.340	0.565	0.955
Gumbel − V_{n1}	0.060	0.052	0.049	0.180	0.334	0.590	0.157	0.328	0.655	0.170	0.304	0.630
Gumbel − R_n-*gumb.*	0.060	0.057	0.059	0.207	0.445	0.802	0.272	0.690	0.962	0.201	0.669	0.942
Gumbel − $R_n^{(perm)}$-*gumb.*	0.046	0.046	0.051	0.176	0.380	0.768	0.273	0.659	0.954	0.220	0.631	0.934
Frank − R_n-*clayt.*	0.154	0.153	0.126	0.375	0.550	0.754	0.180	0.573	0.935	0.148	0.552	0.940
Frank − $R_n^{(perm)}$-*clayt.*	0.051	0.049	0.049	0.194	0.323	0.535	0.199	0.595	0.923	0.156	0.613	0.921
Frank − R_n-*gumb.*	0.070	0.073	0.067	0.357	0.583	0.882	0.260	0.720	0.977	0.186	0.682	0.980
Frank − $R_n^{(perm)}$-*gumb.*	0.048	0.049	0.047	0.271	0.501	0.833	0.332	0.765	0.979	0.275	0.734	0.982
Clayton − R_n-*frank*	0.060	0.060	0.058	0.659	0.979	1.000	0.161	0.447	0.881	0.076	0.459	0.875
Clayton − $R_n^{(perm)}$-*frank*	0.054	0.051	0.051	0.652	0.976	1.000	0.167	0.431	0.866	0.078	0.439	0.864
Clayton − R_n-*gumb.*	0.089	0.087	0.078	0.469	0.725	0.950	0.181	0.304	0.528	0.079	0.219	0.514
Clayton − $R_n^{(perm)}$-*gumb.*	0.047	0.050	0.052	0.334	0.609	0.909	0.086	0.178	0.393	0.042	0.135	0.391
Gumbel − R_n-*frank*	0.053	0.058	0.055	0.214	0.411	0.703	0.387	0.683	0.937	0.319	0.693	0.940
Gumbel − $R_n^{(perm)}$-*frank*	0.050	0.049	0.049	0.180	0.352	0.674	0.297	0.612	0.921	0.240	0.636	0.928
Gumbel − R_n-*clayt.*	0.127	0.132	0.109	0.199	0.349	0.551	0.144	0.428	0.810	0.118	0.430	0.790
Gumbel − $R_n^{(perm)}$-*clayt.*	0.049	0.050	0.050	0.126	0.199	0.403	0.115	0.368	0.725	0.127	0.351	0.718

TABLE 4. *In Models 1—3 the margins are not influenced by X. In Model 3X the margins are influenced by the covariate. Rejection frequencies for sample sizes $n = 50, 100$ and 200.*

sign. level = 0.05	Model 4			Model 5			Model 6		
	50	100	200	50	100	200	50	100	200
Frank - mixt $-$ $\lambda_n(h)$	0.088	0.130	0.146	0.465	0.780	0.990	0.370	0.815	0.985
Frank - mixt $-$ V_{n1}	0.054	0.058	0.053	0.405	0.715	0.930	0.097	0.185	0.445
Frank - mixt $-$ R_n-*frank*	0.096	0.075	0.066	0.467	0.757	0.953	0.386	0.734	0.971
Frank - mixt $-$ $R_n^{(perm)}$-*frank*	0.051	0.050	0.051	0.346	0.683	0.943	0.272	0.627	0.956
Frank - mixt $-$ R_n-*clayt.*	0.053	0.057	0.061	0.164	0.414	0.805	0.267	0.563	0.832
Frank - mixt $-$ $R_n^{(perm)}$-*clayt.*	0.046	0.049	0.047	0.201	0.450	0.802	0.241	0.418	0.728
Frank - mixt $-$ R_n-*gumb.*	0.056	0.055	0.058	0.174	0.490	0.863	0.293	0.650	0.929
Frank - mixt $-$ $R_n^{(perm)}$-*gumb.*	0.046	0.058	0.050	0.242	0.541	0.866	0.258	0.563	0.900
Clayton - mixt $-$ $\lambda_n(h)$	0.363	0.350	0.337	0.540	0.845	0.975	0.580	0.895	1.000
Clayton - mixt $-$ V_{n1}	0.054	0.053	0.052	0.400	0.710	0.935	0.080	0.171	0.425
Clayton - mixt $-$ R_n-*frank*	0.094	0.075	0.060	0.436	0.692	0.955	0.379	0.715	0.966
Clayton - mixt $-$ $R_n^{(perm)}$-*frank*	0.051	0.048	0.047	0.325	0.611	0.940	0.259	0.616	0.952
Clayton - mixt $-$ R_n-*clayt.*	0.055	0.068	0.068	0.190	0.530	0.909	0.419	0.787	0.982
Clayton - mixt $-$ $R_n^{(perm)}$-*clayt.*	0.051	0.056	0.050	0.202	0.555	0.909	0.353	0.629	0.948
Clayton - mixt $-$ R_n-*gumb.*	0.059	0.043	0.047	0.127	0.351	0.765	0.167	0.465	0.855
Clayton - mixt $-$ $R_n^{(perm)}$-*gumb.*	0.056	0.048	0.052	0.164	0.456	0.802	0.180	0.430	0.819
Gumbel - mixt $-$ $\lambda_n(h)$	0.303	0.337	0.198	0.445	0.815	0.980	0.535	0.885	0.995
Gumbel - mixt $-$ V_{n1}	0.049	0.048	0.042	0.417	0.729	0.985	0.083	0.175	0.430
Gumbel - mixt $-$ R_n-*frank*	0.098	0.077	0.067	0.435	0.766	0.960	0.443	0.764	0.981
Gumbel - mixt $-$ $R_n^{(perm)}$-*frank*	0.053	0.051	0.051	0.332	0.692	0.945	0.316	0.687	0.974
Gumbel - mixt $-$ R_n-*clayt.*	0.053	0.058	0.061	0.141	0.351	0.796	0.257	0.534	0.855
Gumbel - mixt $-$ $R_n^{(perm)}$-*clayt.*	0.051	0.056	0.051	0.190	0.395	0.795	0.234	0.416	0.767
Gumbel - mixt $-$ R_n-*gumb.*	0.050	0.053	0.059	0.225	0.545	0.956	0.390	0.784	0.989
Gumbel - mixt $-$ $R_n^{(perm)}$-*gumb.*	0.050	0.052	0.053	0.287	0.611	0.950	0.346	0.673	0.980

TABLE 5. *Models 4—6, the margins are not influenced by X. Rejection frequencies for sample sizes* $n = 50, 100$ *and* 200.

The results are presented in Tables 4 and 5. To facilitate the reading of Table 4 we present results for the test based on correctly specified copula models in the first block, and results for misspecified reference copulas (and parameters/links) in a second major block. The second block thus allows to judge how the test performs in case of reference model misspecification.

Some findings. The asymptotic χ^2-approximation of the distribution of the suggested test statistic R_n under the null hypothesis is often (but not always) appropriate.

If the copula family is correctly specified (see the first block of Table 4) then the performance of the suggested test (using the true copula) is usually comparable with the test of Acar et al. [3]. The advantage of the suggested test is that it continues to work reasonably well even if the copula family is misspecified. This can be seen by comparing results from the second block in Table 4 with the corresponding ones in the first block, as well as by looking at Table 5. Note also from Table 4 that when the margins are influenced by the covariate, and we are working with properly adjusted pseudo-observations, the performances of the test are comparable; see the column ‘Model 3X’.

Relatively to its competitors the suggested tests do a slightly better job for 'quadratic models' (i.e., Models 3 and 6). The reason is that in constructing the proposed test statistic, we assumed that $\theta(x)$ is of the form (13). Thus for 'linear models' (i.e., Models 2 and 5) the quadratic term is (at least in Model 2) useless.

4.2. **Testing for other than a 'linear' effect of the covariate.** In this subsection the copula family and the link function g are always correctly specified. The null hypothesis is that the copula parameter is of the form

$$\theta(x) = g(\psi_0 + \psi_1 \, x), \tag{14}$$

where ψ_0 and ψ_1 are unknown parameters. The alternative is that the copula parameter has a more complex form.

In our framework we test the above hypothesis by assuming for the reference dependence parameter the 'quadratic model'

$$\theta_R(x) = g(\psi_0 + \psi_1 \, x + \alpha \, x^2)$$

and testing the null hypothesis $\mathcal{H}_0 : \alpha = 0$. Note that other ways of modeling the deviation from model (14) are possible depending on what is assumed about this deviation.

As far as we know such a hypothesis was considered only in Acar et al. [3]. For comparison purposes, we use therefore the same design. As in Section 4.1 the covariate X is distributed uniformly on the interval $[2, 5]$ and the margins are not assumed to be influenced by the covariate, except for Model 3X (with a same influence as before). The models are listed in Table 6. We consider all the models used in Acar et al. [3] and also add two new models (see Model 3 for the Clayton and Gumbel families).

TABLE 6. *Models 1–3. The function $\theta(x) = g\{\eta(x)\}$ for the prespecified copula family.*

Family	Model 1	Model 2	Model 3	Link function
Frank	$25 - 4.2\,x$	$1 + 2.5\,(3 - x)^2$	$12 + 8\,\sin(0.4\,x^2)$	$g(x) = x$
Clayton	$\exp(-1.2 + 0.8\,x)$	$\exp\{2 - 0.5\,(x - 3.8)^2\}$	$\exp\{2 - 4\,\sin(0.4\,x^2)\}$	$g(x) = e^x$
Gumbel	$\exp(1.5 - 0.4\,x) + 1$	$\exp\{-1 + 0.5\,(x - 4)^2\} + 1$	$\exp\{-1 + 2\,\sin(x)\} + 1$	$g(x) = e^x + 1$

Some findings. The results are in Table 7. The findings can be summarized as follows.

a) The test suggested in this paper is usually better in keeping the probability of Type I error than the test of Acar et al. [3].
b) If the effect of the covariate is quadratic (and the family and link function are correctly specified), then the suggested test has also slightly better power properties (when taking the level properties into consideration).

sign. level = 0.05	Model 1			Model 2			Model 3			Model 3X		
	50	100	200	50	100	200	50	100	200	50	100	200
Frank $\lambda_n(h)$	0.050	0.044	0.045	0.202	0.320	0.605	0.606	0.931	1.000	0.247	0.905	0.999
Frank R_n-$frank$	0.041	0.043	0.043	0.344	0.614	0.902	0.480	0.763	0.965	0.340	0.789	0.968
Clayton $\lambda_n(h)$	0.160	0.126	0.094	0.345	0.615	0.920	0.965	1.000	1.000	0.861	1.000	1.000
Clayton R_n-*clayt.*	0.038	0.043	0.039	0.303	0.629	0.947	0.481	0.784	0.955	0.398	0.784	0.968
Gumbel $\lambda_n(h)$	0.132	0.102	0.090	0.189	0.257	0.420	0.560	0.820	0.992	0.312	0.776	0.994
Gumbel R_n-*gumb.*	0.042	0.054	0.054	0.063	0.178	0.465	0.580	0.882	0.998	0.336	0.808	0.996

TABLE 7. *Testing for other than a linear effect. In Models 1—3 the margins are not influenced by the covariate. In Model 3X the margins are influenced by the covariate. Rejection frequencies for sample sizes* $n = 50, 100$ *and* 200.

c) If the effect of the covariate is more complex than quadratic, such as in Model 3, then the test of Acar et al. [3] has usually better power properties. The difference in powers can be either substantial if the effect deviates much from the linear effect (Model 3 with Clayton copula) or moderate (Model 3 with Frank copula). In models where the deviation can be well approximated by the quadratic function (Model 3 with Gumbel copula) the difference in powers can be even in favour of the proposed test.

To sum it up, the test of Acar et al. [3] has very good power properties. Nevertheless, the test proposed herein has the following advantages: (i) it is better in keeping the level even under model misspecifications; (ii) it is not so computationally intensive (the computing time is substantially smaller) as the test of Acar et al. [3] that requires to solve many (the sample size times the number of the grid points in the cross-validation procedure) optimisation tasks, due to the involvement of a smoothing parameter choice.

5. Simulation results: A multivariate covariate

In this section we consider a two-dimensional covariate generated by a two-dimensional Gaussian copula with correlation coefficient equal to 0.5. First we consider the problem of testing of the effect of the whole covariate vector. Then we test that only the first component of the covariate influences the conditional copula as described in Section 3.5.

5.1. **Testing for any effect of the (whole) covariate.**

Considered models. In the simulation models we consider Frank copulas or mixtures involving these. A first set of models, Models 1–3, involves a single Frank copula in which the parameter function $\theta(\mathbf{x})$ is as listed in Table 8.

In a second set of models, Models 4–6, the conditional copula is given by

$$C_{\mathbf{x}}(u_1, u_2) = w(\mathbf{x})\, C_1(u_1, u_2) + \{1 - w(\mathbf{x})\}\, C_2(u_1, u_2), \qquad (15)$$

TABLE 8. *First set of models for the multivariate covariate setting, with Frank copula family with parameter function $\theta(\mathbf{x})$ as listed.*

Family	Model 1	Model 2	Model 3
Frank	8	$8\,(x_1 - 2\,x_2 + 1.5)$	$8[2/\left[1 + \exp\{2\,x_1^2 - 4\,x_1 - x_2\}\right] - 1]$

with C_1 and C_2 as in the first line of Table 3 and the weight function $w(\mathbf{x}) = w(x_1, x_2)$ as in Table 9.

TABLE 9. *Second set of models for the multivariate covariate setting, with $\bar{x} = (x_1 + x_2)/2$.*

Model 4	Model 5	Model 6
$w(\mathbf{x}) = 0.5$	$w(\mathbf{x}) = \bar{x}$	$w(\mathbf{x}) = 4\,(\bar{x} - 0.5)^2$

Considered tests. The following test statistics are included in this simulation study:

a) The test of Gijbels et al. [13] given by the test statistic V_{n1} in (2.2.2).

Further the proposed test using the following copula reference models:

b) Frank family assuming that $\theta_R(\mathbf{x}) = \psi + \alpha_1\,x_1 + \alpha_2\,x_2$ – R_n-*frank* and $R_n^{(perm)}$-*frank*;
c) Frank family assuming that $\theta_R(\mathbf{x}) = \psi + \alpha_1\,x_1 + \alpha_2\,x_2 + \alpha_3\,x_1^2 + \alpha_4\,x_2^2$ – R_n-*frank* (*quadr.*) and $R_n^{(perm)}$-*frank* (*quadr.*);
d) Clayton family with $\theta_R(\mathbf{x}) = \exp(\psi + \alpha_1\,x_1 + \alpha_2\,x_2)$ – R_n-*clayt.* and $R_n^{(perm)}$-*clayt.*;
e) Gumbel family with $\theta_R(\mathbf{x}) = \exp\{\psi + \alpha_1\,x_1 + \alpha_2\,x_2\} + 1$ – R_n-*gumb.* and $R_n^{(perm)}$-*gumb.*

Note that for most models (except for Models 1 and 2 with the linear or quadratic form for the dependence parameter) we are in settings of misspecification.

Some findings. The results can be found in Table 10, and can be summarized as follows.

a) The nonparametric test of Gijbels et al. [13] performs very well compared to the suggested tests.
b) Of course, there is some loss of power if the copula family is misspecified.
c) Note that the suggested tests (with the exception of R_n-*frank* (*quadr.*) and $R_n^{(perm)}$-*frank* (*quadr.*)) have practically no power in Model 6. An explanation for this is that in this model the effect of the covariate on the conditional copula is 'quadratic', but these tests assume only a 'linear' effect.

5.2. **Testing for an (partial) effect of the second coordinate of the covariate.** We consider Frank copulas or mixtures involving these. A first set of models, Models 1–3, involves a single multivariate Frank copula in which the parameter function $\theta(\mathbf{x})$ is as listed in Table 11.

sign. level = 0.05	Model 1			Model 2			Model 3		
	50	100	200	50	100	200	50	100	200
Frank - V_{n1}	0.051	0.050	0.048	0.517	0.911	0.995	0.218	0.458	0.800
Frank - R_n-$frank$	0.047	0.047	0.046	0.622	0.949	1.000	0.258	0.458	0.768
Frank - $R_n^{(perm)}$-$frank$	0.053	0.050	0.047	0.610	0.949	1.000	0.217	0.430	0.757
Frank - R_n-$frank$ $(quadr.)$	0.055	0.056	0.058	0.476	0.887	0.999	0.182	0.364	0.707
Frank - $R_n^{(perm)}$-$frank$ $(quadr.)$	0.051	0.049	0.047	0.402	0.848	0.996	0.120	0.287	0.642
Frank - R_n-*clayt.*	0.141	0.144	0.111	0.502	0.771	0.940	0.258	0.371	0.559
Frank - $R_n^{(perm)}$-*clayt.*	0.049	0.052	0.049	0.329	0.622	0.888	0.168	0.243	0.446
Frank - R_n-*gumb.*	0.061	0.065	0.052	0.582	0.854	0.986	0.231	0.410	0.671
Frank - $R_n^{(perm)}$-*gumb.*	0.046	0.051	0.043	0.500	0.813	0.984	0.200	0.371	0.641
	Model 4			Model 5			Model 6		
Frank - mixt - V_{n1}	0.053	0.049	0.054	0.274	0.507	0.820	0.097	0.140	0.240
Frank - mixt - R_n-$frank$	0.081	0.067	0.054	0.349	0.588	0.890	0.068	0.050	0.059
Frank - mixt - $R_n^{(perm)}$-$frank$	0.047	0.050	0.049	0.273	0.531	0.871	0.048	0.047	0.052
Frank - mixt - R_n-$frank$ $(quadr.)$	0.124	0.091	0.072	0.340	0.526	0.830	0.202	0.362	0.681
Frank - mixt - $R_n^{(perm)}$-$frank$ $(quadr.)$	0.051	0.047	0.047	0.189	0.416	0.770	0.107	0.263	0.593
Frank - mixt - R_n-*clayt.*	0.047	0.052	0.056	0.113	0.294	0.657	0.136	0.113	0.100
Frank - mixt - $R_n^{(perm)}$-*clayt.*	0.058	0.054	0.050	0.150	0.349	0.647	0.077	0.049	0.061
Frank - mixt - R_n-*gumb.*	0.032	0.040	0.048	0.128	0.309	0.753	0.084	0.085	0.072
Frank - mixt - $R_n^{(perm)}$-*gumb.*	0.051	0.049	0.049	0.202	0.390	0.767	0.064	0.065	0.054

TABLE 10. *Testing for any effect of the covariate* $(d = 2)$*, the margins are not influenced by the covariate. Rejection frequencies for sample sizes* $n = 50, 100$ *and* 200*.*

TABLE 11. *First set of models for the multivariate covariate setting, with Frank copula family with parameter function* $\theta(\mathbf{x}) = g(\eta(\mathbf{x}))$ *as listed.*

Family	Model 1	Model 2	Model 3	Link function
Frank	8	$8\,(x_1 + 0.5)$	$8\,(x_1 - 2\,x_2 + 1.5)$	$g(x) = x$
Clayton	$e^{1.1}$	$\exp(-1.2 + 5\,x_1)$	$\exp\{-1.2 + 3\,x_1 + 3\,x_2\}$	$g(x) = e^x$
Gumbel	$e^{1/2} + 1$	$\exp(1.5 - 1.4\,x_1)+1$	$\exp\{1.5 - 0.7\,x_1 - 0.7\,x_2\} + 1$	$g(x) = e^x + 1$

In a second set of models, Models 4–6, the conditional copula is given by (15), with C_1 and C_2 as in the first line of Table 3 and the weight function $w(\mathbf{x}) = w(x_1, x_2)$ as in Table 12.

TABLE 12. *Second set of models for the multivariate covariate setting.*

Model 4	Model 5	Model 6
$w(\mathbf{x}) = 0.5$	$w(\mathbf{x}) = x_1$	$w(\mathbf{x}) = (x_1 + x_2)/2$

We want to test the null hypothesis that the conditional copula depends only on the first coordinate of the covariate. Thus Models 1, 2, 4 and 5 represent the null hypothesis.

sign. level = 0.05	Model 1			Model 2			Model 3		
	50	100	200	50	100	200	50	100	200
Frank − R_n-*frank*	0.050	0.052	0.047	0.048	0.053	0.057	0.722	0.977	1.000
Frank − $R_n^{(perm)}$-*frank*	0.054	0.052	0.048	0.048	0.053	0.054	0.715	0.976	1.000
Clayton − R_n-*clayt.*	0.044	0.052	0.052	0.031	0.034	0.033	0.345	0.733	0.984
Clayton − $R_n^{(perm)}$-*clayt.*	0.040	0.044	0.051	0.050	0.056	0.059	0.393	0.802	0.987
Gumbel − R_n-*gumb.*	0.055	0.050	0.052	0.039	0.042	0.049	0.097	0.243	0.479
Gumbel − $R_n^{(perm)}$-*gumb.*	0.051	0.047	0.050	0.046	0.044	0.052	0.107	0.236	0.469
Frank − R_n-*clayt.*	0.092	0.110	0.090	0.087	0.104	0.091	0.506	0.791	0.958
Frank − $R_n^{(perm)}$-*clayt.*	0.044	0.044	0.043	0.045	0.049	0.046	0.399	0.689	0.927
Frank − R_n-*gumb.*	0.056	0.056	0.056	0.052	0.054	0.056	0.598	0.909	0.994
Frank − $R_n^{(perm)}$-*gumb.*	0.048	0.049	0.048	0.050	0.047	0.050	0.575	0.885	0.991
Clayton − R_n-*frank*	0.048	0.053	0.050	0.054	0.053	0.061	0.639	0.948	1.000
Clayton − $R_n^{(perm)}$-*frank*	0.048	0.052	0.049	0.053	0.052	0.059	0.641	0.943	1.000
Clayton − R_n-*gumb.*	0.058	0.061	0.052	0.040	0.054	0.062	0.281	0.576	0.894
Clayton − $R_n^{(perm)}$-*gumb.*	0.043	0.046	0.048	0.045	0.051	0.052	0.282	0.561	0.870
Gumbel − R_n-*frank*	0.042	0.054	0.053	0.044	0.046	0.046	0.103	0.187	0.396
Gumbel − $R_n^{(perm)}$-*frank*	0.045	0.053	0.050	0.052	0.050	0.046	0.115	0.200	0.392
Gumbel − R_n-*clayt.*	0.085	0.091	0.084	0.087	0.086	0.089	0.129	0.204	0.265
Gumbel − $R_n^{(perm)}$-*clayt.*	0.044	0.046	0.050	0.050	0.048	0.060	0.082	0.127	0.192

TABLE 13. *Models 1—3 (testing for a partial effect of the second covariate). The margins are not influenced by the covariate. Rejection frequencies for sample sizes* $n = 50, 100$ *and* 200.

The assumed copula (reference copula for the test) is a Frank copula (with $g_R(x) = x$) and

$$\theta_R(\mathbf{x}) = \psi_1 + \psi_2\, x_1 + \alpha\, x_2.$$

Thus we test the null hypothesis $\mathcal{H}_0 : \alpha = 0$ and the assumed alternative is $\mathcal{H}_1 : \alpha \neq 0$.

The results for margins not influenced by the covariate are in Tables 13 and 14.

Next we have considered also a simulation setting where the margins are influenced by the covariate in the following way

$$Y_{1i} = \sin\{2\,\pi\,(X_{1i} - 1/2)\} + X_{2i} + \varepsilon_{1i}, \quad Y_{2i} = \varepsilon_{2i}, \quad i \in \{1, \dots, n\},$$

where the variables ε_{1i}, ε_{2i} and (X_{1i}, X_{2i}) are independent elements. Further each of the variables ε_{1i} and ε_{2i} has density $1 - |x|$ on the support $[-1, 1]$. For brevity we only present in Table 15 the results for the Frank and Frank mixture models.

Some findings. The findings can be summarized as follows.

a) The suggested tests perform well (i.e., keep the level and have some power) even if the copula family is misspecified. The price to pay is some loss of power.
b) Not surprisingly, there is also some small loss of power if the margins are influenced by the covariate.

sign. level = 0.05	Model 4			Model 5			Model 6		
	50	100	200	50	100	200	50	100	200
Frank-mixt − R_n-$frank$	0.065	0.063	0.062	0.073	0.060	0.051	0.158	0.237	0.382
Frank-mixt − $R_n^{(perm)}$-$frank$	0.046	0.050	0.054	0.058	0.052	0.047	0.121	0.212	0.367
Frank-mixt − R_n-*clayt.*	0.045	0.054	0.061	0.059	0.085	0.081	0.085	0.172	0.268
Frank-mixt − $R_n^{(perm)}$-*clayt.*	0.054	0.050	0.053	0.058	0.060	0.045	0.100	0.146	0.199
Frank-mixt − R_n-*gumb.*	0.038	0.037	0.045	0.048	0.064	0.068	0.086	0.190	0.310
Frank-mixt − $R_n^{(perm)}$-*gumb.*	0.048	0.045	0.048	0.060	0.060	0.061	0.108	0.182	0.293
Clayton-mixt − R_n-$frank$	0.068	0.066	0.058	0.061	0.055	0.051	0.127	0.203	0.386
Clayton-mixt − $R_n^{(perm)}$-$frank$	0.047	0.052	0.050	0.046	0.046	0.045	0.104	0.175	0.353
Clayton-mixt − R_n-*clayt.*	0.043	0.054	0.061	0.055	0.078	0.088	0.084	0.176	0.371
Clayton-mixt − $R_n^{(perm)}$-*clayt.*	0.048	0.046	0.050	0.067	0.061	0.059	0.099	0.162	0.319
Clayton-mixt − R_n-*gumb.*	0.029	0.042	0.057	0.042	0.058	0.064	0.058	0.127	0.254
Clayton-mixt − $R_n^{(perm)}$-*gumb.*	0.046	0.049	0.059	0.056	0.052	0.056	0.078	0.141	0.261
Gumbel-mixt − R_n-$frank$	0.065	0.058	0.053	0.065	0.058	0.058	0.134	0.236	0.390
Gumbel-mixt − $R_n^{(perm)}$-$frank$	0.045	0.048	0.047	0.050	0.049	0.053	0.106	0.215	0.365
Gumbel-mixt − R_n-*clayt.*	0.038	0.053	0.063	0.058	0.075	0.079	0.081	0.160	0.263
Gumbel-mixt − $R_n^{(perm)}$-*clayt.*	0.047	0.050	0.054	0.064	0.055	0.053	0.114	0.148	0.219
Gumbel-mixt − R_n-*gumb.*	0.033	0.048	0.048	0.052	0.066	0.064	0.096	0.217	0.420
Gumbel-mixt − $R_n^{(perm)}$-*gumb.*	0.040	0.046	0.045	0.062	0.058	0.054	0.121	0.211	0.382

TABLE 14. *Models 4—6 (testing for a partial effect of the second covariate). The margins are not influenced by the covariate. Rejection frequencies for sample sizes* $n = 50, 100$ *and* 200.

sign. level = 0.05	Model 1			Model 2			Model 3		
	50	100	200	50	100	200	50	100	200
Frank − R_n-$frank$	0.059	0.044	0.047	0.064	0.047	0.056	0.248	0.715	0.998
Frank − $R_n^{(perm)}$-$frank$	0.054	0.044	0.047	0.059	0.044	0.054	0.224	0.704	0.995
Frank − R_n-*clayt.*	0.069	0.090	0.089	0.065	0.076	0.078	0.220	0.674	0.968
Frank − $R_n^{(perm)}$-*clayt.*	0.046	0.049	0.048	0.050	0.037	0.040	0.206	0.562	0.932
Frank − R_n-*gumb.*	0.040	0.048	0.051	0.044	0.057	0.065	0.218	0.583	0.980
Frank − $R_n^{(perm)}$-*gumb.*	0.048	0.044	0.047	0.049	0.059	0.060	0.229	0.553	0.974
	Model 4			Model 5			Model 6		
Frank-mixt − R_n-$frank$	0.075	0.056	0.059	0.074	0.058	0.058	0.092	0.188	0.366
Frank-mixt − $R_n^{(perm)}$-$frank$	0.054	0.047	0.049	0.054	0.050	0.052	0.070	0.163	0.352
Frank-mixt − R_n-*clayt.*	0.024	0.044	0.063	0.038	0.090	0.091	0.057	0.163	0.326
Frank-mixt − $R_n^{(perm)}$-*clayt.*	0.059	0.049	0.052	0.068	0.086	0.061	0.107	0.174	0.254
Frank-mixt − R_n-*gumb.*	0.025	0.041	0.050	0.035	0.059	0.074	0.050	0.141	0.252
Frank-mixt − $R_n^{(perm)}$-*gumb.*	0.057	0.054	0.057	0.067	0.062	0.066	0.097	0.154	0.236

TABLE 15. *Models 1—6 (testing for a partial effect of the second covariate). The margins are influenced by the covariate. Rejection frequencies for sample sizes* $n = 50, 100$ *and* 200.

6. Real data example

As an illustration we revisit the data on hydro-geochemical stream and sediment reconnaissance from Cook and Johnson [7]. These data consist of the observed log-concentrations of seven chemicals in 655 water samples collected near Grand Junction, Colorado. The data can be found, e.g., as a data set called `uranium` in the R package `copula` [16]. In the analysis we focus on the relationship between Cesium (Cs) and Scandium (Sc), given some of the other chemicals.

For the analysis we use the Frank copula as the reference copula in our testing procedure. This copula can capture both positive and negative dependencies, and it is in conformance, e.g., with the fact that the dependence between these two variables is not extreme-value [5]. The `loess` function in the R software was used to adjust for the possible effect of the covariates on the margins. For testing for the effect of a single covariate (x) we consider the conditional parameter of the form

$$\theta_R(x) = \psi + \alpha_1\, x + \alpha_2\, x^2.$$

For testing any effect of a two-dimensional covariate ($\mathbf{x} = (x_1, x_2)$) we use

$$\theta_R(\mathbf{x}) = \psi + \alpha_1\, x_1 + \alpha_2\, x_1^2 + \alpha_3\, x_2 + \alpha_4\, x_2^2 + \alpha_5\, x_1\, x_2.$$

Finally, for testing a partial effect of the second covariate (x_2) when the first covariate (x_1) is included we consider the model

$$\theta_R(\mathbf{x}) = \psi_1 + \psi_2\, x_1 + \psi_3\, x_1^2 + \alpha_1\, x_2 + \alpha_2\, x_2^2 + \alpha_3\, x_1\, x_2.$$

Further to stabilize the effect of covariate we transform the values of the covariates into (marginal) ranks. For comparison we also included the results based on the statistic V_{n1} in Gijbels et al. [13]; see also Sections 4.1 and 5.1. Results for testing for a partial effect are of course not available (NA) for this nonparametric test.

Some of the results are presented in Table 16. Note that when considering any effect of the univariate or multivariate covariate (lines 1, 2, 3, 6 and 7), our findings are mostly in agreement with the findings based on the statistic V_{n1}. The only exception is the effect of Lithium (Li) on line 6 which is not significant when testing with the help of V_{n1} but it is slightly below the standard level of statistical significance when using the test statistic (5).

The new findings are on lines 4, 5, 8 and 9 concerning testing for partial effects of a bivariate covariate vector. The results suggest that Co has an effect on the conditional copula even if Ti is included (line 4). On the other hand when Co is included the effect of Ti is only borderline (line 5).

Acknowledgments. The authors are grateful to the Editor-in-Chief, Associate Editor and the reviewers for their valuable comments, which led to an improved manuscript.

Appendix

In this appendix we show all the results needed to prove (12) when margins are not influenced by the covariate. First we start with some auxiliary results of independent interest.

TABLE 16. *p-values for the test of the simplifying assumption of Cesium and Scandium when conditioning on one other chemical (lines 1, 2, 6) or two additional chemicals (lines 3 and 7). Lines 4, 5, 8 and 9 contain the p-values of the test that the conditional copula is affected by the first variable provided that the second variable is included.*

Cov. effect	V_{n1}	R_n-$frank$	$R_n^{(perm)}$-$frank$
Ti	0.001	0.001	0.001
Co	0.003	0.001	0.001
Ti and Co	0.001	0.001	0.011
Co when Ti incl.	NA	0.003	0.005
Ti when Co incl.	NA	0.090	0.097
Li	0.151	0.018	0.021
Ti and Li	0.001	0.001	0.001
Li when Ti incl.	NA	0.010	0.011
Ti when Li incl.	NA	0.001	0.001

Then we subsequently prove the asymptotic normality result (9), the asymptotic representation (7), asymptotic normality of (10) and finally the consistency of the estimator of the asymptotic variance estimator (11).

A1 - Auxiliary results. We first introduce the notion of U-shaped functions (see, e.g., Shorack [20]).

Definition 1 (Subclasses of U-shaped functions)**.** A function $r : (0,1) \to (0,\infty)$ is called U-shaped if it is symmetric about $1/2$ and decreasing on $(0,1/2]$.

(i) Let $\mathcal{V}$ be the set of continuous functions v on $[0,1]$, such that $1/v$ is U-shaped and $\int_0^1 \{v(t)\}^{-2} \mathrm{d}t < \infty$.

(ii) For $0 < \beta < 1$ and a U-shaped function r, we define

$$r_\beta(t) = \begin{cases} r(\beta t) & \text{if}\, 0 < t \leq 1/2, \\ r\{1-\beta(1-t)\} & \text{if}\, 1/2 < t < 1. \end{cases}$$

If for every $\beta > 0$ in a neighborhood of 0, there exists a constant M_β, such that $r_\beta \leq M_\beta\, r$ on $(0,1)$, then r is called a reproducing U-shaped function. We denote by $\mathcal{R}$ the set of reproducing U-shaped functions.

Before we proceed, we need to generalize the results of Genest et al. [9] and Tsukahara [24] for a score function, which is also a function of the covariate. Let $\mathbf{R_X} \subset \mathbb{R}^d$ be the support of $\mathbf{X}$ and let J be a function from $[0,1]^2 \times \mathbf{R_X}$ to $\mathbb{R}$.

Suppose we observe independent random vectors $(Y_{11}, Y_{21}, \mathbf{X}_1), \ldots, (Y_{1n}, Y_{2n}, \mathbf{X}_n)$. Recall that $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$ are given by (4) and $(U_{1i}, U_{2i}) = \big(F_1(Y_{1i}), F_2(Y_{2i})\big)$, where F_1 and F_2 are the marginal distribution function of Y_{1i} and Y_{2i}, respectively.

Definition 2 (Class of $\mathcal{J}_1$- and $\mathcal{J}_2$-functions)**.** A function J is called a $\mathcal{J}_1$-function if J is continuous on $(0,1)^2 \times \mathbf{R_X}$ and there exist functions $r_1, r_2 \in \mathcal{R}$ and $M : \mathbf{R_X} \to \mathbb{R}$ such that

$$|J(u_1, u_2, \mathbf{x})| \le M(\mathbf{x})\, r_1(u_1)\, r_2(u_2), \tag{A1}$$

and

$$\mathrm{E}\left\{M(\mathbf{X})\, r_1(U_1) r_2(U_2)\right\} < \infty. \tag{A2}$$

Furthermore, a function J is called a $\mathcal{J}_2$-function if there exist functions $r_1, r_2, \tilde{r}_1, \tilde{r}_2 \in \mathcal{R}$, $v_1, v_2 \in \mathcal{V}$, and $M : \mathbf{R_X} \to \mathbb{R}$ such that

$$\begin{gathered} |J(u_1, u_2, \mathbf{x})| \le M(\mathbf{x})\, r_1(u_1)\, r_2(u_2), \\ \left|J^{(1)}(u_1, u_2, \mathbf{x})\right| \le M(\mathbf{x})\, \tilde{r}_1(u_1)\, r_2(u_2), \ \left|J^{(2)}(u_1, u_2, \mathbf{x})\right| \le M(\mathbf{x})\, r_1(u_1) \tilde{r}_2(u_2), \end{gathered} \tag{A3}$$

where $J^{(j)}(\cdot) = \partial J(\cdot)/\partial u_j$ $(j \in \{1,2\})$ is continuous on $(0,1)^2 \times \mathbf{R_X}$, and

$$\begin{gathered} \mathrm{E}\left\{M(\mathbf{X})\, r_1(U_1)\, r_2(U_2)\right\}^2 < \infty, \\ \mathrm{E}\left\{M(\mathbf{X})\, v_1(U_1) \tilde{r}_1(U_1) r_2(U_2)\right\} < \infty, \qquad \mathrm{E}\left\{M(\mathbf{X})\, v_2(U_2) r_1(U_1) \tilde{r}_2(U_2)\right\} < \infty. \end{gathered} \tag{A4}$$

Lemma 1. *If J is a $\mathcal{J}_1$-function, then*

$$\frac{1}{n} \sum_{i=1}^{n} J(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i) \xrightarrow[n\to\infty]{\mathrm{Pr}} \mathrm{E}\left\{J(U_1, U_2, \mathbf{X})\right\}.$$

Proof. Due to the law of large numbers, it is sufficient to show that

$$D_n = \left| \frac{1}{n} \sum_{i=1}^{n} J(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i) - \frac{1}{n} \sum_{i=1}^{n} J(U_{1i}, U_{2i}, \mathbf{X}_i) \right| \xrightarrow[n\to\infty]{\mathrm{Pr}} 0. \tag{A5}$$

Let $\varepsilon > 0$ be given. For a given $\delta \in (0, 1/2)$ (that will be specified later on), consider the set

$$\mathbf{I}_\delta = \{(u_1, u_2) : (u_1, u_2) \in [\delta, 1-\delta]^2\}. \tag{A6}$$

Thanks to the assumptions of the lemma, one can choose $\delta > 0$ and a compact set $\mathbf{K} \subset \mathbf{R_X}$ such that

$$\mathrm{E}\big[M(\mathbf{X})\, r_1(U_1)\, r_2(U_2)\, \mathbf{1}\big\{(U_1, U_2, \mathbf{X}) \notin \mathbf{I}_\delta \times \mathbf{K}\big\}\big] < \varepsilon. \tag{A7}$$

Now, we introduce the set of indices

$$\mathrm{J}_\delta = \big\{i : (U_{1i}, U_{2i}, \mathbf{X}_i) \in \mathbf{I}_\delta \times \mathbf{K}\big\} \tag{A8}$$

and note that D_n introduced in (A5) can be bounded as

$$D_n \le \left| \frac{1}{n} \sum_{i \in \mathrm{J}_\delta} J(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i) - \frac{1}{n} \sum_{i \in \mathrm{J}_\delta} J(U_{1i}, U_{2i}, \mathbf{X}_i) \right| \tag{A9}$$

$$+ \frac{1}{n} \sum_{i \notin \mathrm{J}_\delta} |J(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i)| + \frac{1}{n} \sum_{i \notin \mathrm{J}_\delta} |J(U_{1i}, U_{2i}, \mathbf{X}_i)|. \tag{A10}$$

Using (A1), (A7), and the law of large numbers, the second term in (A10) can be bounded as

$$\begin{aligned} \frac{1}{n} \sum_{i \notin \mathrm{J}_\delta} |J(U_{1i}, U_{2i}, \mathbf{X}_i)| &\le \frac{1}{n} \sum_{i=1}^{n} M(\mathbf{X}_i)\, r_1(U_{1i})\, r_2(U_{2i})\, \mathbf{1}\big\{(U_{1i}, U_{2i}, \mathbf{X}_i) \notin \mathbf{I}_\delta \times \mathbf{K}\big\} \\ &= \mathrm{E}\big[M(\mathbf{X})\, r_1(U_1)\, r_2(U_2)\, \mathbf{1}\big\{(U_1, U_2, \mathbf{X}) \notin \mathbf{I}_\delta \times \mathbf{K}\big\}\big] + o_P(1) \\ &\le \varepsilon + o_P(1). \end{aligned} \tag{A11}$$

To bound the first term in (A10), Lemma A3 by Shorack [20] assures that there exists $\beta_\varepsilon \in (0,1)$ such that for all sufficiently large n

$$\Pr\big\{\forall_{i \in \{1,\ldots,n\}}\ \forall_{j \in \{1,2\}}\ \beta_\varepsilon\, U_{ji} \le \widetilde{U}_{ji} \le 1 - \beta_\varepsilon\,(1 - U_{ji})\big\} > 1 - \varepsilon. \tag{A12}$$

Taking into account (A12) and (A1) the first term in (A10) can be bounded with probability greater than $1 - \varepsilon$ as follows

$$\begin{aligned} \frac{1}{n} \sum_{i \notin \mathrm{J}_\delta} |J(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i)| &\le \frac{1}{n} \sum_{i=1}^{n} M(\mathbf{X}_i)\, r_{1\beta_\varepsilon}(U_{1i})\, r_{2\beta_\varepsilon}(U_{2i})\, \mathbf{1}\big\{(U_{1i}, U_{2i}, \mathbf{X}_i) \notin \mathbf{I}_\delta \times \mathbf{K}\big\} \\ &= \mathrm{E}\big[M(\mathbf{X})\, r_{1\beta_\varepsilon}(U_1)\, r_{2\beta_\varepsilon}(U_2)\, \mathbf{1}\big\{(U_1, U_2, \mathbf{X}) \notin \mathbf{I}_\delta \times \mathbf{K}\big\}\big] + o_P(1) \\ &\le O(\varepsilon) + o_P(1). \end{aligned} \tag{A13}$$

By the uniform convergence of an empirical distribution function and the continuity of the function J, the term on the right-hand side of (A9) converges to zero in probability. This combined with (A11) and (A13) implies the statement of the lemma. □

Remark 1. Note that in our context $\mathbf{X}$ is independent of (U_1, U_2). Thus assumption (A2) can be replaced with

$$\mathrm{E}\,\{M(\mathbf{X})\} < \infty, \qquad \mathrm{E}\{r_1(U_1) r_2(U_2)\} < \infty.$$

With the help of Hölder's inequality it is easy to show that a sufficient condition for the second assumption is that for some $\delta > 0$

$$r_1(u) = \{u(1-u)\}^{\frac{-1+\delta}{p}}, \qquad r_2(u) = \{u(1-u)\}^{\frac{-1+\delta}{q}},$$

where p, q are positive constants such that $1/p + 1/q = 1$.

Lemma 2. *Let J be a $\mathcal{J}_2$-function, and assume that* $\mathrm{E}\{J(U_1, U_2, \mathbf{X})\} = 0$. *Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} J(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi(U_{1i}, U_{2i}, \mathbf{X}_i) + o_P(1),$$

where

$$\phi(u_1, u_2, \mathbf{x}) = J(u_1, u_2, \mathbf{x}) + \int_{[0,1]^2 \times \mathbb{R}^d} \big[\mathbf{1}\{u_1 \leq v_1\} - v_1\big] J^{(1)}\big(v_1, v_2, \mathbf{x}\big) \,\mathrm{d}H(v_1, v_2, \mathbf{x}) + \int_{[0,1]^2 \times \mathbb{R}^d} \big[\mathbf{1}\{u_2 \leq v_2\} - v_2\big] J^{(2)}\big(v_1, v_2, \mathbf{x}\big) \,\mathrm{d}H(v_1, v_2, \mathbf{x}),$$

where H stands for the joint cumulative distribution function of $(U_1, U_2, \mathbf{X})$.

Proof. The proof will be divided into two steps. First, we will show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} J(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J(U_{1i}, U_{2i}, \mathbf{X}_i) + \sum_{j=1}^{2} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J^{(j)}(U_{1i}, U_{2i}, \mathbf{X}_i)\big(\widetilde{U}_{ji} - U_{ji}\big) + o_P(1), \tag{A14}$$

and second, we will prove that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} J^{(j)}(U_{1i}, U_{2i}, \mathbf{X}_i)\big(\widetilde{U}_{ji} - U_{ji}\big) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \big[\mathbf{1}\{U_{ji} \leq v_j\} - v_j\big] J^{(j)}\big(v_1, v_2, \mathbf{x}\big) \,\mathrm{d}H(v_1, v_2, \mathbf{x}) + o_P(1), \tag{A15}$$

which together with (A14) provides the statement of the lemma.

Step 1: Proving (A14). By the mean value theorem,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} J(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J(U_{1i}, U_{2i}, \mathbf{X}_i) + \sum_{j=1}^{2} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J^{(j)}(U_{1i}^*, U_{2i}^*, \mathbf{X}_i)\big(\widetilde{U}_{ji} - U_{ji}\big), \tag{A16}$$

where U_{ji}^* lies between $\widetilde{U}_{ji}$ and U_{ji}. In what follows, we show that one can indeed replace $J^{(j)}(U_{1i}^*, U_{2i}^*, \mathbf{X}_i)$ with $J^{(j)}(U_{1i}, U_{2i}, \mathbf{X}_i)$ in (A16).

For $\eta \in [0, 1/2)$ and $\delta \in (0, 1/2)$ (that will be specified later) define the processes

$$\mathbb{F}_{jn}(u) = \frac{\sqrt{n}\,\big[F_{jn}\{F_j^{-1}(u)\} - u\big]}{u^\eta (1-u)^\eta}, \quad u \in (0,1). \tag{A17}$$

and

$$\widetilde{\mathbb{F}}_{jn}(u) = \frac{\sqrt{n}\left[\frac{n}{n+1} F_{jn}\{F_j^{-1}(u)\} - u\right]}{u^\eta(1-u)^\eta}\,\mathbf{1}\big\{u \in \big[\delta/n, 1-\delta/n\big]\big\} + \mathbb{F}_{jn}(u)\,\mathbf{1}\{u \notin \big[\delta/n, 1-\delta/n\big]\big\}. \quad \text{(A18)}$$

First of all note that

$$\begin{aligned}\widetilde{\mathbb{F}}_{jn}(u) - \mathbb{F}_{jn}(u) &= \left[\frac{\sqrt{n}\left[\frac{n}{n+1} F_{jn}\{F_j^{-1}(u)\} - u\right]}{u^\eta(1-u)^\eta} - \mathbb{F}_{jn}(u)\right]\mathbf{1}\big\{u \in \big[\delta/n, 1-\delta/n\big]\big\} \\ &= -\frac{\sqrt{n}\left[\frac{1}{n+1} F_{jn}\{F_j^{-1}(u)\}\right]}{u^\eta(1-u)^\eta}\,\mathbf{1}\big\{u \in \big[\delta/n, 1-\delta/n\big]\big\},\end{aligned}$$

which implies that

$$\begin{aligned}\sup_{u\in[0,1]}\left|\widetilde{\mathbb{F}}_{jn}(u) - \mathbb{F}_{jn}(u)\right| &\le \sup_{u\in\left[\delta/n, 1-\delta/n\right]} \frac{\frac{\sqrt{n}}{n+1} F_{jn}\{F_j^{-1}(u)\}}{u^\eta(1-u)^\eta} \\ &\le \frac{\sqrt{n}}{n+1}\,\frac{1}{\left(\delta/n\right)^\eta\left(1-\delta/n\right)^\eta} = o(1). \quad \text{(A19)}\end{aligned}$$

Further note that

$$\widetilde{\mathbb{F}}_{jn}(u) = \frac{\sqrt{n}\left[\frac{n}{n+1} F_{jn}\{F_j^{-1}(u)\} - u\right]}{u^\eta(1-u)^\eta} + R_{jn}^\delta(u),$$

with

$$R_{jn}^\delta(u) = \left[\mathbb{F}_{jn}(u) - \frac{\sqrt{n}\left[\frac{n}{n+1} F_{jn}\{F_j^{-1}(u)\} - u\right]}{u^\eta(1-u)^\eta}\right]\mathbf{1}\{u \notin \big[\delta/n, 1-\delta/n\big]\big\},$$

which is always nonnegative. It can be seen that, for $j \in \{1,2\}$,

$$\begin{aligned}\Pr\left\{\max_{i\in\{1,\dots,n\}} R_{jn}^\delta(U_{ji}) = 0\right\} &= \Pr\Big(\min_{1\le i\le n} U_{ji} > \delta/n \,\text{and}\, \max_{1\le i\le n} U_{ji} < 1-\delta/n\Big) \\ &\ge \Pr\Big(\min_{1\le i\le n} U_{ji} > \delta/n\Big) + \Pr\Big(\max_{1\le i\le n} U_{ji} < 1-\delta/n\Big) - 1.\end{aligned}$$

Since

$$\Pr\Big(\min_{1\le i\le n} U_{ji} \le \delta/n\Big) = \Pr\Big(\max_{1\le i\le n} U_{ji} \ge 1-\delta/n\Big) = (1-\delta/n)^n\,,$$

we obtain

$$\Pr\left\{\max_{i\in\{1,\dots,n\}} R_{jn}^\delta(U_{ji}) = 0\right\} \ge 2\,(1-\delta/n)^n - 1.$$

Hence,

$$\begin{aligned}\Pr\left\{\max_{j\in\{1,2\}}\max_{i\in\{1,\dots,n\}} R_{jn}^\delta(U_{ji}) \ne 0\right\} &\le \Pr\left\{\max_{i\in\{1,\dots,n\}} R_{1n}^\delta(U_{1i}) \ne 0\right\} \\ &\quad + \Pr\left\{\max_{i\in\{1,\dots,n\}} R_{2n}^\delta(U_{2i}) \ne 0\right\} \\ &\le 2\left\{1 - 2\,(1-\delta/n)^n + 1\right\} = 4 - 4\,(1-\delta/n)^n\,.\end{aligned}$$

Consequently,

$$\Pr\left\{\max_{j\in\{1,2\}}\max_{i\in\{1,\ldots,n\}} R_{jn}^{\delta}(U_{ji})=0\right\}\geq -3+4\,(1-\delta/n)^{n}\,. \tag{A20}$$

Note that the right-hand side of (A20) tends to $-3+4e^{-\delta}$, as $n\to\infty$, for any $\delta\in(0,1/2)$; in the limit, for δ going to 0, this bound tends to 1. So, by taking δ sufficiently small we get that the event

$$A_n^{\delta}=\left\{\max_{j\in\{1,2\}}\max_{i\in\{1,\ldots,n\}} R_{jn}^{\delta}(U_{ji})=0\right\}=\left\{\forall_{j\in\{1,2\}}\forall_{i\in\{1,\ldots,n\}}\, U_{ji}\in\big[\delta/n,1-\delta/n\big]\right\}, \tag{A21}$$

Therefore for the rest of the paper, we will work on the intersection with this event. Note that (on A_n^{δ})

$$\max_{j\in\{1,2\}}\max_{i\in\{1,\ldots,n\}}\left|\frac{\sqrt{n}\big(\widetilde{U}_{ji}-U_{ji}\big)}{U_{ji}^{\eta}(1-U_{ji})^{\eta}}\right|=\max_{j\in\{1,2\}}\max_{i\in\{1,\ldots,n\}}\big|\widetilde{\mathbb{F}}_{jn}(U_{ji})\big|.$$

Using this together with the Chibisov–O'Reilly theorem (see, e.g., Shorack and Wellner [21], p. 462) for the process $\mathbb{F}_{jn}(u)$ and (A19) yields that

$$\max_{j\in\{1,2\}}\max_{i\in\{1,\ldots,n\}}\left|\frac{\sqrt{n}\big(\widetilde{U}_{ji}-U_{ji}\big)}{U_{ji}^{\eta}(1-U_{ji})^{\eta}}\right|=O_P(1). \tag{A22}$$

Now, let $\varepsilon>0$ be given. Then, let us introduce $\mathbf{I}_\delta$, $\mathbf{K}$ and J_δ as in (A6) and (A8) with δ chosen such that

$$\begin{aligned}\mathrm{E}\big[M(\mathbf{X})\,v_1(U_1)\tilde{r}_1(U_1)r_2(U_2)\,\mathbf{1}\big\{(U_1,U_2,\mathbf{X})\notin\mathbf{I}_\delta\times\mathbf{K}\big\}\big]<\varepsilon,\\ \mathrm{E}\big[M(\mathbf{X})\,v_2(U_2)r_1(U_1)\tilde{r}_2(U_2)\mathbf{1}\big\{(U_1,U_2,\mathbf{X})\notin\mathbf{I}_\delta\times\mathbf{K}\big\}\big]<\varepsilon.\end{aligned} \tag{A23}$$

With respect to (A22), one can bound (on A_n^{δ})

$$\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}J^{(j)}(U_{1i}^{*},U_{2i}^{*},\mathbf{X}_i)\big(\widetilde{U}_{ji}-U_{ji}\big)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}J^{(j)}(U_{1i},U_{2i},\mathbf{X}_i)\big(\widetilde{U}_{ji}-U_{ji}\big)\right|$$

$$\leq O_P(1)\,\frac{1}{n}\sum_{i=1}^{n}\Big|J^{(j)}(U_{1i}^{*},U_{2i}^{*},\mathbf{X}_i)-J^{(j)}(U_{1i},U_{2i},\mathbf{X}_i)\Big|\,U_{ji}^{\eta}(1-U_{ji})^{\eta}$$

$$\leq O_P(1)\,\Bigg\{\frac{1}{n}\sum_{i\in\mathrm{J}_\delta}\Big|J^{(j)}(U_{1i}^{*},U_{2i}^{*},\mathbf{X}_i)-J^{(j)}(U_{1i},U_{2i},\mathbf{X}_i)\Big| \tag{A24}$$

$$+\frac{1}{n}\sum_{i\notin\mathrm{J}_\delta}\Big|J^{(j)}(U_{1i}^{*},U_{2i}^{*},\mathbf{X}_i)\Big|\,U_{ji}^{\eta}(1-U_{ji})^{\eta} \tag{A25}$$

$$+\frac{1}{n}\sum_{i\notin\mathrm{J}_\delta}\Big|J^{(j)}(U_{1i},U_{2i},\mathbf{X}_i)\Big|\,U_{ji}^{\eta}(1-U_{ji})^{\eta}\Bigg\}. \tag{A26}$$

The term on the right-hand side of (A24) converges in probability to zero by the continuity of $J^{(j)}$. To deal with the term in (A25), one can choose η in (A22) arbitrarily close to 1/2.

Thus similarly as in (A13), using (A12), (A3), (A23), and the law of large numbers, one can bound (on A_n^δ)

$$\frac{1}{n}\sum_{i\notin \mathrm{J}_\delta}\Big|J^{(j)}(U_{1i}^*,U_{2i}^*,\mathbf{X}_i)\Big|\,U_{ji}^\eta(1-U_{ji})^\eta \leq O(1)\,\frac{1}{n}\sum_{i\notin \mathrm{J}_\delta} v_j(U_{ji})\,\tilde{r}_j(U_{ji})\,r_{3-j}(U_{3-j,i})$$
$$\leq O(\varepsilon)+o_P(1).$$

Analogously, one can also treat the term in (A26), which completes the proof of (A14).

Step 2: Proof of (A15). Let $\eta\in[0,1/2)$ be chosen such that $v_j(u)=O\{u^\eta(1-u)^\eta\}$. Similarly as in Step 1 of the proof, we will work on the intersection with the event A_n^δ defined (A21). On A_n^δ one has

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n J^{(j)}(U_{1i},U_{2i},\mathbf{X}_i)\big(\widetilde{U}_{ji}-U_{ji}\big)=\frac{1}{n}\sum_{i=1}^n \widetilde{h}_n(U_{1i},U_{2i},\mathbf{X}_i), \tag{A27}$$

where

$$\widetilde{h}_n(u_1,u_2,\mathbf{x})=J^{(j)}(u_1,u_2,\mathbf{x})\,u_j^\eta(1-u_j)^\eta\,\widetilde{\mathbb{F}}_{jn}(u_j),$$

with $\widetilde{\mathbb{F}}_{jn}$ defined in (A18). Further with the help of (A19) one has (on A_n^δ)

$$\max_{j\in\{1,2\}}\max_{i\in\{1,\ldots,n\}}\big|\widetilde{\mathbb{F}}_{jn}(U_{ji})-\mathbb{F}_{jn}(U_{ji})\big|\leq o(1),$$

where $\mathbb{F}_{jn}$ is defined in (A17). Now using (A27) one gets

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n J^{(j)}(U_{1i},U_{2i},\mathbf{X}_i)\big(\widetilde{U}_{ji}-U_{ji}\big)=\frac{1}{n}\sum_{i=1}^n h_n(U_{1i},U_{2i},\mathbf{X}_i)+o_P(1), \tag{A28}$$

where $h_n(u_1,u_2,\mathbf{x})=J^{(j)}(u_1,u_2,\mathbf{x})\,u_j^\eta(1-u_j)^\eta\,\mathbb{F}_{jn}(u_j)$.

Now, by the Chibisov–O'Reilly theorem the process $\big\{\mathbb{F}_{jn}:u\in(0,1)\big\}$ converges in distribution in the space of bounded function $\ell^\infty([0,1])$ (equipped with the supremum metric ϱ). Hence, by Theorem 1.5.8 of van der Vaart and Wellner [25] the process is asymptotically tight and thus for each $\varepsilon>0$ there exists a compact set $\mathcal{K}\subset\ell^\infty([0,1])$ such that

$$\liminf_{n\to\infty}\Pr(\mathbb{F}_{jn}\in\mathcal{K}_{\varepsilon/2})\geq 1-\varepsilon,$$

where $\mathcal{K}_{\varepsilon/2}=\big\{g\in\ell^\infty([0,1]):\mathrm{dist}(g,\mathcal{K})<\varepsilon/2\big\}$ is an $(\varepsilon/2)$-enlargement around $\mathcal{K}$, with $\mathrm{dist}(g,\mathcal{K})=\inf_{\tilde{g}\in\mathcal{K}}\varrho(g,\tilde{g})$. Thus, $\Pr(h_n\in\mathcal{H}_\varepsilon)\geq 1-\varepsilon$, where

$$\mathcal{H}_\varepsilon=\big\{(u_1,u_2,\mathbf{x})\mapsto J^{(j)}(u_1,u_2,\mathbf{x})\,u_j^\eta(1-u_j)^\eta\,g(u_j),\,g\in\mathcal{K}_{\varepsilon/2}\big\}.$$

Moreover, as the set $\mathcal{K}$ is compact, it can be covered by finitely many balls of radius $\varepsilon/2$. Hence, for each $\varepsilon>0$ the ε-bracketing number of $\mathcal{K}_{\varepsilon/2}$ with respect to the supremum norm is finite. Since $\mathrm{E}\,|J^{(j)}(U_1,U_2,\mathbf{X})\,U_j^\eta(1-U_j)^\eta|<\infty$, it is straightforward to show that for each

$\varepsilon > 0$ the ε-bracketing number of $\mathcal{H}_\varepsilon$ with respect to the L_1-norm is finite. One can proceed as in the proof of Theorem 2.4.1 of van der Vaart and Wellner [25] to show that for each $\varepsilon > 0$

$$\sup_{h\in\mathcal{H}_\varepsilon}\left|\frac{1}{n}\sum_{i=1}^n h(U_{1i},U_{2i},\mathbf{X}_i) - \mathrm{E}\left\{h(U_1,U_2,\mathbf{X})\right\}\right| \leq \varepsilon + o_P(1).$$

Combined with the fact that with an arbitrarily high probability for all sufficiently large n the function h_n lies in $\mathcal{H}_\varepsilon$, this implies that

$$\frac{1}{n}\sum_{i=1}^n h_n(U_{1i},U_{2i},\mathbf{X}_i) = \mathrm{E}_{\mathbf{Z}}\{h_n(U_1,U_2,\mathbf{X})\} + o_P(1), \tag{A29}$$

where $\mathrm{E}_{\mathbf{Z}}$ stands for an expectation taken with respect to $\mathbf{Z} = (U_1,U_2,\mathbf{X})$. We stress this here since the random function h_n (depending on $(U_{1i},U_{2i},\mathbf{X}_i)$, $i\in\{1,\ldots,n\}$) is considered as fixed. It remains to calculate

$$\mathrm{E}_{\mathbf{Z}}\{h_n(U_1,U_2,\mathbf{X})\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n\int\left[\mathbf{1}\{U_{ji}\leq v_j\} - v_j\right]J^{(j)}(v_1,v_2,\mathbf{x})\,\mathrm{d}H(v_1,v_2,\mathbf{x}),$$

which together with (A28) and (A29) provides (A15). □

Remark 2. Analogously as in Remark 1, if $\mathbf{X}$ is independent of (Y_1,Y_2), then (A4) is satisfied if

$$\mathrm{E}\left\{M(\mathbf{X})\right\}^2 < \infty, \quad r_1(u) = \{u(1-u)\}^{\frac{-1/2+\delta}{p}}, \quad r_2(u) = \{u(1-u)\}^{\frac{-1/2+\delta}{q}},$$

for some $\delta > 0$ and p,q positive constants such that $1/p + 1/q = 1$.

A2 - Theorem 1 and proof of (9)**.** To proceed we need to formulate needed regularity assumptions.

Assumptions R:

R1. The function $\rho(u_1,u_2,\mathbf{x};\boldsymbol{\alpha};\boldsymbol{\psi})$ is continuously differentiable with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ for all $(u_1,u_2,\mathbf{x})$.

R2. The function $R(\boldsymbol{\psi}) = \mathrm{E}\left\{\rho(U_{1i},U_{2i},\mathbf{X}_i;\mathbf{0}_p;\boldsymbol{\psi})\right\}$ has under the null hypothesis a unique maximizer $\widetilde{\boldsymbol{\psi}}$.

R3. Under the null hypothesis $\mathrm{E}\left\{\mathbf{s}_{\boldsymbol{\alpha}}\big(U_{1i},U_{2i},\mathbf{X}_i;\mathbf{0}_p;\widetilde{\boldsymbol{\psi}}\big)\right\} = \mathbf{0}_p$.

For each $k\in\{1,\ldots,p+q\}$, let us denote the kth element of $\mathbf{s}(u_1,u_2,\mathbf{x};\boldsymbol{\psi})$ by $s_k(u_1,u_2,\mathbf{x};\boldsymbol{\psi})$.

R4. For each $k\in\{1,\ldots,p+q\}$, the function $s_k(\cdot,\cdot,\cdot;\widetilde{\boldsymbol{\psi}})\in\mathcal{J}_2$.

Theorem 1. *Assume that Assumptions **R1**–**R4** hold. Then, under* $\mathcal{H}_0$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{s}\big(\widetilde{U}_{1i},\widetilde{U}_{2i},\mathbf{X}_i;\widetilde{\boldsymbol{\psi}}\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\phi}(U_{1i},U_{2i},\mathbf{X}_i) + o_P(1), \tag{A30}$$

where

$$\begin{aligned}\boldsymbol{\phi}(u_1,u_2,\mathbf{x}) &= \mathbf{s}(u_1,u_2,\mathbf{x};\widetilde{\boldsymbol{\psi}})\\ &\quad+\int_{[0,1]^2}\int_{\mathbb{R}^d}\big[\mathbf{1}\{u_1\le v_1\}-v_1\big]\,\mathbf{s}^{(1)}\big(v_1,v_2,\mathbf{x};\widetilde{\boldsymbol{\psi}}\big)\,\mathrm{d}C(v_1,v_2)\,\mathrm{d}F_{\mathbf{X}}(\mathbf{x})\\ &\quad+\int_{[0,1]^2}\int_{\mathbb{R}^d}\big[\mathbf{1}\{u_2\le v_2\}-v_2\big]\mathbf{s}^{(2)}\big(v_1,v_2,\mathbf{x};\widetilde{\boldsymbol{\psi}}\big)\,\mathrm{d}C(v_1,v_2)\,\mathrm{d}F_{\mathbf{X}}(\mathbf{x}),\end{aligned} \tag{A31}$$

which further implies (9)*.*

Proof of Theorem 1. The asymptotic representation (A30) follows due to Lemma 2 applied to each element $s_k(u_1,u_2,\mathbf{x};\boldsymbol{\psi})$ of the score function. The central limit theorem for independent identically distributed vectors provides the asymptotic normality result (9). The existence of a finite variance matrix is guaranteed by Assumption **R4**. □

A3 - Proof of representation (7)**.** Let us define the *partial score function* as a function of the nuisance parameter $\boldsymbol{\psi}$, i.e.,

$$\boldsymbol{W}_n(\boldsymbol{\psi})=\frac{1}{n}\sum_{i=1}^{n}\mathbf{s}_{\boldsymbol{\psi}}\big(\widetilde{U}_{1i},\widetilde{U}_{2i},\mathbf{X}_i;\mathbf{0}_p,\boldsymbol{\psi}\big)$$

Thus, the estimator $\widetilde{\boldsymbol{\psi}}_n$ of the nuisance parameter $\boldsymbol{\psi}$ under the null is a solution of the estimating equations $\boldsymbol{W}_n(\widetilde{\boldsymbol{\psi}}_n)=\mathbf{0}_q$. In order to obtain some properties of $\widetilde{\boldsymbol{\psi}}_n$, the following additional assumptions are postulated.

Assumptions I:

I1. The function $\mathbf{s}(u_1,u_2,\mathbf{x};\boldsymbol{\psi})$ is assumed to be continuously differentiable with respect to $\boldsymbol{\psi}$ for all $(u_1,u_2,\mathbf{x})$. Further there exists an open neighborhood $\mathcal{U}$ of $\widetilde{\boldsymbol{\psi}}$ such that $\partial\mathbf{s}(u_1,u_2,\mathbf{x};\boldsymbol{\psi})/\partial\boldsymbol{\psi}$ is continuous in $(0,1)^2\times\mathbf{R}_{\mathbf{X}}\times\mathcal{U}$ and there exists a dominating function $h(u_1,u_2,\mathbf{x})\in\mathcal{J}_1$ such that

$$\sup_{\boldsymbol{\psi}\in\mathcal{U}}\left\|\frac{\partial\mathbf{s}}{\partial\boldsymbol{\psi}}(u_1,u_2,\mathbf{x};\boldsymbol{\psi})\right\|\le h(u_1,u_2,\mathbf{x}).$$

I2. The $q\times q$ matrix $I_{\boldsymbol{\psi}\boldsymbol{\psi}}=-\mathrm{E}\left\{\partial\mathbf{s}_{\boldsymbol{\psi}}(U_1,U_2,\mathbf{X};\mathbf{0}_p,\widetilde{\boldsymbol{\psi}})/\partial\boldsymbol{\psi}^{\top}\right\}$ is nonsingular.

Lemma 3. *Suppose that Assumptions **I1**, **I2**, and **R1**–**R4** are satisfied, then* (7) *holds.*

Proof of Lemma 3. Similarly as in the first part of the proof of Theorem 1 by Tsukahara [24], we employ Theorem A.10.2 of Bickel et al. [6] for $\boldsymbol{W}_n(\boldsymbol{\psi})$ and the corresponding estimator $\widetilde{\boldsymbol{\psi}}_n$. Note that Assumption (GM0) in that theorem is trivially satisfied due to Assumption **R2**. Moreover, Assumptions **I1** and **I2** imply Assumption (GM3). Assumption (GM2) is also satisfied as by Theorem 1

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{s}_{\boldsymbol{\psi}}\big(\widetilde{U}_{1i},\widetilde{U}_{2i},\mathbf{X}_i;\widetilde{\boldsymbol{\psi}}\big)=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\phi}_{\boldsymbol{\psi}}(U_{1i},U_{2i},\mathbf{X}_i)+o_P(n^{-1/2}),$$

where $\boldsymbol{\phi}_{\boldsymbol{\psi}}$ are the corresponding components of $\boldsymbol{\phi}$ which is given in (A31). Thus, it remains to check Assumption (U) from Theorem A.10.2. Therefore for each $\varepsilon > 0$ and for each $j, \ell \in \{1, \dots, q\}$, it is sufficient to find a neighborhood $\mathcal{U}_\varepsilon = \{\boldsymbol{\psi} \in \mathcal{U} : \|\boldsymbol{\psi} - \widetilde{\boldsymbol{\psi}}\| < \varepsilon\}$ such that

$$\sup_{\boldsymbol{\psi} \in \mathcal{U}_\varepsilon} \left| \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \psi_\ell} s_{\boldsymbol{\psi},j}(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p, \boldsymbol{\psi}) - I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{(j,\ell)} \right| \le \varepsilon + o_P(1),$$

where $I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{(j,\ell)}$ stands for the (j, ℓ) element of $I_{\boldsymbol{\psi}\boldsymbol{\psi}}$.

For simplicity of notation, let us put $g_{j,\ell}(u_1, u_2, \mathbf{x}; \boldsymbol{\psi}) = \partial s_{\boldsymbol{\psi},j}(u_1, u_2, \mathbf{x}; \mathbf{0}_p, \boldsymbol{\psi})/\partial \psi_\ell$. Assumption **I1** allows to adapt Lemma 1, which gives

$$\frac{1}{n} \sum_{i=1}^n g_{j,\ell}(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \widetilde{\boldsymbol{\psi}}) - I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{(j,\ell)} = o_P(1).$$

Hence, it remains to show

$$D_n = \sup_{\boldsymbol{\psi} \in \mathcal{U}_\varepsilon} \left| \frac{1}{n} \sum_{i=1}^n g_{j,\ell}(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \boldsymbol{\psi}) - \frac{1}{n} \sum_{i=1}^n g_{j,\ell}(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \widetilde{\boldsymbol{\psi}}) \right| \le \varepsilon + o_P(1). \tag{A32}$$

For a given $\delta \in (0, 1/4)$ (that will be specified later on), let us introduce the sets $\mathbf{I}_\delta$ and J_δ as in (A6) and (A8). Then the left-hand side of (A32) can be bounded by

$$D_n \le \sup_{\boldsymbol{\psi} \in \mathcal{U}_\varepsilon} \left| \frac{1}{n} \sum_{i \in \mathrm{J}_\delta} g_{j,\ell}(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \boldsymbol{\psi}) - \frac{1}{n} \sum_{i \in \mathrm{J}_\delta} g_{j,\ell}(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \widetilde{\boldsymbol{\psi}}) \right| \tag{A33}$$

$$+ \frac{2}{n} \sum_{i \notin \mathrm{J}_\delta} h(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i). \tag{A34}$$

With probability going to 1 for each sufficiently large n, if $(U_{1i}, U_{2i}) \in \mathbf{I}_\delta$, then $(\widetilde{U}_{1i}, \widetilde{U}_{2i}) \in \mathbf{I}_{\delta/2}$. Thus for each $\delta \in (0, 1/4)$ the right-hand side of (A33) can be made arbitrarily small (Assumption **I1**) up to $o_P(1)$ term by considering a sufficiently small neighbourhood $\mathcal{U}_\varepsilon$.

Finally, analogously as in the proof of Lemma 1, one can show that

$$\frac{1}{n} \sum_{i \notin \mathrm{J}_\delta} h(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i) \le O(1)\, \mathrm{E}\left[M(\mathbf{X})\, r_1(U_1)\, r_2(U_2)\, \mathbf{1}\{(U_1, U_2, \mathbf{X}) \notin \mathbf{I}_\delta \times \mathbf{K}\}\right] + o_P(1), \tag{A35}$$

where $\mathbf{K}$ is a compact subset of $\mathbf{R}_{\mathbf{X}}$. The right-hand side of the inequality (A35) can be made arbitrarily small by decreasing δ and enlarging K. Combining (A33), (A34), and (A35) verifies (A32). Thus by Theorem A.10.2 of Bickel et al. [6] one gets

$$\sqrt{n}\,(\widetilde{\boldsymbol{\psi}}_n - \widetilde{\boldsymbol{\psi}}) = \{I_{\boldsymbol{\psi}\boldsymbol{\psi}}(\mathbf{0}_p, \widetilde{\boldsymbol{\psi}})\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\phi}_{\boldsymbol{\psi}}(U_{1i}, U_{2i}, \mathbf{X}_i; \mathbf{0}_p, \widetilde{\boldsymbol{\psi}}) + o_P(1),$$

which together with Theorem 1 gives (7). □

A4 - Theorem 2 and proof of (10)**.**

Theorem 2. *Under Assumptions **I1**, **I2**, and **R1**–**R4**,* (10) *holds.*

Proof of Theorem 2. Let $s_{\boldsymbol{\alpha},j}$ be the jth element of the vector function $\mathbf{s}_{\boldsymbol{\alpha}}$ and $S_{n,j}$ the jth coordinate of $\mathbf{S}_n$. By the mean value theorem and using Assumption **I1**, there exists $\widetilde{\boldsymbol{\psi}}_n^*$, which lies between $\widetilde{\boldsymbol{\psi}}_n$ and $\widetilde{\boldsymbol{\psi}}$, such that

$$\begin{aligned} S_{n,j} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_{\boldsymbol{\alpha},j}\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p; \widetilde{\boldsymbol{\psi}}_n\big) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_{\boldsymbol{\alpha},j}\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p, \widetilde{\boldsymbol{\psi}}\big) \\ &\quad + \sum_{\ell=1}^{q} \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \psi_\ell} s_{\boldsymbol{\alpha},j}\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p, \widetilde{\boldsymbol{\psi}}_n^*\big) \sqrt{n}\,(\widetilde{\psi}_{n,\ell} - \widetilde{\psi}_\ell), \end{aligned} \tag{A36}$$

where $\widetilde{\psi}_{n,\ell}$ and $\widetilde{\psi}_\ell$ are the ℓth elements of $\widetilde{\boldsymbol{\psi}}_n$ and $\widetilde{\boldsymbol{\psi}}$, respectively.

Analogously as in the proof of Lemma 3, one can show that

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \psi_\ell} s_{\boldsymbol{\alpha},j}\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p, \widetilde{\boldsymbol{\psi}}_n^*\big) = I_{\boldsymbol{\alpha},\boldsymbol{\psi}}^{(j,\ell)} + o_P(1),$$

for all $j \in \{1, \ldots, p\}$ and $\ell \in \{1, \ldots, q\}$. The above equation combined with (A36) and Lemma 3 yields that, for all $j \in \{1, \ldots, q\}$,

$$\begin{aligned} S_{n,j} = &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_{\boldsymbol{\alpha},j}\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p, \widetilde{\boldsymbol{\psi}}\big) \\ &+ I_{\boldsymbol{\alpha},\boldsymbol{\psi}}^{(j,)}\, I_{\boldsymbol{\psi},\boldsymbol{\psi}}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{s}_{\boldsymbol{\psi}}\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \mathbf{0}_p; \widetilde{\boldsymbol{\psi}}_n\big) + o_P(1), \end{aligned} \tag{A37}$$

where $I_{\boldsymbol{\alpha},\boldsymbol{\psi}}^{(j,)}$ is the jth row of the matrix $I_{\boldsymbol{\alpha},\boldsymbol{\psi}}$. Hence, Theorem 1 (requiring Assumptions **R1**–**R4**) and (A37) imply

$$\begin{aligned} \mathbf{S}_n &= (\mathbb{I}_{p\times p}, -I_{\boldsymbol{\alpha}\boldsymbol{\psi}} I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{s}\big(\widetilde{U}_{1i}, \widetilde{U}_{2i}, \mathbf{X}_i; \widetilde{\boldsymbol{\psi}}\big) + o_P(1) \\ &\rightsquigarrow \mathcal{N}_p\big(\mathbf{0}_p, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} - I_{\boldsymbol{\alpha}\boldsymbol{\psi}}\, I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\, \boldsymbol{\Sigma}_{\boldsymbol{\psi}\boldsymbol{\alpha}} - \boldsymbol{\Sigma}_{\boldsymbol{\alpha}\boldsymbol{\psi}}\, I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\, I_{\boldsymbol{\psi}\boldsymbol{\alpha}} + I_{\boldsymbol{\alpha}\boldsymbol{\psi}}\, I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\, \boldsymbol{\Sigma}_{\boldsymbol{\psi}\boldsymbol{\psi}}\, I_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1}\, I_{\boldsymbol{\psi}\boldsymbol{\alpha}}\big), \end{aligned}$$

where $\mathbb{I}_{p\times p}$ is the identity matrix of the dimension (p, p). □

A5 - Proving consistency of the asymptotic variance estimator (11)**.** To proceed it is useful to generalize the notion of $\mathcal{J}_2$-functions introduced in Definition 2 so that it includes also the parameter $\boldsymbol{\psi}$ from a parameter space $\Theta_{\boldsymbol{\psi}}$.

Definition 3 ($\widetilde{\mathcal{J}}_2$-functions)**.** A function $\widetilde{J} : (0,1)^2 \times \mathbf{R_X} \times \Theta_{\boldsymbol{\psi}}$ is called a $\widetilde{\mathcal{J}}_2$-function if there exist functions $r_1, r_2, \tilde{r}_1, \tilde{r}_2 \in \mathcal{R}$, $v_1, v_2 \in \mathcal{V}$, and $M : \mathbf{R_X} \to \mathbb{R}$ such that

$$\sup_{\boldsymbol{\psi}\in\Theta_{\boldsymbol{\psi}}} \left|\widetilde{J}(u_1,u_2,\mathbf{x};\boldsymbol{\psi})\right| \leq M(\mathbf{x})\, r_1(u_1)\, r_2(u_2),$$

$$\sup_{\boldsymbol{\psi}\in\Theta_{\boldsymbol{\psi}}} \left|\widetilde{J}^{(1)}(u_1,u_2,\mathbf{x};\boldsymbol{\psi})\right| \leq M(\mathbf{x})\, \tilde{r}_1(u_1)\, r_2(u_2),\ \sup_{\boldsymbol{\psi}\in\Theta_{\boldsymbol{\psi}}} \left|\widetilde{J}^{(2)}(u_1,u_2,\mathbf{x};\boldsymbol{\psi})\right| \leq M(\mathbf{x})\, r_1(u_1)\tilde{r}_2(u_2),$$

where $\widetilde{J}^{(j)}(\cdot) = \partial J(\cdot)/\partial u_j$, $j \in \{1,2\}$, is continuous on $(0,1)^2 \times \mathbf{R_X} \times \Theta_{\boldsymbol{\psi}}$, and (A4) holds.

Theorem 3. *Let Assumptions **I1**, **I2**, and **R1**–**R4** hold. Suppose that there exists a neighborhood $\mathcal{U}$ of $\widetilde{\boldsymbol{\psi}}$ such that for each $k \in \{1,\dots,p+q\}$, the function $s_k(\cdot,\cdot,\cdot,\cdot) \in \widetilde{\mathcal{J}}_2$ (with $\mathcal{U}$ taken in place of $\Theta_{\boldsymbol{\psi}}$). Then under the null hypothesis $\widehat{\mathrm{avar}(\mathbf{S}_n)} \xrightarrow[n\to\infty]{\Pr} \mathbb{V}$.*

Proof. Note that from the proof of Lemma 3 it follows that $\widehat{I}_{\boldsymbol{\psi}\boldsymbol{\psi}} \xrightarrow[n\to\infty]{\Pr} I_{\boldsymbol{\psi}\boldsymbol{\psi}}$ and completely analogously one can show that $\widehat{I}_{\boldsymbol{\alpha}\boldsymbol{\psi}} \xrightarrow[n\to\infty]{\Pr} I_{\boldsymbol{\alpha}\boldsymbol{\psi}}$ and also that

$$\frac{1}{n}\sum_{i=1}^{n} s_k(\widetilde{U}_{1i},\widetilde{U}_{2i},\mathbf{X}_i;\widetilde{\boldsymbol{\psi}}_n)\, s_{k'}(\widetilde{U}_{1i},\widetilde{U}_{2i},\mathbf{X}_i;\widetilde{\boldsymbol{\psi}}_n) \xrightarrow[n\to\infty]{\Pr} \mathrm{E}\,\{s_k(U_1,U_2,\mathbf{X};\widetilde{\boldsymbol{\psi}})\, s_{k'}(U_1,U_2,\mathbf{X};\widetilde{\boldsymbol{\psi}})\}$$

for each $k, k' \in \{1,\dots,p+q\}$. Now the proof is finished provided that one shows that for each $k, k' \in \{1,\dots,p\}$

$$\frac{1}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{\vartheta}}^{(1)}_{\boldsymbol{\alpha},k}(\widetilde{U}_{1i})\widehat{\boldsymbol{\vartheta}}^{(1)}_{\boldsymbol{\alpha},k'}(\widetilde{U}_{1i}) = \mathrm{E}\,\{\boldsymbol{\vartheta}^{(1)}_{\boldsymbol{\alpha},k}(U_1)\,\boldsymbol{\vartheta}^{(1)}_{\boldsymbol{\alpha},k}(U_1)\} + o_P(1),$$

where

$$\boldsymbol{\vartheta}^{(j)}_{\boldsymbol{\alpha},k}(u) = \int \big[\mathbf{1}\{u \leq v_j\} - v_j\big]\, s^{(j)}_k(v_1,v_2,\mathbf{x};\widetilde{\boldsymbol{\psi}})\,\mathrm{d}C(v_1,v_2)\,\mathrm{d}F_{\mathbf{X}}(\mathbf{x}), \quad j \in \{1,2\}.$$

All the remaining terms can be handled analogously. Note that one can choose $\eta \in [0,1/2)$ such that uniformly in $u \in (0,1)$

$$\begin{aligned}\big|\widehat{\boldsymbol{\vartheta}}^{(j)}_{\boldsymbol{\alpha},k}(u)\big| &\leq \frac{1}{n}\sum_{j=1}^{n}\left|\frac{\mathbf{1}\{u \leq \widetilde{U}_{1j}\} - \widetilde{U}_{1j}}{\widetilde{U}^{\eta}_{1j}(1-\widetilde{U}_{1j})^{\eta}}\,\widetilde{U}^{\eta}_{1j}(1-\widetilde{U}_{1j})^{\eta}\, s^{(j)}_k(\widetilde{U}_{1j},\widetilde{U}_{2j},\mathbf{X}_j;\widetilde{\boldsymbol{\psi}})\right| \\ &\leq \frac{1}{u^{\eta}(1-u)^{\eta}}\frac{1}{n}\sum_{j=1}^{n} v_1(\widetilde{U}_{1j})\,\tilde{r}_1(\widetilde{U}_{1j})\,r_2(\widetilde{U}_{1j}) = \frac{1}{u^{\eta}(1-u)^{\eta}}\,O_P(1).\end{aligned}$$

This implies that (in probability) one has an integrable majorant $1/(u^{2\eta}(1-u)^{2\eta})$ for the set of the functions $\big\{g_n(u) = \widehat{\boldsymbol{\vartheta}}^{(j)}_{\boldsymbol{\alpha},k}(u)\widehat{\boldsymbol{\vartheta}}^{(j)}_{\boldsymbol{\alpha},k'}(u) : n \in \mathbb{N}\big\}$. Thus one can proceed analogously as in the proof of Lemma 3. □

N.B.: The references have been typeset according to JMVA standards.

References

[1] F. Abegaz, I. Gijbels, N. Veraverbeke, Semiparametric estimation of conditional copulas, J. Multivariate Anal. 110 (2012) 43–73.

[2] E.F. Acar, R.V. Craiu, F. Yao, Dependence calibration in conditional copulas: A nonparametric approach, Biometrics 67 (2011) 445–453.

[3] E.F. Acar, R.V. Craiu, F. Yao, Statistical testing of covariate effects in conditional copula models, Electron. J. Stat. 7 (2013) 2822–2850.

[4] E.F. Acar, C. Genest, J. Nešlehová, Beyond simplified pair-copula constructions, J. Multivariate Anal. 110 (2012) 74–90.

[5] 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.

[6] P.J. Bickel, C.A.J. Klaassen, Y. Ritov, J.A. Wellner, Efficient and Adaptive Estimation for Semiparametric Models, Johns Hopkins University Press, Baltimore, 1993.

[7] R.D. Cook, M.E. Johnson, A family of distributions for modelling non-elliptically symmetric multivariate data, J. R. Stat. Soc. Ser. B 43 (1981) 210–218.

[8] A.C. Davison, D.V. Hinkley, Bootstrap Methods and their Application, Cambridge University Press, New York, 1997.

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

[10] I. Gijbels, M. Omelka, N. Veraverbeke, Multivariate and functional covariates and conditional copulas, Electron. J. Stat. 6 (2012) 1273–1306.

[11] I. Gijbels, M. Omelka, N. Veraverbeke, Estimation of a copula when a covariate affects only marginal distributions, Scand. J. Statist. 42 (2015) 1109–1126.

[12] I. Gijbels, M. Omelka, N. Veraverbeke, Partial and average copulas and association measures, Electron. J. Stat. 9 (2015) 2420–2474.

[13] I. Gijbels, M. Omelka, N. Veraverbeke, Nonparametric testing for no covariate effects in conditional copulas, Statistics 51 (2017) 475–509.

[14] I. Gijbels, N. Veraverbeke, M. Omelka, Conditional copulas, association measures and their application, Comput. Stat. Data Anal. 55 (2011) 1919–1932.

[15] 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.

[16] I. Kojadinovic, J. Yan, Modeling multivariate distributions with continuous margins using the copula R package, J. Statist. Software 34 (2010) 1–20.

[17] R.B. Nelsen, An Introduction to Copulas, Second Edition, Springer, New York, 2006.

[18] J.A. Patton, Modeling asymmetric exchange rate dependence, Internat. Econom. Rev. 47 (2006) 527–556.

[19] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2016.

[20] G.R. Shorack, Functions of order statistics, Ann. Math. Statist. 43 (1972) 412–427.

[21] G.R. Shorack, J.A. Wellner, Empirical Processes with Applications to Statistics, Wiley, New York, 1986.

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

[23] J. Stöber, H. Joe, C. Czado, Simplified pair copula constructions: Limitations and extensions, J. Multivariate Anal. 119 (2013) 101–118.

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

[25] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, Springer, New York, 1996.

[26] N. Veraverbeke, M. Omelka, I. Gijbels, Estimation of a conditional copula and association measures, Scand. J. Statist. 38 (2011) 766–780.