IMPROVED KERNEL ESTIMATION OF COPULAS: WEAK CONVERGENCE AND GOODNESS-OF-FIT TESTING*

MAREK OMELKA[1,2], IRÈNE GIJBELS[3], NOËL VERAVERBEKE[1]

[1] Center for Statistics, Hasselt University, Agoralaan -building D, B-3590 Diepenbeek, Belgium;

[2] Jaroslav Hájek Center for Theoretical and Applied Statistics, Charles University Prague, Sokolovská 83, 186 75 Praha 8, Czech Republic;

[3] Department of Mathematics and Leuven Statistics Research Center (LStat), Katholieke Universiteit Leuven, Celestijnenlaan 200B, B-3001 Leuven (Heverlee), Belgium.

Abstract. We reconsider the existing kernel estimators for a copula function, as proposed in (1) Gijbels and Mielniczuk (1990), (2) Fermanian *et al.* (2004) and (3) Chen and Huang (2007). All these estimators have as a drawback that they can suffer from a corner bias problem. A way to deal with this is to impose rather stringent conditions on the copula, outruling as such many classical families of copulas. In this paper we propose improved estimators that take care of the typical corner bias problem. For (1) and (3), the improvement involves shrinking the bandwidth with an appropriate functional factor, and for (2) this is done by using a transformation. The theoretical contribution of the paper is a weak convergence result for the three improved estimators under conditions that are met for most copula families. We also discuss the choice of bandwidth parameters, theoretically and practically, and illustrate the finite-sample behaviour of the estimators in a simulation study. The improved estimators are applied to goodness-of-fit testing for copulas.

Keywords and phrases: copula; Cramér-von Mises statistics; Gaussian process; goodness-of-fit; Kendall's tau; Kolmogorov-Smirnov statistics; parametric bootstrap; pseudo-observations; weak convergence.

1. Introduction

Consider a random vector $\mathbf{X} = (X_1, \ldots, X_d)^\mathsf{T}$ with joint cumulative distribution function H and marginal distribution functions $F_1, \ldots, F_d$. According to Sklar's theorem (see e.g. Nelsen (2006)) there exists a d-variate function C such that

$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)) \,. \tag{1}$$

*This work was supported by the IAP Research Network P6/03 of the Belgian State (Belgian Science Policy). This work was done while the first author was a postdoctoral researcher at the Katholieke Universiteit Leuven and the Universiteit Hasselt within the IAP Research Network. Support of the Research Project LC06024 is also highly appreciated.

The function C is called a copula and it is in itself a joint cumulative distribution function on $[0,1]^d$ with uniform marginals. If the marginal distribution functions $F_1, \dots, F_d$ are continuous, then the function C is unique and $C(u_1, \dots, u_d) = H(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))$, where, for $j = 1, \dots, d$, $F_j^{-1}(u) = \inf\{x : F_j(x) \geq u\}$, with $u \in [0,1]$, is the quantile function of F_j. The copula C 'couples' the joint distribution function H to its univariate marginals, capturing as such the dependence structure between the components of $\mathbf{X} = (X_1, \dots, X_d)^\mathsf{T}$.

Methods for estimation of copulas usually depend on how much we are willing to assume about the joint distribution function H. In fully parametric approaches with parametric models for both the copula and the marginals, maximum likelihood estimation may be used. Nowadays semiparametric estimation is quite popular, in which one specifies a parametric copula and estimates the marginals nonparametrically. In this paper we focus on nonparametric estimation of the copula making as such no restrictive distributional assumptions on the copula, nor on the marginals.

For simplicity of the presentation we will restrict to the case $d = 2$, and consider an independent and identically distributed sample $(X_1, Y_1)^\mathsf{T}, \dots, (X_n, Y_n)^\mathsf{T}$ of a bivariate random vector $(X, Y)^\mathsf{T}$ with joint distribution function H and marginal distribution functions F and G.

Nonparametric estimation of copulas goes back to Deheuvels (1979) who proposed, in order to test for independence, the following empirical copula estimator

$$C_n(u,v) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{\hat{U}_i \leq u, \hat{V}_i \leq v\}, \qquad \text{with } \hat{U}_i = F_n(X_i),\ \hat{V}_i = G_n(Y_i),$$

where F_n and G_n are the empirical cumulative distribution functions of the marginals, and where $\mathbb{I}\{A\}$ denotes the indicator of a set A. This estimator is asymptotically equivalent (up to a term $O(n^{-1})$) with the estimator based directly on Sklar's Theorem given by

$$C_n(u,v) = H_n(F_n^{-1}(u), G_n^{-1}(v)), \tag{2}$$

with H_n the empirical joint distribution function. Weak convergence studies of this estimator can be found in Gänssler and Stute (1987), Fermanian et al. (2004) and Tsukuhara (2005). Our Monte-Carlo experiments showed that it is better to use the following (asymptotically equivalent) modification of the Empirical copula

$$C_n^{(\mathrm{E})}(u,v) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{\hat{U}_i^{(\mathrm{E})} \leq u, \hat{V}_i^{(\mathrm{E})} \leq v\},\ \text{with } \hat{U}_i^{(\mathrm{E})} = \tfrac{n}{n+1}\, F_n(X_i),\ \hat{V}_i^{(\mathrm{E})} = \tfrac{n}{n+1}\, G_n(Y_i)\ , \tag{3}$$

which shifts the pseudo-observations $F_n(X_i)$ and $G_n(Y_i)$ a bit closer to the left corner of the unit interval $[0,1]$. See for example Genest et al. (1995).

Fermanian et al. (2004) also proposed a Smoothed version of the Empirical copula. Their proposal is a straightforward modification of (3) and the estimator is defined as

$$\hat{C}_n^{(\mathrm{SE})}(u,v) = \hat{H}_n(\hat{F}_n^{-1}(u), \hat{G}_n^{-1}(v)) \ , \tag{4}$$

where the quantities $\hat{H}_n$, $\hat{F}_n$ and $\hat{G}_n$ are given by

$$\hat{H}_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} K_n(x - X_i, y - Y_i), \quad \hat{F}_n(x) = \hat{H}_n(x, +\infty), \quad \hat{G}_n(x) = \hat{H}_n(+\infty, y) \ , \tag{5}$$

with

$$K_n(x,y) = K(\tfrac{x}{b_n}, \tfrac{y}{b_n}), \quad K(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} k(s,t)\, ds\, dt \ ,$$

where $k(s,t)$ is a given bivariate kernel density function, and b_n is a bandwidth sequence tending to zero with n. Fermanian et al. (2004) proved weak convergence of this estimator.

There are two kernel type estimators in the literature that pay special attention to the correction of the boundary bias. This typical bias associated with kernel estimation is present since a copula has its support on the bounded set $[0,1]^2$. The first reference is the mirror-reflection type estimator originating from the work of Gijbels and Mielniczuk (1990) on copula density estimation. They take care of boundary bias correction through data-augmentation obtained by reflecting the original data with respect to the edges and the corners of the unit square. The second reference is the estimator of Chen and Huang (2007) who proposed to use a local linear kernel in order to deal with the bias near the boundaries of the unit square.

A first goal of the present paper is to prove the weak convergence of the estimators of Gijbels and Mielniczuk (1990) and Chen and Huang (2007) under the assumption that C has bounded second order partial derivatives on $[0,1]^2$. See Theorem 1 in Section 2. It turns out however that for many commonly-used families of copulas (e.g. Clayton, Gumbel, normal, Student), the latter condition is not satisfied and the bias behaviour at the corners of the unit square precludes the weak convergence on the whole $[0,1]^2$. We therefore propose improved 'shrinked' versions of the estimators of Gijbels and Mielniczuk (1990) and Chen and Huang (2007). This shrinking is done by including a weight function which removes the corner bias. In a same spirit we also suggest a modification of the copula estimator (4) of Fermanian *et al.* (2004). In Theorem 2 we establish weak convergence for all newly-proposed estimators. The finite-sample performance of the estimators is demonstrated via a simulation study. We discuss optimal bandwidth selection and compare the performances of the estimators using various well-known distance measures.

The second goal of the paper is to discuss the use of the various estimators of copulas in goodness-of-fit testing problems.

The paper is organized as follows. In Section 2 we introduce the improved kernel estimators and state the main theoretical results on weak convergence. In Section 3, we investigate the finite-sample performance of the newly-proposed estimators, and compare these with performances of existing estimators. In Section 4 simulation results are reported for goodness-of-fit testing. The proofs of the weak convergence results are given in the Appendix.

2. Nonparametric kernel estimators of a copula

In this section we briefly discuss existing kernel estimators and propose important modifications. We also state the weak convergence results.

2.1. **Local linear kernel estimator.** Chen and Huang (2007) constructed their estimator in the following way. In the first stage they estimate marginals by

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K\big(\tfrac{x-X_i}{b_{n1}}\big), \qquad \hat{G}_n(y) = \frac{1}{n}\sum_{i=1}^{n} K\big(\tfrac{y-Y_i}{b_{n2}}\big), \tag{6}$$

with K the integral of a symmetric bounded kernel function k supported on $[-1,1]$. In the second stage the pseudo-observations $\hat{U}_i = \hat{F}_n(X_i)$ and $\hat{V}_i = \hat{G}_n(Y_i)$ are used to estimate the joint distribution function of the unobserved $F(X_i)$ and $G(Y_i)$ which gives the estimate of the unknown copula C. To prevent for boundary bias, Chen and Huang (2007) suggested using a local linear version of the kernel k given by

$$k_{u,h}(x) = \frac{k(x)\{a_2(u,h) - a_1(u,h)\,x\}}{a_0(u,h)a_2(u,h) - a_1^2(u,h)}\,\mathbb{I}\{\tfrac{u-1}{h} < x < \tfrac{u}{h}\}, \tag{7}$$

where

$$a_l(u,h) = \int_{\frac{u-1}{h}}^{\frac{u}{h}} t^l k(t)\,dt \qquad \text{for } l = 0,1,2\ .$$

Finally the Local Linear type estimator of the copula is given by

$$\hat{C}_n^{(\mathrm{LL})}(u,v) = \frac{1}{n}\sum_{i=1}^{n} K_{u,h_n}\big(\tfrac{u-\hat{U}_i}{h_n}\big) K_{v,h_n}\big(\tfrac{v-\hat{V}_i}{h_n}\big)\,, \tag{8}$$

where $K_{u,h}(x) = \int_{-\infty}^{x} k_{u,h}(s)\,ds$. Chen and Huang (2007) derived expressions for asymptotic bias, variance and mean squared error for this estimator and showed that a proper choice of the second stage smoothing constants $h = h_n$ may considerably decrease variance, as well as mean squared error of the copula estimate. Moreover their Monte Carlo experiments showed that the estimator $\hat{C}_n^{(\mathrm{LL})}$ is quite insensitive to the choice of the constants b_{1n} and b_{2n} used for smoothing the marginals in the first stage. Variance considerations provided by the authors even showed that it is reasonable to take b_{1n} and b_{2n} as small as possible. Note that strong undersmoothing in the first stage, recommended in Chen and Huang (2007), results in using the pseudo-observations $(\hat{U}_i, \hat{V}_i)^{\mathsf{T}} = \big(\frac{2\,n\,F_n(X_i)-1}{2\,n}, \frac{2\,n\,G_n(Y_i)-1}{2\,n}\big)^{\mathsf{T}}$, which is asymptotically equivalent to the mostly-used pseudo-observations defined in (3).

As already mentioned in the introduction, the theoretical inconvenience of the estimator (8) is that for many common families of copulas (e.g. Clayton, Gumbel, normal, Student) the bias of the estimator at some of the corners of the unit square is only of order $O(h_n)$. As the optimal bandwidth for distribution function estimation is of order $O(n^{-1/3})$, this violates the $n^{1/2}$-order weak convergence on the whole $[0,1]^2$.

The problem is caused by unboundedness of second order partial derivatives of many copula families. Although parametric models with unbounded densities are rather rare in 'standard' parametric models, copula families with unbounded densities are quite common. As a benchmark we can take the normal bivariate density which is usually supposed to be a well-behaved model. But the resulting normal copula density is unbounded.

To overcome this difficulty we propose a method of shrinking the bandwidth when coming close to the borders of the unit square. The proposed method is based on the observation that when calculating the bias of the estimator (8) we have to deal with terms of the form $h^2\,C_{uu}(u,v)$, $h^2\,C_{uv}(u,v)$ and $h^2\,C_{vv}(u,v)$, where $C_{uu}(u,v)$, $C_{uv}(u,v)$ and $C_{vv}(u,v)$ are the second order partial derivatives of C, that is $C_{uu}(u,v) = \partial^2 C(u,v)/\partial u^2$ and similarly for $C_{uv}(u,v)$, $C_{vv}(u,v)$. The problem is that for many common families of copulas these second order partial derivatives are not bounded, and in fact a closer inspection of them shows that

$$(9) \qquad C_{uu}(u,v) = O\left(\tfrac{1}{u(1-u)}\right),\ C_{vv}(u,v) = O\left(\tfrac{1}{v(1-v)}\right),\ C_{uv}(u,v) = O\left(\tfrac{1}{\sqrt{uv(1-u)(1-v)}}\right).$$

This is shown in Appendix D for Clayton, Gumbel, normal and Student copulas. In order to keep the bias bounded we suggest an improved 'Shrinked' version of (8) given by

$$(10) \quad \hat{C}_n^{(\mathrm{LLS})}(u,v) = \frac{1}{n}\sum_{i=1}^{n} K_{u,h_n}\left(\tfrac{u-\hat{U}_i}{b(u)\,h_n}\right) K_{v,h_n}\left(\tfrac{v-\hat{V}_i}{b(v)\,h_n}\right), \text{ with } \ b(w) = \min(\sqrt{w}, \sqrt{1-w}),$$

in which the constant bandwidth h_n is replaced by a bandwidth function $b(u)h_n$ that 'shrinks' the value of the bandwidth close to zero at the corners of the unit square. A straightforward adaptation of the result of Chen and Huang (2007) gives that for $(u/b(u), v/b(v)) \in [h_n, 1-h_n]^2$ (and no smoothing of the marginals in the first stage)

$$(11) \qquad \mathrm{Bias}\left\{\hat{C}_n^{(\mathrm{LLS})}(u,v)\right\} = \tfrac{\sigma_K^2}{2}\,h_n^2\left\{b^2(u)\,C_{uu}(u,v) + b^2(v)\,C_{vv}(u,v)\right\} + o(h_n^2)$$

$$(12) \quad \mathrm{Var}\left\{\hat{C}_n^{(\mathrm{LLS})}(u,v)\right\} = \tfrac{1}{n}\mathrm{Var}\left[\mathbb{I}\{U \le u, V \le v\} - C_u(u,v)\mathbb{I}\{U \le u\} - C_v(u,v)\mathbb{I}\{V \le v\}\right] - \tfrac{h_n\,b_K}{n}\left[b(u)\,C_u(u,v)(1-C_u(u,v)) + b(v)\,C_v(u,v)(1-C_v(u,v))\right] + o(\tfrac{h_n}{n}),$$

with $\sigma_K^2 = \int_{-1}^{1} t^2 k(t)\,dt$, $b_K = 2\int_{-1}^{1} t\,k(t)\,K(t)\,dt$ and $b(\cdot)$ as defined in (10). Taking $b(w) = 1$ gives back the bias and variance expressions for $\hat{C}_n^{(\mathrm{LL})}$ in Chen and Huang (2007) (in case of no smoothing at the first stage).

The improvements are obtained by shrinking the bandwidth through the function $b(\alpha, w) = \min\{w^\alpha, (1-w)^\alpha\}$. Different choices of α or different choices of shrinking factors are possible, but our extensive investigations showed that $b(w) = \min\{\sqrt{w}, \sqrt{1-w}\}$ is overall a very good choice. The choice of a possible optimal shrinking factor is an open question.

2.2. **Mirror-reflection kernel estimator.** Another version of a kernel estimator for the copula might be obtained by integration of the estimator of the density of the copula introduced and studied in Gijbels and Mielniczuk (1990). This estimator deals with the boundary problem by the technique known as mirror-reflection. If a multiplicative kernel $k(x,y) = k(x)\,k(y)$ is used, then the Mirror-Reflection estimate of the copula has a simple form

$$\hat{C}_n^{(\mathrm{MR})}(u,v) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{9}\left[K\left(\tfrac{u-\hat{U}_i^{(\ell)}}{h_n}\right) - K\left(\tfrac{-\hat{U}_i^{(\ell)}}{h_n}\right)\right]\left[K\left(\tfrac{v-\hat{V}_i^{(\ell)}}{h_n}\right) - K\left(\tfrac{-\hat{V}_i^{(\ell)}}{h_n}\right)\right], \tag{13}$$

where $\{(\hat{U}_i^{(\ell)}, \hat{V}_i^{(\ell)}),\, i=1,\ldots,n,\, \ell=1,\ldots,9\} = \{(\pm\hat{U}_i, \pm\hat{V}_i),\, (\pm\hat{U}_i, 2-\hat{V}_i),\, (2-\hat{U}_i, \pm\hat{V}_i),\, (2-\hat{U}_i, 2-\hat{V}_i),\, i=1,\ldots,n\}$.

The mirror-type estimator (13) faces the same 'corner bias' problem as the local linear estimator (8). To prevent this problem we can 'shrink' the bandwidth similarly as in (10) and propose

$$\hat{C}_n^{(\mathrm{MRS})}(u,v) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{9}\left[K\left(\tfrac{u-\hat{U}_i^{(\ell)}}{b(u)\,h_n}\right) - K\left(\tfrac{-\hat{U}_i^{(\ell)}}{b(u)\,h_n}\right)\right]\left[K\left(\tfrac{v-\hat{V}_i^{(\ell)}}{b(v)\,h_n}\right) - K\left(\tfrac{-\hat{V}_i^{(\ell)}}{b(v)\,h_n}\right)\right]. \tag{14}$$

2.3. **Transformation estimator.** The unboundedness of the densities of many copula families brings us back to Sklar's theorem in (1) and to the estimator (4) proposed in Fermanian et al. (2004).

To control the bias of this estimator in order to achieve weak convergence, we need the boundedness of the second order partial derivatives of the original joint distribution H. As the bivariate normal benchmark example shows, this condition may be considerably weaker than the requirement of the bounded second order derivatives of the underlying copula C.

A possible methodological objection to the estimator $\hat{C}_n^{(\mathrm{SE})}$, defined in (4), may be its dependence on the marginal distributions. This is confirmed by Monte-Carlo simulations which show that for a given copula the success of this estimator depends on the marginals crucially.

As the copula function is invariant to increasing transformations of the margins, it is possible to transform the original data to $X_i' = T_1(X_i)$ and $Y_i' = T_2(Y_i)$, where T_1 and T_2 are increasing functions, and then use (X_i', Y_i') instead of the original observations (X_i, Y_i) in the estimator $\hat{C}_n^{(\mathrm{SE})}$. The aim of the transformation is to simplify the kernel estimation of the joint distribution. As the direct choice of functions T_1, T_2 is difficult, we propose the following procedure. Let us first construct the uniform pseudo-observations $\hat{U}_i^{(\mathrm{E})} = \frac{n}{n+1}\,F_n(X_i)$ and $\hat{V}_i^{(\mathrm{E})} = \frac{n}{n+1}\,G_n(Y_i)$. Then for a given distribution function Φ put $\hat{S}_i = \Phi^{-1}(\hat{U}_i^{(\mathrm{E})})$ and $\hat{T}_i = \Phi^{-1}(\hat{V}_i^{(\mathrm{E})})$. Finally use these transformed pseudo-observations $(\hat{S}_i, \hat{T}_i)$ instead of the original observations (X_i, Y_i) in the estimator (5) of the joint distribution function. As we know the marginals to be given by the function Φ, the suggested estimator has in the case of multiplicative kernel the following simple formula

$$\hat{C}_n^{(\mathrm{T})}(u,v) = \frac{1}{n}\sum_{i=1}^{n} K\left(\tfrac{\Phi^{-1}(u)-\Phi^{-1}(\hat{U}_i^{(\mathrm{E})})}{h_n}\right) K\left(\tfrac{\Phi^{-1}(v)-\Phi^{-1}(\hat{V}_i^{(\mathrm{E})})}{h_n}\right) . \tag{15}$$

The advantage of this estimator is that it is not affected by the marginal distributions. Further bias calculations show that if we choose Φ, such that $\frac{\Phi'(x)^2}{\Phi(x)}$ is bounded, we take care of the 'corner bias problem' which is present if we try to estimate the joint distribution of pseudo-observations directly. The above condition is satisfied e.g. for Φ the normal cumulative distribution function.

2.4. **Main results.** The main theoretical contribution of this paper is the weak convergence of the kernel estimators $\hat{C}_n^{(\mathrm{LL})}$, $\hat{C}_n^{(\mathrm{LLS})}$, $\hat{C}_n^{(\mathrm{MR})}$, $\hat{C}_n^{(\mathrm{MRS})}$ and $\hat{C}_n^{(\mathrm{T})}$.

For notational convenience, let us denote $\hat{F}_n$ and $\hat{G}_n$ the estimates of the marginals which are used to construct pseudo-observations, that is in the following we will write $\hat{U}_i = \hat{F}_n(X_i)$ and $\hat{V}_i = \hat{G}_n(Y_i)$. For the weak convergence results we need these functions to be asymptotically equivalent to the empirical cumulative distribution functions F_n, G_n, i.e.

$$\sup_x |\hat{F}_n(x) - F_n(x)| = o_p(\tfrac{1}{\sqrt{n}}), \qquad \sup_y |\hat{G}_n(y) - G_n(y)| = o_p(\tfrac{1}{\sqrt{n}}), \tag{16}$$

which further implies the standard weak convergence of the processes $\sqrt{n}(\hat{F}_n - F)$ and $\sqrt{n}(\hat{G}_n - G)$ to particular Brownian bridges. For technical reason we will also suppose that the functions $\hat{F}_n$ and $\hat{G}_n$ are nondecreasing, which excludes higher order kernels (taking negative values) for the estimation of the marginals.

It is easy to see that (16) is satisfied if we define pseudo-observations as $\hat{U}_i = \frac{2\,n\,F_n(X_i)-1}{2\,n}$, $\hat{V}_i = \frac{2\,n\,G_n(Y_i)-1}{2\,n}$, or in a way given in (3).

If we decide for kernel smoothing of the marginals given in (6), then it is well known (see e.g. Lemma 7 of Fermanian et al. (2004)) that assumption (16) is met if there exists $\alpha > 0$ such that, uniformly in x,

$$F(x + b_{1n}) = F(x) + b_{1n}\, f(x) + o(b_{1n}^{1+\alpha}) \quad \text{with} \quad \sqrt{n}\, b_{1n}^{1+\alpha} \to 0 \, ,$$

where f denotes the derivative of F, and similarly for G involving b_{2n}.

Let $\mathbb{C}_n^{(\mathrm{LL})}$, $\mathbb{C}_n^{(\mathrm{LLS})}$, $\mathbb{C}_n^{(\mathrm{MR})}$, $\mathbb{C}_n^{(\mathrm{MRS})}$, $\mathbb{C}_n^{(\mathrm{T})}$ be suitably normalized empirical copula processes, i.e. for $(u,v) \in [0,1]^2$

$$\mathbb{C}_n^{(\cdot)} = \sqrt{n}\left\{ C_n^{(\cdot)} - C(u,v) \right\}.$$

The proof of the following theorem is given in Appendix A. The termininology on stochastic processes (such as for example pinned C-Brownian sheet) is taken from Tsukuhara (2005). We refer the reader to this reference for details on the concepts used.

Theorem 1. *Suppose that H has continuous marginal distribution functions and that the underlying copula function C has bounded second order partial derivatives on $[0,1]^2$. If $h_n =$*

$O(n^{-1/3})$ and (16) is satisfied, then the (kernel) copula processes $\mathbb{C}_n^{(LL)}$, $\mathbb{C}_n^{(MR)}$ converge weakly to the Gaussian process G_C in $\ell^\infty([0,1]^2)$ having representation

$$G_C(u,v) = B_C(u,v) - C_u(u,v)\, B_C(u,1) - C_v(u,v)\, B_C(1,v)\ , \tag{17}$$

where C_u and C_v denote the first order partial derivatives of C, and B_C is a two-dimensional pinned C-Brownian sheet on $[0,1]^2$, i.e. it is a centered Gaussian process with covariance function

$$\mathrm{E}\left[B_C(u,v)B_C(u',v')\right] = C(u \wedge u', v \wedge v') - C(u,v)C(u',v'). \tag{18}$$

While Theorem 1 requires boundedness of the second order partial derivatives of the copula C, the weak convergence result of Fermanian et al. (2004) for the estimator $\mathbb{C}_n^{(\mathrm{SE})}$ given by (4) requires boundedness of the second order derivatives of the original joint distribution function H. This may or may not be more stringent depending on the marginals. Unfortunately, Theorem 1 excludes many commonly-used families of copulas. The next theorem and Appendix D guarantee that the weak convergence of the proposed improved estimators $\mathbb{C}_n^{(\mathrm{LLS})}$, $\mathbb{C}_n^{(\mathrm{MRS})}$, $\mathbb{C}_n^{(\mathrm{T})}$ holds for commonly-used copulas such as Clayton, Gumbel, normal and Student copulas.

Remark. A careful reader may find out that all the published weak convergence results for the empirical estimator (2) or the smoothed empirical estimator (4) require smoothness of the first order partial derivatives C_u and C_v of the copulas C on $[0,1]^2$. But this smoothness assumption usually is not true for the families which do not have bounded second order partial derivatives (e.g. Clayton, Gumbel, normal, Student). For instance, the first order partial derivatives of the Clayton copula are not continuous in the corner point $(0,0)$. The second step of our proof given in Appendix B shows that it is sufficient to assume:

$$C_u, C_v \text{ are continuous in } [0,1]^2 \backslash \{(0,0),(0,1),(1,0),(1,1)\}. \tag{19}$$

Theorem 2. *Suppose that H has continuous marginal distribution functions and that the copula C has bounded second order partial derivatives on $(0,1)^2$ and satisfies (9) and (19). If $h_n = O(n^{-1/3})$ and (16) is satisfied, then the (kernel) copula processes $\mathbb{C}_n^{(LLS)}$ and $\mathbb{C}_n^{(MRS)}$ converge weakly to the Gaussian process G_C in $\ell^\infty([0,1]^2)$ given in Theorem 1.*

Moreover, if the functions Φ' and $\frac{\Phi'(x)^2}{\Phi(x)}$ are bounded then the above statement holds also for the process $\mathbb{C}_n^{(T)}$.

3. Finite sample comparisons

3.1. **Set up and performed comparisons.** In our simulation study we always use the Epanechnikov kernel $k(x) = \frac{3}{4}(1-x^2)\,\mathbb{I}\{|x| \leq 1\}$ and the bivariate multiplicative kernel $k(x,y) = k(x)\,k(y)$. The optimality of the Epanechnikov kernel in kernel density estimation was proven in Epanechnikov (1969). For background information on multivariate kernels see for example Wand and Jones (1995) and Fan and Gijbels (1996).

We investigate the performances of the estimators $C_n^{(\mathrm{E})}, \hat{C}_n^{(\mathrm{T})}, \hat{C}_n^{(\mathrm{LL})}, \hat{C}_n^{(\mathrm{MR})}$ and $\hat{C}_n^{(\mathrm{LLS})}$. We do not include the estimator $\hat{C}_n^{(\mathrm{SE})}$, defined in (4), because this estimator is too strongly influenced by the marginals, which makes the comparison difficult. For example, for a normal copula with normal marginals, the estimator $\hat{C}_n^{(\mathrm{SE})}$ usually does slightly better than its competitors. But for a normal copula with for example exponential marginals, the performance of $\hat{C}_n^{(\mathrm{SE})}$ is considerably worse than its competitors. We do not present results for the modification of the mirror-type estimator $\hat{C}_n^{(\mathrm{MRS})}$ either, since its performance was found to be close to that of the estimator $\hat{C}_n^{(\mathrm{LLS})}$.

The performances of the various estimators were evaluated using two criteria: a Kolmogorov-Smirnov distance KS_n and a Cramér-von Mises distance CM_n , i.e.

$$\mathrm{KS}_n = \sup_{u,v} |\hat{C}_n(u,v) - C(u,v)|, \qquad \mathrm{CM}_n = \sum_{i=1}^{n} \left[\hat{C}_n(\hat{U}_i,\hat{V}_i) - C(\hat{U}_i,\hat{V}_i)\right]^2 ,$$

where $\hat{C}_n$ stands for any of the investigated estimators, e.g. $\hat{C}_n^{(\mathrm{E})}$. The corresponding statistics are denoted accordingly, e.g. $\mathrm{KS}_n^{(\mathrm{E})}$ and $\mathrm{CM}_n^{(\mathrm{E})}$. Originally we included the mean integrated asymptotic error $Q_n = n \iint [\hat{C}_n(u,v) - C(u,v)]^2 du\, dv$ as well. Not surprisingly, this measure behaves similarly to the Cramér-von Mises distance, since $\mathrm{CM}_n \approx n \int\int (\hat{C}_n - C)^2 dC$, but it is not so sensitive to the bias of the underlying copula estimator. See also Section 4.

For computational reasons, the supremum in the Kolmogorov-Smirnov distance KS_n was replaced by a maximum over a grid of 101×101 points.

3.2. **Bandwidth choice.** The estimator $\hat{C}_n^{(\mathrm{LL})}$ involves bandwidths b_{n1} and b_{n2} (for estimation of the marginals) as well as a bandwidth h_n when using local linear fitting to estimate the copula. Preliminary simulation results confirmed the results of Chen and Huang (2007), that the estimator $\hat{C}_n^{(\mathrm{LL})}$ (as well as its modification $\hat{C}_n^{(\mathrm{LLS})}$) cannot be improved by smoothing the marginals. Therefore we simply work with the pseudo-observations $\hat{U}_i^{(\mathrm{E})} = \frac{n}{n+1} F_n(X_i)$ and $\hat{V}_i^{(\mathrm{E})} = \frac{n}{n+1} G_n(Y_i)$, with F_n and G_n the empirical cumulative distribution functions. This slightly differs from the strategy of strong undersmoothing recommended in Chen and Huang (2007), which more or less results in taking $\hat{U}_i = \frac{2n\, F_n(X_i)-1}{2n}$ and $\hat{V}_i = \frac{2n\, G_n(Y_i)-1}{2n}$. Nevertheless, the behavior of the resulting estimators is very similar.

For choosing the bandwidth h_n for $\hat{C}_n^{(\mathrm{LL})}$ and $\hat{C}_n^{(\mathrm{LLS})}$ we rely on the expressions for asymptotic bias, variance and MISE derived in Chen and Huang (2007). From the main (Asymptotic) terms in (11) and (12) we derive the asymptotic mean squared error of the copula estimator in a given point (u, v)

$$\mathrm{AMSE}\left\{\hat{C}_n(u,v)\right\} = \mathrm{AVar}\left\{\hat{C}_n(u,v)\right\} + \left[\mathrm{ABias}\left\{\hat{C}_n(u,v)\right\}\right]^2 . \tag{20}$$

An optimal bandwidth is obtained by minimization of $\iint \mathrm{AMSE}\{C_n(u,v)\}\, dC(u,v)$. As the true copula is unknown, this minimization cannot be carried out. A possible approach is then to consider a so-called reference copula. Chen and Huang (2007) proposed to use a t-copula as a reference copula. But as the second derivatives of the t-copula are not bounded, we experienced numerical difficulties and instabilities trying to apply this reference rule. We therefore decided to use Frank's copula, which has bounded second derivatives. The unknown parameter in Frank's copula family is estimated by inversion of Kendall's tau. The computational simplicity of this approach makes also the goodness-of-fit testing procedures, presented in Section 4, much more feasible.

Since the shrinkage of the bandwidth in the estimator $\hat{C}_n^{(\mathrm{LLS})}$ removes the problem of possible unboundedness of the second order partial derivatives, there are plenty of families of copulas to use as a reference copula for this estimator. For simplicity and for more appropriate comparisons we also use Frank's copula as a reference for $\hat{C}_n^{(\mathrm{LLS})}$.

The asymptotic expansions (11) and (12) hold for the mirror-type kernel estimators $\hat{C}_n^{(\mathrm{MR})}$ and $\hat{C}_n^{(\mathrm{MRS})}$ as well, and hence also here we rely on the same choice for h_n.

For the two improved estimators such a Frank copula based reference selection rule seems to give quite good performance (see Sections 3 and 4). A normal copula based reference rule tends to result in a too large bandwidth, whereas a Clayton copula based reference rule tends to give, on average, too small bandwidths. Thus Frank's reference rule seems to be a good compromise.

More problematic is the bandwidth choice for $\hat{C}_n^{(\mathrm{T})}$ as we do not have asymptotic expressions for bias and variance here. We tried to minimize the expected mean squared integrated error

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Big[\hat{H}_n(x,y) - H(x,y)\Big]^2 h(x,y)\, dx\, dy \tag{21}$$

taking a bivariate normal distribution H, with corresponding density h, as a reference distribution (see Jin and Shao (1999)). The resulting bandwidth selector however turned out to be too big. A possible explanation is that such a selection rule does not take into account that we rely on pseudo-observations $(\hat{U}_i, \hat{V}_i)$ instead of on the unobservable (U_i, V_i). In our simulation study we then used the above mentioned bandwidth divided by a factor two. This seems to be a reasonable ad-hoc solution. To further investigate the bandwidth selection problem for $\hat{C}_n^{(\mathrm{T})}$ we calculated the ratio of the bandwidth selected via (21) to the one selected via searching for a bandwidth that minimizes the criterion $\mathrm{KS}_n(h)$ (respectively $\mathrm{CM}_n(h)$) over a grid of h-values. Table 1 summarizes the obtained average ratios, for various values of Kendall's tau from 2 000 simulated samples. Note that the ratios stay quite stable across different families of copulas as well as for different sample sizes. This suggests that it may be possible to find a reliable reference-based rule for $\hat{C}_n^{(\mathrm{T})}$ as well.

TABLE 1. Average ratios of bandwidths for $\hat{C}_n^{(\mathrm{T})}$ selected from minimizing (21) and from criteria $\mathrm{KS}_n(h)$ and $CM_n(h)$ respectively, for different Kendall's τ and sample sizes n.

	Clayton				Frank				Normal			
	$\tau = 0.25$		$\tau = 0.75$		$\tau = 0.25$		$\tau = 0.75$		$\tau = 0.25$		$\tau = 0.75$	
	KS_n	CM_n	KS_n	CM_n	KS_n	CM_n	KS_n	CM_n	KS_n	CM_n	KS_n	CM_n
$n = 50$	1.20	1.28	1.53	2.00	1.23	1.38	1.53	2.04	1.28	1.37	1.26	1.62
$n = 150$	1.13	1.27	1.38	2.08	1.21	1.47	1.55	2.07	1.20	1.36	1.24	1.60

The simulation studies, reported below, showed a promising performance for the transformation estimator $\hat{C}_n^{(\mathrm{T})}$. A good bandwidth selection rule is missing for the moment, and is subject of further research.

3.3. **Simulation Results.** An extensive simulation study was carried out to compare the performances of all estimators using the performance measures KS_n (the Kolmogorov-Smirnov distance) and CM_n (the Cramér-von Mises distance). To illustrate our main findings we only report on results obtained for two simulation models

Model 1: Frank copula with Kendall's $\tau = 0.25$.
Model 2: Clayton copula with Kendall's $\tau = 0.75$.

Models 1 and 2 represent very different copula functions. The copula in Model 1 has bounded second order partial derivatives and presents a case of mild dependence, whereas the copula in Model 2 has unbounded second order partial derivatives and shows a strong dependence between X and Y. From each model we simulated 10 000 samples of sample size $n = 150$.

Figure 1 shows the boxplots of the performance measures KS_n and CM_n for Model 1 (top panels) and Model 2 (bottom panels). Note that for Model 1, the estimators $\hat{C}_n^{(\mathrm{LL})}$, $\hat{C}_n^{(\mathrm{MR})}$, $\hat{C}_n^{(\mathrm{LLS})}$ and $\hat{C}_n^{(\mathrm{T})}$ perform very comparable for the Cramér-von Mises distance measure. For the Kolmogorov-Smirnov performance measure, the estimator $\hat{C}_n^{(\mathrm{T})}$ performs slightly but significantly worse than the three other estimators. The latter is likely caused by the usage of a too small bandwidth, as can be anticipated by looking at the different results for the two performance measures and Table 1. For Model 2 (bottom panels) one clearly sees a better performance of the improved estimators $\hat{C}_n^{(\mathrm{LLS})}$ and $\hat{C}_n^{(\mathrm{T})}$ especially when looking at the performance measure CM_n. This is as to be expected since the copula in Model 2 has unbounded second order partial derivatives and the measure $\mathrm{CM}_n \approx n \int\int (\hat{C}_n - C)^2 dC$ is most affected by points with higher values of the copula density $c(u, v)$ (which usually correspond with points with higher values of the second order partial derivatives C_{uu} and C_{vv}). In other words the performance measure CM_n is more sensitive to the corner bias problem than the measure KS_n. For Model 1, there was no need for 'shrinking' the bandwidth since the copula function has bounded second order partial derivatives. Nevertheless, the 'shrinked-bandwidth' (improved) local linear estimator performs very well also for this model. The very promising performance of the transformation estimator $\hat{C}_n^{(\mathrm{T})}$ in view of the Cramér-von Mises distance CM_n may be partially understood by the fact that, in view of Table 1, our ad-hoc rule of bandwidth choice for the estimator $\hat{C}_n^{(\mathrm{T})}$ is almost optimal in that situation.

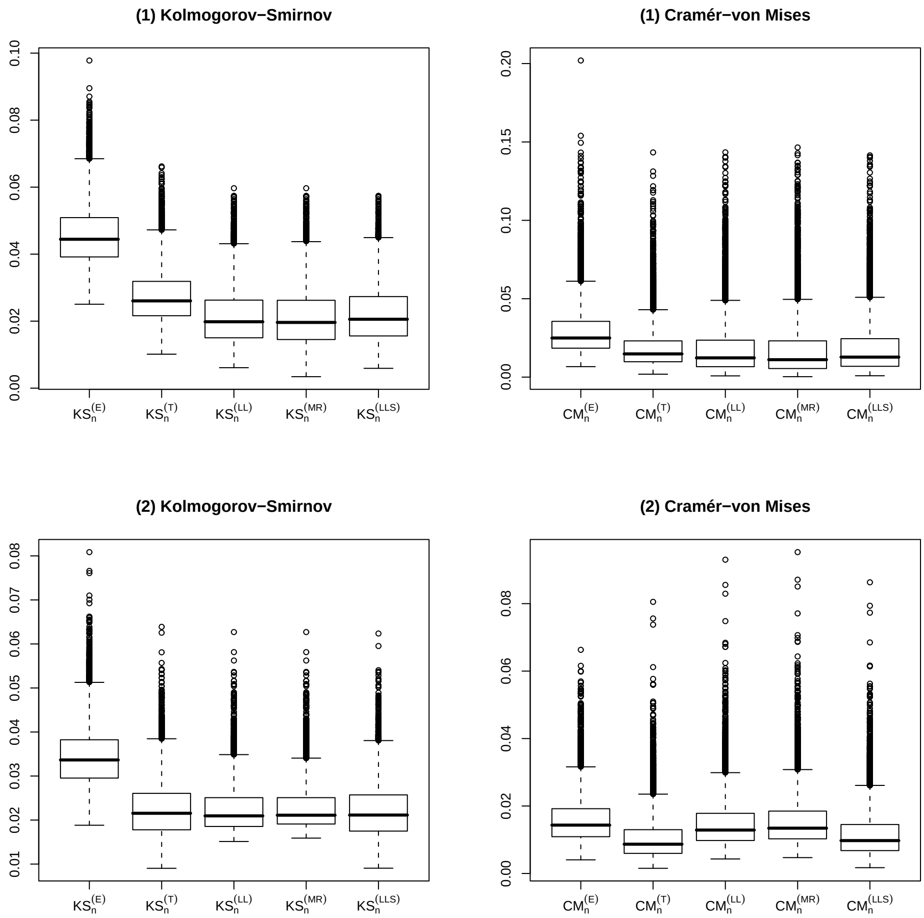

FIGURE 1. Boxplots of quantities KS_n and CM_n for different copula estimators. Top panels: Model 1; Bottom panels: Model 2.

From the more extensive simulation study we further report the following observations. All kernel copula estimators usually improve upon the empirical estimate $\hat{C}_n^{(\mathrm{E})}$. For copulas with bounded second order partial derivatives, the performances of the estimators $\hat{C}_n^{(\mathrm{LL})}$, $\hat{C}_n^{(\mathrm{MR})}$, and $C_n^{(\mathrm{E})}$ become very comparable, especially with increasing sample size. Overall, $\hat{C}_n^{(\mathrm{MR})}$ works slightly better for copulas with bounded second order partial derivatives (e.g. Frank, Farlie-Gumbel-Morgenstern, Ali-Mikhail-Haq, see Nelson (2006)) and mild dependence, with a significant improvement for copulas very close to independence copulas. On the other hand

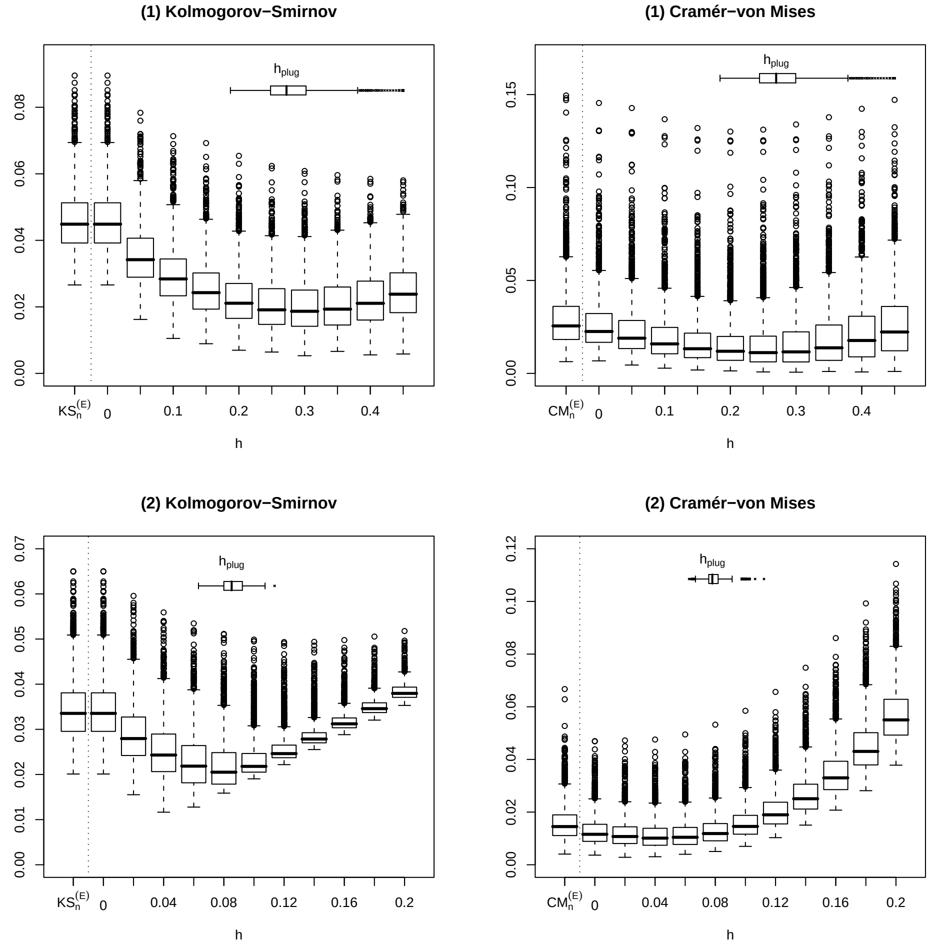

FIGURE 2. Boxplots of the quantities KS_n and CM_n for different values of fixed bandwidths for the estimator $\hat{C}_n^{(\mathrm{LL})}$, and boxplot (far left) for the estimator $\hat{C}_n^{(\mathrm{E})}$. Top panels: Model 1; Bottom panels: Model 2.

the local linear kernel estimator $\hat{C}_n^{(\mathrm{LL})}$ is preferable (compared to $\hat{C}_n^{(\mathrm{MR})}$) in the remaining cases.

To gain further insights in the kernel estimators we examined the dependence of these estimators on the bandwidth. We again use Models 1 and 2 to illustrate our findings. For brevity we present results only for the estimator $\hat{C}_n^{(\mathrm{LL})}$, since similar findings can be reported on for the other kernel estimators.

Figure 2 illustrates the performance of the copula $\hat{C}_n^{(\mathrm{LL})}$ with a fixed bandwidth h in view of the performance measures KS_n and CM_n for Models 1 and 2 (respectively top and bottom panels). For comparison purpose we also include (at the far left of the horizontal axis) the boxplot summarizing the results for the empirical copula $\hat{C}_n^{(\mathrm{E})}$. In addition, we provide in each picture a (vertical) boxplot that indicates the bandwidths selected for $\hat{C}_n^{(\mathrm{LL})}$ via (20).

Note that the effect of bandwidth choice is most noticeable from the Kolmogorov-Smirnov quantity KS_n. This is particularly true for the Clayton copula, Model 2. For Model 1, the estimator $\hat{C}_n^{(\mathrm{LL})}$ improves upon the empirical copula $\hat{C}_n^{(\mathrm{E})}$ for almost all h-values in the considered range of values. For Model 2 however, which presents a case of stronger dependence, a kernel estimator comes with a gain but only for a carefully selected bandwidth.

From the vertically displayed boxplots of bandwidths selected, we can further remark that a bandwidth selected via (20) works in fact quite satisfactory. This is particularly true in case of mild dependence and for copulas with bounded second derivatives (such as Model 1). It may lead to a slight oversmoothing in a situation of strong dependence and for copulas with unbounded second derivatives (cfr Model 2). It is worth mentioning though that the presented results for Model 2 are almost among the 'worst-case' scenarios here.

4. Goodness-of-Fit tests for copulas

When modelling multivariate data using copulas, a popular method is to estimate marginals nonparametrically and the copula in a parametric way. This requires choosing a suitable family of copulas for the data at hand, which is not an easy task. In this section we focus on testing the null hypothesis

$$H_0 : C \in \mathcal{C}_0,$$

where $\mathcal{C}_0 = \{C_\theta, \theta \in \Theta\}$ is a given parametric family of copulas.

Many testing methods have been proposed. See for example Chen and Huang (2007) and the review paper of Genest et al. (2008). The latter paper included a simulation study on classical goodness-of-fit measures such as the Kolmogorov-Smirnov and the Cramér-von Mises statistics, which we denote by (allowing a small abuse of previous notations)

$$\text{(22)} \qquad \mathrm{KS}_n^{(\mathrm{E})} = \sup_{u,v} |C_n^{(\mathrm{E})}(u,v) - C_{\hat{\theta}_n}(u,v)|, \qquad \mathrm{CM}_n^{(\mathrm{E})} = \sum_{i=1}^{n} [C_n^{(\mathrm{E})}(\hat{U}_i, \hat{V}_i) - C_{\hat{\theta}_n}(\hat{U}_i, \hat{V}_i)]^2$$

where $\hat{\theta}_n$ is an estimate of the unknown parameter θ_0 based on the inversion of the observed Kendall's tau.

The aim of this section is to investigate the size and power properties for testing procedures based on the test statistics $\mathrm{KS}_n^{(\mathrm{LL})}$, $\mathrm{KS}_n^{(\mathrm{LLS})}$, $\mathrm{CM}_n^{(\mathrm{LL})}$ $\mathrm{CM}_n^{(\mathrm{LLS})}$ computed by replacing $C_n^{(\mathrm{E})}$ in (22) with $\hat{C}_n^{(\mathrm{LL})}$ or $\hat{C}_n^{(\mathrm{LLS})}$. In addition we consider here the test statistic

$$\mathrm{Q}_n^{(\mathrm{E})} = \iint [C_n^{(\mathrm{E})}(u,v) - C_{\hat{\theta}_n}(u,v)]^2 \, du \, dv \ ,$$

and its $\hat{C}_n^{(\mathrm{LL})}$ and $\hat{C}_n^{(\mathrm{LLS})}$ versions. The double integral in the definition of $\mathrm{Q}_n^{(\cdot)}$ was approximated by a double sum over a grid of 101×101 points.

Since the asymptotic distributions of these test statistics are too complex, a parametric bootstrap is used. This procedure runs as follows

TABLE 2. Percentage of rejection of H_0 by various tests for samples of size $n = 150$ arising from different copula models with $\tau = 0.25$.

Copula	True	Cramér-von Mises			Kolmogorov-Smirnov			MISE		
under H_0	copula	$\mathrm{CM}_n^{(\mathrm{E})}$	$\mathrm{CM}_n^{(\mathrm{LL})}$	$\mathrm{CM}_n^{(\mathrm{LLS})}$	$\mathrm{KS}_n^{(\mathrm{E})}$	$\mathrm{KS}_n^{(\mathrm{LL})}$	$\mathrm{KS}_n^{(\mathrm{LLS})}$	$\mathrm{Q}_n^{(\mathrm{E})}$	$\mathrm{Q}_n^{(\mathrm{LL})}$	$\mathrm{Q}_n^{(\mathrm{LLS})}$
Clayton	Clayton	*4.9*	*4.3*	*4.9*	*5.1*	*3.8*	*4.5*	*5.0*	*4.9*	*5.6*
	Gumbel	86.7	91.3	**93.3**	61.1	81.8	**84.6**	84.1	91.7	**93.6**
	Frank	54.4	**63.5**	61.4	33.5	**64.1**	61.2	51.6	**62.2**	59.7
	Plackett	56.3	61.0	**63.5**	34.2	**57.5**	57.2	53.0	61.8	**62.7**
	Normal	49.9	59.0	**61.6**	28.2	**54.8**	52.8	46.6	59.6	**62.5**
	Student, 4 df	55.9	65.5	**67.0**	34.6	**40.8**	40.7	52.1	68.0	**69.5**
Gumbel	Clayton	72.2	85.2	**86.7**	48.3	77.5	**78.1**	74.0	83.3	**85.6**
	Gumbel	*4.9*	*5.1*	*5.6*	*5.0*	*4.9*	*4.7*	*4.5*	*4.9*	*5.4*
	Frank	14.7	18.2	**20.1**	11.0	18.2	**22.8**	**18.4**	15.7	17.3
	Plackett	13.7	16.7	**17.4**	9.5	17.7	**17.9**	16.2	15.7	**16.3**
	Normal	10.0	**15.6**	15.5	8.2	**16.8**	16.1	12.6	**15.5**	14.1
	Student, 4 df	12.8	21.5	**22.6**	7.3	11.3	**13.3**	14.3	23.3	**24.5**
Frank	Clayton	41.3	49.9	**52.1**	25.5	37.6	**39.1**	42.1	48.2	**50.8**
	Gumbel	31.8	47.5	**50.1**	18.1	**28.5**	28.3	25.4	45.7	**47.5**
	Frank	*4.7*	*4.9*	*4.9*	*5.1*	*4.9*	*4.5*	*4.6*	*4.6*	*5.1*
	Plackett	5.7	**6.9**	6.2	4.7	5.0	**5.8**	5.4	6.7	**7.0**
	Normal	8.5	**13.4**	12.6	9.7	**12.1**	11.5	7.6	**14.4**	12.8
	Student, 4 df	18.1	**34.5**	33.9	10.1	**18.1**	17.1	15.4	**36.6**	35.2
Normal	Clayton	34.3	36.9	**40.1**	18.3	25.6	**28.7**	36.2	36.0	**38.9**
	Gumbel	26.1	**35.1**	33.7	12.0	**20.0**	18.7	21.0	**34.0**	32.6
	Frank	**7.6**	3.7	3.9	**7.1**	3.4	4.3	**7.9**	4.2	4.3
	Plackett	**8.1**	4.5	5.4	**8.0**	3.6	5.1	**8.6**	5.7	6.1
	Normal	*4.8*	*5.2*	*5.5*	*5.0*	*5.2*	*4.5*	*5.8*	*4.9*	*5.1*
	Student, 4 df	11.7	**14.8**	14.5	6.5	**9.2**	9.0	9.5	17.9	**19.1**
Student	Clayton	29.0	29.5	**34.1**	22.1	33.2	**36.3**	**32.8**	26.2	30.2
	Gumbel	20.7	26.1	**26.6**	12.1	22.7	**25.5**	17.9	17.9	**22.7**
	Frank	**9.0**	7.1	6.1	9.1	8.0	**9.2**	**9.7**	3.6	3.4
	Plackett	**7.6**	6.4	4.5	**8.6**	6.6	7.7	**7.6**	3.9	2.9
	Normal	**4.7**	4.3	3.5	6.4	**6.9**	6.7	**4.9**	3.9	2.9
	Student, 4 df	*4.7*	*4.7*	*4.9*	*5.3*	*4.8*	*4.7*	*4.4*	*4.7*	*5.4*

(1) By inversion of the empirical Kendall's τ, estimate the unknown parameter θ of the null hypothesis family by $\hat{\theta}_n$ and compute the test statistic $\mathrm{KS}_n^{(\cdot)}$ (where the superscript $(\cdot)$ refers to any of the considered estimators of the copula).

(2) Generate $\{(U_i^*, V_i^*)\}_{i=1}^n$ from the copula $C_{\hat{\theta}_n}$ and use them as original observations to compute $\hat{\theta}_n^*$ and $\mathrm{KS}_n^{*(\cdot)}$.

(3) Repeat Step (2) B-times.

(4) Estimate the p-values as

$$p_{\mathrm{KS}_n^{(\cdot)}} = \frac{1 + \#\left\{\mathrm{KS}_n^{*(\cdot)} \geq \mathrm{KS}_n^{(\cdot)}\right\}}{B+1}.$$

See Davison and Hinkley (1997).

For any of the other test statistics we proceed similarly replacing $\mathrm{KS}_n^{(\cdot)}$ by $\mathrm{CM}_n^{(\cdot)}$ or $Q_n^{(\cdot)}$.

According to Genest and Rémillard (2008), the validity of this bootstrap procedure requires the weak convergence of the copula processes $\mathbb{C}_n^{(\mathrm{LL})}$ and $\mathbb{C}_n^{(\mathrm{LLS})}$. For the latter process the

TABLE 3. Percentage of rejection of H_0 by various tests for samples of size $n = 150$ arising from different copula models with $\tau = 0.50$.

Copula under H_0	True copula	Cramér-von Mises $\mathrm{CM}_n^{(\mathrm{E})}$	Cramér-von Mises $\mathrm{CM}_n^{(\mathrm{LL})}$	Cramér-von Mises $\mathrm{CM}_n^{(\mathrm{LLS})}$	Kolmogorov-Smirnov $\mathrm{KS}_n^{(\mathrm{E})}$	Kolmogorov-Smirnov $\mathrm{KS}_n^{(\mathrm{LL})}$	Kolmogorov-Smirnov $\mathrm{KS}_n^{(\mathrm{LLS})}$	MISE $\mathrm{Q}_n^{(\mathrm{E})}$	MISE $\mathrm{Q}_n^{(\mathrm{LL})}$	MISE $\mathrm{Q}_n^{(\mathrm{LLS})}$
Clayton	Clayton	*5.3*	*5.2*	*5.4*	*5.5*	*5.3*	*5.7*	*4.8*	*5.6*	*5.8*
	Gumbel	99.9	100.0	99.9	98.9	99.5	**99.9**	100.0	100.0	99.9
	Frank	95.9	**96.4**	96.2	82.5	**98.3**	95.6	91.1	**93.3**	92.3
	Plackett	95.6	**96.7**	96.0	75.3	**94.7**	89.7	92.3	**95.7**	94.5
	Normal	94.4	96.3	**96.9**	75.0	**91.9**	91.1	89.9	94.5	**94.9**
	Student, 4 df	94.9	96.7	**97.2**	77.9	86.4	**88.2**	92.7	96.2	**96.3**
Gumbel	Clayton	99.6	99.7	99.7	94.3	98.7	**98.8**	**98.9**	99.3	99.1
	Gumbel	*4.5*	*4.7*	**4.9**	*5.0*	*4.9*	*5.4*	*4.8*	*3.9*	*4.6*
	Frank	40.5	**48.2**	39.7	29.6	**48.1**	40.4	**41.1**	35.3	30.6
	Plackett	29.4	**33.3**	30.7	18.9	**26.5**	22.8	**31.0**	30.0	27.6
	Normal	18.8	**26.4**	25.1	14.6	**25.3**	23.8	**22.3**	22.1	21.8
	Student, 4 df	22.3	27.8	**29.2**	11.7	**19.1**	16.5	23.6	28.5	**29.0**
Frank	Clayton	89.6	88.9	**91.9**	68.1	72.7	**75.5**	85.1	**86.3**	85.5
	Gumbel	63.8	71.0	**74.1**	39.3	44.6	**47.3**	50.5	**68.9**	65.7
	Frank	*5.3*	*4.9*	*5.2*	*5.1*	*5.2*	*5.0*	*5.1*	*4.9*	*5.0*
	Plackett	8.4	10.4	**12.1**	5.4	**6.9**	6.7	8.3	**15.2**	8.5
	Normal	19.6	26.0	**29.5**	17.6	**26.9**	25.5	16.5	**25.7**	17.5
	Student, 4 df	35.0	44.9	**52.8**	17.9	27.8	**29.4**	29.0	**51.0**	46.2
Normal	Clayton	**83.0**	78.3	82.9	55.8	**66.6**	66.1	**79.6**	76.1	79.4
	Gumbel	41.7	39.5	**44.3**	18.3	22.7	**26.6**	32.4	39.6	**41.2**
	Frank	**21.2**	20.1	14.9	**15.1**	11.8	11.0	**19.3**	14.4	10.5
	Plackett	**12.0**	7.4	7.8	**7.7**	5.8	4.5	**13.4**	11.9	9.9
	Normal	*4.8*	*5.1*	*5.5*	*4.7*	*5.0*	*5.4*	*5.8*	*4.5*	*4.6*
	Student, 4 df	8.1	6.4	**8.3**	4.0	4.5	**4.8**	7.6	11.5	**12.5**
Student	Clayton	80.6	78.4	**83.7**	62.1	74.9	**76.5**	**79.0**	72.3	75.0
	Gumbel	36.5	39.1	**39.6**	20.5	31.3	**32.4**	25.7	25.7	**31.0**
	Frank	28.5	**30.5**	23.1	18.8	18.0	**20.8**	**25.0**	15.0	11.3
	Plackett	**13.5**	**13.5**	8.9	**10.0**	6.9	7.3	**11.0**	6.8	6.6
	Normal	5.1	**6.4**	5.5	7.6	**7.8**	7.5	**5.1**	3.4	4.0
	Student, 4 df	*4.7*	*4.9*	*5.1*	*4.9*	*5.0*	*4.9*	*5.2*	*4.1*	*4.7*

weak convergence is justified for all copula families considered in our simulation study, by Theorem 2. In contrast, only for Frank's copula the condition of Theorem 1 is satisfied when dealing with the weak convergence of the process $\mathbb{C}_n^{(\mathrm{LL})}$. Lemma 5 of Appendix C shows that the test based on $\hat{C}_n^{(\mathrm{LL})}$ holds asymptotically the level even for the other families $\mathcal{C}_0$ appearing in the simulation study.

The setup of our simulation study closely follows this of Genest et al. (2008). The sample size is $n = 150$ and we take 999 number of bootstrap samples. Three values of Kendal's tau are considered, namely $\tau = 0.25, 0.50, 0.75$, for the following copula families: Clayton, Gumbel, Frank, Normal and Student with four degrees of freedom (df). We use the R-computing environment, version 2.5.0 (see R Development Core Team (2007)), with copula package (see Yan (2007)). For approximating the level of the test (i.e. under the null hypothesis) we use 6 000 repetitions. The estimated powers of the test statistics are based on 1 500 repetitions.

The results of the simulations are presented in Tables 2, 3 and 4. For ease of the reader the estimated values for the size of the test statistics are presented in italics. Furthermore,

TABLE 4. Percentage of rejection of H_0 by various tests for samples of size $n = 150$ arising from different copula models with $\tau = 0.75$.

Copula under H_0	True copula	Cramér-von Mises $\mathrm{CM}_n^{(\mathrm{E})}$	$\mathrm{CM}_n^{(\mathrm{LL})}$	$\mathrm{CM}_n^{(\mathrm{LLS})}$	Kolmogorov-Smirnov $\mathrm{KS}_n^{(\mathrm{E})}$	$\mathrm{KS}_n^{(\mathrm{LL})}$	$\mathrm{KS}_n^{(\mathrm{LLS})}$	MISE $\mathrm{Q}_n^{(\mathrm{E})}$	$\mathrm{Q}_n^{(\mathrm{LL})}$	$\mathrm{Q}_n^{(\mathrm{LLS})}$
Clayton	Clayton	*5.3*	*5.6*	*5.3*	*5.1*	*4.9*	*4.6*	*3.3*	*4.6*	*4.0*
	Gumbel	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0
	Frank	**98.8**	98.3	98.5	83.7	**98.1**	97.2	94.0	**95.4**	95.1
	Plackett	99.5	99.2	**99.9**	85.1	90.6	**93.9**	97.0	98.3	**99.1**
	Normal	99.8	99.5	**99.9**	90.5	93.6	**96.8**	98.0	98.4	**99.2**
	Student, 4 df	99.9	99.7	**100.0**	92.7	91.4	**97.7**	98.5	**99.5**	**99.5**
Gumbel	Clayton	**99.9**	99.5	**99.9**	95.8	**98.5**	**98.5**	**99.6**	99.1	99.0
	Gumbel	*4.5*	*4.6*	*4.8*	*4.7*	*4.6*	*4.9*	*5.0*	*3.2*	*3.7*
	Frank	53.3	**54.5**	47.0	25.6	38.1	**38.3**	**40.7**	26.6	23.7
	Plackett	**24.3**	23.5	19.1	6.6	**9.3**	**9.3**	**29.5**	24.9	25.1
	Normal	12.4	13.1	**13.9**	11.3	**13.6**	13.3	**12.3**	8.3	7.7
	Student, 4 df	15.6	15.7	**20.1**	8.5	10.4	**10.8**	**16.2**	15.1	14.9
Frank	Clayton	96.7	91.3	**96.9**	57.6	63.7	**64.9**	**90.8**	86.2	89.0
	Gumbel	81.6	80.5	**87.8**	36.9	**41.7**	39.7	61.1	73.7	**75.3**
	Frank	*4.4*	*4.6*	*4.4*	*4.7*	*4.5*	*4.8*	*4.7*	*3.0*	*3.7*
	Plackett	19.6	20.7	**27.5**	5.9	**8.7**	7.5	34.5	45.7	**48.3**
	Normal	40.7	41.1	**52.7**	28.5	**33.3**	30.2	31.9	38.7	**42.9**
	Student, 4 df	58.4	57.9	**72.6**	26.7	**33.6**	30.3	50.6	63.1	**64.6**
Normal	Clayton	**93.4**	88.9	90.5	66.7	77.9	**79.7**	**88.4**	83.3	82.6
	Gumbel	41.1	38.3	**43.4**	13.0	19.6	**20.9**	24.3	33.5	**35.0**
	Frank	46.1	**46.7**	37.7	17.5	18.2	**22.9**	**31.0**	24.1	18.4
	Plackett	**15.2**	11.6	9.2	3.1	3.1	**4.0**	24.4	**27.7**	23.7
	Normal	*4.7*	*4.8*	*4.4*	*4.4*	*4.6*	*4.7*	*5.2*	*3.5*	*3.5*
	Student, 4 df	**6.9**	6.3	6.8	**4.7**	3.9	4.2	7.0	**10.1**	9.6
Student	Clayton	**92.8**	89.4	89.9	73.7	84.4	**86.7**	**85.8**	75.6	74.8
	Gumbel	**37.3**	34.5	37.0	17.3	26.3	**26.9**	18.4	18.4	**21.6**
	Frank	**52.2**	51.8	45.5	24.5	28.1	**33.6**	**30.4**	20.7	14.5
	Plackett	**16.4**	15.7	10.3	4.3	3.9	**5.8**	5.8	**12.7**	11.6
	Normal	4.5	**4.7**	3.5	6.3	6.9	**7.8**	**2.5**	2.1	1.9
	Student, 4 df	*4.3*	*4.9*	*4.4*	*4.7*	*5.1*	*4.9*	*4.8*	*3.6*	*3.1*

for each testing problem we highlighted the 'best' power performances using bold characters. Readers should be aware of the fact that these estimated powers and sizes (using respectively 1 500 and 6 000 repetitions) are of course subject to Monte Carlo approximation errors. A conservative upper bound (relying on a binomial distribution with parameters B and p) for these approximations errors (in terms of standard deviation) is for the size estimates 0.28% (using $B = 6\,000$ and $p = 0.05$) and for the power estimates 1.29% (using $B = 1\,500$ and $p = 0.5$, for getting to an upper bound).

A summary of conclusions from the simulations results is as follows.

- The use of a kernel estimator (e.g $\hat{C}_n^{(\mathrm{LLS})}$) in goodness-of-fit testing seems to be promising in case true copulas are in the Clayton, Gumbel and Frank families. We consistently improve upon the power for the Kolmogorov-Smirnov test.
- If a kernel estimator improves upon the power, it is most noticeable when the dependence is weaker, and is greatest for the Kolmogorov-Smirnov test.

- The power of the test statistics Q_n is usually somewhere between the power of Kolmogorov-Smirnov and the Cramér-von Mises test statistics.
- The power of the test statistics based on the improved estimator $\hat{C}_n^{(\mathrm{LLS})}$ is usually higher than this for test statistics based on $\hat{C}_n^{(\mathrm{LL})}$ for alternatives with unbounded second order partial derivatives.
- For a true Frank copula and a Cramér-von Mises test statistic, the estimator $\hat{C}_n^{(\mathrm{LLS})}$ is usually the best choice.
- The use of kernel estimators seems to be promising for Archimedean families of copulas (Clayton, Frank, Gumbel), but is somewhat questionable for elliptical families of copulas (Normal, Student). Although kernel estimators may improve the power against Clayton and Gumbel alternatives, a loss in power is noticed for Frank alternatives. This holds in particular for $\hat{C}_n^{(\mathrm{LLS})}$-based statistics.

Appendix A – Proof of Theorem 1

For simplicity we will suppress the dependence on n in the notation of pseudo-observations $(\hat{U}_i, \hat{V}_i)^{\mathsf{T}}$ and write simply $\hat{U}_i = \hat{F}_n(X_i) = \hat{F}_n(F^{-1}(U_i))$ and $\hat{V}_i = \hat{G}_n(Y_i) = \hat{G}_n(G^{-1}(V_i))$, where (U_i, V_i) have a joint distribution function given by the copula C.

As our proof is a straightforward adaptation of the ideas used in van der Vaart and Wellner (2007), we would like to clarify one point. In the following we will encounter the expectations of the form $\mathrm{E}\, g(\hat{U}, \hat{V})$, where g is a measurable function on $[0,1]^2$ and $\hat{U} = \hat{F}_n(F^{-1}(U))$, $\hat{V} = \hat{G}_n(G^{-1}(V))$. In these type of expectations the estimators of the marginal distribution functions $\hat{F}_n, \hat{G}_n$ are considered to be fixed (nonrandom) functions and the expectation is taken only with respect to (U, V) with joint distribution given by the copula C. Formally

$$\mathrm{E}\, g(\hat{U}, \hat{V}) = \mathrm{E}_{U,V} \left[g(\hat{U}, \hat{V}) \Big| (X_1, Y_1), \ldots, (X_n, Y_n) \right] ,$$

whenever the integral on the right-hand side exists.

A1. **Weak convergence of the process $\mathbb{C}_n^{(\mathbf{LL})}$.** In view of the previous remark we decompose

$$\text{(A1)} \qquad \begin{aligned} \sqrt{n}(\hat{C}_n^{(\mathrm{LL})}(u,v) - C(u,v)) &= \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{n} K_{u,h_n}(\tfrac{u-\hat{U}_i}{h_n}) K_{v,h_n}(\tfrac{v-\hat{V}_i}{h_n}) - C(u,v) \right] \\ &= A_n^{h_n}(u,v) + B_n(u,v) + C_n^{h_n}(u,v), \end{aligned}$$

where

$$\text{(A2)} \quad \begin{aligned} A_n^{h_n}(u,v) = \frac{1}{\sqrt{n}} \Bigg[\sum_{i=1}^{n} & K_{u,h_n}(\tfrac{u-\hat{U}_i}{h_n}) K_{v,h_n}(\tfrac{v-\hat{V}_i}{h_n}) - \mathbb{I}\{U_i \leq u, V_i \leq v\} \\ & - \mathrm{E}\left(K_{u,h_n}(\tfrac{u-\hat{U}}{h_n}) K_{v,h_n}(\tfrac{v-\hat{V}}{h_n}) - C(u,v) \right) \Bigg] \end{aligned}$$

and

$$B_n(u,v) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\mathbb{I}\{U_i \le u, V_i \le v\} - C(u,v)\right], \tag{A3}$$

$$C_n^{h_n}(u,v) = \sqrt{n}\,\mathrm{E}\left[K_{u,h_n}(\tfrac{u-\hat{U}}{h_n})K_{v,h_n}(\tfrac{v-\hat{V}}{h_n}) - C(u,v)\right]. \tag{A4}$$

Our proof will be divided into two steps. Firstly we will show, in Step 1, that $\sup_{u,v}|A_n^{h_n}| = o_p(1)$. Then we will prove, in Step 2, that

$$\sup_{u,v}\left|C_n^{h_n}(u,v) - \partial_1 C(u,v)\sqrt{n}[F_n^*(u) - u] - \partial_2 C(u,v)\sqrt{n}[G_n^*(v) - v]\right| = o_P(1),$$

where $F_n^*(u) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{U_i \le u\}$ and $G_n^*(v) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{V_i \le v\}$. The convergence of the smoothed copula process $\mathbb{C}_n^{(\mathrm{LL})} = \sqrt{n}(\hat{C}_n^{(\mathrm{LL})} - C)$ to a Gaussian process given by (17) will now follow by the results of Fermanian et al. (2004).

Step 1

We consider the following class of functions from $[0,1]^2$ to $[0,1]$

$$\mathcal{F} = \Big\{(w_1,w_2) \mapsto K_{u,h}\left(\tfrac{u-\zeta_1(w_1)}{h}\right) K_{v,h}\left(\tfrac{v-\zeta_2(w_2)}{h}\right), \quad (u,v)\in[0,1]^2, h\in[0,\tfrac{1}{4}],\ \zeta_1,\zeta_2 : [0,1]\to[0,1]\ \text{nondecreasing}.\Big\} \tag{A5}$$

As each function f from $\mathcal{F}$ is characterized by a quintuple (u,v,h,ζ_1,ζ_2), the empirical process indexed by $\mathcal{F}$ can be written as

$$Z_n(f) = Z_n(u,v,h,\zeta_1,\zeta_2) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} K_{u,h}\left(\tfrac{u-\zeta_1(U_i)}{h}\right) K_{v,h}\left(\tfrac{v-\zeta_2(V_i)}{h}\right).$$

Put $\bar{Z}_n = Z_n - \mathrm{E}\,Z_n$ and note that

(A6)

$$A_n^{h_n}(u,v) = \bar{Z}_n(f_1^n) - \bar{Z}_n(f_2), \quad \text{where } f_1^n = (u,v,h_n,F_n(F^{-1}),G_n(F^{-1})),\ f_2 = (u,v,0,\mathbb{I},\mathbb{I})$$

with $\mathbb{I}$ being the identity function on the interval $[0,1]$.

Lemma 1, which is given below, states that the set of functions $\mathcal{F}$ is Donsker. Indeed, $\mathcal{F}$ is a subset of $\mathcal{F}^*$ in Lemma 1, taking $b(\cdot) = 1$ and $u_0 = u$ and $v_0 = v$. This implies the weak convergence of the process $\bar{Z}_n(f), f \in \mathcal{F}$, which further implies that the process $\bar{Z}_n$ is asymptotically uniformly ρ-equicontinuous in probability (see pp.37-41 of van der Vaart and Wellner (1996)) with semimetric ρ given by

$$\rho(f,f') = \mathrm{E}\left[K_{u,h}\left(\tfrac{u-\zeta_1(U)}{h}\right) K_{v,h}\left(\tfrac{v-\zeta_2(V)}{h}\right) - K_{u',h'}\left(\tfrac{u'-\zeta_1'(U)}{h'}\right) K_{v',h'}\left(\tfrac{v'-\zeta_2'(V)}{h'}\right)\right]^2.$$

Using this asymptotic uniform ρ-equicontinuity and (A6) we get that $\sup_{u,v}|A_n^{h_n}| = o_p(1)$ provided that $\sup_{u,v}\rho(f_1^n, f_2)$ converges to zero in probability, where f_1^n and f_2 are given in (A6) (for details consult the proof in van der Vaart (1994)).

Put $M = \sup_{u,h,x} |k_{u,h}(x)|$, where $k_{u,h}$ is defined in (7) and denote

$$A_\varepsilon = \Big\{ |\hat{U} - U| > \varepsilon \text{ or } |\hat{V} - V| > \varepsilon \Big\}.$$

The consistency of $\hat{F}_n$ and $\hat{G}_n$ yields that for every $\varepsilon > 0$, for all sufficiently large n,

$$P\left[\max\left\{ \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)|, \sup_{y\in\mathbb{R}} |\hat{G}_n(y) - G(y)| \right\} > \varepsilon \right] < \varepsilon,$$

which further implies that for all sufficiently large n

$$P(A_\varepsilon) = P\left[\max\Big\{ |\hat{F}_n(X) - F(X)|, |\hat{G}_n(Y) - G(Y)| \Big\} > \varepsilon \right] < \varepsilon. \tag{A7}$$

Now we can bound

$$\begin{aligned}
\rho(f_1^n, f_2) &= \mathrm{E}\left[K_{u,h_n}(\tfrac{u-\hat{U}}{h_n})\, K_{v,h_n}(\tfrac{v-\hat{V}}{h_n}) - \mathbb{I}\{U \le u, V \le v\} \right]^2 \\
&\le \mathrm{E}\left[K_{u,h}(\tfrac{u-\hat{U}}{h})\, K_{v,h}(\tfrac{v-\hat{V}}{h}) - \mathbb{I}\{U \le u, V \le v\} \right]^2 \mathbb{I}_{A_\varepsilon^c} + \mathbb{I}_{A_\varepsilon} \\
&\le M^4\, \mathrm{E}\Big[|\mathbb{I}\{\hat{U} \le u - h_n\} - \mathbb{I}\{U \le u\}| + |\mathbb{I}\{\hat{U} \le u + h_n\} - \mathbb{I}\{U \le u\}| \\
&\qquad + |\mathbb{I}\{\hat{V} \le v - h_n\} - \mathbb{I}\{V \le v\}| + |\mathbb{I}\{\hat{V} \le v + h_n\} - \mathbb{I}\{V \le v\}| \Big] \mathbb{I}_{A_\varepsilon^c} + \mathbb{I}_{A_\varepsilon} \\
&\le 4\, M^4\, (\varepsilon + h_n) + \mathbb{I}_{A_\varepsilon}.
\end{aligned}$$

As the above bound holds uniformly in (u, v) and by (A7) for all sufficiently large n we have $P(A_\varepsilon) < \varepsilon$. Since ε can be arbitrarily small, this implies $\sup_{u,v} \rho(f_1^n, f_2) = o_p(1)$.

Lemma 1. *Suppose that the function k is of bounded variation and $\int k(x)\, dx = 1$. Then the set of functions from $[0,1]^2$ to $[0,1]$*

$$\begin{aligned}
\mathcal{F}^* = \Big\{ (w_1, w_2) \mapsto K_{u_0,h}\left(\tfrac{u-\zeta_1(w_1)}{b(u_0)\, h} \right) K_{v_0,h}\left(\tfrac{v-\zeta_2(w_2)}{b(v_0)\, h} \right), \\
(u_0, v_0), (u, v) \in [0,1]^2, h \in [0, \tfrac{1}{4}],\ \zeta_1, \zeta_2 : [0,1] \to [0,1] \textit{ nondecreasing}. \Big\}
\end{aligned}$$

where $b(w) = 1$ or $b(w) = \min\{\sqrt{w}, \sqrt{1-w}\}$, is Donsker.
Consequently, the family $\mathcal{F}$ in (A5) *is Donsker.*

Proof. Note that the class of functions

$$\mathcal{G}_1 = \{(w_1, w_2) \mapsto \mathbb{I}\{\zeta_1(w_1) \le a, \zeta_2(w_2) \le b\},\ a, b \in \mathbb{R}, \zeta_1, \zeta_2 : [0,1] \to [0,1]\, \text{nondecreasing}\}$$

is a subset of the class of indicators

$$\mathcal{G}_2 = \{(w_1, w_2) \mapsto \mathbb{I}\{w_1 < (\le) a, w_2 < (\le) b\},\ a, b \in \mathbb{R}\}.$$

But this implies that $\mathcal{G}_1$ is a Donsker class (see Example 2.5.4 of van der Vaart and Wellner (1996)).

As the set $\mathcal{G}_1$ is closed under translation, then by the beginning of the proof of van der Vaart (1994) we know that the set of functions

$$\mathcal{H} = \left\{ \int f(x+y)\, d\mu(y), f \in \mathcal{G}_1, \mu \in \mathcal{M}_B \right\}, \tag{A8}$$

is a Donsker class, where $\mathcal{M}_B$ is a family of all signed measures (on $\mathbb{R}^2$) of total mass bounded by a fixed constant B.

Let us introduce the set of signed measures

$$\mathcal{M}_0 = \left\{ (-\infty, w_1] \times (-\infty, w_2] \mapsto K_{u_0,h}(\tfrac{w_1}{b(u_0)\,h})\, K_{v_0,h}(\tfrac{w_2}{b(v_0)\,h}),\ (u_0, v_0) \in [0,1]^2, h \in [0, \tfrac{1}{4}] \right\}.$$

If k is of bounded variation, then by taking sufficiently large B we ensure that $\mathcal{M}_0 \subset \mathcal{M}_B$. Further, if x stands for (w_1, w_2) and y for (y_1, y_2), then for $f \in \mathcal{G}_1$ and $\mu \in \mathcal{M}_0$ we get

$$\begin{aligned}\int f(x+y) d\,\mu(y) &= \iint \mathbb{I}\{\zeta_1(w_1) + y_1 \le u, \zeta_2(w_2) + y_2 \le v\}\, d\, \left(K_{u_0,h}(\tfrac{y_1}{b(u_0)\,h})\, K_{v_0,h}(\tfrac{y_2}{b(v_0)\,h}) \right) \\ &= K_{u_0,h}\left(\tfrac{u-\zeta_1(w_1)}{b(u_0)\,h}\right) K_{v_0,h}\left(\tfrac{v-\zeta_2(w_2)}{b(v_0)\,h}\right).\end{aligned}$$

is a Donsker class. As $\mathcal{F}^*$ includes $\mathcal{F}$ of (A5) (consider $u_0 = u$ and $v_0 = v$), the family $\mathcal{F}$ is a Donsker class as well. □

Step 2

Now we can turn our attention to the process $C_n^{h_n}$ given by (A4). For $(u, v) \in \mathbb{R}^2$ define

$$C^*_{\tilde{F},\tilde{G}}(u, v) = C(F(\tilde{F}^{-1}(u^*)), G(\tilde{G}^{-1}(v^*))),$$

where $w^* = \max\{\min\{w, 1\}, 0\}$. Note that

$$\begin{aligned}\mathrm{E}\,\mathbb{I}\{\hat{U} \le u, \hat{V} \le v\} &= \mathrm{E}\,\mathbb{I}\{U \le F(\hat{F}_n^{-1}(u)), \hat{V} \le G(\hat{G}_n^{-1}(u))\} + O(n^{-1}) \\ &= C^*_{\hat{F}_n,\hat{G}_n}(u, v) + O(n^{-1}),\end{aligned}$$

where the remainder term $O(n^{-1})$ disappears if we do some smoothing on the first stage, that is if $b_{1n}, b_{2n} > 0$. As

$$\begin{aligned}\mathrm{E}\, K_{u,h}(\tfrac{u-\hat{U}}{h}) K_{v,h}(\tfrac{v-\hat{V}}{h}) &= \mathrm{E} \int_{-1}^{1}\int_{-1}^{1} \mathbb{I}\{\hat{U}_i \le u - t\,h_n, \hat{V}_i \le u - s\,h_n\}\, k_{u,h}(s) k_{v,h}(t)\, dt\, ds \\ &= \int_{-1}^{1}\int_{-1}^{1} C^*_{\hat{F}_n,\hat{G}_n}(u - t\,h_n, v - s\,h_n)\, k_{u,h}(s) k_{v,h}(t)\, dt\, ds + O(n^{-1}),\end{aligned}$$

it will be useful to have a closer look at the process $\{C^*_{\hat{F}_n,\hat{G}_n}(u, v) \in [0,1]^2\}$. In the following we will prove that uniformly in (u, v)

$$\begin{aligned}&\sqrt{n}(C^*_{\hat{F}_n,\hat{G}_n}(u, v) - C(u, v)) \\ &\quad = -C_u(u, v) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\mathbb{I}\{U_i \le u\} - u] - C_v(u, v) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\mathbb{I}\{V_i \le v\} - v] + o_p(1).\end{aligned} \tag{A9}$$

Let $D[a,b]$ be the Banach space of all cadlag functions on an interval $[a,b]$ equipped with the uniform norm.

Lemma 2. *Let F be a continuous distribution function. Then the map $\tilde{F} \mapsto F \circ \tilde{F}^{-1}$ as a map $D[0,1] \mapsto \ell^\infty[0,1]$ is Hadamard-differentiable at $\tilde{F} = F$ tangentially to the set of functions*

$$\alpha \in E_F = \{\alpha(x) = \beta(F(x)),\, x \in \mathbb{R},\, \beta \in C[0,1], \beta(0) = \beta(1) = 0\}\,.$$

The derivative is given by $-\alpha \circ F^{-1}$.

Proof. Let α_t converge uniformly to $\alpha \in E_F$. Put $F_t = F + t\,\alpha_t$. As

$$\sup_{u \in [0,1]} \left| F_t(F_t^{-1}(u)) - u \right| = \sup_{u \in [0,1]} |F_t(u+) - F_t(u-)|,$$

the function F is continuous and the function α_t converges uniformly to a bounded and continuous function, we get that $F_t(F_t^{-1}(u)) - u = o(t)$ uniformly in u.

Thus we can calculate

$$\begin{aligned}
&\frac{1}{t}\left[F(F_t^{-1})(u) - F(F^{-1}(u)) + t\,\alpha(F^{-1}(u))\right] \\
&\quad = \alpha(F^{-1}(u)) - \alpha_t(F_t^{-1}(u)) + o(1) \\
&\quad = \left[\alpha(F^{-1}(u)) - \alpha(F_t^{-1}(u))\right] + \left[\alpha(F_t^{-1}(u)) - \alpha_t(F_t^{-1}(u))\right] + o(1)\,.
\end{aligned}$$

As $\alpha_t \to \alpha$ uniformly, the second term converges to zero uniformly in u. By using the representation $\alpha(x) = \beta(F(x))$ we see, that to ensure a uniform convergence of the first term to zero we need to show that $F(F_t^{-1}(u)) \to u$ uniformly. But this follows by a simple calculation, which yields

$$\begin{aligned}
|F(F_t^{-1}(u)) - F(F^{-1}(u))| &= |t\,\alpha_t(F_t^{-1}(u))| + o(t) \\
&\leq |t|\,|\alpha_t(F_t^{-1}(u)) - \alpha(F_t^{-1}(u))| + |t|\,|\alpha(F_t^{-1}(u))| + o(t) = O(t)
\end{aligned}$$

uniformly in u. □

Remark. As (16) implies that $\sqrt{n}(\hat{F}_n - F)$ converges in distribution to a Gaussian process $B \circ F$, where B is a standard Brownian motion on the interval $[0,1]$, the Hadamard-differentiability tangentially to E_F given in Lemma 2 is exactly what is needed to derive asymptotic distribution of the process $\sqrt{n}(F(F_n^{-1}(u)) - u)$.

Similarly we can prove that the mapping $\tilde{G} \mapsto G \circ \tilde{G}^{-1}$ is Hadamard-differentiable at $\tilde{G} = G$ tangentially to the set of functions

$$E_G = \{\alpha(x) = \beta(G(x)),\, x \in \mathbb{R},\, \beta \in C[0,1], \beta(0) = \beta(1) = 0\}\,.$$

The proof of the following lemma follows easily by applying Lemma 2 and the chain rule.

Lemma 3. *Let the copula C have continuous partial derivatives on $[a,b]\times[c,d]\subset[0,1]^2$ and F,G are continuous, then the map $(\tilde F,\tilde G)\mapsto C^*_{\tilde F,\tilde G}$ as a map $D[a,b]\times D[a,b]\mapsto \ell^\infty([a,b]\times[c,d])$ is Hadamard-differentiable at the point $(\tilde F,\tilde G)=(F,G)$ tangentially to the set of functions $(\alpha_1,\alpha_2)\in E_F\times E_G$. The derivative is given by*

$$\phi'(\alpha_1,\alpha_2) = -C_u\circ\alpha_1\circ F^{-1} - C_v\circ\alpha_2\circ G^{-1} = -C_u\circ\beta_1 - C_v\circ\beta_2, \tag{A10}$$

where $\beta_1=\alpha_1\circ F^{-1}$ and $\beta_2=\alpha_2\circ G^{-1}$.

Lemma 3 together with Theorem 3.9.4 of van der Vaart and Wellner (1996) imply that representation (A9) holds uniformly for $(u,v)\in[a,b]\times[c,d]$. Unfortunately, many of the most popular families (e.g. Clayton, Gumbel, Normal) do not have continuous C_u and C_v at some of the points $\{(0,0),(0,1),(1,0),(1,1)\}$. The following lemma takes care of this situation.

Lemma 4. *Let the distribution H have continuous margins F,G and a copula function whose first derivatives are continuous on $[0,1]^2\backslash\{(0,0),(0,1),(1,0),(1,1)\}$. Then representation* (A9) *holds uniformly in $(u,v)\in[0,1]^2$.*

Proof. Suppose for simplicity that the point of discontinuity is only at $(0,0)$ (other points $(0,1)$, $(1,0)$, $(1,1)$ might be handled in a similar way). Let us denote

$$\begin{aligned} Z_n(u,v) =&\sqrt{n}(C^*_{\hat F_n,\hat G_n}(u,v)-C(u,v)) \\ &+C_u(u,v)\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\mathbb{I}\{U_i\le u\}-u\right]+C_v(u,v)\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\mathbb{I}\{V_i\le v\}-v\right] . \end{aligned}\tag{A11}$$

Let $\varepsilon>0$ be given. As all the process

$$\begin{aligned} X_n^1(u) &= \sqrt{n}[F(\hat F_n^{-1}(u))-u], & X_n^3(u) &= \tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n\left[\mathbb{I}\{U_i\le u\}-u\right], \\ X_n^2(u) &= \sqrt{n}[G(\hat G_n^{-1}(u))-u], & X_n^4(u) &= \tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n\left[\mathbb{I}\{V_i\le u\}-u\right] \end{aligned}$$

converge to a Brownian motion, we can find δ_ε and n_ε such that for all $n>n_\varepsilon$

$$P\left(\sup_{u\le\delta_\varepsilon}|X_n^j(u)|\ge\tfrac{\varepsilon}{4}\right)<\tfrac{\varepsilon}{4},\quad j=1,\dots,4.$$

As C_u,C_v are bounded by 1, the triangular inequality implies that for all $n>n_\varepsilon$

$$P\left(\sup_{u,v\le\delta_\varepsilon}|Z_n(u,v)|\ge\varepsilon\right)\le\sum_{j=1}^4 P\left(\sup_{u\le\delta_\varepsilon}|X_n^j(u)|\ge\frac{\varepsilon}{4}\right)<\varepsilon\ .$$

Next, the existence of n'_ε such that for all $n>n'_\varepsilon$:

$$P\left(\sup_{u,v\in A_\varepsilon}|Z_n(u,v)|\ge\varepsilon\right)<\varepsilon,\quad \text{with } A_\varepsilon=[0,1]^2\backslash[0,\delta_\varepsilon]^2,$$

follows by Lemma 3 applied to rectangles $[0,\delta_\varepsilon]\times[\delta_\varepsilon,1]$ and $[\delta_\varepsilon,1]\times[0,1]$. Thus for $n>\max\{n_\varepsilon,n'_\varepsilon\}$: $P\left(\sup_{u,v}|Z_n(u,v)|\ge\varepsilon\right)<\varepsilon$, which proves the lemma. □

Combining (A9), Lemma 4, the fact that $h_n \to 0$ and asymptotic equicontinuity of the processes $\mathbb{U}_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\mathbb{I}\{U_i \leq u\} - u]$, $\mathbb{V}_n(v) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\mathbb{I}\{V_i \leq v\} - v]$ yields

$$\begin{aligned} C_n^{h_n}(u,v) &= \sqrt{n}\,\mathrm{E}\left[K_{u,h_n}(\tfrac{u-\hat{U}}{h_n})K_{v,h_n}(\tfrac{v-\hat{V}}{h_n}) - C(u,v)\right] \\ &= -C_u(u,v)\mathbb{U}_n(u) - C_v(u,v)\mathbb{V}_n(v) + \sqrt{n}\,D_n(u,v) + o_P(1), \end{aligned} \tag{A12}$$

where the bias term D_n is given by

$$D_n(u,v) = \int_{-1}^{1}\int_{-1}^{1} \sqrt{n}\,[C(u - t\,h_n, v - s\,h_n) - C(u,v)]\; k_{u,h}(s)\,k_{v,h}(t)\,dt\,ds. \tag{A13}$$

If copula C has bounded second order partial derivatives on $[0,1]^2$, then

$$\sqrt{n}\sup_{u,v} |D_n(u,v)| = O(n^{1/2}\,h_n^2) = o(1). \tag{A14}$$

Finally combining (A1), (A3), (A4), (A12) and (A14) yields

$$\begin{aligned} \sqrt{n}(\hat{C}_n^{(\mathrm{LL})}(u,v) - C(u,v)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n [\mathbb{I}\{U_i \leq u, V_i \leq v\} - C(u,v)] \\ &- C_u(u,v)\frac{1}{\sqrt{n}}\sum_{i=1}^n [\mathbb{I}\{U_i \leq u\} - u] - C_v(u,v)\frac{1}{\sqrt{n}}\sum_{i=1}^n [\mathbb{I}\{V_i \leq v\} - v] + o_P(1)\,. \end{aligned} \tag{A15}$$

A2. **Weak convergence of the process $\mathbb{C}_n^{(\mathbf{MR})}$.** Here we adapt the foregoing proof for the mirrored-type kernel estimator $\hat{C}_n^{(\mathrm{MR})}$ given in (13).

Step 1

At first we rewrite

$$\hat{C}_n^{(\mathrm{MR})} = \sum_{\ell=1}^{9} [Z_n(\ell,u,v) - Z_n(\ell,u,0) - Z_n(\ell,0,v) + Z_n(\ell,0,0),] \tag{A16}$$

where

$$Z_n(\ell,u,v) = \frac{1}{n}\sum_{i=1}^n K\left(\tfrac{u-\hat{U}_i^{(\ell)}}{h_n}\right)\,K\left(\tfrac{v-\hat{V}_i^{(\ell)}}{h_n}\right).$$

Let us define

$$Z_{0n}(\ell,u,v) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{U_i^{(\ell)} \leq u, V_i^{(\ell)} \leq v\}.$$

Similarly as in Step 1 of the proof of Appendix A1 (weak convergence of $\mathbb{C}_n^{(\mathrm{LL})}$) we can show that for each $\ell = 1,\ldots,9$:

$$\sup_{u,v}\left|\sqrt{n}(Z_n(\ell,u,v) - \mathrm{E}\,Z_n(\ell,u,v)) - \sqrt{n}(Z_{0n}(\ell,u,v) - \mathrm{E}\,Z_{0n}(\ell,u,v))\right| = o_p(1). \tag{A17}$$

Further note that

$$\sum_{\ell=1}^{9} Z_{0n}(\ell,u,v) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{U_i \leq u, V_i \leq v\} + F_n^*(u) + G_n^*(v) + 1. \tag{A18}$$

Combining (A16), (A17) and (A18), implies that uniformly in (u,v)

$$\sqrt{n}(\hat{C}_n^{(\mathrm{MR})}(u,v) - C(u,v)) = B_n(u,v) + C_n^{h_n}(u,v) + o_p(1),$$

where B_n is given by (A3) and

$$C_n^{h_n}(u,v) = \sqrt{n}\left\{\sum_{\ell=1}^{9} \mathrm{E}\left[K\left(\tfrac{u-\hat{U}^{(\ell)}}{h_n}\right) - K\left(\tfrac{-\hat{U}^{(\ell)}}{h_n}\right)\right]\left[K\left(\tfrac{v-\hat{V}^{(\ell)}}{h_n}\right) - K\left(\tfrac{-\hat{V}^{(\ell)}}{h_n}\right)\right] - C(u,v)\right\}.$$

Step 2

Now similarly as in Step 2 of the proof in Appendix A1 we derive that uniformly in (u,v)

$$C_n^{h_n}(u,v) = -C_u(u,v)\mathbb{U}_n(u) - C_v(u,v)\mathbb{V}_n(v) + \sqrt{n}\, D_n(u,v) + o_P(1), \tag{A19}$$

with the bias term $D_n(u,v)$ given by

$$D_n(u,v) = \sum_{\ell=1}^{9} \mathrm{E}\left[K\left(\frac{u-U_1^{(\ell)}}{h_n}\right) - K\left(\frac{-U_1^{(\ell)}}{h_n}\right)\right]\left[K\left(\frac{v-V_1^{(\ell)}}{h_n}\right) - K\left(\frac{-V_1^{(\ell)}}{h_n}\right)\right] - C(u,v). \tag{A20}$$

To show that this bias term is uniformly $O(h_n^2)$ is straightforward but tedious. The most simple case is if $(u,v) \in [h_n, 1-h_n]^2$. Then (A20) boils down to

$$\begin{aligned} D_n(u,v) &= \mathrm{E}\, K\left(\tfrac{u-U_1}{h_n}\right) K\left(\tfrac{v-V_1}{h_n}\right) - C(u,v) \\ &= \int_{-1}^{1}\int_{-1}^{1} C(u - t\, h_n, v - s\, h_n)\, k(t)k(s)\, dt\, ds - C(u,v) \end{aligned}$$

and the assertion follows simply by Taylor expansion.

Regarding the remaining cases we will be dealing explicitly only with $(u,v) \in [1-h_n, 1]^2$. The other cases may be handled in a similar way.

Note that Taylor expansion together with the assumptions of the theorem imply $C(u,v) = u+v-1+O(h_n^2)$ uniformly in $(u,v) \in [1-2\,h_n, 1]^2$. Further, routine algebra shows that (A20) simplifies to

$$\begin{aligned} D_n(u,v) = &\,\mathrm{E}\, K\left(\tfrac{u-U_1}{h_n}\right) K\left(\tfrac{v-V_1}{h_n}\right) + \mathrm{E}\, K\left(\tfrac{u+U_1-2}{h_n}\right) K\left(\tfrac{v-V_1}{h_n}\right) \\ &+ \mathrm{E}\, K\left(\tfrac{u-U_1}{h_n}\right) K\left(\tfrac{v+V_1-2}{h_n}\right) + \mathrm{E}\, K\left(\tfrac{u+U_1-2}{h_n}\right) K\left(\tfrac{v+V_1-2}{h_n}\right) - C(u,v)\ . \end{aligned} \tag{A21}$$

Let us compute

$$\begin{aligned} &\mathrm{E}\, K\left(\tfrac{u-U_1}{h_n}\right) K\left(\tfrac{v-V_1}{h_n}\right) \qquad \text{(A22)} \\ &= \int_{\frac{u-1}{h_n}}^{1}\int_{\frac{v-1}{h_n}}^{1} C(u - t\, h_n, v - s\, h_n)\, k(t)\, k(s)\, dt\, ds + \int_{\frac{u-1}{h_n}}^{1}\int_{-1}^{\frac{v-1}{h_n}} (u - t\, h_n)\, k(t)\, k(s)\, dt\, ds \\ &+ \int_{-1}^{\frac{u-1}{h_n}}\int_{\frac{v-1}{h_n}}^{1} (v - s\, h_n)\, k(t)\, k(s)\, dt\, ds + \int_{-1}^{\frac{u-1}{h_n}}\int_{-1}^{\frac{v-1}{h_n}} 1\, k(t)\, k(s)\, dt\, ds \end{aligned}$$

$$
\begin{aligned}
&= \int_{\frac{u-1}{h_n}}^{1}\int_{\frac{v-1}{h_n}}^{1}(u - t\,h_n + v - s\,h_n - 1)\,k(t)\,k(s)\,dt\,ds + \int_{\frac{u-1}{h_n}}^{1}(u - t\,h_n)\,k(t)\,dt\,K(\tfrac{u-1}{h_n})\\
&\quad + \int_{\frac{v-1}{h_n}}^{1}(v - s\,h_n)\,k(s)\,ds\,K(\tfrac{v-1}{h_n}) + K(\tfrac{u-1}{h_n})\,K(\tfrac{v-1}{h_n}) + O(h_n^2)\\
&= \dots\\
&= (u+v-1) + K(\tfrac{u-1}{h_n})(1-u) + K(\tfrac{v-1}{h_n})(1-v)\\
&\quad - h_n \int_{\frac{u-1}{h_n}}^{1} t\,k(t)\,dt - h_n \int_{\frac{v-1}{h_n}}^{1} t\,k(t)\,dt + O(h_n^2).
\end{aligned}
$$

Similarly

$$
\begin{aligned}
\text{(A23)}\qquad & \mathrm{E}\,K\left(\tfrac{u+U_1-2}{h_n}\right)\,K\left(\tfrac{v-V_1}{h_n}\right)\\
&= \int_{-1}^{\frac{u-1}{h_n}}\int_{\frac{v-1}{h_n}}^{1} P(U_1 > 2 + t\,h_n - u, V_1 \le v - s\,h_n)\,k(t)\,k(s)\,dt\,ds\\
&\quad + \int_{-1}^{\frac{u-1}{h_n}}\int_{-1}^{\frac{v-1}{h_n}}(1 - 2 - t\,h_n + u)\,k(t)\,k(s)\,dt\,ds\\
&= \dots = (u-1)K(\tfrac{u-1}{h_n}) - h_n \int_{-1}^{\frac{u-1}{h_n}} t\,k(t)\,dt + O(h_n^2),
\end{aligned}
$$

$$
\text{(A24)}\qquad \mathrm{E}\,K\left(\tfrac{v+V_1-2}{h_n}\right)\,K\left(\tfrac{u-U_1}{h_n}\right) = (v-1)\,K(\tfrac{v-1}{h_n}) - h_n \int_{-1}^{\frac{v-1}{h_n}} t\,k(t)\,dt + O(h_n^2)
$$

$$
\text{(A25)}\qquad \mathrm{E}\,K\left(\tfrac{u+U_1-2}{h_n}\right)\,K\left(\tfrac{v+V_1-2}{h_n}\right) = O(h_n^2).
$$

Combining (A21) with (A22), (A23), (A24) and (A25) gives us

$$
D_n(u,v) = u + v - 1 + O(h_n^2) - C(u,v) = O(h_n^2),
$$

which was to be proved.

Appendix B – Proof of Theorem 2

B1. **Weak convergence of the processes $\mathbb{C}_n^{(\mathbf{LLS})}$ and $\mathbb{C}_n^{(\mathbf{MRS})}$.** The proof of Theorem 2 for these estimators goes completely along the lines of the proof of Theorem 1, apart from a small difference in Step 2.

This difference is in calculating the bias term D_n given by (A13) for $\hat{C}_n^{(\mathrm{LL})}$ and by (A20) for $\hat{C}_n^{(\mathrm{MR})}$. The shrinking of the bandwidths by the function $b(w) = \min\{\sqrt{w}, \sqrt{1-w}\}$ together with condition (9) guarantees that the Taylor expansion

$$
C(u - s\,h_n\,b(u), v - t\,h_n\,b(v)) = C(u,v) - s\,h_n\,b(u)\,C_u(u,v) - t\,h_n\,b(v)\,C_v(u,v) + O(h_n^2)
$$

holds uniformly in $(u,v) \in [0,1]^2$ and $(s,t) \in [-1,1]^2$. Applying the above expansion in the bias calculations completes the proof.

B2. **Weak convergence of the process $\mathbb{C}_n^{(\mathbf{T})}$.** The proof is completely analogous to (and simpler than) the proof of Theorem 1 for $\hat{C}_n^{(\mathrm{LL})}$. The only difference is in calculating the bias D_n which is for the estimator $\hat{C}_n^{(\mathrm{T})}$ given by

$$D_n(u,v) = \int_{-1}^{1}\int_{-1}^{1} C(\Phi(\Phi^{-1}(u) - s\,h_n), \Phi(\Phi^{-1}(v) - t\,h_n))\,k(s)\,k(t)\,ds\,dt\,.$$

As all the second order partial derivatives of $C(\Phi(\Phi^{-1}(u) - s\,h_n), \Phi(\Phi^{-1}(v) - t\,h_n))$ taken as a function of (s,t) are bounded by the assumptions of the theorem, Taylor expansion gives us $D_n(u,v) = O(h_n^2)$ uniformly in (u,v) which proves the statement.

Appendix C – Justification of bootstrap tests based on $\hat{C}_n^{(\mathrm{LL})}$

Denote the process underlying the goodness-of-fit statistics as $G_n = \sqrt{n}(\hat{C}_n^{(\mathrm{LL})} - C_{\hat{\theta}_n})$ and $G_n^* = \sqrt{n}(\hat{C}_n^{(\mathrm{LL})*} - C_{\hat{\theta}_n^*})$ for its bootstrap version. In the following lemma we will suppose that the true copula C belongs to a known parametric family of copulas $\mathcal{C}_0 = \{C_\theta, \theta \in \Theta\}$.

Lemma 5. *Assume that the parametric family of copulas $\mathcal{C}_0$ satisfies the assumptions of Theorem 1 of Genest and Rémillard (2008) and moreover the derivative of the true copula C_{θ_0} with respect to θ is continuous as a function of (u,v) in $[0,1]^2$. Then there exists a 'nonstochastic' sequence of functions a_n from $[0,1]^2$ to $[0,1]$ such that $(G_n - a_n, G_n^* - a_n)$ converges in distribution to two independent copies of the same process.*

Proof. From the proof of Theorem 1 of this paper it follows that

$$\sqrt{n}\,\hat{C}_n^{(\mathrm{LL})} = \sqrt{n}\,C_n^{(\mathrm{E})} + \sqrt{n}\,D_n^{\theta_0} + o_p(1),$$

where

$$D_n^{\theta}(u,v) = \int_{-1}^{1}\int_{-1}^{1} [C_\theta(u - t\,h_n, v - s\,h_n) - C_\theta(u,v)]\;k_{u,h}(s)\,k_{v,h}(t)\,dt\,ds. \tag{C1}$$

Thus the processes G_n and G_n^* might be rewritten as

$$G_n = \sqrt{n}(C_n^{(\mathrm{E})} - C_{\hat{\theta}_n}) + \sqrt{n}\,D_n^{\theta_0} + o_p(1), \tag{C2}$$

$$G_n^* = \sqrt{n}(C_n^{(E)*} - C_{\hat{\theta}_n^*}) + \sqrt{n}\,D_n^{\hat{\theta}_n} + o_p(1). \tag{C3}$$

As the empirical copula process $\sqrt{n}(C_n^{(\mathrm{E})} - C_{\theta_0})$ converges weakly, Theorem 1 of Genest and Rémillard (2008) implies, that the first terms on the right hand sides of (C2) and (C3) converge jointly in distribution to independent copies of the same process. Thus defining $a_n(u,v)$ as $\sqrt{n}\,D_n^{\theta_0}(u,v)$ it remains to show that

$$\sup_{u,v}\left|D_n^{\theta_0}(u,v) - D_n^{\hat{\theta}_n}(u,v)\right| = o_p(\tfrac{1}{\sqrt{n}}).$$

But this follows directly from (C1), the first order Taylor expansion of $C_{\hat{\theta}_n}$ around the true value of the parameter θ_0 and the assumptions of the lemma. □

Appendix D – Verification of (9) for some families of copulas

We will verify assumption (9) only for C_{uu}. The assumptions about C_{uv}, C_{vv} may be checked analogously.

Clayton and Gumbel copulas. Clayton and Gumbel copulas belong to an Archimedean family of copulas, given by

$$C(u,v) = \phi^{-1}\left(\phi(u) + \phi(v)\right), \tag{D1}$$

where the function ϕ is called a generator of the copula. The generator of a Clayton copula is given by $\phi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$ with $\theta \geq 0$ and that of a Gumbel copula by $\phi(t) = (-\log t)^{\theta}$ with $\theta \geq 1$.

Direct differentiaton of (D1) yields

$$C_u(u,v) = \frac{\phi'(u)}{\phi'(C(u,v))}, \qquad C_{uu}(u,v) = \frac{\phi''(u)}{\phi'(C(u,v))} - \frac{[\phi'(u)]^2 \phi''(C(u,v))}{[\phi'(C(u,v))]^3}. \tag{D2}$$

For a Clayton and a Gumbel copula it is easy to verify that $\frac{\phi''(u)}{\phi'(u)} = O(\frac{1}{u(1-u)})$. Hence, we can bound the first term on the right-hand side of the expression for $C_{uu}(u,v)$ in (D2) uniformly in v by

$$\left|\frac{\phi''(u)}{\phi'(C(u,v))}\right| \leq \left|\frac{\phi''(u)}{\phi'(u)}\right| \left|\frac{\phi'(u)}{\phi'(C(u,v))}\right| \leq \left|\frac{\phi''(u)}{\phi'(u)}\right| |C_u(u,v)| = O\left(\tfrac{1}{u(1-u)}\right). \tag{D3}$$

The second term on the right-hand side of the expression for $C_{uu}(u,v)$ in (D2) is a more delicate one. For a Clayton copula we have $\phi'(t) = -t^{-\theta-1}$ and $\phi''(t) = (\theta+1)t^{-\theta-2}$, which implies

(D4)

$$\left|\frac{[\phi'(u)]^2 \phi''(C(u,v))}{[\phi'(C(u,v))]^3}\right| = \left|\frac{(\theta+1)[C(u,v)]^{3\theta+3}}{u^{2\theta+2}[C(u,v)]^{\theta+2}}\right| = \left|\frac{(\theta+1)[C(u,v)]^{2\theta+1}}{u^{2\theta+2}}\right| \leq \frac{\theta+1}{u} = O\left(\tfrac{1}{u}\right),$$

using the Fréchet-Hoeffding upper bound for a copula. See Nelsen (2006). Combining (D3) and (D4) verifies (9) for C_{uu} of a Clayton copula.

For a Gumbel copula we have

$$\phi'(u) = \theta(-\log u)^{\theta-1}\left(\tfrac{-1}{u}\right), \qquad \phi''(u) = \theta(\theta-1)(-\log u)^{\theta-2}\left(\tfrac{1}{u^2}\right) + \theta(-\log u)^{\theta-1}\left(\tfrac{1}{u^2}\right),$$

which implies

$$\begin{aligned}
&\frac{[\phi'(u)]^2 \phi''(C(u,v))}{[\phi'(C(u,v))]^3} \\
&\quad = \frac{\theta^2(-\log u)^{2\theta-2}\left[\theta(\theta-1)(-\log C(u,v))^{\theta-2}\frac{1}{[C(u,v)]^2} + \theta(-\log C(u,v))^{\theta-1}\frac{1}{[C(u,v)]^2}\right]}{u^2\left[\theta(-\log C(u,v))^{\theta-1}\frac{-1}{C(u,v)}\right]^3} \\
&\quad = \frac{-(-\log u)^{2\theta-2}C(u,v)}{u^2}\left[(\theta-1)(-\log C(u,v))^{1-2\theta} + (-\log C(u,v))^{2-2\theta}\right]. \qquad \text{(D5)}
\end{aligned}$$

When $u \to 0_+$ the key fact is that

$$\text{(D6)} \qquad \frac{C(u,v)}{u^2}\,\frac{(-\log u)^{2\theta-2}}{(-\log C(u,v))^{2\theta-2}} \le \frac{u}{u^2}\,\frac{(-\log u)^{2\theta-2}}{(-\log u)^{2\theta-2}} = \frac{1}{u},$$

and when $u \to 1_-$

$$\text{(D7)} \qquad \frac{(-\log u)^{2\theta-2}}{(-\log C(u,v))^{2\theta-1}} \le \frac{(-\log u)^{2\theta-2}}{(-\log u)^{2\theta-1}} = \frac{1}{-\log u} = O(\tfrac{1}{1-u}).$$

Combining (D3), (D5), (D6) and (D6) verifies (9) for C_{uu} of a Gumbel copula.

Normal copula. The normal copula is given by

$$C(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \tfrac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{\tfrac{s^2-2\rho st+t^2}{2(1-\rho^2)}\right\} ds\,dt, \qquad \rho \in (-1,1),$$

where Φ is the cumulative distribution function of a standard normal variable.

By a direct computation (or with the help of properties of a conditional normal distribution) we get

$$C_u(u,v) = \Phi\left(\tfrac{\Phi^{-1}(v)-\rho\,\Phi^{-1}(u)}{\sqrt{1-\rho^2}}\right)$$

whose derivative with respect to u is given by

$$\text{(D8)} \qquad C_{uu}(u,v) = \frac{-\rho}{\sqrt{1-\rho^2}}\,\phi\left(\tfrac{\Phi^{-1}(v)-\rho\;\Phi^{-1}(u)}{\sqrt{1-\rho^2}}\right)\frac{1}{\phi(\Phi^{-1}(u))}.$$

where $\phi = \Phi'$. As ϕ is bounded, it is sufficient to deal with $[\phi(\Phi^{-1}(u))]^{-1}$. L'Hôpital's rule yields

$$\frac{u(1-u)}{\phi(\Phi^{-1}(u))} \sim \frac{1-2u}{\Phi^{-1}(u)} = o(1), \qquad \text{for } u \to 0_+ \ (u \to 1_-),$$

which together with (D8) verifies (9) for C_{uu} of a normal copula.

Student copula. The Student copula (with m degrees of freedom) is given by

$$C(u,v) = \int_{-\infty}^{t_m^{-1}(u)} \int_{-\infty}^{t_m^{-1}(v)} \tfrac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \tfrac{s^2-2\rho st+t^2}{m(1-\rho^2)}\right)^{-(m+2)/2} ds\,dt, \qquad \rho \in (-1,1),$$

where $t_m^{-1}(\cdot)$ is the quantile function of the Student distribution with m degrees of freedom.

Direct calculation shows that

$$\text{(D9)} \qquad C_u(u,v) = \frac{d^{(m+2)/2}}{f_m(t_m^{-1}(u))}\,\frac{1}{(c+d)^{(m+1)/2}} \int_{-\infty}^{\frac{t_m^{-1}(v)-\rho t_m^{-1}(u)}{\sqrt{d+c}}} (1+x^2)^{-(m+2)/2}\,dx,$$

where $c = [t_m^{-1}(u)]^2$, $d = m(1-\rho^2)$ and f_m is the density of the Student distribution with m degrees of freedom. Assumption (9) for C_{uu} of a Student copula can be verified by differentiating (D9) with respect to u. The useful facts (which follow by l'Hôpital's rule or properties of the density f_m) are

$$\frac{t_m^{-1}(u)}{u} \sim \frac{1}{f_m(t_m^{-1}(u))}, \qquad \frac{f_m'(t_m^{-1}(u))}{f_m(t_m^{-1}(u))} = O\left(\frac{1}{t_m^{-1}(u)}\right), \qquad \text{for } u \to 0_+ \ (u \to 1_-),$$

where f'_m is the derivative of f_m.

Acknowledgement

The authors are grateful to the co-Editor, an Associate-Editor and two reviewers for their very helpful comments which led to a considerable improvement of the original version of the paper.

References

Chen, S. X. and Huang, T.-M. (2007). Nonparametric estimation of copula functions for dependence modelling. *Canad. J. of Statist.*, 35:265–282.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application.* Cambridge University Press, New York.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. *Acad. Roy. Belg. Bull. Cl. Sci.*, 65:274–292.

Epanechnikov, V. (1969). Nonparametric estimation of a multidimensional probability density. *Theory Probab. Appl.*, 13:153–158.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman & Hall.

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

Gänssler, P. and Stute, W. (1987). *Seminar on Empirical Processes.* Basil Birkhäuser.

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

Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann. Inst. Henri Poincaré-PR.* In press.

Genest, C., Rémillard, B., and Beaudoin, D. (2008). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics.* In press. doi: 10.1016/j.insmatheco.2007.10.005.

Gijbels, I. and Mielniczuk, J. (1990). Estimating the density of a copula function. *Commun. Statist. Theory Meth.*, 19:445–464.

Jin, Z. and Shao, Y. (1999). On kernel estimation of a multivariate distribution function. *Statist. Probab. Lett.*, 41:163–168.

Nelsen, R. B. (2006). *An Introduction to Copulas.* Springer, New York. Second edition.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Tsukuhara, H. (2005). Semiparametric estimation in copula models. *Canad. J. of Statist.*, 33:357–375.

van der Vaart, A. W. (1994). Weak convergence of smoothed empirical processes. *Scand. J. Statist.*, 21:501–504.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

van der Vaart, A. W. and Wellner, J. A. (2007). Empirical processes indexed by estimated functions. *IMS Lecture Notes Monograph Series 2007*, 55:234–252.

Wand, M. and Jones, M. (1995). *Kernel Smoothing.* Chapman & Hall.

Yan, J. (2007). Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software*, 21(4):1–21.