Ann Inst Stat Math (2012) 64:811–834
DOI 10.1007/s10463-011-0337-6

Tests of symmetry for bivariate copulas

Christian Genest · Johanna Nešlehová · Jean-François Quessy

Received: 30 July 2010 / Revised: 10 March 2011 / Published online: 15 September 2011

Abstract Tests are proposed for the hypothesis that the underlying copula of a continuous random pair is symmetric. The procedures are based on Cramér–von Mises and Kolmogorov–Smirnov functionals of a rank-based empirical process whose large-sample behaviour is obtained. The asymptotic validity of a re-sampling method to compute P values is also established. The technical arguments supporting the use of a Chi-squared test due to Jasson are also presented. A power study suggests that the proposed tests are more powerful than Jasson's procedure under many scenarios of copula asymmetry. The methods are illustrated on a nutrient data set.

Keywords Empirical copula process · Exchangeability · Multiplier Central Limit Theorem · Ranks · Symmetry

1 Introduction

Consider a continuous random pair (X, Y) the joint distribution function of which is defined for all $x, y \in \mathbb{R}$ by

C. Genest (✉) · J. Nešlehová
Department of Mathematics and Statistics, McGill University, 805, rue Sherbrooke ouest, Montréal, Québec H3A 2K6, Canada
e-mail: cgenest@math.mcgill.ca

J. Nešlehová
e-mail: neslehova@math.mcgill.ca

J.-F. Quessy
Département de mathématiques et d'informatique, Université du Québec à Trois-Rivières, C.P. 500, Trois-Rivières, Québec G9A 5H7, Canada
e-mail: jean-francois.quessy@uqtr.ca

$$H(x, y) = \Pr(X \leqslant x, Y \leqslant y).$$

The variables X and Y are said to be exchangeable if the following hypothesis holds:

$$\mathscr{H}_0^\star : \quad \forall_{(x,y)\in\mathbb{R}^2} \quad H(x, y) = H(y, x).$$

The problem of testing this hypothesis is of general interest. Early contributors to the subject include Bell and Haller (1969), who addressed the issue within the bivariate Gaussian model, and Hollander (1971), who proposed a test that is consistent against a broad class of alternatives. Rank tests were also designed for restricted alternatives by Sen (1967), Yanagimoto and Sibuya (1976), and Snijders (1981). Hollander's test, which is based on the bivariate empirical distribution function, was further considered by Koziol (1979), Hilton and Gee (1997), and Hilton (2000).

Testing $\mathscr{H}_0^\star$ amounts to checking that the variables X and Y are identically distributed *and* that their dependence structure is symmetric. In other words, suppose that the marginal distribution functions of X and Y are defined for all $x, y \in \mathbb{R}$ by

$$F(x) = \Pr(X \leqslant x), \quad G(y) = \Pr(Y \leqslant y)$$

and let C be the unique copula corresponding to H, i.e., the joint distribution of the pair $(U, V) = (F(X), G(Y))$, which has uniform margins on $[0, 1]$. As implied, e.g., by the work of Sklar (1959), the identity

$$H(x, y) = C\{F(x), G(y)\} \tag{1}$$

holds for all $x, y \in \mathbb{R}$ and hence $\mathscr{H}_0^\star$ is verified if and only if

(i) $F(x) = G(x)$ for all $x \in \mathbb{R}$;
(ii) $C(u, v) = C(v, u)$ for all $(u, v) \in [0, 1]^2$.

Thus $\mathscr{H}_0^\star$ may be rejected either because (i) or (ii) fails, or both. While Condition (i) could be validated through standard graphical or formal statistical procedures, it is not immediately clear how to test for Condition (ii). Indeed, unless the marginal distributions F and G are known, data from C cannot be observed directly.

The purpose of this paper is to propose tests of the hypothesis

$$\mathscr{H}_0 : \quad \forall_{(u,v)\in[0,1]^2} \quad C(u, v) = C(v, u)$$

against the general alternative

$$\mathscr{H}_1 : \quad \exists_{(u,v)\in[0,1]^2} \quad C(u, v) \neq C(v, u).$$

This issue is of immediate relevance for the construction of copula models, in which H is assumed to be of the form (1) with F, G and C taken from parametric classes $(F_\alpha), (G_\beta)$, and (C_θ), respectively. In recent years, copula modelling has experienced rapid growth, e.g., in finance (Cherubini et al. 2004), risk management (McNeil et al. 2005), and hydrology (Salvadori et al. 2007).

This approach is particularly useful when the variables X and Y have different behaviour, i.e., when Condition (i) fails. Equation (1) yields a bivariate distribution with margins F and G (possibly involving covariates) for *any* choice of copula C.

At present, most bivariate copula families used in practice are symmetric, i.e., they satisfy Condition (ii). Such is the case, e.g., for all meta-elliptical and Archimedean copulas (Fang et al. 2002; Nelsen 2006). Before such dependence structures are fitted and used for prediction purposes, it would be wise to test the validity of hypothesis $\mathscr{H}_0$. This issue is formally addressed here for the first time; for an informal treatment involving an adaptation of the Chi-squared test, see Jasson (2005).

Three rank-based tests of the hypothesis $\mathscr{H}_0$ are proposed in Sect. 2, and their large-sample properties are discussed in Sect. 3. Although the procedures are not distribution-free under $\mathscr{H}_0$, it is shown in Sect. 4 that they can be implemented effectively using the Multiplier Central Limit Theorem (van der Vaart and Wellner 1996, Section 2.9). In Sect. 5, the behaviour of Jasson's statistic is examined. A power study comparing its merits relative to the current proposals is then reported in Sect. 6, and a small illustration is presented in Sect. 7. Technical arguments are grouped in the Appendix.

2 Description of the test statistics

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a bivariate distribution H with continuous margins F and G. Let C be the associated copula implicitly defined by (1). When F and G are known, one can construct a random sample from C by setting, for all $i \in \{1, \ldots, n\}$,

$$(U_i, V_i) = (F(X_i), G(Y_i)).$$

A consistent estimator of C is then defined for all $(u, v) \in [0, 1]^2$ by

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(U_i \leqslant u, V_i \leqslant v). \tag{2}$$

In contrast, data from C are not directly observable when F and G are unknown. However, a simple analogue of the pair (U_i, V_i) is then provided by

$$(\hat{U}_i, \hat{V}_i) = (F_n(X_i), G_n(Y_i)),$$

where F_n and G_n are the margins of the empirical analogue H_n of H. Note that $n\,\hat{U}_i$ is the rank of X_i among $X_1, \ldots, X_n$ and that similarly, $n\,\hat{V}_i$ is the rank of Y_i among $Y_1, \ldots, Y_n$.

As originally shown by Rüschendorf (1976) under weak regularity conditions, a consistent estimate of C is defined for all $(u, v) \in [0, 1]^2$ by

$$\hat{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(\hat{U}_i \leqslant u, \hat{V}_i \leqslant v). \tag{3}$$

Although it is not a copula per se (as its margins are discontinuous), it makes sense to call $\hat{C}_n$ the "empirical copula," given the analogy between (2) and (3). This definition differs slightly from the original one given by Deheuvels (1979), who set $\hat{C}_n^D(u, v) = H_n\{F_n^{\leftarrow}(u), G_n^{\leftarrow}(v)\}$ in terms of the generalized inverses $F_n^{\leftarrow}$ and $G_n^{\leftarrow}$ of F_n and G_n, respectively. As $|\hat{C}_n^D(u, v) - \hat{C}_n(u, v)| \leqslant 2/n$ for all $(u, v) \in [0, 1]^2$, the two estimators are asymptotically equivalent.

In addition to being consistent, the rank-based estimator (3) shares with C an invariance with respect to strictly increasing transformations of the variables X and Y. Furthermore, Genest and Segers (2010) showed that the asymptotic covariance of $\hat{C}_n$ is smaller than the asymptotic covariance of C_n under broad positive dependence conditions.

To test the hypothesis of exchangeability, it is natural to compare the values taken by $\hat{C}_n$ at (u, v) and (v, u) for all possible choices of $u, v \in [0, 1]$. Intuitively, $\hat{C}_n(u, v)$ and $\hat{C}_n^\top(u, v) = \hat{C}_n(v, u)$ should be close under $\mathscr{H}_0$. Three global measures of the discrepancy between $\hat{C}_n$ and $\hat{C}_n^\top$ are given by

$$\begin{aligned} R_n &= \int_0^1 \int_0^1 \{\hat{C}_n(u, v) - \hat{C}_n(v, u)\}^2 \, \mathrm{d}v \, \mathrm{d}u \\ S_n &= \int_0^1 \int_0^1 \{\hat{C}_n(u, v) - \hat{C}_n(v, u)\}^2 \, \mathrm{d}\hat{C}_n(u, v), \\ T_n &= \sup_{(u,v)\in[0,1]^2} |\hat{C}_n(u, v) - \hat{C}_n(v, u)|. \end{aligned} \tag{4}$$

In particular, S_n is a rank-based analogue of a Cramér–von Mises type statistic due to Hollander (1971), viz.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{H_n(x, y) - H_n(y, x)\}^2 \, \mathrm{d}H_n(x, y).$$

The following result, proved in the Appendix, gives alternative expressions for R_n, S_n, and T_n that are convenient for computation purposes.

Proposition 1 *The statistics defined in* (4) *may be expressed as*

$$\begin{aligned} R_n &= \frac{1}{n^2} \mathbf{1}^\top \mathbf{A} \mathbf{1}, \quad S_n = \frac{1}{n^3} \sum_{k=1}^{n} \mathbf{1}^\top \mathbf{B}_k \mathbf{1}, \\ T_n &= \max_{i,j\in\{1,\ldots,n\}} \left| \hat{C}_n\left(\frac{i}{n}, \frac{j}{n}\right) - \hat{C}_n\left(\frac{j}{n}, \frac{i}{n}\right) \right|, \end{aligned}$$

where $\mathbf{1}$ *is an* $n \times 1$ *vector of* 1*'s and* $\mathbf{A}, \mathbf{B}_1, \ldots, \mathbf{B}_n$ *are* $n \times n$ *matrices with entry at position* (i, j) *given by*

$$\begin{aligned} A_{ij} &= 2(1 - \hat{U}_i \vee \hat{U}_j)(1 - \hat{V}_i \vee \hat{V}_j) - 2(1 - \hat{U}_i \vee \hat{V}_j)(1 - \hat{U}_j \vee \hat{V}_i), \\ B_{kij} &= \mathbb{I}(\hat{U}_i \vee \hat{U}_j \leqslant \hat{U}_k, \hat{V}_i \vee \hat{V}_j \leqslant \hat{V}_k) - \mathbb{I}(\hat{U}_i \vee \hat{V}_j \leqslant \hat{U}_k, \hat{V}_i \vee \hat{U}_j \leqslant \hat{V}_k) \\ &\quad - \mathbb{I}(\hat{U}_i \vee \hat{V}_j \leqslant \hat{V}_k, \hat{V}_i \vee \hat{U}_j \leqslant \hat{U}_k) + \mathbb{I}(\hat{U}_i \vee \hat{U}_j \leqslant \hat{V}_k, \hat{V}_i \vee \hat{V}_j \leqslant \hat{U}_k), \end{aligned}$$

where for arbitrary $a, b \in \mathbb{R}$, $a \vee b = \max(a, b)$.

3 Asymptotic behaviour

Weak limits of the test statistics given in (4) can be derived from the asymptotic behaviour of the empirical copula process, defined for all $(u, v) \in [0, 1]^2$ by

$$\hat{\mathbb{C}}_n(u, v) = n^{1/2}\{\hat{C}_n(u, v) - C(u, v)\}.$$

This process may be viewed as a random element of the space $\ell^\infty[0, 1]^2$ of bounded functions $f : [0, 1]^2 \to \mathbb{R}$ equipped with the uniform norm. From the work of Rüschendorf (1976) and Segers (2012), the sequence $(\hat{\mathbb{C}}_n)$ admits a weak limit $\hat{\mathbb{C}}$, denoted $\hat{\mathbb{C}}_n \rightsquigarrow \hat{\mathbb{C}}$, whenever C is regular in the following sense:

Definition 1 A bivariate copula C is said to be regular if

(i) the partial derivatives $\dot{C}_1(u, v) = \partial C(u, v)/\partial u$ and $\dot{C}_2(u, v) = \partial C(u, v)/\partial v$ exist everywhere on $[0, 1]^2$, where by convention, one-sided derivatives are used at the boundary points;

(ii) $\dot{C}_1$ is continuous on $(0, 1) \times [0, 1]$ and $\dot{C}_2$ is continuous on $[0, 1] \times (0, 1)$.

The limit $\hat{\mathbb{C}}$ is a centred Gaussian process defined for all $(u, v) \in [0, 1]^2$ by

$$\hat{\mathbb{C}}(u, v) = \mathbb{C}(u, v) - \dot{C}_1(u, v)\mathbb{C}(u, 1) - \dot{C}_2(u, v)\mathbb{C}(1, v)$$

in terms of a tucked C-Brownian sheet $\mathbb{C}$, i.e., a centred Gaussian random field whose covariance function is given for all $u, v, s, t \in [0, 1]$ by

$$\Gamma_{\mathbb{C}}(u, v, s, t) = C(u \wedge s, v \wedge t) - C(u, v)\, C(s, t),$$

where for all $a, b \in \mathbb{R}$, $a \wedge b = \min(a, b)$. For variants on Rüschendorf's result, see Gänßler and Stute (1987), Fermanian et al. (2004) or Tsukahara (2005).

Now consider the symmetrised empirical process $\hat{\mathbb{D}}_n = n^{1/2}(\hat{C}_n - \hat{C}_n^\top)$, explicitly defined for all $(u, v) \in [0, 1]^2$ by

$$\hat{\mathbb{D}}_n(u, v) = n^{1/2}\{\hat{C}_n(u, v) - \hat{C}_n(v, u)\}.$$

Under $\mathscr{H}_0$, the limit of this process can be deduced easily from the asymptotic behaviour of $\hat{\mathbb{C}}_n$, because one then has $\hat{\mathbb{D}}_n(u, v) = \hat{\mathbb{C}}_n(u, v) - \hat{\mathbb{C}}_n(v, u)$ for all $(u, v) \in [0, 1]^2$. The result is stated below; see the Appendix for a proof.

Proposition 2 *If C is a regular symmetric copula, then $\hat{\mathbb{D}}_n$ converges weakly, as $n \to \infty$, to a Gaussian random field $\hat{\mathbb{D}}$ defined for all $(u, v) \in [0, 1]^2$ by*

$$\hat{\mathbb{D}}(u, v) = \mathbb{D}(u, v) - \dot{C}_1(u, v)\,\mathbb{D}(u, 1) - \dot{C}_2(u, v)\,\mathbb{D}(1, v),$$

in terms of a centred Gaussian random field $\mathbb{D}$ with covariance function given at each $u, v, s, t \in [0, 1]$ by $\Gamma_{\mathbb{D}}(u, v, s, t) = 2\,\{\Gamma_{\mathbb{C}}(u, v, s, t) - \Gamma_{\mathbb{C}}(u, v, t, s)\}$.

Note that the covariance function $\Gamma_{\hat{\mathbb{D}}}$ of $\hat{\mathbb{D}}$ can be expressed, for all $u, v, s, t \in [0, 1]$, in terms of the covariance function $\Gamma_{\hat{\mathbb{C}}}$ of the process $\hat{\mathbb{C}}$, viz.

$$\Gamma_{\hat{\mathbb{D}}}(u, v, s, t) = 2\{\Gamma_{\hat{\mathbb{C}}}(u, v, s, t) - \Gamma_{\hat{\mathbb{C}}}(u, v, t, s)\}.$$

Closed form expressions for $\Gamma_{\hat{\mathbb{D}}}$ are rare and intricate. One exception is when $C(u, v) = uv$ for all $(u, v) \in [0, 1]^2$, i.e., the independence copula. In that case,

$$\Gamma_{\hat{\mathbb{C}}}(u, v, s, t) = (u \wedge s - us)\,(v \wedge t - vt)$$

for all $u, v, s, t \in [0, 1]$, and hence

$$\Gamma_{\hat{\mathbb{D}}}(u, v, s, t) = 2\,(u \wedge s - us)\,(v \wedge t - vt) - 2\,(u \wedge t - ut)\,(v \wedge s - vs)\,.$$

The asymptotic behaviour of the statistics R_n, S_n and T_n under $\mathscr{H}_0$ can be deduced from Proposition 2, in combination with the Continuous Mapping Theorem and the Functional Delta Method; see the Appendix for details.

Proposition 3 *If C is a regular symmetric copula, then as $n \to \infty$,*

$$nR_n = \int_0^1 \int_0^1 \{\hat{\mathbb{D}}_n(u, v)\}^2\,\mathrm{d}v\,\mathrm{d}u \rightsquigarrow \mathbb{D}_R = \int_0^1 \int_0^1 \{\hat{\mathbb{D}}(u, v)\}^2\,\mathrm{d}v\,\mathrm{d}u,$$
$$nS_n = \int_0^1 \int_0^1 \{\hat{\mathbb{D}}_n(u, v)\}^2\,\mathrm{d}\hat{C}_n(u, v) \rightsquigarrow \mathbb{D}_S = \int_0^1 \int_0^1 \{\hat{\mathbb{D}}(u, v)\}^2\,\mathrm{d}C(u, v),$$
$$n^{1/2}\,T_n = \sup_{(u,v)\in[0,1]^2} |\hat{\mathbb{D}}_n(u, v)| \rightsquigarrow \mathbb{D}_T = \sup_{(u,v)\in[0,1]^2} |\hat{\mathbb{D}}(u, v)|.$$

If C is regular, it also follows from the Continuous Mapping Theorem that the statistics R_n and T_n converge in probability to the measures of asymmetry proposed by Nelsen (2007). With some additional effort, the limit of S_n can also be identified. This is stated formally below and proved in the Appendix.

Proposition 4 *If C is a regular copula, then as $n \to \infty$,*

$$R_n \xrightarrow{p} R(C) = \int_0^1 \int_0^1 \{C(u,v) - C(v,u)\}^2 \, \mathrm{d}v \, \mathrm{d}u,$$
$$S_n \xrightarrow{p} S(C) = \int_0^1 \int_0^1 \{C(u,v) - C(v,u)\}^2 \, \mathrm{d}C(u,v), \tag{5}$$
$$T_n \xrightarrow{p} T(C) = \sup_{(u,v)\in[0,1]^2} |C(u,v) - C(v,u)|.$$

4 Procedures based on the Multiplier Central Limit Theorem

As seen in Sect. 3, the null distributions of the test statistics nR_n, nS_n, and $n^{1/2}T_n$ depend on the underlying form of the copula, which is generally unknown. It is thus impossible to compute valid P values from standard Monte Carlo simulations. The Multiplier Central Limit Theorem (van der Vaart and Wellner 1996, Section 2.9) provides a solution to this problem. The resulting bootstrap approximation of the empirical copula process was successfully applied in other testing situations involving copulas, e.g., by Scaillet (2005), Rémillard and Scaillet (2009), Kojadinovic and Yan (2011), and Kojadinovic et al. (2011).

Following Bücher and Dette (2010), fix $M \in \mathbb{N}$ and for each $h \in \{1, \dots, M\}$, let $\xi^{(h)} = (\xi_1^{(h)}, \dots, \xi_n^{(h)})$ be a vector of independent non-negative random variables with unit mean and unit variance. These variables should also be completely independent from the data. Further write

$$\bar{\xi}_n^{(h)} = \frac{1}{n}\left(\xi_1^{(h)} + \cdots + \xi_n^{(h)}\right)$$

and

$$\Xi_n^{(h)} = \left(\frac{\xi_1^{(h)}}{\bar{\xi}_n^{(h)}} - 1, \dots, \frac{\xi_n^{(h)}}{\bar{\xi}_n^{(h)}} - 1\right).$$

Given $(u, v) \in [0, 1]^2$, let $P_n(u, v)$ be an $n \times 1$ vector with ith component

$$P_{in}(u,v) = \mathbb{I}(\hat{U}_i \leqslant u, \hat{V}_i \leqslant v) - \mathbb{I}(\hat{U}_i \leqslant v, \hat{V}_i \leqslant u). \tag{6}$$

For each $h \in \{1, \dots, M\}$, a weighted bootstrap version $\mathbb{D}_n^{(h)}$ of $\mathbb{D}$ may then be defined by setting, for all $(u, v) \in [0, 1]^2$,

$$\mathbb{D}_n^{(h)}(u,v) = n^{-1/2} \Xi_n^{(h)} P_n(u,v). \tag{7}$$

From the asymptotic representation of $\hat{\mathbb{D}}$ under $\mathscr{H}_0$ stated in Proposition 2, a bootstrap replicate of this process is then given, for all $(u, v) \in [0, 1]^2$, by

$$\begin{aligned}\hat{\mathbb{D}}_n^{(h)}(u,v) &= \mathbb{D}_n^{(h)}(u,v) - \dot{C}_{1n}(u,v)\mathbb{D}_n^{(h)}(u,1) - \dot{C}_{2n}(u,v)\mathbb{D}_n^{(h)}(1,v)\\ &= n^{-1/2}\Xi_n^{(h)}\{P_n(u,v) - \dot{C}_{1n}(u,v)P_n(u,1) - \dot{C}_{2n}(u,v)P_n(1,v)\}.\end{aligned}$$

Here, $\dot{C}_{1n}$ and $\dot{C}_{2n}$ are the estimates of the partial derivatives $\dot{C}_1$ and $\dot{C}_2$ defined by Segers (2012). Specifically, let $\ell_n \in (0, 1/2)$ be a bandwidth parameter and for arbitrary $v \in [0, 1]$, set

$$\dot{C}_{1n}(u,v) = \begin{cases} \dfrac{\hat{C}_n(2\ell_n, v)}{2\ell_n} & \text{if } u \in [0, \ell_n),\\ \dfrac{\hat{C}_n(u+\ell_n, v) - \hat{C}_n(u-\ell_n, v)}{2\ell_n} & \text{if } u \in [\ell_n, 1-\ell_n],\\ \dfrac{\hat{C}_n(1, v) - \hat{C}_n(1-2\ell_n, v)}{2\ell_n} & \text{if } u \in (1-\ell_n, 1].\end{cases}$$

Similarly, for arbitrary $u \in [0, 1]$, set

$$\dot{C}_{2n}(u,v) = \begin{cases} \dfrac{\hat{C}_n(u, 2\ell_n)}{2\ell_n} & \text{if } v \in [0, \ell_n),\\ \dfrac{\hat{C}_n(u, v+\ell_n) - \hat{C}_n(u, v-\ell_n)}{2\ell_n} & \text{if } v \in [\ell_n, 1-\ell_n],\\ \dfrac{\hat{C}_n(u, 1) - \hat{C}_n(u, 1-2\ell_n)}{2\ell_n} & \text{if } v \in (1-\ell_n, 1].\end{cases}$$

The following proposition implies that asymptotically, $\hat{\mathbb{D}}_n^{(1)}, \ldots, \hat{\mathbb{D}}_n^{(M)}$ are independent copies of $\hat{\mathbb{D}}$ under $\mathscr{H}_0$:

Proposition 5 *Let C be a regular symmetric copula. Suppose that*

$$\lim_{n\to\infty} \ell_n = 0, \quad \inf_{n\in\mathbb{N}} n^{1/2}\ell_n > 0.$$

Then for all $M \in \mathbb{N}$, $(\hat{\mathbb{D}}_n, \hat{\mathbb{D}}_n^{(1)}, \ldots, \hat{\mathbb{D}}_n^{(M)}) \rightsquigarrow (\hat{\mathbb{D}}, \hat{\mathbb{D}}^{(1)}, \ldots, \hat{\mathbb{D}}^{(M)})$ *as* $n \to \infty$*, where the processes* $\hat{\mathbb{D}}^{(1)}, \ldots, \hat{\mathbb{D}}^{(M)}$ *are independent copies of* $\hat{\mathbb{D}}$*.*

Bootstrap replicates of the three test statistics can now be defined, for each $n \in \mathbb{N}$ and $h \in \{1, \ldots, M\}$, by

$$\begin{aligned} n\hat{R}_n^{(h)} &= \int_0^1\int_0^1 \{\hat{\mathbb{D}}_n^{(h)}(u,v)\}^2\,\mathrm{d}v\,\mathrm{d}u,\\ n\hat{S}_n^{(h)} &= \int_0^1\int_0^1 \{\hat{\mathbb{D}}_n^{(h)}(u,v)\}^2\,\mathrm{d}\hat{C}_n(u,v),\\ n^{1/2}\,\hat{T}_n^{(h)} &= \sup_{(u,v)\in[0,1]^2} |\hat{\mathbb{D}}_n^{(h)}(u,v)|.\end{aligned}$$

The following result is an immediate consequence of Proposition 5:

Corollary 1 *Suppose that the conditions of Proposition 5 hold. Then for all $M \in \mathbb{R}$ and as $n \to \infty$,*

$$(n\hat{R}_n, n\hat{R}_n^{(1)}, \ldots, n\hat{R}_n^{(M)}) \rightsquigarrow (\mathbb{D}_R, \mathbb{D}_R^{(1)}, \ldots, \mathbb{D}_R^{(M)}),$$
$$(n\hat{S}_n, n\hat{S}_n^{(1)}, \ldots, n\hat{S}_n^{(M)}) \rightsquigarrow (\mathbb{D}_S, \mathbb{D}_S^{(1)}, \ldots, \mathbb{D}_S^{(M)}),$$
$$(n^{1/2}\,\hat{T}_n, n^{1/2}\,\hat{T}_n^{(1)}, \ldots, n^{1/2}\,\hat{T}_n^{(M)}) \rightsquigarrow (\mathbb{D}_T, \mathbb{D}_T^{(1)}, \ldots, \mathbb{D}_T^{(M)}),$$

where each limit consists of independent, identically distributed processes.

It follows from Corollary 1 that approximate P values for the tests of $\mathscr{H}_0$ based on R_n, S_n, and T_n are given, respectively, by

$$\frac{1}{M}\sum_{h=1}^{M}\mathbb{I}(\hat{R}_n^{(h)} > R_n), \quad \frac{1}{M}\sum_{h=1}^{M}\mathbb{I}(\hat{S}_n^{(h)} > S_n), \quad \frac{1}{M}\sum_{h=1}^{M}\mathbb{I}(\hat{T}_n^{(h)} > T_n).$$

For convenience, let $\hat{\mathbb{D}}_n^{(h)} = n^{-1/2}\,\Xi_n^{(h)} Q_n$ for each $h \in \{1, \ldots, M\}$, where for all $(u, v) \in [0, 1]^2$,

$$Q_n(u, v) = P_n(u, v) - \dot{C}_{1n}(u, v)P_n(u, 1) - \dot{C}_{2n}(u, v)P_n(1, v)$$

depends only on the pseudo-observations retrieved from the data. While

$$\hat{S}_n^{(h)} = \frac{1}{n^3}\sum_{i=1}^{n}\{\Xi_n^{(h)} Q_n(\hat{U}_i, \hat{V}_i)\}^2,$$

explicit expressions for the bootstrap replicates $\hat{R}_n^{(h)}$ and $\hat{T}_n^{(h)}$ are difficult to obtain. The following approximations can be used in practice:

$$\begin{aligned}
\hat{R}_n^{(h)} &\approx \frac{1}{nN^2}\sum_{k=1}^{N}\sum_{\ell=1}^{N}\left\{\hat{\mathbb{D}}_n^{(h)}\left(\frac{k}{N}, \frac{\ell}{N}\right)\right\}^2 \\
&= \frac{1}{n^2N^2}\sum_{k=1}^{N}\sum_{\ell=1}^{N}\left\{\Xi_n^{(h)} Q_n\left(\frac{k}{N}, \frac{\ell}{N}\right)\right\}^2, \\
\hat{T}_n^{(h)} &\approx \frac{1}{n^{1/2}}\max_{k,\ell\in\{1,\ldots,N\}}\left|\hat{\mathbb{D}}_n^{(h)}\left(\frac{k}{N}, \frac{\ell}{N}\right)\right| \\
&= \frac{1}{n}\max_{k,\ell\in\{1,\ldots,N\}}\left|\Xi_n^{(h)} Q_n\left(\frac{k}{N}, \frac{\ell}{N}\right)\right|.
\end{aligned} \tag{8}$$

5 Jasson's test

To test whether a copula is symmetric, Jasson (2005) suggested that the set $[0, 1]^2$ be partitioned into squares of width $1/L$ for some integer $L > 2$ and that a contingency table be derived from the pseudo-observations $(\hat{U}_1, \hat{V}_1), \ldots, (\hat{U}_n, \hat{V}_n)$ by counting how many of them fall in each of these squares. If $\mathscr{H}_0$ is true, the counts in cells (i, j) and (j, i) should be roughly the same. The traditional Chi-square statistic could thus be used to check whether the observed differences are within sampling error or not.

To be more specific, consider a general mapping $f : [0, 1]^2 \to \mathbb{R}$ and for arbitrary $k, \ell \in \{1, \ldots, L\}$, introduce the notation

$$p^L_{k\ell}(f) = f\left(\frac{k}{L}, \frac{\ell}{L}\right) - f\left(\frac{k-1}{L}, \frac{\ell}{L}\right) - f\left(\frac{k}{L}, \frac{\ell-1}{L}\right) + f\left(\frac{k-1}{L}, \frac{\ell-1}{L}\right).$$

If a random pair (U, V) is distributed according to copula C, then

$$p^L_{k\ell}(C) = \Pr\left\{(U, V) \in \left(\frac{k-1}{L}, \frac{k}{L}\right] \times \left(\frac{\ell-1}{L}, \frac{\ell}{L}\right]\right\},$$

and hence $p^L_{k\ell}(C) = p^L_{\ell k}(C)$ for all $k, \ell \in \{1, \ldots, L\}$ whenever C is symmetric.

Following Jasson (2005), a "local" test of $\mathscr{H}_0$ could be based on

$$W^L_{n,(k,\ell)} = p^L_{k\ell}(\hat{C}_n) - p^L_{\ell k}(\hat{C}_n),$$

i.e., the difference in the proportion of counts observed in cells (k, ℓ) and (ℓ, k) for fixed $k, \ell \in \{1, \ldots, L\}$. Clearly, one has $W^L_{n,(k,\ell)} = -W^L_{n,(\ell,k)}$ and hence in particular $W^L_{n,(k,k)} \equiv 0$ for all $k, \ell \in \{1, \ldots, L\}$. Therefore, it suffices to restrict attention to the pairs (k, ℓ) with $1 \leqslant k < \ell \leqslant L$. In his paper, Jasson considers the ratio

$$Z_{n,(k,\ell)} = n^{1/2} \frac{W^L_{n,(k,\ell)}}{\{p^L_{k\ell}(\hat{C}_n) + p^L_{\ell k}(\hat{C}_n)\}^{1/2}},$$

and suggests as a "global" test statistic

$$Z_n = \sum_{k<\ell} Z^2_{n,(k,\ell)}.$$

A basic problem with this approach, however, is that the asymptotic distribution of Z_n is not necessarily Chi-square with $\nu = (L-1)(L-2)/2$ degrees of freedom under $\mathscr{H}_0$, as claimed by Jasson (2005). To see why, first observe that $n^{1/2}\, W^L_{n,(k,\ell)} = p^L_{k\ell}(\hat{\mathbb{D}}_n)$, i.e.,

$$n^{1/2}\, W^L_{n,(k,\ell)} = \hat{\mathbb{D}}_n\left(\frac{k}{L},\frac{\ell}{L}\right) - \hat{\mathbb{D}}_n\left(\frac{k-1}{L},\frac{\ell}{L}\right) - \hat{\mathbb{D}}_n\left(\frac{k}{L},\frac{\ell-1}{L}\right) + \hat{\mathbb{D}}_n\left(\frac{k-1}{L},\frac{\ell-1}{L}\right)$$

for all $k, \ell \in \{1, \ldots, L\}$. Thus if $\mathscr{H}_0$ is true, one finds that as $n \to \infty$,

$$n^{1/2}\, \mathbf{W}^L_n = n^{1/2}\, (W^L_{n,(1,2)}, \ldots, W^L_{n,(L-1,L)})^\top \rightsquigarrow \mathbf{W}^L,$$

where $\mathbf{W}^L$ is a centred Gaussian vector. Observe, however, that its covariance matrix Σ_L depends on the unknown value of the underlying copula C at points $(k/L, \ell/L)$ for $k, \ell \in \{1, \ldots, L\}$. Clearly, therefore, the limiting distributions of the random variables $Z_{n,(k,\ell)}$ depend on C and are not necessarily independent of one another. As a result, it is far from obvious that the limiting behaviour of Z_n could be distribution-free.

To circumvent this problem, one can again call on the bootstrap replicates of $\hat{\mathbb{D}}$. For all $h \in \{1, \ldots, M\}$ and $k, \ell \in \{1, \ldots, L\}$ with $k < \ell$, set $n^{1/2}\, W^{L(h)}_{n,(k,\ell)} = p^L_{k\ell}(\hat{\mathbb{D}}^{(h)}_n)$, i.e.,

$$n^{1/2}\, W^{L(h)}_{n,(k,\ell)} = \hat{\mathbb{D}}^{(h)}_n\left(\frac{k}{L},\frac{\ell}{L}\right) - \hat{\mathbb{D}}^{(h)}_n\left(\frac{k-1}{L},\frac{\ell}{L}\right) - \hat{\mathbb{D}}^{(h)}_n\left(\frac{k}{L},\frac{\ell-1}{L}\right) + \hat{\mathbb{D}}^{(h)}_n\left(\frac{k-1}{L},\frac{\ell-1}{L}\right)$$

and for each $h \in \{1, \ldots, M\}$, write

$$n^{1/2}\, \mathbf{W}^{L(h)}_n = (n^{1/2}\, W^{L(h)}_{n,(1,2)}, \ldots, n^{1/2}\, W^{L(h)}_{n,(L-1,L)})^\top.$$

It then follows from Proposition 5 and the Continuous Mapping Theorem that under $\mathscr{H}_0$, the vectors $n^{1/2}\, \mathbf{W}^{L(1)}_n, \ldots, n^{1/2}\, \mathbf{W}^{L(M)}_n$ are asymptotically independent copies of $\mathbf{W}^L$. Accordingly, the empirical covariance matrix based on $\mathbf{W}^{L(1)}_n, \ldots, \mathbf{W}^{L(M)}_n$ provides a consistent estimate $\hat{\Sigma}_L$ of Σ_L. The proposed test statistic is then

$$J^L_n = (\mathbf{W}^L_n)^\top \hat{\Sigma}^{-1}_L \mathbf{W}^L_n, \tag{9}$$

and its asymptotic distribution is $\chi^2_{(\nu)}$ with $\nu = \mathrm{rank}(\Sigma_L)$ degrees of freedom. From numerical experimentation it seems that Σ_L is of full rank for many classical copula models. It may thus be conjectured that $\nu = (L-1)(L-2)/2$.

6 Finite-sample performance

A Monte Carlo experiment was designed to study the finite-sample performance of the tests of $\mathscr{H}_0$ based on the statistics R_n, S_n, T_n defined in (4). For comparison purposes, the statistic J^L_n given in (9) was also included; values of $L \in \{3, 4, 5, 6\}$ serve to show the effect of increasingly finer partitions of $[0, 1]^2$. All tests were carried out at the 5% nominal level.

Table 1 Level of the tests of $\mathscr{H}_0$ based on R_n, S_n, T_n, and J_n^L with $L \in \{3, 4, 5, 6\}$, as estimated from 1,000 replicates from four symmetric copulas using $M = 250$ bootstrap replicates and $N = 50$ in approximation (8): independence (IN), Clayton (CL), Gaussian (GA), and Gumbel–Hougaard (GH)

Copula	τ	R_n	S_n	T_n	J_n^3	J_n^4	J_n^5	J_n^6
$n = 100$								
IN	0	2.4	3.0	3.0	4.3	5.7	4.6	9.6
	1/4	2.0	3.2	4.5	5.8	4.9	6.3	9.3
CL	1/2	1.5	2.0	6.1	3.7	3.5	4.9	6.1
	3/4	0.5	2.0	5.7	0.0	5.5	9.9	9.4
	1/4	2.0	3.1	3.9	3.6	4.6	5.7	7.3
GA	1/2	1.9	1.7	5.1	3.4	3.7	3.9	4.9
	3/4	0.1	1.7	4.9	0.0	3.9	8.1	10.8
	1/4	1.9	3.1	4.2	4.8	4.9	5.1	8.2
GH	1/2	1.2	1.9	4.6	4.3	3.5	4.1	5.5
	3/4	0.7	2.4	5.1	0.0	5.3	3.4	5.8
$n = 250$								
IN	0	3.9	3.7	2.7	3.8	4.9	4.6	6.8
	1/4	3.4	3.7	3.3	4.5	4.8	4.7	6.0
CL	1/2	2.1	3.4	4.0	5.1	2.7	3.3	4.1
	3/4	1.0	2.7	4.4	0.1	4.2	7.0	9.9
	1/4	4.1	4.2	4.0	4.0	3.1	4.0	6.1
GA	1/2	2.0	3.5	4.5	3.8	2.8	3.3	3.7
	3/4	0.9	2.2	4.8	0.0	1.5	4.7	6.4
	1/4	2.7	3.2	3.5	4.0	4.5	5.3	6.3
GH	1/2	3.2	3.3	4.7	4.6	3.7	2.7	4.7
	3/4	0.6	2.3	3.9	0.2	0.3	3.0	3.1

P values were computed on the basis of $M = 250$ bootstrap replicates and $N = n/5$ was used in approximations (8). The variables $\xi_1^{(h)}, \ldots, \xi_n^{(h)}$ were taken to be independent exponential random variables with unit mean. The resulting scheme is sometimes referred to as the Bayesian bootstrap (van der Vaart and Wellner 1996, Example 3.6.9). Other choices of distribution did not affect the conclusions presented below.

Table 1 shows the empirical level of the tests based on 1,000 random samples of size $n = 100$ and 250 from the following copula models:

(i) the Clayton (CL) copula is given for all $(u, v) \in (0, 1]^2$ by

$$C_\theta^{\mathrm{CL}}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad \theta \in (0, \infty);$$

(ii) the Gaussian (GA) copula is given for all $(u, v) \in (0, 1]^2$ by

$$C_r^{\mathrm{GA}}(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1 - r^2)^{1/2}} \exp\left\{-\frac{s^2 + t^2 - 2rst}{2(1 - r^2)}\right\} \mathrm{d}t\, \mathrm{d}s,$$

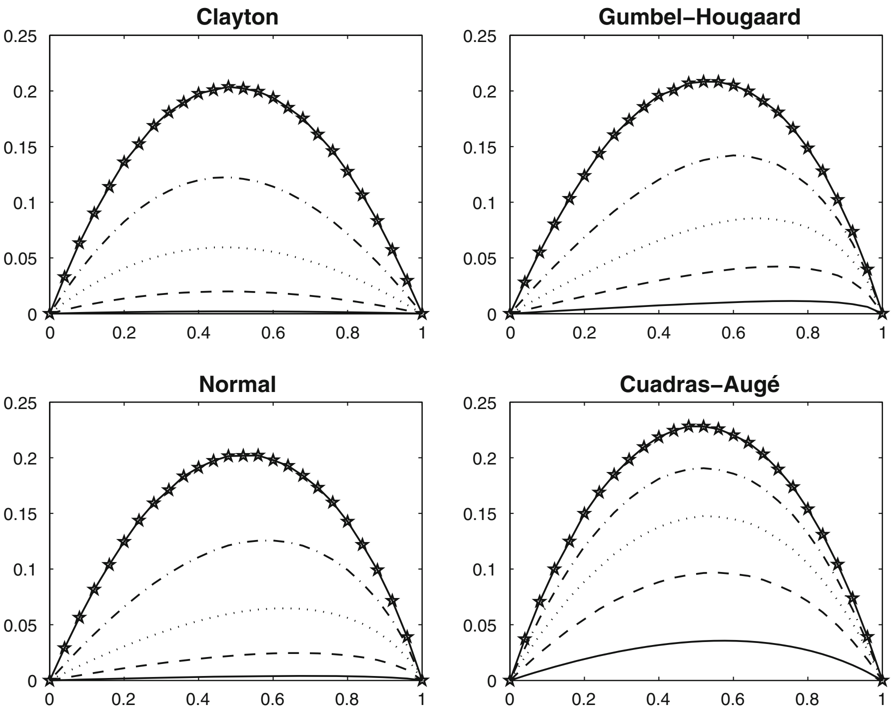

Fig. 1 Value of $3 \times T(C)$ as a function of δ for copulas arising from Khoudraji's device. *Continuous line* $\tau = 0.1$; *dashed line* $\tau = 0.3$; *dotted line* $\tau = 0.5$; *dashed dotted line* $\tau = 0.7$; *stared line* $\tau = 0.9$

where $r \in (-1, 1)$ and Φ denotes the distribution function of a Gaussian random variable with zero mean and unit variance;

(iii) the Gumbel–Hougaard (GH) copula is given for all $(u, v) \in (0, 1]^2$ by

$$C_\theta^{\text{GH}}(u, v) = \exp\left\{ -(|\ln u|^\theta + |\ln v|^\theta)^{1/\theta}\right\}, \quad \theta \in (1, \infty).$$

To assess the effect of the degree of dependence, three values of Kendall's tau were used: 1/4, 1/2, and 3/4. All tests except those based on T_n and J_n^4 appear to be either too liberal or conservative when $n = 100$. There are general signs of improvement when $n = 250$, although important discrepancies remain for some tests, especially when $\tau = 3/4$. This is likely due to the fact that pairs of normalized ranks are sparse in the vicinity of $(0, 1)$ and $(1, 0)$, resulting in many cells with low counts, among others.

To study the power of the tests, the three copula models given above were made asymmetric by Khoudraji's device (Khoudraji 1995; Genest et al. 1998; Liebscher 2008). Specifically, an asymmetric version of a copula C was defined at all $(u, v) \in [0, 1]^2$ by

$$K_\delta(u, v) = u^\delta C(u^{1-\delta}, v) \tag{10}$$

for various choices of $\delta \in (0,1)$. The symbols K_δ^{CL}, K_δ^{GA}, and K_δ^{GH} denote the resulting asymmetric versions of the Clayton, Gaussian, and Gumbel–Hougaard copulas, respectively. Further, the asymmetric versions K_δ^{CA} of the Cuadras–Augé (CA) copulas C_θ^{CA} were considered, where for all $\theta, u, v \in [0,1]$,

$$C_\theta^{\mathrm{CA}}(u,v) = \min(u^{1-\theta}v, vu^{1-\theta}).$$

Note that each K_δ^{CA} belongs to the Marshall–Olkin copula family.

Figure 1 shows how Nelsen's asymmetry index $T(C)$ defined in (5) varies as a function of δ for various choices of Kendall's τ in families K_δ^{CA}, K_δ^{CL}, K_δ^{GA}, and K_δ^{GH}. As can be seen, Khoudraji's device (10) provides little asymmetry for small and moderate values of τ (say $\tau \leqslant 1/2$). In all cases, maximum asymmetry occurs at (or near) $\delta = 1/2$. Nevertheless, Table 2 shows that the tests generally achieve reasonable power against these various alternatives, even when $n = 100$. The results for $n = 250$, presented in Table 3, are more encouraging still, although not uniformly good. More specifically

(i) the test based on the Cramér–von Mises statistic S_n is almost systematically more powerful than its competitors;
(ii) in accordance with Fig. 1, the power of the tests tends to increase with τ and is generally highest at $\delta = 1/2$ (the only exception occurs for the test based on J_n^3 when $\delta = 1/4$);
(iii) also in accordance with Fig. 1, asymmetry is particularly difficult to detect for K_δ^{CL} when $\tau = 1/2$, even for $\delta = 1/2$ and $n = 250$;
(iv) the test based on J_n^L generally gains in power with increasing L.

Additional evidence is provided by Fig. 2, which shows the power of the various tests for mixture alternatives defined for all $\delta, u, v \in [0,1]$ by

$$K_{\delta,C_{\mathrm{CL}}}^{I}(u,v) = (1-\delta)\, C_2^{\mathrm{CL}}(u,v) + \delta\, C^{\mathrm{NE}}(u,v), \tag{11}$$

in terms of Clayton's copula C_2^{CL} with $\tau = 1/2$, and the copula C^{NE} defined at all $(u,v) \in [0,1]^2$ by

$$C^{\mathrm{NE}}(u,v) = \min\{u, v, (u-2/3)_+ + (v-1/3)_+\},$$

where for arbitrary $a \in \mathbb{R}$, $a_+ = a \vee 0$. The latter is one of two copulas identified by Nelsen (2007) as maximizing the asymmetry measure $T(C)$ defined in (5). For the $K_{\delta,C_{\mathrm{CL}}}^{I}$ alternatives, the best power is achieved by the two Cramér–von Mises statistics R_n and S_n. The power of the test based on J_n^L also seems to be increasing in L, the power of J_n^6 being of the same order as that of the Kolmogorov–Smirnov statistic T_n.

7 Data application

As a simple illustration of the procedures described herein, the hypothesis $\mathscr{H}_0$ was tested for data from a survey of nutritional habits commissioned in 1985 by the United

Table 2 Power of the tests of $\mathscr{H}_0$ based on R_n, S_n, T_n, and J_n^L with $L \in \{3, 4, 5, 6\}$, as estimated from 1,000 samples of size $n = 100$ from four asymmetric copulas using $M = 250$ bootstrap replicates and $N = 20$ in approximation (8)

Model	δ	τ	R_n	S_n	T_n	J_n^3	J_n^4	J_n^5	J_n^6
$n = 100$									
		0.5	21.0	55.8	17.3	2.8	4.1	12.8	30.3
	1/4	0.7	41.0	89.6	30.7	2.2	5.4	21.1	48.8
		0.9	73.0	99.9	61.5	0.6	4.4	30.0	68.0
		0.5	47.5	71.4	35.7	21.4	35.4	47.8	54.1
K_δ^{CA}	1/2	0.7	77.0	96.2	61.8	31.3	60.7	73.6	83.1
		0.9	97.7	100.0	82.0	39.6	81.2	93.7	94.5
		0.5	32.0	39.0	20.1	17.7	19.5	22.0	27.2
	3/4	0.7	54.8	67.2	31.0	27.1	33.6	39.0	45.0
		0.9	75.3	85.2	46.2	38.6	47.3	55.3	61.6
		0.5	6.6	7.7	6.1	6.9	6.6	6.5	11.1
	1/4	0.7	40.7	54.3	23.7	9.3	12.0	16.8	20.1
		0.9	92.8	99.9	70.0	0.6	7.7	47.8	76.0
		0.5	11.4	10.9	8.5	7.6	7.6	9.3	12.8
K_δ^{CL}	1/2	0.7	57.3	61.0	38.0	27.6	27.4	29.9	30.6
		0.9	99.8	100.0	83.8	56.7	84.9	88.4	89.8
		0.5	6.4	5.7	6.4	6.4	6.0	6.4	9.4
	3/4	0.7	23.4	24.1	14.1	15.8	15.1	14.4	17.9
		0.9	69.5	73.0	42.9	38.2	44.3	43.8	49.7
		0.5	7.3	7.0	7.0	6.4	6.1	5.2	8.8
	1/4	0.7	35.9	41.4	22.9	14.2	11.9	11.3	15.7
		0.9	92.1	99.2	66.5	0.5	9.6	49.7	72.4
		0.5	13.8	15.4	9.5	9.0	8.5	11.9	14.9
K_δ^{GA}	1/2	0.7	72.9	74.4	43.5	36.2	31.1	32.7	34.5
		0.9	99.8	100.0	85.5	51.9	81.0	90.9	93.7
		0.5	11.8	13.1	12.3	9.9	9.4	11.5	14.2
	3/4	0.7	49.2	49.8	25.9	24.9	25.8	25.5	24.5
		0.9	80.9	83.3	47.8	42.8	51.9	55.8	60.3
		0.5	8.4	8.4	8.1	9.4	6.8	7.1	9.3
	1/4	0.7	40.3	46.0	22.7	13.0	13.4	12.8	17.3
		0.9	93.3	99.6	67.8	0.7	9.8	48.2	72.1
		0.5	23.8	26.4	16.4	12.0	13.2	11.4	16.7
K_δ^{GH}	1/2	0.7	78.4	85.3	47.0	37.8	39.5	41.0	46.8
		0.9	99.9	100.0	86.9	53.9	84.8	91.9	92.7
		0.5	23.9	23.1	13.2	13.6	13.6	14.2	18.0
	3/4	0.7	60.1	59.7	34.3	27.2	30.3	33.4	37.9
		0.9	81.7	86.1	47.3	43.0	55.3	57.5	59.7

Table 3 Power of the tests of $\mathscr{H}_0$ based on R_n, S_n, T_n, and J_n^L with $L \in \{3, 4, 5, 6\}$, as estimated from 1,000 samples of size $n = 250$ from four asymmetric copulas using $M = 250$ bootstrap replicates and $N = 50$ in approximation (8)

Model	δ	τ	R_n	S_n	T_n	J_n^3	J_n^4	J_n^5	J_n^6
$n = 250$									
		0.5	54.5	94.5	65.0	5.8	8.4	14.2	26.6
	1/4	0.7	84.2	99.7	91.3	6.9	9.6	19.1	40.1
		0.9	99.5	100.0	99.9	7.4	22.9	38.7	65.2
		0.5	92.2	99.9	93.0	44.7	77.2	91.3	91.8
K_δ^{CA}	1/2	0.7	99.9	100.0	100.0	67.3	93.5	99.5	99.8
		0.9	100.0	100.0	100.0	90.5	99.5	100.0	100.0
		0.5	76.8	84.4	69.0	34.6	40.6	56.0	57.6
	3/4	0.7	97.0	98.7	91.1	57.5	70.6	81.6	86.5
		0.9	99.6	99.9	97.6	71.7	93.6	96.4	98.1
		0.5	22.8	25.5	24.3	10.4	10.0	12.3	13.2
	1/4	0.7	95.3	97.1	87.0	38.0	42.5	49.3	56.9
		0.9	100.0	100.0	100.0	8.8	48.9	83.2	96.8
		0.5	31.4	32.3	28.9	13.3	14.5	16.5	15.7
K_δ^{CL}	1/2	0.7	98.7	98.8	93.4	62.9	70.4	74.9	75.1
		0.9	100.0	100.0	100.0	97.4	100.0	100.0	100.0
		0.5	14.4	14.9	16.1	8.4	8.6	8.6	10.2
	3/4	0.7	67.6	68.5	57.4	33.7	34.3	34.5	31.3
		0.9	99.7	99.9	95.5	77.5	92.7	92.6	93.8
		0.5	23.7	25.1	24.9	11.4	8.6	8.2	9.5
	1/4	0.7	91.0	91.2	78.1	53.5	48.5	45.5	39.9
		0.9	100.0	100.0	100.0	15.9	61.4	90.7	97.4
		0.5	50.1	50.4	40.1	23.3	21.8	22.5	19.6
K_δ^{GA}	1/2	0.7	99.8	99.8	96.6	75.3	84.7	84.3	83.1
		0.9	100.0	100.0	100.0	97.7	99.9	100.0	100.0
		0.5	41.4	40.9	34.0	17.3	18.4	19.7	17.4
	3/4	0.7	94.1	93.5	80.4	54.5	62.9	66.4	63.3
		0.9	100.0	100.0	98.3	81.3	95.1	97.8	98.8
		0.5	28.4	29.3	28.5	15.8	14.2	13.6	13.0
	1/4	0.7	94.5	95.5	82.7	47.2	46.8	46.2	41.8
		0.9	100.0	100.0	100.0	19.1	53.9	83.9	94.0
		0.5	70.7	72.5	58.4	32.3	37.3	38.2	30.5
K_δ^{GH}	1/2	0.7	99.9	100.0	99.2	84.8	90.9	94.0	91.6
		0.9	100.0	100.0	100.0	96.5	100.0	100.0	100.0
		0.5	67.8	68.0	54.4	28.6	33.5	35.3	34.0
	3/4	0.7	98.2	98.4	90.5	61.8	75.2	81.2	80.2
		0.9	99.8	99.9	99.0	76.7	95.8	97.9	98.5

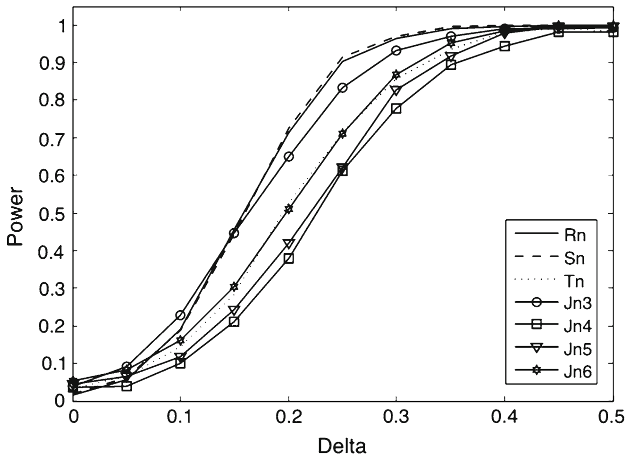

Fig. 2 Graphical representation of the data from Table 2 showing the power of the tests of $\mathscr{H}_0$ based on R_n, S_n, T_n, and J_n^L with $L \in \{3, 4, 5, 6\}$, as estimated from 1,000 samples of size $n = 100$ from the mixture copula (11)

States Department of Agriculture. Five variables were measured on a sample of 747 women aged between 25 and 50 years, namely daily calcium intake (in mg); daily iron intake (in mg); daily protein intake (in g); daily vitamin A intake (in µg), and daily vitamin C intake (in mg).

McNeil and Nešlehová (2010) used this data set to illustrate the so-called Liouville copulas, which constitute an asymmetric extension of the Archimedean class of dependence models. These authors found that the subsample of the daily intake of calcium, iron, and protein was best described by the Clayton–Liouville copula with parameter $(1, 3, 4)$, which is an asymmetric generalization of the Clayton model. This result suggests a strongly asymmetric dependence structure between the intakes of calcium and iron, and between the intakes of calcium and protein; the copula of the intakes of iron and protein appears only mildly asymmetric.

Here, the hypothesis $\mathscr{H}_0$ of symmetry was tested for every pair using the Cramér–von Mises statistic S_n. The latter emerged as the most powerful test statistic from the simulations discussed in Sect. 6. The resulting P values, computed on the basis of $M = 1{,}000$ bootstrap replicates, are reported in Table 4. Rank plots of pairs that were

Table 4 P values (in percentage) of the test based on S_n for the nutrient data

Variable	Calcium	Iron	Protein	Vitamin A	Vitamin C
Calcium		**0.3**	**0.0**	**0.0**	18.1
Iron			40.2	**0.2**	**0.4**
Protein				**0.4**	13.6
Vitamin A					62.0

Values that lead to rejection of the symmetry hypothesis $\mathscr{H}_0$ at the 5% level are highlighted in bold

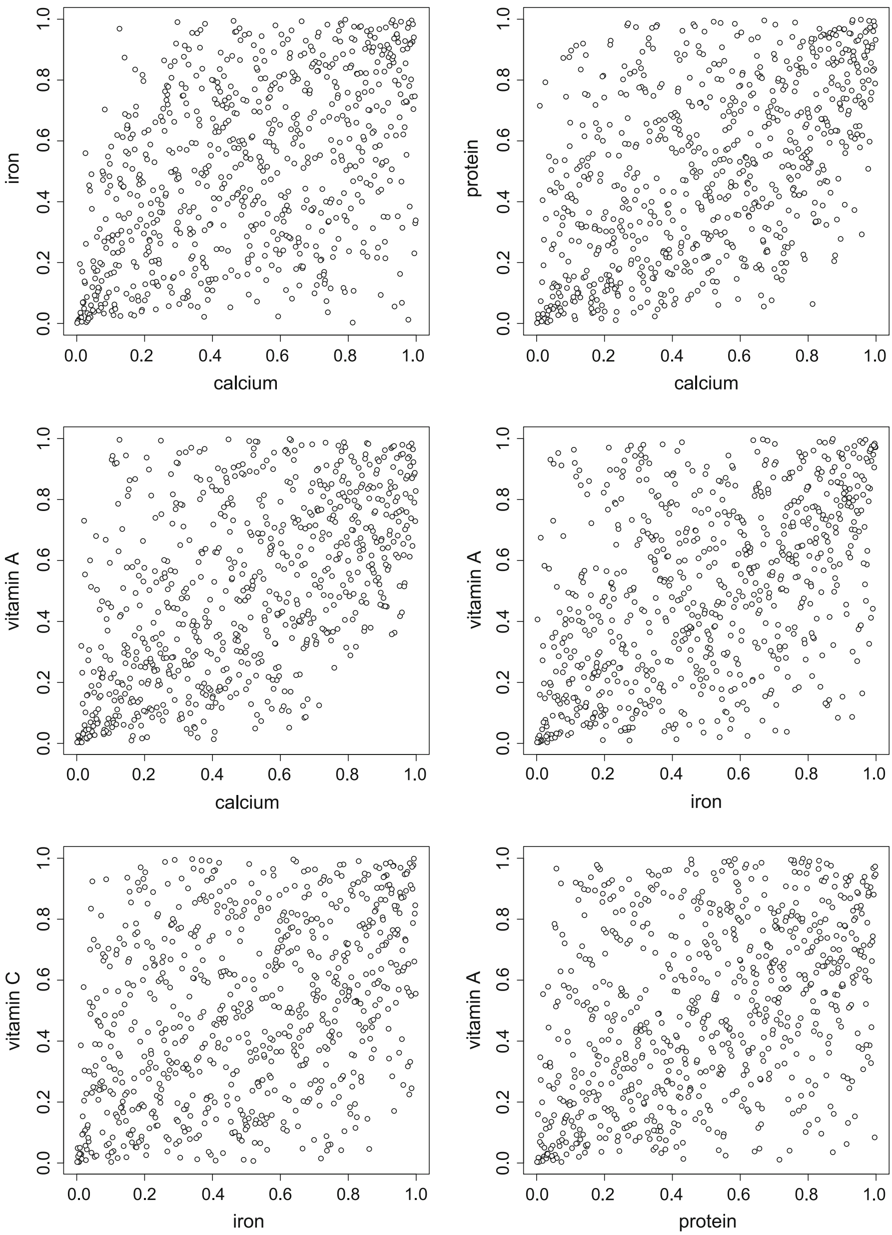

Fig. 3 Rank plots for the pairs identified as asymmetric ($P < 5\%$) in the nutrient data set

identified as asymmetric ($P < 5\%$) are displayed in Fig. 3, whereas Fig. 4 shows pairs for which the dependence appears to be symmetric.

These findings confirm the observations of McNeil and Nešlehová (2010) concerning the asymmetric dependence in the pairs (calcium, iron) and (calcium, protein). In contrast, the hypothesis of a symmetric dependence between the intakes of iron and

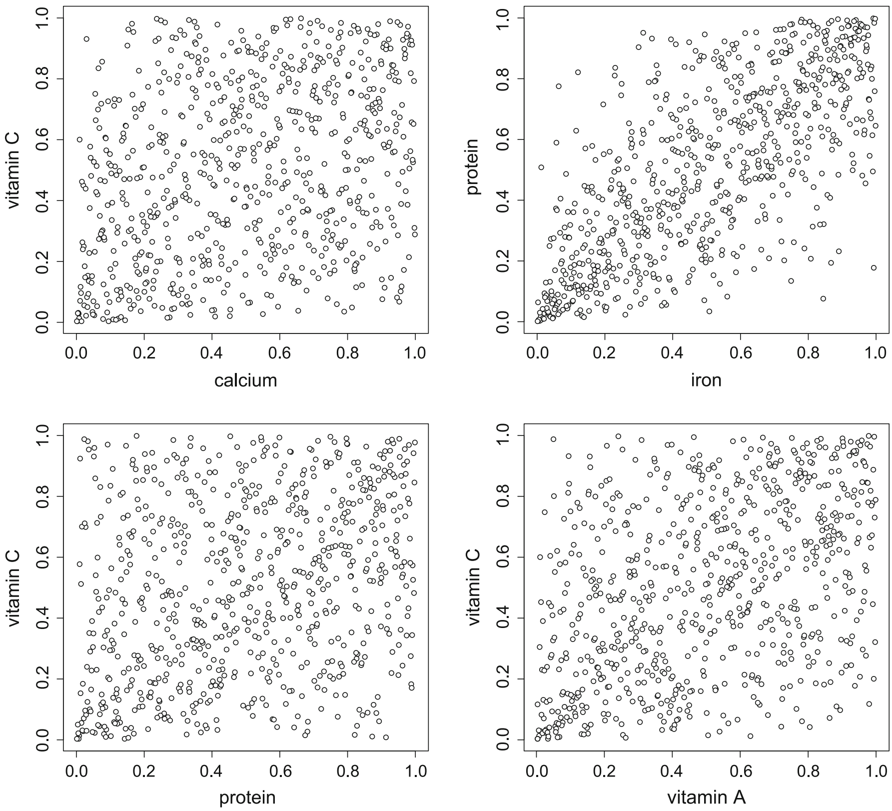

Fig. 4 Rank plots for the pairs identified as symmetric ($P > 5\%$) in the nutrient data set

protein was rejected neither by the test based on S_n nor by the tests based on R_n, T_n, and J_n^L; the corresponding P values ranged from 23.1 to 97.0%.

Appendix: Technical arguments

Proof of Proposition 1 For arbitrary $(u, v) \in [0, 1]^2$, let

$$\hat{\mathbb{D}}_n(u, v) = n^{-1/2}\,\mathbf{1}^\top P_n(u, v) \quad \text{and} \quad \hat{\mathbb{D}}_n^2(u, v) = \frac{1}{n}\,\mathbf{1}^\top P_n(u, v) P_n(u, v)^\top \mathbf{1},$$

where $\mathbf{1}$ is an $n \times 1$ vector of 1's and $P_n(u, v)$ is the $n \times 1$ vector with ith element $P_{in}(u, v)$ defined in Equation (6). Further introduce the $n \times n$ matrix

$$\mathbf{A} = \int_0^1 \int_0^1 P_n(u, v) P_n(u, v)^\top \,\mathrm{d}v\,\mathrm{d}u;$$

its element in position (i, j) can be expressed in the form

$$A_{ij} = 2(1 - \hat{U}_i \vee \hat{U}_j)(1 - \hat{V}_i \vee \hat{V}_j) - 2(1 - \hat{U}_i \vee \hat{V}_j)(1 - \hat{U}_j \vee \hat{V}_i).$$

One then has

$$R_n = \frac{1}{n} \int_0^1 \int_0^1 \{\hat{\mathbb{D}}_n(u, v)\}^2 \, \mathrm{d}v \, \mathrm{d}u = \frac{1}{n^2} \, \mathbf{1}^\top \mathbf{A} \mathbf{1}.$$

The expression for S_n follows upon noting that the measure $\hat{C}_n$ assigns a weight of $1/n$ to each pair of pseudo-observations $(\hat{U}_1, \hat{V}_1), \ldots, (\hat{U}_n, \hat{V}_n)$. One gets

$$S_n = \frac{1}{n} \int_0^1 \int_0^1 \{\hat{\mathbb{D}}_n(u, v)\}^2 \, \mathrm{d}\hat{C}_n(u, v) = \frac{1}{n^2} \sum_{k=1}^n \{\hat{\mathbb{D}}_n(\hat{U}_k, \hat{V}_k)\}^2$$
$$= \frac{1}{n^3} \sum_{k=1}^n \{\mathbf{1}^\top P_n(\hat{U}_k, \hat{V}_k)\}^2.$$

Finally, let $I_{ij} = [i/n, (i+1)/n) \times [j/n, (j+1)/n)$ for all $i, j \in \{1, \ldots, n\}$. Note that

$$(u, v) \in I_{ij} \quad \Rightarrow \quad (v, u) \in I_{ji} \Rightarrow \begin{cases} \hat{C}_n(u, v) = \hat{C}_n(i/n, j/n), \\ \hat{C}_n(v, u) = \hat{C}_n(j/n, i/n). \end{cases}$$

Consequently,

$$T_n = \sup_{(u,v)\in[0,1]^2} |\hat{C}_n(u, v) - \hat{C}_n(v, u)| = \max_{i,j\in\{1,\ldots,n\}} \sup_{(u,v)\in I_{ij}} |\hat{C}_n(u, v) - \hat{C}_n(v, u)|$$
$$= \max_{1\leqslant i<j\leqslant n} \left| \hat{C}_n\left(\frac{i}{n}, \frac{j}{n}\right) - \hat{C}_n\left(\frac{j}{n}, \frac{i}{n}\right) \right|,$$

as claimed. □

Proof of Proposition 2 It follows from the Continuous Mapping Theorem that as $n \to \infty$, one has $\hat{\mathbb{D}}_n \rightsquigarrow \hat{\mathbb{D}}$, where $\hat{\mathbb{D}}(u, v) = \hat{\mathbb{C}}(u, v) - \hat{\mathbb{C}}(v, u)$ for all $(u, v) \in [0, 1]^2$. Note that under $\mathscr{H}_0$, one has $\dot{C}_1(u, v) = \dot{C}_2(v, u)$ for all $(u, v) \in (0, 1)^2$. Accordingly, the limiting process $\hat{\mathbb{D}}$ can be expressed in the alternative form

$$\hat{\mathbb{D}}(u, v) = \mathbb{D}(u, v) - \dot{C}_1(u, v) \, \mathbb{D}(u, 1) - \dot{C}_2(u, v) \, \mathbb{D}(1, v),$$

where $\mathbb{D}(u, v) = \mathbb{C}(u, v) - \mathbb{C}(v, u)$ for all $(u, v) \in [0, 1]^2$. The process $\mathbb{D}$ is a centred Gaussian random field whose covariance function at any $u, v, s, t \in [0, 1]$ is given by

$$\mathrm{cov}\{\mathbb{D}(u, v), \mathbb{D}(s, t)\} = \mathrm{cov}\{\mathbb{C}(u, v) - \mathbb{C}(v, u), \mathbb{C}(s, t) - \mathbb{C}(t, s)\}$$
$$= \Gamma_{\mathbb{C}}(u, v, s, t) - \Gamma_{\mathbb{C}}(v, u, s, t) - \Gamma_{\mathbb{C}}(u, v, t, s) + \Gamma_{\mathbb{C}}(v, u, t, s).$$

Now under $\mathscr{H}_0$, one has $\Gamma_{\mathbb{C}}(u, v, s, t) = \Gamma_{\mathbb{C}}(v, u, t, s)$ for all $u, v, s, t \in [0, 1]$. Hence

$$\Gamma_{\mathbb{D}}(u, v, s, t) = \mathrm{cov}\{\mathbb{D}(u, v), \mathbb{D}(s, t)\} = 2\Gamma_{\mathbb{C}}(u, v, s, t) - 2\Gamma_{\mathbb{C}}(u, v, t, s)$$

for all $u, v, s, t \in [0, 1]$, as claimed. □

Proof of Proposition 3 Let C be a regular symmetric copula. Given that R_n and T_n are continuous functionals of $\hat{\mathbb{D}}_n$, their weak limits can easily be deduced from Proposition 2, as a direct application of the Continuous Mapping Theorem. The argument for S_n is more subtle.

First introduce some notation. Let $\mathscr{C}[0,1]^2$ be the space of functions $f:[0,1]^2 \to \mathbb{R}$ that are continuous, and write $\mathscr{D}[0,1]^2$ for the space of functions f that are continuous from the upper right quadrant and have limits from the other quadrants; here, both spaces are equipped with the uniform norm. Furthermore, denote by $\mathrm{BV}_1[0,1]^2$ the subspace of $\mathscr{D}[0,1]^2$ consisting of functions with total variation bounded by 1.

A direct application of the Continuous Mapping Theorem implies that as $n \to \infty$, $(\hat{\mathbb{D}}_n^2, \hat{\mathbb{C}}_n) \rightsquigarrow (\hat{\mathbb{D}}^2, \hat{\mathbb{C}})$ on $\ell^\infty[0,1]^2 \times \ell^\infty[0,1]^2$. Write

$$(\hat{\mathbb{D}}_n^2, \hat{\mathbb{C}}_n) = n^{1/2}\{(A_n, \hat{C}_n) - (A, C)\},$$

where $A \equiv 0$ and $A_n = n^{1/2}(\hat{C}_n - \hat{C}_n^\top)^2$. Now consider the map $\Phi : \ell^\infty[0,1]^2 \times \mathrm{BV}_1[0,1]^2 \to \mathbb{R}$ defined by

$$\Phi(\alpha, \beta) = \int_{(0,1]^2} \alpha \, \mathrm{d}\beta.$$

One then has

$$nS_n = n^{1/2}\{\Phi(A_n, \hat{C}_n) - \Phi(A, C)\}.$$

In view of Lemma 4.3 of Carabarin-Aguirre and Ivanoff (2010), the map Φ is Hadamard differentiable tangentially to $\mathscr{C}[0,1]^2 \times \mathscr{D}[0,1]^2$ at each (α, β) in $\ell^\infty[0,1]^2 \times \mathrm{BV}_1[0,1]^2$ such that $\int |\mathrm{d}\alpha| < \infty$. An application of the Functional Delta Method (van der Vaart and Wellner 1996, Theorem 3.9.4) implies that $nS_n \rightsquigarrow \Phi'_{(A,C)}(\hat{\mathbb{D}}^2, \hat{\mathbb{C}})$ as $n \to \infty$, where

$$\Phi'_{(A,C)}(\hat{\mathbb{D}}^2, \hat{\mathbb{C}}) = \int_{(0,1]^2} A \, \mathrm{d}\hat{\mathbb{C}} + \int_{(0,1]^2} \hat{\mathbb{D}}^2 \, \mathrm{d}C = \int_{(0,1]^2} \hat{\mathbb{D}}^2 \, \mathrm{d}C.$$

This is the desired conclusion. □

Proof of Proposition 4 Let C be a regular copula, which may or may not be symmetric. As stated in Sect. 3, it follows from the Continuous Mapping Theorem that the statistics R_n and T_n converge in probability to the measures of asymmetry $R(C)$ and $T(C)$, respectively. To determine the limit of S_n, write

$$|S_n - S(C)| \leqslant |\gamma_n| + |\zeta_n|,$$

where

$$\gamma_n = \int_0^1 \int_0^1 \{\hat{C}_n(u, v) - \hat{C}_n(v, u)\}^2 \, \mathrm{d}\hat{C}_n(u, v) - \int_0^1 \int_0^1 \{C(u, v) - C(v, u)\}^2 \, \mathrm{d}\hat{C}_n(u, v)$$

and

$$\zeta_n = \int_0^1 \int_0^1 \{C(u, v) - C(v, u)\}^2 \, \mathrm{d}\hat{C}_n(u, v) - \int_0^1 \int_0^1 \{C(u, v) - C(v, u)\}^2 \, \mathrm{d}C(u, v).$$

Because

$$\begin{aligned} &|\{\hat{C}_n(u, v) - \hat{C}_n(v, u)\}^2 - \{C(u, v) - C(v, u)\}^2| \\ &\quad = |\hat{C}_n(u, v) - C(u, v) - \hat{C}_n(v, u) + C(v, u)| \\ &\qquad \times |\hat{C}_n(u, v) + C(u, v) - \hat{C}_n(v, u) - C(v, u)| \\ &\quad \leqslant 8 \sup_{(u,v)\in[0,1]^2} |\hat{C}_n(u, v) - C(u, v)|, \end{aligned}$$

one gets

$$|\gamma_n| \leqslant 8 \sup_{(u,v)\in[0,1]^2} |\hat{C}_n(u, v) - C(u, v)| \xrightarrow{p} 0.$$

Turning to ζ_n, set $\psi = (C - C^\top)^2$ and observe that

$$\int_0^1 \int_0^1 \{C(u, v) - C(v, u)\}^2 \, \mathrm{d}\hat{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^n \psi\left(\frac{R_i}{n}, \frac{S_i}{n}\right).$$

The formula on the right-hand side is a bivariate linear rank statistics, which converges to $\int_0^1 \int_0^1 \psi(u, v) \, \mathrm{d}C(u, v)$ almost surely by Proposition A.1 (i) in Genest et al. (1995). This means that $\zeta_n \to 0$ almost surely, and hence $|\zeta_n| \xrightarrow{p} 0$ as $n \to \infty$. □

Proof of Proposition 5 For $(u, v) \in [0, 1]^2$, let

$$\mathbb{C}_n^{(h)}(u, v) = n^{-1/2} \sum_{i=1}^n \left(\frac{\xi_i^{(h)}}{\bar{\xi}_n^{(h)}} - 1\right) \mathbb{I}(\hat{U}_i \leqslant u, \hat{V}_i \leqslant v)$$

and observe that in view of formula (7), one has

$$\begin{aligned} \hat{\mathbb{D}}_n^{(h)}(u, v) = &\{\mathbb{C}_n^{(h)}(u, v) - \dot{C}_{1n}(u, v)\, \mathbb{C}_n^{(h)}(u, 1) - \dot{C}_{2n}(u, v)\, \mathbb{C}_n^{(h)}(1, v)\} \\ &- \{\mathbb{C}_n^{(h)}(v, u) - \dot{C}_{1n}(u, v)\, \mathbb{C}_n^{(h)}(1, u) - \dot{C}_{2n}(u, v)\, \mathbb{C}_n^{(h)}(v, 1)\}. \end{aligned}$$

If $\mathscr{H}_0$ holds, then for all $(u, v) \in (0, 1)^2$, one has

$$\dot{C}_1(u, v) = \dot{C}_2(v, u) \quad \text{and} \quad \dot{C}_2(u, v) = \dot{C}_1(v, u).$$

Invoking Proposition 4.2 of Segers (2012) and the fact that as $n \to \infty$, $\bar{\xi}_n^{(h)} \to 1$ almost surely by the Law of Large Numbers, one can conclude that

$$(\hat{\mathbb{D}}_n, \hat{\mathbb{D}}_n^{(1)}, \ldots, \hat{\mathbb{D}}_n^{(M)}) \rightsquigarrow (\hat{\mathbb{D}}, \hat{\mathbb{D}}^{(1)}, \ldots, \hat{\mathbb{D}}^{(M)}),$$

where $\hat{\mathbb{D}}^{(1)}, \ldots, \hat{\mathbb{D}}^{(M)}$ are independent copies of $\hat{\mathbb{D}}$, as claimed. □

Acknowledgments Funding from the Canada Research Chairs Program, the Natural Sciences and Engineering Research Council of Canada, the Fonds québécois de la recherche sur la nature et les technologies, and the Fonds institutionnel de la recherche de l'Université du Québec à Trois-Rivières is gratefully acknowledged.

References

Bell, C. B., Haller, H. S. (1969). Bivariate symmetry tests: Parametric and nonparametric. *The Annals of Mathematical Statistics, 40*, 259–269.

Bücher, A., Dette, H. (2010). A note on bootstrap approximations for the empirical copula process. *Statistics and Probability Letters, 80*, 1925–1932.

Carabarin-Aguirre, A., Ivanoff, B. G. (2010). Estimation of a distribution under generalized censoring. *Journal of Multivariate Analysis, 101*, 1501–1519.

Cherubini, U., Luciano, E., Vecchiato, W. (2004). *Copula methods in finance*. Wiley: Chichester.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés: Un test non paramétrique d'indépendance. *Académie royale de Belgique, Bulletin de la classe des sciences* (5)*, 65*, 274–292.

Fang, H.-B., Fang, K.-T., Kotz, S. (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis, 82*, 1–16 (Corr: *Journal of Multivariate Analysis, 94*, 222–223.).

Fermanian, J.-D., Radulović, D., Wegkamp, M. H. (2004). Weak convergence of empirical copula processes. *Bernoulli, 10*, 847–860.

Gänßler, P., Stute, W. (1987). *Seminar on empirical processes*. Birkhäuser: Basel.

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

Genest, C., Ghoudi, K., Rivest, L.-P. (1998). Discussion of "Understanding relationships using copulas," by E. W. Frees and E. A. Valdez. *North American Actuarial Journal, 2*, 143–149.

Genest, C., Segers, J. (2010). On the covariance of the asymptotic empirical copula process. *Journal of Multivariate Analysis, 101*, 1837–1845.

Hilton, J. F. (2000). A new asymptotic distribution for Hollander's bivariate symmetry statistic. *Computational Statistics and Data Analysis, 32*, 455–463.

Hilton, J. F., Gee, L. (1997). The size and power of the exact bivariate symmetry test. *Computational Statistics and Data Analysis, 26*, 53–69.

Hollander, M. (1971). A nonparametric test for bivariate symmetry. *Biometrika, 58*, 203–212.

Jasson, S. (2005). L'asymétrie de la dépendance, quel impact sur la tarification? Technical report, AXA Group Risk Management, Paris, France.

Khoudraji, A. (1995). *Contributions à l'étude des copules et à la modélisation de valeurs extrêmes bivariées*. PhD thesis, Université Laval, Québec, Canada.

Kojadinovic, I., Yan, J. (2011). A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing, 21*, 17–30.

Kojadinovic, I., Yan, J., Holmes, M. (2011). Fast large-sample goodness-of-fit tests for copulas. *Statistica Sinica, 21*, 841–871.

Koziol, J. A. (1979). A test for bivariate symmetry based on the empirical distribution function. *Communications in Statistics A, Theory and Methods, 8*, 207–221.

Liebscher, E. (2008). Construction of asymmetric multivariate copulas. *Journal of Multivariate Analysis, 99*, 2234–2250 (Corr: *Journal of Multivariate Analysis, 102*, 869–870.)

McNeil, A. J., Nešlehová, J. (2010). From Archimedean to Liouville copulas. *Journal of Multivariate Analysis, 101*, 1772–1790.

McNeil, A. J., Frey, R., Embrechts, P. (2005). *Quantitative risk management: Concepts, techniques and tools*. Princeton University Press: Princeton, NJ.

Nelsen, R. B. (2006). *An introduction to copulas* (2nd ed.). Springer: New York.

Nelsen, R. B. (2007). Extremes of nonexchangeability. *Statistical Papers, 48*, 329–336.

Rémillard, B., Scaillet, O. (2009). Testing for equality between two copulas. *Journal of Multivariate Analysis, 100*, 377–386.

Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics. *The Annals of Statistics, 4*, 912–923.

Salvadori, G., de Michele, C., Kottegoda, N. T., Rosso, R. (2007). *Extremes in nature: An approach using copulas*. Springer: New York.

Scaillet, O. (2005). A Kolmogorov–Smirnov type test for positive quadrant dependence. *The Canadian Journal of Statistics, 33*, 415–427.

Segers, J. (2012). Weak convergence of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli, 18* (in press).

Sen, P. K. (1967). Nonparametric tests for multivariate interchangeability. I. Problems of location and scale in bivariate distributions. *Sankhyā Series A, 29*, 351–372.

Sklar, A. (1959). Fonctions de répartition à *n* dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris, 8*, 229–231.

Snijders, T. (1981). Rank tests for bivariate symmetry. *The Annals of Statistics, 9*, 1087–1095.

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

van der Vaart, A. W., Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer: New York.

Yanagimoto, T., Sibuya, M. (1976). Test of symmetry of a bivariate distribution. *Sankhyā Series A, 38*, 105–115.