The Annals of Statistics
2011, Vol. 39, No. 4, 1963–2006
DOI: 10.1214/11-AOS890

NEW ESTIMATORS OF THE PICKANDS DEPENDENCE FUNCTION AND A TEST FOR EXTREME-VALUE DEPENDENCE[1]

BY AXEL BÜCHER, HOLGER DETTE AND STANISLAV VOLGUSHEV

Ruhr-Universität Bochum

We propose a new class of estimators for Pickands dependence function which is based on the concept of minimum distance estimation. An explicit integral representation of the function $A^*(t)$, which minimizes a weighted L^2-distance between the logarithm of the copula $C(y^{1-t}, y^t)$ and functions of the form $A(t)\log(y)$ is derived. If the unknown copula is an extreme-value copula, the function $A^*(t)$ coincides with Pickands dependence function. Moreover, even if this is not the case, the function $A^*(t)$ always satisfies the boundary conditions of a Pickands dependence function. The estimators are obtained by replacing the unknown copula by its empirical counterpart and weak convergence of the corresponding process is shown. A comparison with the commonly used estimators is performed from a theoretical point of view and by means of a simulation study. Our asymptotic and numerical results indicate that some of the new estimators outperform the estimators, which were recently proposed by Genest and Segers [*Ann. Statist.* **37** (2009) 2990–3022]. As a by-product of our results, we obtain a simple test for the hypothesis of an extreme-value copula, which is consistent against all positive quadrant dependent alternatives satisfying weak differentiability assumptions of first order.

1. Introduction. The copula provides an elegant margin-free description of the dependence structure of a random variable. By the famous theorem of Sklar (1959), it follows that the distribution function H of a bivariate random variable (X, Y) can be represented in terms of the marginal distributions F and G of X and Y, that is,

$$H(x, y) = C(F(x), G(y)),$$

where C denotes the copula, which characterizes the dependence between X and Y. Extreme-value copulas arise naturally as the possible limits of copulas of component-wise maxima of independent, identically distributed or strongly mixing stationary sequences [see Deheuvels (1984) and Hsing (1989)]. These copulas provide flexible tools for modeling joint extremes in risk management. An important application of extreme-value copulas appears in the modeling of data with

Received December 2010.

[1]Supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG).

MSC2010 subject classifications. Primary 62G05, 60G32; secondary 62G20.

Key words and phrases. Extreme-value copula, minimum distance estimation, Pickands dependence function, weak convergence, empirical copula process, test for extreme-value dependence.

positive dependence, and in contrast to the more popular class of Archimedean copulas they are not symmetric [see Tawn (1988) or Ghoudi, Khoudraji and Rivest (1998)]. Further applications can be found in Coles, Heffernan and Tawn (1999) or Cebrian, Denuit and Lambert (2003) among others. A copula C is an extreme-value copula if and only if it has a representation of the form

$$C(y^{1-t}, y^t) = y^{A(t)} \qquad \forall y, t \in [0, 1], \tag{1.1}$$

where $A : [0, 1] \to [1/2, 1]$ is a convex function satisfying $\max\{s, 1-s\} \leq A(s) \leq 1$, which is called Pickands dependence function. The representation of (1.1) of the extreme-value copula C depends only on the one-dimensional function A and statistical inference on a bivariate extreme-value copula C may now be reduced to inference on its Pickands dependence function A.

The problem of estimating Pickands dependence function nonparametrically has found considerable attention in the literature. Roughly speaking, there exist two classes of estimators. The classical nonparametric estimator is that of Pickands (1981) [see Deheuvels (1991) for its asymptotic properties] and several variants have been discussed. Alternative estimators have been proposed and investigated in the papers by Capéraà, Fougères and Genest (1997), Jiménez, Villa-Diharce and Flores (2001), Hall and Tajvidi (2000), Segers (2007) and Zhang, Wells and Peng (2008), where the last-named authors also discussed the multivariate case. In most references, the estimators of Pickands dependence function are constructed assuming knowledge of the marginal distributions. Recently Genest and Segers (2009) proposed rank-based versions of the estimators of Pickands (1981) and Capéraà, Fougères and Genest (1997), which do not require knowledge of the marginal distributions. In general, all of these estimators are neither convex nor do they satisfy the boundary restriction $\max\{t, 1-t\} \leq A(t) \leq 1$, in particular the endpoint constrains $A(0) = A(1) = 1$. However, the estimators can be modified without changing their asymptotic properties in such a way that these constraints are satisfied, see, for example, Fils-Villetard, Guillou and Segers (2008).

Before the specific model of an extreme-value copula is selected, it is necessary to check this assumption by a statistical test, that is a test for the hypotheses

$$H_0 : C \in \mathcal{C} \quad \text{vs.} \quad H_1 : C \notin \mathcal{C}, \tag{1.2}$$

where $\mathcal{C}$ denotes the class of all copulas satisfying (1.1). Throughout this paper, we call (1.2) the hypothesis of extreme-value dependence. The problem of testing this hypothesis has found much less attention in the literature. To our best knowledge, only two tests of extremeness are currently available in the literature. The first one was proposed by Ghoudi, Khoudraji and Rivest (1998). It exploits the fact that for an extreme-value copula the random variable $W = H(X, Y) = C(F(X), G(Y))$ satisfies the identity

$$-1 + 8\mathbb{E}[W] - 9\mathbb{E}[W^2] = 0. \tag{1.3}$$

The properties of this test have been studied by Ben Ghorbal, Genest and Nešlehová (2009), who determined the finite- and large-sample variance of the test statistic. In particular, the test proposed by Ghoudi, Khoudraji and Rivest (1998) is not consistent against alternatives satisfying (1.3). The second class of tests was recently introduced by Kojadinovic and Yan (2010) who proposed to compare the empirical copula and a copula estimator which is constructed from the estimators proposed by Genest and Segers (2009) under the assumption of an extreme-value copula. These tests are only consistent against alternatives that are left tail decreasing in both arguments and satisfy strong smoothness assumptions on the copula and convexity assumptions on an analogue of Pickands dependence function, which are hard to verify analytically.

The present paper has two purposes. The first is the development of some alternative estimators of Pickands dependence function using the principle of minimum distance estimation. We propose to consider the best approximation of the logarithm of the empirical copula $\hat{C}$ evaluated in the point (y^{1-t}, y^t), that is, $\log \hat{C}(y^{1-t}, y^t)$, by functions of the form

$$\log(y)A(t) \tag{1.4}$$

with respect to a weighted L^2-distance. It turns out that the minimal distance and the corresponding optimal function can be determined explicitly. On the basis of this result, and by choosing various weight functions in the L^2-distance, we obtain an infinite-dimensional class of estimators for the function A. Our approach is closely related to the theory of Z-estimation and in Section 3 we indicate how this point of view provides several interesting relationships between the different concepts for constructing estimates of Pickands dependence function.

The second purpose of the paper is to present a new test for the hypothesis of extreme-value dependence, which is consistent against a much broader class of alternatives than the tests which have been proposed so far. Here our approach is based on an estimator of a weighted minimum L^2-distance between the true copula and the class of functions satisfying (1.4) and the corresponding tests are consistent with respect to all positive quadrant dependent alternatives satisfying weak differentiability assumptions of first order. To our best knowledge, this method provides the first test in this context which is consistent against such a general class of alternatives. Moreover, in contrast to Ghoudi, Khoudraji and Rivest (1998) and Kojadinovic and Yan (2010) we also provide a weak convergence result under fixed alternative which can be used for studying the power of the test.

The remaining part of the paper is organized as follows. In Section 2, we consider the approximation problem from a theoretical point of view. In particular, we derive explicit representations for the minimal L^2-distance between the logarithm of the copula and its best approximation by a function of the form (1.4), which will be the basis for all statistical applications in this paper. The new estimators, say $\hat{A}_n$, are defined in Section 3, where we also prove weak convergence of the process $\{\sqrt{n}(\hat{A}_n(t) - A(t))\}_{t\in[0,1]}$ in the space of uniformly bounded functions on the interval [0, 1] under appropriate assumptions on the weight function used in the L^2-distance. Furthermore, we give a theoretical and empirical comparison of the new estimators with the estimators proposed in Genest and Segers (2009). We will also determine "optimal" estimators in the proposed class by minimizing the asymptotic MSE with respect to the choice of the weight function used in the L^2-distance. In particular, we demonstrate that some of the new estimators have a substantially smaller asymptotic variance than the estimators proposed by the last-named authors. We also provide a simulation study in order to investigate the finite sample properties of the different estimates. In Section 4, we introduce and investigate the new test of extreme-value dependence. In particular, we derive the asymptotic distribution of the test statistic under the null hypothesis as well as under the alternative. In order to approximate the critical values of the test, we introduce a multiplier bootstrap procedure, prove its consistency and study its finite sample properties by means of a simulation study. Finally, most of the technical details are deferred to the Appendix.

2. A measure of extreme-value dependence. Let $\mathcal{A}$ denote the set of all functions $A:[0,1]\to[1/2,1]$, and define Π as the copula corresponding to independent random variables, that is, $\Pi(u,v)=uv$. Throughout this paper, we assume that the copula C satisfies $C\ge\Pi$ which holds for any extreme-value copula due to the lower bound for the function A. As pointed out by Scaillet (2005), this property is equivalent to the concept of positive quadrant dependence, that is,

$$\mathbb{P}(X\le x, Y\le y)\ge \mathbb{P}(X\le x)\mathbb{P}(Y\le y) \qquad \forall (x,y)\in\mathbb{R}^2. \tag{2.1}$$

For a copula with this property, we define the weighted L^2-distance

$$M_h(C,A)=\int_{(0,1)^2}\bigl(\log C(y^{1-t},y^t)-\log(y)A(t)\bigr)^2 h(y)\,d(y,t), \tag{2.2}$$

where $h:[0,1]\to\mathbb{R}^+$ is a continuous weight function.

The following result is essential for our approach and provides an explicit expression for the best L^2-approximation of the logarithm of the copula by the logarithm of a function of the form (1.1) and as a by-product characterizes the function A^* minimizing $M_h(C,A)$.

THEOREM 2.1. *Assume that the given copula satisfies $C\ge\Pi^\kappa$ for some $\kappa\ge 1$ and that the weight function h satisfies $\int_0^1(\log y)^2h(y)\,dy<\infty$. Then the function*

$$A^*=\arg\min\{M_h(C,A)|A\in\mathcal{A}\}$$

is unique and given by

$$A^*(t)=B_h^{-1}\int_0^1\frac{\log C(y^{1-t},y^t)}{\log y}h^*(y)\,dy, \tag{2.3}$$

where the associated weight function h^* is defined by

$$h^*(y) = \log^2(y)h(y), \qquad y \in (0,1), \tag{2.4}$$

and

$$B_h = \int_0^1 (\log y)^2 h(y)\,dy = \int_0^1 h^*(y)\,dy. \tag{2.5}$$

Moreover, the minimal L^2-distance between the logarithms of the given copula and the class of functions of the form (1.4) *is given by*

$$M_h(C, A^*) = \int_{(0,1)^2} \left(\frac{\log C(y^{1-t}, y^t)}{\log y}\right)^2 h^*(y)\,d(y,t) - B_h \int_0^1 (A^*(t))^2\,dt. \tag{2.6}$$

PROOF. Since $C \geq \Pi^\kappa$, we get $0 \geq \log C(y^{1-t}, y^t) \geq \kappa \log y$ and thus $|\log C(y^{1-t}, y^t)| \leq \kappa |\log y|$ and all integrals exist. Rewriting the L^2 distance in (2.2) gives

$$M_h(C, A) = \int_0^1 \int_0^1 \left(\frac{\log C(y^{1-t}, y^t)}{\log y} - A(t)\right)^2 (\log y)^2 h(y)\,dy\,dt$$

and the assertion is now obvious. □

Note that $A^*(t) = A(t)$ if C is an extreme-value copula of the form (1.1) with Pickands dependence function A. Furthermore, the following lemma shows that the minimizing function A^* defined in (2.3) satisfies the boundary conditions of Pickands dependence functions.

LEMMA 2.2. *Assume that C is a copula satisfying $C \geq \Pi$. Then the function A^* defined in* (2.3) *has the following properties*:

(i) $A^*(0) = A^*(1) = 1$,
(ii) $A^*(t) \geq t \vee (1-t)$,
(iii) $A^*(t) \leq 1$.

PROOF. Assertion (i) is obvious. For a proof of (ii), one uses the Fréchet–Hoeffding bound $C(u, v) \leq u \wedge v$ [see, e.g., Nelsen (2006)] and obtains the assertion by a direct calculation. Similarly, assertion (iii) follows from the inequality $C \geq \Pi$. □

Unfortunately, the function A^* is in general not convex for every copula satisfying $C \geq \Pi$. A counterexample can be derived from Theorem 3.2.2 in Nelsen

(2006) and is given by the following shuffle of the copula $u \wedge v$:

$$
C(u,v) = \begin{cases} \min\{u, v, 1/2\}, & (u,v) \in [0, \sqrt{1/2}]^2, \\ \min\{u, v + 1/2 - \sqrt{1/2}\}, & \\ & (u,v) \in [0, \sqrt{1/2}] \times [\sqrt{1/2}, 1], \\ \min\{u + 1/2 - \sqrt{1/2}, v\}, & \\ & (u,v) \in [\sqrt{1/2}, 1] \times [0, \sqrt{1/2}], \\ \min\{u, v, u + v + 1/2 - 2\sqrt{1/2}\}, & \\ & (u,v) \in [\sqrt{1/2}, 1]^2, \end{cases} \tag{2.7}
$$

for which an easy calculation shows that the mapping $t \mapsto -\log C(1/2^{1-t}, 1/2^t)$ is not convex. Consequently, one can find a weight function h such that the corresponding best approximating function A^* is not convex.

With the notation

$$
f_y(t) = C(y^{1-t}, y^t), \tag{2.8}
$$

the function A^* is convex (for every weight function h) if and only if the function $g_y(t) = -\log f_y(t)$ is convex for every $y \in (0,1)$. The following lemma is now obvious.

LEMMA 2.3. *If the function $t \to f_y(t) = C(y^{1-t}, y^t)$ is twice differentiable and the inequality*

$$
[f_y'(t)]^2 \geq f_y''(t) f_y(t)
$$

holds for every $(y,t) \in (0,1)^2$, then the best approximation A^* defined by (2.3) is convex.

It is worthwhile to mention that the function A^* is convex for some frequently considered classes of copulas, which will be illustrated in the following examples.

EXAMPLE 2.4. Consider the Clayton copula

$$
C_{\text{Clayton}}(u, v; \theta) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \qquad \theta > 0. \tag{2.9}
$$

Then a tedious calculation yields

$$
\begin{aligned}
&[f_y'(t)]^2 - f_y''(t) f_y(t) \\
&\quad = \theta \log^2(y) \{C_{\text{Clayton}}(y^{1-t}, y^t; \theta)\}^{2+2\theta} \big(4y^{-\theta} - y^{-\theta t} - y^{-\theta(1-t)}\big) \\
&\quad \geq \theta \log^2(y) \{C_{\text{Clayton}}(y^{1-t}, y^t; \theta)\}^{2+2\theta} (3y^{-\theta} - 1) \geq 0,
\end{aligned}
$$

where the inequalities follow observing that $m(t) = y^{-\theta t} + y^{-\theta(1-t)} \leq m(0) = 1 + y^{-\theta}$ and $y^{-\theta} \geq 1$. Therefore, we obtain from Lemma 2.3 that the best approximation A^* is convex and corresponds to an extreme-value copula.

EXAMPLE 2.5. In the following, we discuss the weight function $h_k(y) = -y^k/\log y$ ($k \geq 0$) with associated function $h_k^*(y) = -y^k \log y$, which will be used later for the construction of the new estimators of Pickands dependence function. On the one hand this choice is made for mathematical convenience, because it allows an explicit calculations of the asymptotic variance A^* in specific examples. On the other hand, estimates constructed on the basis of this weight function turn out to have good asymptotic and finite sample properties (see the discussion in Section 3.7). It follows that

$$B_{h_k} = -\int_0^1 y^k \log y \, dy = (k+1)^{-2}$$

and

$$A^*(t) = -(k+1)^2 \int_0^1 \log C(y^{1-t}, y^t) y^k \, dy, \tag{2.10}$$

which simplifies in the case $k = 0$ to the representation

$$A^*(t) = -\int_0^1 \log C(y^{1-t}, y^t) \, dy. \tag{2.11}$$

EXAMPLE 2.6. In the following, we calculate the minimal distance $M_h(C, A^*)$ and its corresponding best approximation A^* for two copula families and the associated weight function $h_1^*(y) = -y \log y$ from Example 2.5. First, we investigate the Gaussian copula defined by

$$C_\rho(u, v) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v), \rho),$$

where Φ is the standard normal distribution function and $\Phi_2(\cdot, \cdot, \rho)$ is the distribution function of a bivariate normal random variable with standard normally distributed margins and correlation $\rho \in [0, 1]$. For the limiting cases $\rho = 0$ and $\rho = 1$, we obtain the independence and perfect dependence copula, respectively, while for $\rho \in (0, 1)$ the copula C_ρ is not an extreme-value copula. The minimal distances are plotted as a function of ρ in the left part of the first line of Figure 1. In the right part, we show some functions A^* corresponding to the best approximation of the logarithm of the Gaussian copula by a function of the form (1.4). We note that all functions A^* are convex although C_ρ is only an extreme value copula in the case $\rho = 0$.

In the second example, we consider a convex combination of a Gumbel copula with parameter $\theta_1 = \log 2/\log 1.5$ (corresponding to a coefficient of tail dependence of 0.5) and a Clayton copula with parameter $\theta_2 = 2$, that is,

$$C_\alpha(u, v) = \alpha C_{\text{Clayton}}(u, v; \theta_2) + (1-\alpha) C_{\text{Gumbel}}(u, v; \theta_1), \qquad \alpha \in [0, 1],$$

where the Clayton copula is given in (2.9) and the Gumbel copula is defined by

$$C_{\text{Gumbel}}(u, v; \theta) = \exp\bigl(-\{(-\log u)^\theta + (-\log v)^\theta\}^{1/\theta}\bigr), \qquad \theta > 1.$$

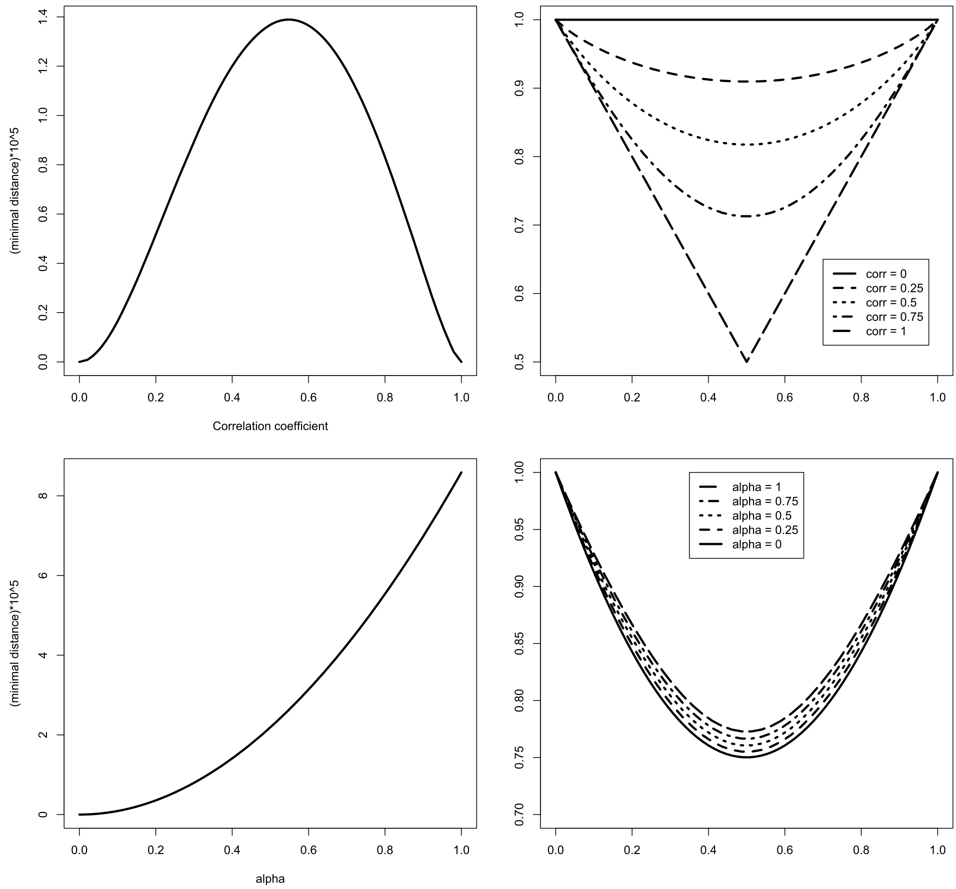

FIG. 1. *Left*: *minimal distances* $M_h(C, A^*) \times 10^5$ *for the Gaussian copula* (*as a function of its correlation coefficient*) *and for the convex combination of a Gumbel and a Clayton copula* (*as a function of the parameter* α *in the convex combination*). *Right*: *the functions* A^* *corresponding to the best approximations by functions of the form* (1.4).

Note that only the Gumbel copula is an extreme-value copula and obtained for $\alpha = 0$. The minimal distances are depicted in the left part of the lower panel of Figure 1 as a function of α. In the right part, we show the functions A^* corresponding to the best approximation of the logarithm of C_α by a function of the form (1.4). Again all approximations are convex, which means that A^* corresponds in fact to an extreme value copula.

3. A class of minimum distance estimators.

3.1. *Pickands and CFG estimators*. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ denote a sample of independent identically distributed bivariate random variables with copula

C and marginals F and G. Most of the estimates which have been proposed in the literature so far are based on the fact that the random variable

$$\xi(t) = \frac{-\log F(X)}{1-t} \wedge \frac{-\log G(Y)}{t}$$

is exponentially distributed with parameter $A(t)$. In particular, we have $E[\xi(t)] = 1/A(t)$. If the marginal distributions would be known, an estimate of $A(t)$ could be obtained by the method of moments. In the case of unknown marginals, Genest and Segers (2009) proposed to replace F and G by their empirical counterparts and obtained

$$\hat{A}^P_{n,r}(t) = \left(\frac{1}{n}\sum_{i=1}^{n}\hat{\xi}_i(t)\right)^{-1}$$

as a rank-based version of Pickands estimate, where

$$\hat{\xi}_i(t) = \frac{-\log \hat{F}_n(X_i)}{1-t} \wedge \frac{-\log \hat{G}_n(Y_i)}{t}, \qquad i = 1, \dots, n,$$

and

$$\hat{F}_n(X_i) = \frac{1}{n+1}\sum_{j=1}^{n} I\{X_j \le X_i\} \quad \text{and} \quad \hat{G}_n(Y_i) = \frac{1}{n+1}\sum_{j=1}^{n} I\{Y_j \le Y_i\} \tag{3.1}$$

denote the (slightly modified) empirical distribution functions of the samples $\{X_j\}_{j=1}^n$ and $\{Y_j\}_{j=1}^n$ at the points X_i and Y_i, respectively. Similarly, observing the identity $E[\log \xi(t)] = -\log A(t) - \gamma$ (here $\gamma = -\int_0^\infty \log x e^{-x}\,dx$ denotes Euler's constant), they obtained a rank-based version of the estimate proposed by Capéraà, Fougères and Genest (1997), that is,

$$\hat{A}^{\mathrm{CFG}}_{n,r}(t) = \exp\left(-\gamma - \frac{1}{n}\sum_{i=1}^{n}\log \hat{\xi}_i(t)\right).$$

For illustrative purposes, we finally recall two integral representations for the rank-based version of Pickands and CFG estimate, which we use in Section 3.6 to put all estimates considered in this paper in a general context, that is,

$$\frac{1}{\hat{A}^P_{n,r}(t)} = \int_0^1 \frac{\hat{C}_n(y^{1-t}, y^t)}{y}\,dy, \tag{3.2}$$

$$\gamma + \log \hat{A}^{\mathrm{CFG}}_{n,r}(t) = \int_0^1 \frac{\hat{C}_n(y^{1-t}, y^t) - I\{y > e^{-1}\}}{\log y}\,dy, \tag{3.3}$$

where

$$\hat{C}_n(u, v) = \frac{1}{n}\sum_{i=1}^{n} I\{\hat{F}_n(X_i) \le u, \hat{G}_n(Y_i) \le v\} \tag{3.4}$$

denotes the empirical copula and $\hat{F}_n(X_i)$, $\hat{G}_n(Y_i)$ are defined in (3.1) [see Genest and Segers (2009) for more details].

3.2. *New estimators and weak convergence.* Theorem 2.1 suggests to define a class of new estimators for Pickands dependence function by replacing the unknown copula in (2.3) through the empirical copula defined in (3.4). The asymptotic properties of the corresponding estimators will be investigated in this section. For technical reasons, we require that the argument in the logarithm in the representation (2.3) is positive and propose to use the estimator

$$\tilde{C}_n = \hat{C}_n \vee n^{-\gamma}, \tag{3.5}$$

where the constant γ satisfies $\gamma > 1/2$ and the empirical copula $\hat{C}_n$ is defined in (3.4).

For the subsequent proofs, we will need a result on the weak convergence of the empirical copula process with estimated margins. While this problem has been considered by many authors [see, e.g., Rüschendorf (1976), Fermanian, Radulović and Wegkamp (2004) or Tsukahara (2005) among others], all of them assume that the copula has continuous partial derivatives on the whole unit square $[0, 1]^2$. However, as was pointed out by Segers (2010), there is only one extreme-value copula that has this property. Luckily, in a remarkable paper Segers (2010) was able to show that the following condition is sufficient for weak convergence of the empirical copula process

$$\partial_j C \text{ exists and is continuous on } \{(u_1, u_2) \in [0, 1]^2 | u_j \in (0, 1)\} \tag{3.6}$$

$(j = 1, 2)$. This condition can be shown to hold for any extreme-value copula with continuously differentiable Pickands function A [see Segers (2010)]. Moreover, under this assumption, the process $\sqrt{n}(\tilde{C}_n - C)$ shows the same limiting behavior as the empirical copula process $\sqrt{n}(\hat{C}_n - C)$, that is,

$$\sqrt{n}(\tilde{C}_n - C) \overset{w}{\rightsquigarrow} \mathbb{G}_C, \tag{3.7}$$

where the symbol $\overset{w}{\rightsquigarrow}$ denotes weak convergence in $l^\infty[0, 1]^2$. Here, $\mathbb{G}_C$ is a Gaussian field on the square $[0, 1]^2$ which admits the representation

$$\mathbb{G}_C(\mathbf{x}) = \mathbb{B}_C(\mathbf{x}) - \partial_1 C(\mathbf{x})\mathbb{B}_C(x_1, 1) - \partial_2 C(\mathbf{x})\mathbb{B}_C(1, x_2),$$

where $\mathbf{x} = (x_1, x_2)$, $\mathbb{B}_C$ is a bivariate pinned C-Brownian sheet on the square $[0, 1]^2$ with covariance kernel given by

$$\mathrm{Cov}(\mathbb{B}_C(\mathbf{x}), \mathbb{B}_C(\mathbf{y})) = C(\mathbf{x} \wedge \mathbf{y}) - C(\mathbf{x})C(\mathbf{y}),$$

and the minimum $\mathbf{x} \wedge \mathbf{y}$ is understood component-wise. Observing the representation (2.3), we obtain the estimator

$$\hat{A}_{n,h}(t) = B_h^{-1} \int_0^1 \frac{\log \tilde{C}_n(y^{1-t}, y^t)}{\log y} h^*(y)\, dy \tag{3.8}$$

for Pickands dependence function, where $\tilde{C}_n$ is defined in (3.5). Note that this relation specifies an infinite-dimensional class of estimators indexed by the set of all admissible weight functions. The following results specify the asymptotic properties of these estimators. We begin with a slightly more general statement, which shows weak convergence for the weighted integrated process

$$\sqrt{n}\mathbb{W}_{n,w}(t) = \sqrt{n}\int_0^1 \log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} w(y,t)\,dy,$$

where the weight function $w:[0,1]^2 \to \bar{\mathbb{R}}$ depends on y and t. The result (and some arguments in its proof) are also needed in Section 4.

THEOREM 3.1. *Assume that for the weight function $w:[0,1]^2 \to \bar{\mathbb{R}}$ there exists a function $\bar{w}:[0,1] \to \bar{\mathbb{R}}_0^+$ such that*

$$\forall (y,t) \in [0,1]^2 \qquad |w(y,t)| \le \bar{w}(y), \tag{3.9}$$

$$\forall \varepsilon > 0 \qquad \sup_{y\in[\varepsilon,1]} \bar{w}(y) < \infty, \tag{3.10}$$

$$\int_0^1 \bar{w}(y) y^{-\lambda}\,dy < \infty \tag{3.11}$$

for some $\lambda > 1$. If the copula C satisfies (3.6) *and $C \ge \Pi$, then we have for any $\gamma \in (1/2, \lambda/2)$ as $n \to \infty$*

$$\begin{aligned} \sqrt{n}\mathbb{W}_{n,w}(t) &= \sqrt{n}\int_0^1 \log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} w(y,t)\,dy \\ &\overset{w}{\rightsquigarrow} \quad \mathbb{W}_{C,w}(t) = \int_0^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} w(y,t)\,dy \end{aligned} \tag{3.12}$$

in $l^\infty[0,1]$.

The following result is now an immediate consequence of Theorem 3.1 using $w(y,t) := -B_h^{-1} h^*(y)$ [recall the definition of the associated weight function h^* in (2.4)] and yields the weak convergence of the process $\sqrt{n}(\hat{A}_{n,h} - A^*)$ for a broad class of weight functions.

THEOREM 3.2. *If the copula $C \ge \Pi$ satisfies condition* (3.6) *and the weight function h satisfies the conditions*

$$\text{for all } \varepsilon > 0 \qquad \sup_{y\in[\varepsilon,1]} \left| \frac{h^*(y)}{\log y} \right| < \infty, \tag{3.13}$$

$$\int_0^1 h^*(y)(-\log y)^{-1} y^{-\lambda}\,dy < \infty \tag{3.14}$$

for some $\lambda > 1$, then we have for any $\gamma \in (1/2, \lambda/2)$ as $n \to \infty$

$$\mathbb{A}_{n,h} = \sqrt{n}(\hat{A}_{n,h} - A^*) \quad \overset{w}{\rightsquigarrow} \quad \mathbb{A}_{C,h} \text{ in } l^\infty[0,1],$$

where the process $\mathbb{A}_{C,h}$ *is given by*

$$\mathbb{A}_{C,h}(t) = B_h^{-1} \int_0^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} \frac{h^*(y)}{\log y} \, dy. \tag{3.15}$$

REMARK 3.3. (a) Conditions (3.13) and (3.14) restrict the behavior of the function h^* near the boundary of the interval [0, 1]. A simple sufficient condition for (3.13) and (3.14) is given by

$$\sup_{x \in [0,1]} \left| \frac{h^*(x)}{x^\alpha (1-x)^\beta} \right| < \infty$$

for some $\alpha > 0$, $\beta \geq 1$. In this case, λ can be chosen as $1 + \alpha/2$.

(b) In the construction discussed so far, it is also possible to use weight functions that depend on t, that is, functions of the form $\tilde{h}^*(y, t)$. As long as $\tilde{h}^*(y, t) > 0$ for $(y, t) \in (0, 1) \times [0, 1]$, the corresponding best approximation A^* will still be well defined and correspond to the Pickands dependence function if C is an extreme-value copula. Theorem 3.1 provides the asymptotic properties of the corresponding estimator A if we set $w(y, t) := \tilde{h}^*(y, t)/(-\log y)$ and assume that $\int_0^1 \tilde{h}^*(y, t) \, dy = 1$ for all t. However, for the sake of a clear presentation, we will only use weight functions that do not depend on t.

Note that Theorem 3.2 is also correct if the given copula is not an extreme-value copula. In other words: it establishes weak convergence of the process $\sqrt{n}(\hat{A}_{n,h} - A^*)$ to a centered Gaussian process, where A^* denotes the function corresponding to the best approximation of the logarithm of the copula C by a function of the form (1.4). If A^* is convex, it corresponds to an extreme-value copula and coincides with Pickands dependence function. Note also that Theorem 3.2 excludes the case $h_0^*(y) = -\log y$, because condition (3.14) is not satisfied for this weight function. Nevertheless, under the additional assumption that C is an extreme-value copula with twice continuously differentiable Pickands dependence function A, the assertion of the preceding theorem is still valid.

THEOREM 3.4. *Assume that C is an extreme-value copula with twice continuously differentiable Pickands dependence function A. For the weight function $h_0^*(y) = -\log y$, we have for any $\gamma \in (1/2, 3/4)$ as $n \to \infty$*

$$\mathbb{A}_{n,h_0}(t) = \sqrt{n}(\hat{A}_{n,h_0} - A)(t) \quad \overset{w}{\rightsquigarrow} \quad \mathbb{A}_{C,h_0}(t) = -\int_0^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} \, dy$$

in $l^\infty[0, 1]$, *where* $\hat{A}_{n,h_0}(t) = -\int_0^1 \log \tilde{C}_n(y^{1-t}, y^t) \, dy$.

REMARK 3.5. (a) If the marginals of (X, Y) are independent the distribution of the random variable $\mathbb{A}_{\Pi,h_0}$ coincides with the distribution of the random variable $\mathbb{A}_r^P = -\int_0^1 \mathbb{G}_\Pi(y^{1-t}, y^t) y^{-1}\,dy$, which appears as the weak limit of the appropriately standardized Pickands estimator; see Genest and Segers (2009). In fact, a much more general statement is true: by using weight functions $\tilde{h}^*(y,t)$ depending on t it is possible to obtain for any extreme-value copula estimators of the form (3.8) which show the same limiting behavior as the estimators proposed by Genest and Segers (2009). This already indicates that for any extreme-value copula it is possible to find weight functions which will make the new minimum distance estimators asymptotically at least as efficient (in fact better, as will be shown in Section 3.4) as the estimators introduced by Genest and Segers (2009).

(b) A careful inspection of the proof of Theorem 3.1 reveals that the condition $C \geq \Pi$ can be relaxed to $C \geq \Pi^\kappa$ for some $\kappa > 1$, if one imposes stronger conditions on the weight function.

(c) The estimator depends on the parameter γ which is used for the construction of the statistic $\tilde{C}_n = \hat{C}_n \vee n^{-\gamma}$. This modification is only made for technical purposes and from a practical point of view the behavior of the estimators does not change substantially provided that γ is chosen larger than $2/3$.

REMARK 3.6. The new estimators can be alternatively motivated observing that the identity (1.1) yields the representation $A(t) = \log C(y^{1-t}, y^t)/\log y$ for any $y \in (0,1)$. This leads to a simple class of estimators, that is,

$$\tilde{A}_{n,\delta_y}(t) = \frac{\log \tilde{C}_n(y^{1-t}, y^t)}{\log y}; \qquad y \in (0,1),$$

where δ_y is the Dirac measure at the point y and $\tilde{C}_n$ is defined in (3.5). By averaging these estimators with respect to a distribution, say π, we obtain estimators of the form

$$\tilde{A}_{n,\pi}(t) = \int_0^1 \frac{\log \tilde{C}_n(y^{1-t}, y^t)}{\log y} \pi(dy),$$

which coincide with the estimators obtained by the concept of best L^2-approximation.

3.3. *A special class of weight functions.* In this subsection, we illustrate the results investigating Example 2.5 discussed at the end of Section 2. For the associated weight function $h_k^*(x) = -y^k \log y$ with $k \geq 0$, we obtain

$$\hat{A}_{n,h_k}(t) = -(k+1)^2 \int_0^1 \log \tilde{C}_n(y^{1-t}, y^t) y^k\,dy. \tag{3.16}$$

The process $\{\mathbb{A}_{n,h_k}(t)\}_{t\in[0,1]}$ converge weakly in $l^\infty[0,1]$ to the process $\{\mathbb{A}_{C,h_k}\}_{t\in[0,1]}$, which is given by

$$\mathbb{A}_{C,h_k}(t) = -(k+1)^2 \int_0^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} y^k\,dy. \tag{3.17}$$

Consequently, for $C \in \mathcal{C}$, the asymptotic variance of $\hat{A}_{n,h_k}$ is obtained as

$$\operatorname{Var}(\mathbb{A}_{C,h_k}(t)) = (k+1)^4 \int_0^1 \int_0^1 \sigma(u,v;t)(uv)^{k-A(t)}\,du\,dv, \tag{3.18}$$

where the function σ is given by

$$\sigma(u,v;t) = \operatorname{Cov}(\mathbb{G}_C(u^{1-t},u^t), \mathbb{G}_C(v^{1-t},v^t)).$$

In order to find an explicit expression for these variances, we assume that the function A is differentiable and introduce the notation

$$\mu(t) = A(t) - tA'(t), \qquad \nu(t) = A(t) + (1-t)A'(t),$$

where A' denotes the derivative of A. The following results can be shown by similar arguments as given in Genest and Segers (2009); for details, see Bücher, Dette and Volgushev (2010).

PROPOSITION 3.7. *For $t \in [0,1]$, let $\bar{\mu}(t) = 1 - \mu(t)$ and $\bar{\nu}(t) = 1 - \nu(t)$. If C is an extreme-value copula with Pickands dependence function A, then the variance of the random variable $\mathbb{A}_{C,h_k}(t)$ is given by*

$$\begin{aligned}
(k+1)^2 \Biggl\{ & \frac{2(k+1)}{2k+2-A(t)} - \bigl(\mu(t)+\nu(t)-1\bigr)^2 \\
& - \frac{2\mu(t)\bar{\mu}(t)(k+1)}{2k+1+t} - \frac{2\nu(t)\bar{\nu}(t)(k+1)}{2k+2-t} \\
& + 2\mu(t)\nu(t)\frac{(k+1)^2}{(1-t)t} \int_0^1 \biggl(A(s) + (k+1)\Bigl(\frac{1-s}{1-t} + \frac{s}{t}\Bigr) - 1\biggr)^{-2} ds \\
& - 2\mu(t)\frac{(k+1)^2}{(1-t)t} \int_0^t \biggl(A(s) + (k+t)\frac{1-s}{1-t} + \bigl(k+1-A(t)\bigr)\frac{s}{t}\biggr)^{-2} ds \\
& - 2\nu(t)\frac{(k+1)^2}{(1-t)t} \int_t^1 \biggl(A(s) + \bigl(k+1-A(t)\bigr)\frac{1-s}{1-t} \\
& \qquad\qquad + (k+1-t)\frac{s}{t}\biggr)^{-2} ds \Biggr\}.
\end{aligned}$$

Note that the limiting process in (3.15) is a centered Gaussian process. This means that, asymptotically, the quality of the new estimators [as well as of the estimators of Genest and Segers (2009), which show a similar limiting behavior] is determined by the variance. Based on these observations, we will now provide an asymptotic comparison of the new estimators $\hat{A}_{n,h_k}(t)$ with the estimators investigated by Genest and Segers (2009). Some finite sample results will be presented in the following section for various families of copulas. For the sake of brevity, we restrict ourselves to the independence copula Π, for which $A(t) \equiv 1$. In the case

$k = 0$, we obtain from Proposition 3.7 the same variance as for the rank-based version of Pickands estimator, that is,

$$\mathrm{Var}(\mathbb{A}_{\Pi,h_0}) = \frac{3t(1-t)}{(2-t)(1+t)} = \mathrm{Var}(\mathbb{A}_r^P)$$

[see Corollary 3.4 in Genest and Segers (2009)] while the case $k > 0$ yields

$$\mathrm{Var}(\mathbb{A}_{\Pi,h_k}) = \frac{(3+4k)(k+1)^2}{2k+1} \frac{t(1-t)}{(2k+2-t)(2k+1+t)}.$$

Investigating the derivative in k, it is easy to see that $\mathrm{Var}(\mathbb{A}_{\Pi,h_k})$ is strictly decreasing in k with

$$\lim_{k\to\infty} \mathrm{Var}(\mathbb{A}_{\Pi,h_k}) = \frac{t(1-t)}{2}.$$

Therefore, we have

$$\mathrm{Var}(\mathbb{A}_r^P) = \mathrm{Var}(\mathbb{A}_{\Pi,h_0}) \geq \mathrm{Var}(\mathbb{A}_{\Pi,h_k})$$

for all $k \geq 0$ with strict inequality for all $k > 0$. This means that for the independence copula all estimators obtained by our approach with associated weight function $h_k^*(y) = -y^k \log y$, $k > 0$, have a smaller asymptotic variance than the rank-based version of Pickands estimator. On the other hand, a comparison with the CFG estimator proposed by Genest and Segers (2009) does not provide a clear picture about the superiority of one estimator and we defer this comparison to the following section, where optimal weight functions for the new estimates $\hat{A}_{n,h}$ are introduced.

3.4. *Optimal weight functions.* In this section, we discuss asymptotically optimal weight functions corresponding to the class of estimates introduced in Section 3.2. As pointed out in the previous section, from an asymptotic point of view the mean squared error of the estimates is dominated by the variance and therefore we concentrate on weight functions minimizing the asymptotic variance of the estimate $\hat{A}_{n,h}$. The finite sample properties of the mean squared error of the various estimates will be investigated by means of a simulation study in Section 3.7.

Note that an optimal weight function depends on the point t where Pickands dependence function has to be estimated and on the unknown copula. Therefore, an estimator with an optimal weight function cannot be implemented in concrete applications without preliminary knowledge about the copula. However, it can serve as a benchmark for user-specified weight functions. To be precise, observe that by Theorem 3.2 the variance of the limiting process $\mathbb{A}_{C,h}$ is of the form

$$V(\xi) = \int_0^1 \int_0^1 k_t(x,y)\, d\xi(x)\, d\xi(y), \tag{3.19}$$

where ξ denotes a probability measure on the interval $[0, 1]$ defined by $d\xi(x) = B_h^{-1} h^*(x)\,dx$ and the kernel $k_t(x, y)$ is given by

$$k_t(x, y) = E\left[\frac{\mathbb{G}_C(x^{1-t}, x^t)}{C(x^{1-t}, x^t)\log x}\frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)\log y}\right].$$

It is easy to see that V defines a convex function on the space of all probability measures on the interval $[0, 1]$ and the existence of a minimizing measure follows if the kernel k_t is continuous on $[0, 1]^2$. The following result characterizes the minimizer of V and is proved in the Appendix.

THEOREM 3.8. *A probability measure η on the interval $[0, 1]$ minimizes V if and only if the inequality*

$$\int_0^1 k_t(x, y)\,d\eta(y) \geq \int_0^1 \int_0^1 k_t(x, y)\,d\eta(x)\,d\eta(y) \tag{3.20}$$

is satisfied for all $x \in [0, 1]$.

Theorem 3.8 can be used to check the optimality of a given weight function. For example, if the copula C is given by the independence copula Π we have

$$k_t(x, y) = \frac{(x^t \wedge y^t - (xy)^t)(x^{1-t} \wedge y^{1-t} - (xy)^{1-t})}{x \log x\, y \log y},$$

and it is easy to see that none of the associated weight functions $h_k^*(y) = -y^k \log y$ with $k \geq 0$ is optimal in the sense that it minimizes the asymptotic variance of the estimate $\hat{A}_{n,h}$ with respect to the choice of the weight function. On the other hand, the result is less useful for an explicit computation of optimal weight functions. Deriving an analytical expression for the optimal weight function seems to be impossible, even for the simple case of the independence copula.

However, approximations to the optimal weight function can easily be computed numerically. To be precise we approximate the double integral appearing in the representation of $\operatorname{Var}(\mathbb{A}_{C,h}(t))$ by the finite sum

$$V(\xi) \approx \sum_{i=1}^N \sum_{j=1}^N \xi_{i,N}\xi_{j,N} k_t(i/N, j/N) = \Xi^T K_t \Xi, \tag{3.21}$$

where $N \in \mathbb{N}$, $K_t = (k_t(i/N, j/N))_{i,j=1}^N$ denotes an $N \times N$ matrix, $\Xi = (\xi_{i,N})_{i=1}^N$ is an vector of length N and $\xi_{i,N} = \xi((i-1)/N, i/N]$ represents the mass of ξ allocated to the interval $((i-1)/N, i/N]$ $(i = 1, \dots, N)$. Minimizing the right-hand side of the above equation with respect to Ξ under the constrains $\xi_{i,N} \geq 0$, $\sum_{i=1}^N \xi_{i,N} = 1$ is a quadratic (convex) optimization problem which can be solved by standard methods; see, for example, Nocedal and Wright (2006) and approximations of the optimal weight function can be calculated with arbitrary precision by increasing N.

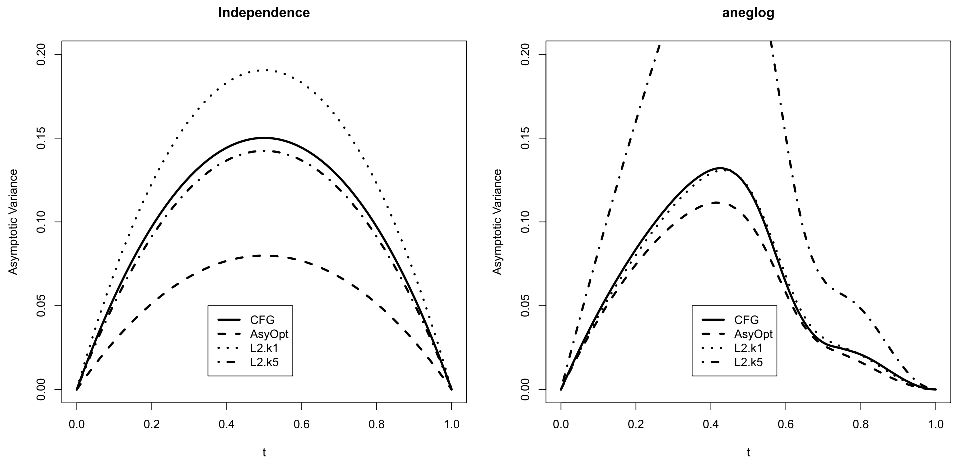

FIG. 2. *Asymptotic variances of various estimators of the Pickands dependence function. Left panel: independence copula; right panel: asymmetric negative logistic model.*

In the remaining part of this section, we will compare the asymptotic variance of the Pickands-, the CFG-estimator proposed by Genest and Segers (2009) and the new estimates, where the new estimators are based on the weight functions h_k^* discussed in Section 3.3 for two values of k as well as on the optimal weights minimizing the right-hand side of (3.21), where we set $N = 100$. In order to compute the solution Ξ_{opt}, we used the routine *ipop* from the R-package *kernlab* by Karatzoglou et al. (2004). In the left part of Figure 2, we show the asymptotic variances of the different estimators for the independence copula. We observe that Pickands estimator has the largest asymptotic variances (this curve is not displayed in the figure), while the CFG estimator of Genest and Segers (2009) yields smaller variances than the estimator $\hat{A}_{n,h_1}$, but larger asymptotic variances than the estimators $\hat{A}_{n,h_5}$. On the other hand, the estimate $\hat{A}_{n,h_{\mathrm{opt}}}$ corresponding to the numerically determined optimal weight function yields a substantially smaller variance than all other estimates under consideration. In the right-hand part of Figure 2, we display the corresponding results for the asymmetric negative logistic model [see Joe (1990)]

$$A(t) = 1 - \{(\psi_1(1-t))^{-\theta} + (\psi_2 t)^{-\theta}\}^{-1/\theta} \tag{3.22}$$

with parameters $\psi_1 = 1$, $\psi_2 = 2/3$ and $\theta \in (0, \infty)$ chosen such that the coefficient of tail dependence is 0.6. We observe that the estimate $\hat{A}_{n,h_5}$ yields the largest asymptotic variance. The CFG estimate proposed by Genest and Segers (2009) and the estimate $\hat{A}_{n,h_1}$ show a similar behavior (with minor advantages for the latter), while the best results are obtained for the new estimate corresponding to the optimal weight function.

We conclude this section with the remark that we have presented a comparison of the different estimators based on the asymptotic variance which determines the mean squared error asymptotically. For finite samples, minimizing only the variance might increase the bias and therefore the asymptotic results cannot directly be transferred to applications. In the finite sample study presented in Section 3.7, we will demonstrate that not all of the asymptotic results yield good predictions for the finite-sample behavior of the corresponding estimators.

3.5. *Convex estimates and endpoint corrections.* In general, all of the estimates discussed so far [including those proposed by Genest and Segers (2009)] will neither be convex, nor will they satisfy the other characterizing properties of Pickands dependence functions. However, the literature provides many proposals on how to enforce these conditions. Various endpoint corrections have been proposed by Deheuvels (1991), Segers (2007) or Hall and Tajvidi (2000) among others. Fils-Villetard, Guillou and Segers (2008) proposed an L^2-projection of the estimate of Pickands dependence function on a space of partially linear functions which is arbitrarily close to the space of all convex functions in $\mathcal{A}$ satisfying the conditions of Lemma 2.2. They also showed that this transformation decreases the L^2-distance between the "true" dependence function and the estimate. An alternative concept of constructing convex estimators is based on the greatest convex minorant, which yields a decrease in the sup-norm, that is,

$$\sup_{0<t<1} |\hat{A}_n^{\mathrm{gcm}}(t) - A(t)| \leq \sup_{0<t<1} |\hat{A}_n(t) - A(t)|,$$

where $\hat{A}_n$ is any initial estimate of Pickands dependence function and $\hat{A}_n^{\mathrm{gcm}}$ its greatest convex minorant [see, e.g., Marshall (1970), Wang (1986), Robertson, Wright and Dykstra (1996) among others]. It is also possible to combine this concept with an endpoint correction calculating the greatest convex minorant of the function

$$t \longrightarrow \big(\hat{A}_n(t) \wedge 1\big) \vee t \vee (1-t)$$

[see Genest and Segers (2009) who also proposed alternative special endpoint corrections for their estimators]. All these methods can be used to produce an estimate of A which has the characterizing properties of a Pickands dependence function.

3.6. *M- and Z-estimates.* As mentioned in the Introduction, a broader class of estimates could be obtained by minimizing more general distances between the given copula and the class of functions defined by (1.1) and in this paragraph we briefly indicate this principle. Consider the best approximation of the copula C by functions of the form (1.1) with respect to the distance

$$D_w(C, A) = \int_0^1 \int_0^1 \Phi\big(C(y^{1-t}, y^t), y^{A(t)}\big) w(y,t)\, dy\, dt, \tag{3.23}$$

where $\Phi : [0,1] \times [0,1] \to \mathbb{R}_0^+$ denotes a "distance" and w is a given weight function. Note that the minimization in (3.23) can be carried out by separately minimizing the inner integral for every value of t. Consequently, the problem reduces to a one-dimensional minimization problem and assuming differentiability it follows that for fixed t the optimal value $A^*(t)$ minimizing the interior integral in (3.23) is obtained as a solution of the equation

$$\frac{\partial}{\partial a}\int_0^1 \Phi(C(y^{1-t}, y^t), y^a)w(y,t)\,dy\bigg|_{a=A^*(t)} = 0.$$

Under suitable assumptions, integration and differentiation can be exchanged and we have

$$\int_0^1 \Psi(C(y^{1-t}, y^t), y^a)(\log y)y^a w(y,t)\bigg|_{a=A^*(t)} dy = 0, \tag{3.24}$$

where $\Psi = \partial_2\Phi$ denotes the derivative of Φ with respect to the second argument. In general, the solution of (3.24) is only defined implicitly as a functional of the copula C. Therefore, if C is replaced through the empirical copula the analysis of the stochastic properties of the corresponding process turns out to be extremely difficult because in many cases one has to control improper integrals (see the proofs of Theorems 3.1 and 3.4 in the Appendix). For the sake of a clear exposition, we do not discuss details in this paper and defer these considerations to future research.

Nevertheless, equation (3.24) yields a different view on the estimation problem of Pickands dependence function. Note that the estimate introduced in Section 3.2 is obtained by the choice $w(y,t) = h(y)B_h^{-1}$ and

$$\Phi(z_1, z_2) = (\log z_1 - \log z_2)^2; \qquad \Psi(z_1, z_2) = -2(\log z_1 - \log z_2)/z_2$$

in (3.24). This estimate corresponds to a minimum distance estimate. Similarly, an estimate corresponding to the classical L^2-distance is obtained for the choice

$$\Phi(z_1, z_2) = (z_1 - z_2)^2; \qquad \Psi(z_1, z_2) = -2(z_1 - z_2).$$

This yields for (3.24) the equation

$$\int_0^1 \big(C(y^{1-t}, y^t) - y^a\big)(\log y)^2 y^a h(-\log y)\bigg|_{a=A^*(t)} dy = 0,$$

which cannot be solved analytically. The rank-based versions of Pickands and the CFG estimator proposed by Genest and Segers (2009) do not correspond to M-estimates, but could be considered as Z-estimates obtained from (3.24) for the function

$$\Psi(z_1, z_2) = (z_1 - z_2)/z_2$$

with $w_{\mu,\nu}(y) = y^{\mu-1}/(-\log y)^{1+\nu}$ with $\mu = \nu = 0$ and $\mu = 0$, $\nu = 1$, respectively. In fact, this choice leads to a general class of estimators which relates the Pickands and the CFG estimate in an interesting way. To be precise, note that for $\nu \in [0, 1)$ equation (3.24) yields

$$
\begin{aligned}
&\int_0^1 \frac{(C(y^{1-t}, y^t) - I\{y > e^{-1}\})y^{\mu-1}}{(-\log y)^\nu}\, dy \\
&\quad = \int_0^1 \frac{(y^{A(t)} - I\{y > e^{-1}\})y^{\mu-1}}{(-\log y)^\nu}\, dy \qquad (3.25)\\
&\quad = \frac{\Gamma(1-\nu)}{(A(t)+\mu)^{1-\nu}} - \int_0^1 \frac{e^{-\mu x}}{x^\nu}\, dx.
\end{aligned}
$$

Here the case $\nu = 1$ has to be interpreted as the limit $\nu \to 1$, which yields a generalization of the defining equation for the CFG estimate, that is,

$$
-\log \mu - \int_\mu^\infty \frac{e^{-t}}{t}\, dt + \log(A(t)+\mu) = \int_0^1 \frac{(C(y^{1-t}, y^t) - I\{y > e^{-1}\})y^{\mu-1}}{\log y}\, dy.
$$

Observing the relation

$$
\lim_{\mu\to 0} \log \mu + \int_\mu^\infty \frac{e^{-t}}{t}\, dt = -\gamma
$$

we obtain the defining equation for the estimate proposed by Genest and Segers (2009) [see (3.3)]. Similarly, if $\nu \in [0, 1)$ it follows from (3.25)

$$
\int_0^1 \frac{C(y^{1-t}, y^t)y^{\mu-1}}{(-\log y)^\nu}\, dy = \frac{\Gamma(1-\nu)}{(A(t)+\mu)^{1-\nu}} \qquad (3.26)
$$

and we obtain a defining equation for a generalization of the Pickands estimate. The classical case is obtained for $\mu = \nu = 0$ [see Genest and Segers (2009) or equation (3.2)], but (3.26) defines many other estimates of this type. Therefore, the Pickands and the CFG estimate correspond to the extreme cases in the class $\{w_{\mu,\nu} | \mu \geq 0, \nu \in [0, 1]\}$.

We finally note that there are numerous other functions Ψ, which could be used for the construction of alternative Z-estimates, but most of them do not lead to an explicit solution for $A^*(t)$. In this sense the CFG-estimator, Pickands-estimator and the estimates proposed in this paper could be considered as attractive special cases, which can be explicitly represented in terms of an integral of the empirical copula.

3.7. *Finite sample properties.* In this subsection, we investigate the small sample properties of the new estimators by means of a simulation study. Especially, we compare the new estimators with the rank-based estimators suggested by Genest and Segers (2009), which are most similar in spirit with the method proposed in this paper. We study the finite sample behavior of the greatest convex minorants of the endpoint corrected versions of the various estimators. The new estimators are corrected in a first step by

$$\hat{A}^{\mathrm{corr}}_{n,h}(t) := \max\bigl(t, 1-t, \min(\hat{A}_{n,h}, 1)\bigr) \tag{3.27}$$

and in a second step the greatest convex minorant of $\hat{A}^{\mathrm{corr}}_{n,h}$ is calculated. For the rank-based CFG and Pickands estimators, we first used the endpoint corrections proposed in Genest and Segers (2009), then applied (3.27) and finally calculated the greatest convex minorant. Hereby, we compare the performance of the different statistical procedures which will be used in concrete applications and apply the corrections, that are most favorable for the respective estimators. The greatest convex minorants are computed using the routine *gcmlcm* from the package *fdrtool* by Strimmer (2009). All results presented here are based on 5,000 simulation runs and the sample size is $n = 100$.

As estimators, we consider the statistics defined in (3.8) with the weight function h_k and the optimal weight function determined in Section 3.4. An important question is the choice of the parameter k for the statistic $\hat{A}_{n,h_k}$ in order to achieve a balance between bias and variance. For this purpose, we first study the performance of the estimator $\hat{A}_{n,h_k}$ with respect to different choices for the parameter k and consider the asymmetric negative logistic model defined in (3.22) and the symmetric mixed model [see Tawn (1988)] defined by

$$A(t) = 1 - \theta t + \theta t^2, \qquad \theta \in [0, 1]. \tag{3.28}$$

The results for other copula models are similar and are omitted for the sake of brevity. For the Pickands dependence function (3.22), we used the parameters $\psi_1 = 1$ and $\psi_2 = 2/3$ such that the coefficient of tail dependence is given by $\rho = 2(3^\theta + 2^\theta)^{-1/\theta}$ and varies in the interval $(0, 2/3)$, while the parameter $\theta \in [0, 1]$ used in (3.28) yields $\rho = \theta/2 \in [0, 1/2]$.

The quality of an estimator $\hat{A}$ is measured with respect to mean integrated squared error

$$\mathrm{MISE}(\hat{A}) = \mathbb{E}\biggl[\int_0^1 \bigl(\hat{A}(t) - A(t)\bigr)^2\,dt\biggr],$$

which was computed by taking the average over 5,000 simulated samples. The new estimators turned out to be rather robust with respect to the choice of the parameter γ in the definition of the process $\tilde{C}_n = \hat{C}_n \vee n^{-\gamma}$ provided that $\gamma \geq 2/3$. For this reason, we use $\gamma = 0.95$ throughout this section. Analyzing the impact of choosing different values for k, in Figure 3 we display simulated curves

$$k \mapsto \frac{\mathrm{MISE}(\hat{A}_{n,h_k})}{\min_{\ell \geq 0} \mathrm{MISE}(\hat{A}_{n,h_\ell})} \tag{3.29}$$

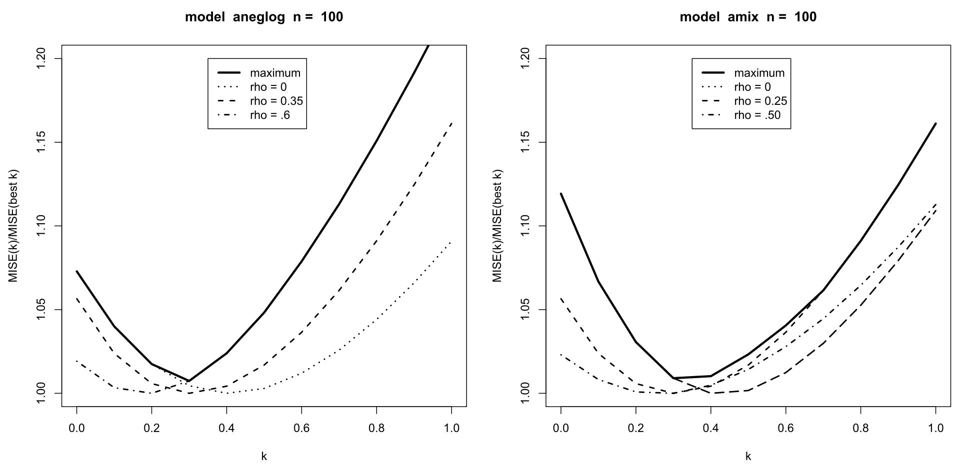

FIG. 3. *The function defined in* (3.29) *for various models and coefficients of tail dependence. The minimum corresponds to the optimal value of* k *in the weight function* h_k. *The solid curve corresponds to the worst case defined by* (3.30). *The sample size is* $n = 100$ *and the MISE is calculated by* 5,000 *simulation runs. Left panel*: *asymmetric negative logistic model. Right panel*: *mixed model.*

for the asymmetric negative logistic and the mixed models with different coefficients of tail dependence ρ, as well as the maximum over such curves for different values of ρ (solid curves), that is,

$$k \mapsto \max_{\rho} \frac{\mathrm{MISE}_\rho(\hat{A}_{n,h_k})}{\min_{\ell \geq 0} \mathrm{MISE}_\rho(\hat{A}_{n,h_\ell})}, \tag{3.30}$$

where by MISE_ρ we denote the MISE for the tail dependence coefficient ρ. The curves in (3.29) attain their minima in the optimal k for the respective ρ, and their shapes provide information about the performance of the estimators for nonoptimal values of k. The solid curve gives an impression about the "worst case" scenario (with respect to ρ) in every model. The simulations indicate, that for $n = 100$ the optimal values of k for different models and tail dependence coefficients lie in the interval $[0.2, 0.6]$. Moreover, for values of k in this interval the quality of the estimators remains very stable. For $n = 200$, $n = 500$ and additional models the picture remains quite similar and these results are not depicted for the sake of brevity. We thus recommend using $k = 0.4$ in practical applications. Note that the asymptotic analysis in Section 3.4 suggests that the asymptotically optimal k should differ substantially for various models. However, this effect is not visible for sample size up to $n = 500$. In these cases, the optimal values for k usually varies in the interval $[0.2, 0.8]$.

Next, we compare the new estimators with rank-based versions of Pickands and the CFG estimator proposed by Genest and Segers (2009). In Figure 4, the

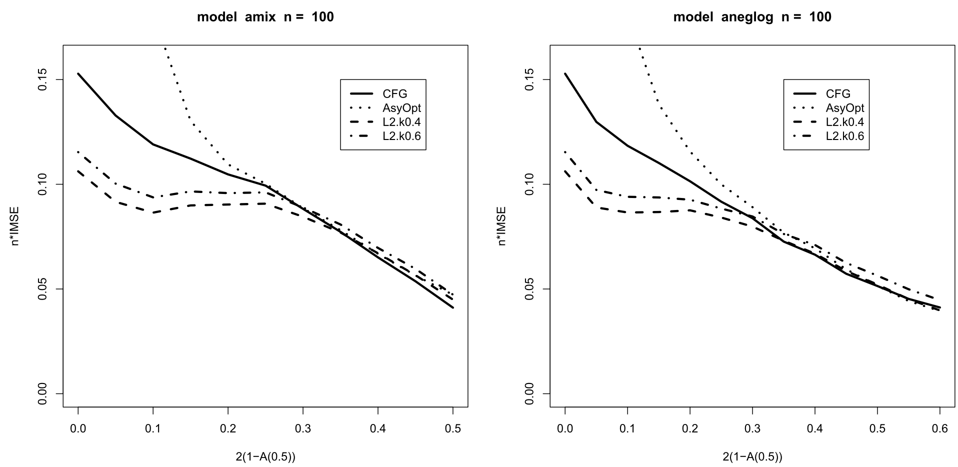

FIG. 4. $100 \times$ MISE *for various estimators*, *models and coefficients of tail dependence*, *based on* 5,000 *samples of size* $n = 100$.

normalized MISE is plotted as a function of the tail dependence parameter ρ for the asymmetric negative logistic and the mixed model, where the parameter θ is chosen in such a way, that the coefficient of tail dependence $\rho = 2(1 - A(0.5))$ varies over the specific range of the corresponding model. For each sample, we computed the rank-based versions of Pickands estimator, the CFG estimator [see Genest and Segers (2009)] and two of the new estimators $\hat{A}_{n,h_k}$ $(k = 0.4,\ 0.6)$. In this comparison, we also include the estimator $\hat{A}_{n,h_{\text{opt}}}$ which uses the optimal weight function determined in Section 3.4.

Summarizing the results, one can conclude that in general the best performance is obtained for our new estimator based on the weight function h_k with $k = 0.4$ and $k = 0.6$, in particular if the coefficient of tail dependence is small. A comparison of the two estimators $\hat{A}_{n,h_{0.4}}$ and $\hat{A}_{n,h_{0.6}}$ shows that the choice $k = 0.4$ performs slightly better than the choice $k = 0.6$ in both models. In both settings, the MISE obtained by $\hat{A}_{n,h_{0.4}}$ and $\hat{A}_{n,h_{0.6}}$ is smaller than the MISE of the CFG estimator proposed in Genest and Segers (2009) if the coefficient of tail dependence is small. On the other hand, the latter estimators yield sightly better results for a large coefficient of tail dependence. The results for rank-based version of the Pickands estimator are not depicted, because this estimator yields a uniformly larger MISE. Simulations of other scenarios show similar results and are also not displayed for the sake of brevity. It is remarkable that the optimal weight function usually yields an estimator with a substantially larger MISE than all other estimates if the coefficient of tail dependence is small. Similar results can be observed for the sample size $n = 500$ (these results are not depicted). This indicates that the advantages of the asymptotically optimal weight function only start to play a role for rather large sample sizes.

4. A test for an extreme-value dependence.

4.1. *The test statistic and its weak convergence.* From the definition of the functional $M_h(C, A)$ in (2.2) it is easy to see that, for a strictly positive weight function h with $h^* \in L^1(0, 1)$, a copula function C is an extreme-value copula if and only if

$$\min_{A \in \mathcal{A}} M_h(C, A) = M_h(C, A^*) = 0,$$

where A^* denotes the best approximation defined in (2.3). This suggests to use $M_h(\tilde{C}_n, \hat{A}_{n,h})$ as a test statistic for the hypothesis (1.2), that is,

$$H_0 : C \text{ is an extreme-value copula.}$$

Recalling the representation (2.6)

$$M_h(C, A^*) = \int_0^1 \int_0^1 \bar{C}^2(y,t) h^*(y)\, dy\, dt - B_h \int_0^1 (A^*(t))^2\, dt$$

with $\bar{C}(y,t) = -\log C(y^{1-t}, y^t)$ and defining $\bar{C}_n(y,t) := -\log \tilde{C}_n(y^{1-t}, y^t)$ we obtain the decomposition

$$\begin{aligned}
M_h(\tilde{C}_n, &\hat{A}_{n,h}) - M_h(C, A^*) \\
&= \int_0^1 \int_0^1 \bigl(\bar{C}_n^2(y,t) - \bar{C}^2(y,t)\bigr) \frac{h^*(y)}{(\log y)^2}\, dy\, dt \\
&\quad - B_h \int_0^1 \hat{A}_{n,h}^2(t) - (A^*(t))^2\, dt \\
&= 2 \int_0^1 \int_0^1 \bigl(\bar{C}_n(y,t) - \bar{C}(y,t)\bigr) \bar{C}(y,t) \frac{h^*(y)}{(\log y)^2}\, dy\, dt \\
&\quad - 2B_h \int_0^1 \bigl(\hat{A}_{n,h}(t) - A^*(t)\bigr) A^*(t)\, dt \\
&\quad + \int_0^1 \int_0^1 \bigl(\bar{C}_n(y,t) - \bar{C}(y,t)\bigr)^2 \frac{h^*(y)}{(\log y)^2}\, dy\, dt \qquad (4.1) \\
&\quad - B_h \int_0^1 \bigl(\hat{A}_{n,h}(t) - A^*(t)\bigr)^2\, dt \\
&= 2 \int_0^1 \int_0^1 \bigl(\bar{C}_n(y,t) - \bar{C}(y,t)\bigr) \bigl(\bar{C}(y,t) - A^*(t)(-\log y)\bigr) \frac{h^*(y)}{(\log y)^2}\, dy\, dt \\
&\quad + \int_0^1 \int_0^1 \bigl(\bar{C}_n(y,t) - \bar{C}(y,t)\bigr)^2 \frac{h^*(y)}{(\log y)^2}\, dy\, dt \\
&\quad - B_h \int_0^1 \bigl(\hat{A}_{n,h}(t) - A^*(t)\bigr)^2\, dt \\
&=: S_1 + S_2 + S_3,
\end{aligned}$$

where the last identity defines the terms S_1, S_2 and S_3 in an obvious manner. Note that under the null hypothesis of extreme-value dependence we have $A^* = A$ and thus $\bar{C}(y,t) = A^*(t)(-\log y)$. This means that under H_0 the term S_1 will vanish and the asymptotic distribution will be determined by the large sample properties of the random variable $S_2 + S_3$. Under the alternative, the equality $\bar{C}(y,t) = A^*(t)(-\log y)$ will not hold anymore and it turns out that in this case the statistic is asymptotically dominated by the random variable S_1. With the following results, we will derive the limiting distribution of the proposed test statistic under the null hypothesis and the alternative.

THEOREM 4.1. *Assume that the given copula C satisfies condition (3.6) and is an extreme-value copula with Pickands dependence function A^*. If the function $\bar{w}(y) := h^*(y)/(\log y)^2$ fulfills conditions (3.10) and (3.11) for some $\lambda > 2$ and the weight function h is strictly positive and satisfies assumptions (3.13), (3.14) for $\tilde{\lambda} := \lambda/2 > 1$, then we have for any $\gamma \in (1/2, \lambda/4)$ and $n \to \infty$*

$$nM_h(\tilde{C}_n, \hat{A}_{n,h}) \overset{w}{\rightsquigarrow} Z_0,$$

where the random variable Z_0 is defined by

$$Z_0 := \int_0^1 \int_0^1 \left(\frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right)^2 \bar{w}(y)\,dy\,dt - B_h \int_0^1 \mathbb{A}_{C,h}^2(t)\,dt$$

with $B_h = \int_0^1 h^*(y)\,dy$ and the process $\{\mathbb{A}_{C,h}(t)\}_{t\in[0,1]}$ is defined in Theorem 3.2.

The next theorem gives the distribution of the test statistic $M_h(\tilde{C}_n, \hat{A}_{n,h})$ under the alternative. Note that in this case we have $M_h(C, A^*) > 0$.

THEOREM 4.2. *Assume that the given copula C satisfies $C \geq \Pi$, condition (3.6) and that $M_h(C, A^*) > 0$. If additionally the weight function h is strictly positive and h and the function $\bar{w}(y) := h^*(y)/(\log y)^2$ satisfy the assumptions (3.13), (3.14) and (3.10), (3.11) for some $\lambda > 1$, respectively, then we have for any $\gamma \in (1/2, (1+\lambda)/4 \wedge \lambda/2)$ and $n \to \infty$*

$$\sqrt{n}\big(M_h(\tilde{C}_n, \hat{A}) - M_h(C, A^*)\big) \overset{w}{\rightsquigarrow} Z_1,$$

where the random variable Z_1 is defined as

$$Z_1 = 2\int_0^1 \int_0^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} v(y,t)\,dy\,dt$$

with

$$v(y,t) = \big(\log C(y^{1-t}, y^t) - \log(y)A^*(t)\big)\frac{h^*(y)}{(\log y)^2}.$$

REMARK 4.3. (a) Note that the weight functions $h_k^*(y) = -y^k \log y$ satisfy the assumptions of Theorems 4.1 and 4.2 for $k > 1$ and $k > 0$, respectively.

(b) The preceding two theorems yield a consistent asymptotic level α test for the hypothesis of extreme-value dependence by rejecting the null hypothesis H_0 if

$$nM_h(\tilde{C}_n, \hat{A}_{n,h}) > z_{1-\alpha}, \tag{4.2}$$

where $z_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the distribution of the random variable Z_0.

(c) By Theorem 4.2, the power of the test (4.2) is approximately given by

$$\mathbb{P}\big(nM_h(\tilde{C}_n, \hat{A}_{n,h}) > z_{1-\alpha}\big) \approx 1 - \Phi\left(\frac{z_{1-\alpha}}{\sqrt{n}\sigma} - \sqrt{n}\frac{M_h(C, A^*)}{\sigma}\right)$$
$$\approx \Phi\left(\sqrt{n}\frac{M_h(C, A^*)}{\sigma}\right),$$

where the function A^* is defined in (2.3) corresponding to the best approximation of the logarithm of the copula C by a function of the form (1.4), σ is the standard deviation of the distribution of the random variable Z_1 and Φ is the standard normal distribution function. Thus, the power of the test (4.2) is an increasing function of the quantity $M_h(C, A^*)\sigma^{-1}$.

4.2. *Multiplier bootstrap*. In general, the distribution of the random variable Z_0 cannot be determined explicitly, because of its complicated dependence on the (unknown) copula C. We hence propose to determine the quantiles by the multiplier bootstrap approach as described in Bücher and Dette (2010). To be precise, let $\xi_1, \dots, \xi_n$ denote independent identically distributed random variables with

$$\mathbb{P}(\xi_1 = 0) = \mathbb{P}(\xi_1 = 2) = 1/2.$$

We define $\bar{\xi}_n = n^{-1}\sum_{i=1}^n \xi_i$ as the mean of $\xi_1, \dots, \xi_n$ and consider the multiplier statistics

$$\hat{C}_n^*(u, v) = \hat{F}_n^*(\hat{F}_{n1}^-(u), \hat{F}_{n2}^-(v)),$$

where

$$\hat{F}_n^*(x_1, x_2) = \frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\bar{\xi}_n}\mathbb{I}\{X_{i1} \le x_1, X_{i2} \le x_2\},$$

and $\hat{F}_{nj}$ denotes the marginal empirical distribution functions. If we estimate the partial derivatives of the copula C by

$$\widehat{\partial_1 C}(u, v) := \frac{\hat{C}_n(u+h, v) - \hat{C}_n(u-h, v)}{2h},$$
$$\widehat{\partial_2 C}(u, v) := \frac{\hat{C}_n(u, v+h) - \hat{C}_n(u, v-h)}{2h},$$

where $h = n^{-1/2} \to 0$, we can approximate the distribution of $\mathbb{G}_C$ by the distribution of the process

$$\hat{\alpha}_n^{\mathrm{pdm}}(u, v) := \hat{\beta}_n(u, v) - \widehat{\partial_1 C}(u, v)\hat{\beta}_n(u, 1) - \widehat{\partial_2 C}(u, v)\hat{\beta}_n(1, v), \tag{4.3}$$

where $\hat{\beta}_n(u, v) = \sqrt{n}(\hat{C}_n^*(u, v) - \hat{C}_n(u, v))$. More precisely, it was shown by Bücher and Dette (2010) that we have weak convergence conditional on the data in probability toward $\mathbb{G}_C$, that is,

$$\hat{\alpha}_n^{\mathrm{pdm}} \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{G}_C \qquad \text{in } l^\infty[0, 1]^2, \tag{4.4}$$

where the symbol $\underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}}$ denotes weak convergence conditional on the data in probability as defined by Kosorok (2008), that is, $\alpha_n^{\mathrm{pdm}} \rightsquigarrow_\xi^{\mathbb{P}} \mathbb{G}_C$ if

$$\sup_{h \in BL_1(l^\infty[0,1]^2)} |\mathbb{E}_\xi h(\alpha_n^{\mathrm{pdm}}) - \mathbb{E}h(\mathbb{G}_C)| \xrightarrow{\mathbb{P}} 0 \tag{4.5}$$

and

$$\mathbb{E}_\xi h(\alpha_n^{\mathrm{pdm}})^* - \mathbb{E}_\xi h(\alpha_n^{\mathrm{pdm}})_* \xrightarrow{\mathbb{P}} 0 \qquad \text{for every } h \in BL_1(l^\infty[0, 1]^2). \tag{4.6}$$

Here

$$\begin{aligned} &BL_1(l^\infty[0, 1]^2) \\ &\quad = \{f : l^\infty[0, 1]^2 \to \mathbb{R} : \|f\|_\infty \le 1, |f(\beta) - f(\gamma)| \le \|\beta - \gamma\|_\infty \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall \gamma, \beta \in l^\infty[0, 1]^2\} \end{aligned}$$

is the class of all uniformly bounded functions which are Lipschitz continuous with constant smaller one, and $\mathbb{E}_\xi$ denotes the conditional expectation with respect to the weights ξ_n given the data $(X_1, Y_1) \cdots (X_n, Y_n)$. As a consequence, we obtain the following bootstrap approximation for Z_0.

THEOREM 4.4. *If condition* (3.6) *is satisfied*, *the weight function h satisfies the conditions of Theorem* 4.1 *and $h^*(y)(y \log y)^{-2}$ is uniformly bounded then*

$$\begin{aligned} \hat{Z}_0^* = &\int_0^1 \int_0^1 \left(\frac{\hat{\alpha}_n^{\mathrm{pdm}}(y^{1-t}, y^t)}{\tilde{C}_n(y^{1-t}, y^t)} \right)^2 \frac{h^*(y)}{(\log y)^2}\, dy\, dt \\ &- B_h^{-1} \int_0^1 \left(\int_0^1 \frac{\hat{\alpha}_n^{\mathrm{pdm}}(y^{1-t}, y^t)}{\tilde{C}_n(y^{1-t}, y^t)} \frac{h^*(y)}{\log y}\, dy \right)^2 dt \end{aligned}$$

converges weakly to Z_0 conditional on the data, *that is*,

$$\hat{Z}_0^* \underset{\xi}{\overset{\mathbb{P}}{\rightsquigarrow}} Z_0 \qquad \textit{in } l^\infty[0, 1].$$

By Theorem 4.4, $\hat{Z}_0^*$ is a valid bootstrap approximation for the distribution of Z_0. Consequently, repeating the procedure B times yields a sample $\hat{Z}_0^*(1), \dots, \hat{Z}_0^*(B)$ that is approximately distributed according to Z_0 and we can use the empirical $(1-\alpha)$-quantile of this sample, say $z_{1-\alpha}^*$, as an approximation for $z_{1-\alpha}$. Therefore, rejecting the null hypothesis if

$$nM_h(\tilde{C}_n, \hat{A}_{n,h}) > z_{1-\alpha}^* \tag{4.7}$$

yields a consistent asymptotic level α test for extreme-value dependence.

Note that the condition on the boundedness of the function $h^*(y)(y \log y)^2$ is not satisfied for any member of the class $h_k^*(y) = -y^k/\log(y)$ from Example 2.5. Nevertheless, mimicking the procedure from Kojadinovic and Yan (2010) and using $h_k^*(y)\mathbb{I}_{[\varepsilon,1-\varepsilon]}(y)$ instead of $h_k^*(y)$ is sufficient for the boundedness. Since this is the procedure being usually performed in practical applications, Theorem 4.4 is still valuable for the weight functions investigated in this paper.

4.3. *Finite sample properties.* In this subsection, we investigate the finite sample properties of the test for extreme-value dependence. We consider the asymmetric negative logistic model (3.22), the symmetric mixed model (3.28) and additionally the symmetric model of Gumbel

$$A(t) = \left(t^\theta + (1-t)^\theta\right)^{1/\theta} \tag{4.8}$$

with parameter $\theta \in [1, \infty)$ [see Gumbel (1960)] and the model of Hüsler and Reiss

$$A(t) = (1-t)\Phi\left(\theta + \frac{1}{2\theta}\log\frac{1-t}{t}\right) + t\Phi\left(\theta + \frac{1}{2\theta}\log\frac{t}{1-t}\right), \tag{4.9}$$

where $\theta \in (0, \infty)$ and Φ is the standard normal distribution function [see Hüsler and Reiss (1989)]. The coefficient of tail dependence in (4.9) is given by $\rho = 2(1-\Phi(\theta))$, that is, independence is obtained for $\theta \to \infty$ and complete dependence for $\theta \to 0$. For the Gumbel model (4.8), complete dependence is obtained in the limit as θ approaches infinity while independence corresponds to $\theta = 1$. The coefficient of tail dependence $\rho = 2(1 - A(0.5))$ is given by $\rho = 2 - 2^{1/\theta}$.

We generated 1,000 random samples of sample size $n = 200$ from various copula models and calculated the probability of rejecting the null hypothesis. Under the null hypothesis, we chose the model parameters in such a way that the coefficient of tail dependence ρ varies over the specific range of the corresponding model. Under the alternative, the coefficient of tail dependence does not need to exist and we therefore chose the model parameters, such that Kendall's τ is an element of the set $\{1/4, 1/2, 3/4\}$. The weight function is chosen as $h_{0.4}(y) = -y^{0.4}/\log(y)$ and the critical values are determined by the multiplier bootstrap approach as described in Section 4.2 with $B = 200$ Bootstrap replications. The results are stated in Table 1.

We observe from the left part of Table 1 that the level of test is accurately approximated for most of the models, if the tail dependence is not too strong. For

TABLE 1

Simulated rejection probabilities of the test (4.7) *for the null hypothesis of an extreme-value copula for various models. The first four columns deal with models under the null hypothesis, while the last four are from the alternative*

H_0-model	ρ	0.05	0.1	H_1-model	τ	0.05	0.1
Independence	0	0.031	0.075	Clayton	0.25	0.874	0.916
Gumbel	0.25	0.045	0.098		0.5	1	1
	0.5	0.029	0.066		0.75	0.999	1
	0.75	0.025	0.065	Frank	0.25	0.291	0.396
Mixed model	0.25	0.043	0.09		0.5	0.73	0.822
	0.5	0.047	0.10		0.75	0.783	0.898
Asy. Neg. Log.	0.25	0.041	0.09	Gaussian	0.25	0.168	0.240
	0.5	0.038	0.077		0.5	0.237	0.336
Hüsler–Reiß	0.25	0.04	0.091		0.75	0.084	0.156
	0.5	0.045	0.089	t_4	0.25	0.105	0.187
	0.75	0.009	0.053		0.5	0.158	0.263
					0.75	0.046	0.092

a large tail dependence coefficient the bootstrap test is conservative. This phenomenon can be explained by the fact that for the limiting case of random variables distributed according to the upper Fréchet–Hoeffding the empirical copula $\hat{C}_n$ does not converge weakly to a nondegenerate process at a rate $1/\sqrt{n}$, rather in this case it follows that $\|\hat{C}_n - C\| = O(1/n)$. Consequently, the approximations proposed in this paper, which are based on the weak convergence of $\sqrt{n}(\hat{C}_n - C)$ to a nondegenerate process, are not appropriate for small samples, if the tail dependence coefficient is large. Considering the alternative, we observe reasonably good power for the Frank and Clayton copulas, while for the Gaussian or t-copula deviations from an extreme-value copula are not detected well with a sample size $n = 200$. In some cases, the power of the test (4.7) is close the nominal level. This observation can be again explained by the closeness to the upper Fréchet–Hoeffding bound.

Indeed, we can use the minimal distance $M_h(C, A^*)$ as a measure of deviation from an extreme-value copula. Calculating the minimal distance $M_h(C, A^*)$ (with Kendall's $\tau = 0.5$ and $h = h_{0.4}$), we observe that the minimal distances are about ten times smaller for the Gaussian and t_4 than for the Frank and Clayton copula, that is,

$$M_h(C, A^*_{\text{Clayton}}) = 1.65 \times 10^{-3}, \qquad M_h(C, A^*_{\text{Frank}}) = 5.87 \times 10^{-4},$$

$$M_h(C, A^*_{\text{Gaussian}}) = 2.08 \times 10^{-4}, \qquad M_h(C, A^*_{t_4}) = 1.18 \times 10^{-4}.$$

Moreover, as explained in Remark 4.3(b) the power of the tests (4.2) and (4.7) is an increasing function of the quantity $p(\text{copula}) = M_h(C, A^*)\sigma^{-1}$. For the four copulas considered in the simulation study (with $\tau = 0.5$), the corresponding ratios are approximately given by

$$p(\text{Clayton}) = 0.230, \qquad p(\text{Frank}) = 0.134,$$
$$p(\text{Gaussian}) = 0.083, \qquad p(t_4) = 0.064,$$

which provides some theoretical explanation of the findings presented in Table 1. Loosely speaking, if the value $M_h(C, A^*)\sigma^{-1}$ is very small a larger sample size is required to detect a deviation from an extreme-value copula. This statement is confirmed by further simulations results. For example, for the Gaussian and t_4 copula (with Kendall's $\tau = 0.75$) we obtain for the sample size $n = 500$ the rejection probabilities 0.766 (0.629) and 0.40 (0.544) for the bootstrap test with level 5% (10%), respectively.

APPENDIX A: PROOFS

PROOF OF THEOREM 3.1. Fix $\lambda > 1$ as in (3.11) and $\gamma \in (1/2, \lambda/2)$. Due to Lemma 1.10.2(i) in Van der Vaart and Wellner (1996), the process $\sqrt{n}(\tilde{C}_n - C)$ will have the same weak limit (with respect to the $\overset{w}{\rightsquigarrow}$ convergence) as $\sqrt{n}(\hat{C}_n - C)$.

For $i = 2, 3, \dots,$ we consider the following random functions in $l^\infty[0, 1]$:

$$W_n(t) = \int_0^1 \sqrt{n}\big(\log \tilde{C}_n(y^{1-t}, y^t) - \log C(y^{1-t}, y^t)\big) w(y, t)\, dy,$$

$$W_{i,n}(t) = \int_{1/i}^1 \sqrt{n}\big(\log \tilde{C}_n(y^{1-t}, y^t) - \log C(y^{1-t}, y^t)\big) w(y, t)\, dy,$$

$$W(t) = \int_0^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} w(y, t)\, dy,$$

$$W_i(t) = \int_{1/i}^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} w(y, t)\, dy.$$

We prove the theorem by an application of Theorem 4.2 in Billingsley (1968), adapted to the concept of weak convergence in the sense of Hoffmann–Jørgensen, see, for example, Van der Vaart and Wellner (1996). More precisely, we will show in Lemma B.1 in Appendix B that the weak convergence $W_n \overset{w}{\rightsquigarrow} W$ in $l^\infty[0, 1]$ follows from the following three assertions:

$$\text{(A.1)} \quad \begin{aligned} &\text{(i)} \quad \text{For every } i \geq 2 \qquad W_{i,n} \overset{w}{\rightsquigarrow} W_i \qquad \text{for } n \to \infty \text{ in } l^\infty[0, 1], \\ &\text{(ii)} \quad W_i \overset{w}{\rightsquigarrow} W \qquad \text{for } i \to \infty \text{ in } l^\infty[0, 1], \\ &\text{(iii)} \quad \text{For every } \varepsilon > 0 \\ &\qquad \lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*\Big(\sup_{t\in[0,1]} |W_{i,n}(t) - W_n(t)| > \varepsilon\Big) = 0. \end{aligned}$$

The main part of the proof now consists in the verification assertion (iii).

We begin by proving assertion (i). For this purpose, set $T_i = [1/i, 1]^2$ and consider the mapping

$$\Phi_1 : \begin{cases} \mathbb{D}_{\Phi_1} \to l^\infty(T_i), \\ f \mapsto \log \circ f, \end{cases}$$

where its domain $\mathbb{D}_{\Phi_1}$ is defined by $\mathbb{D}_{\Phi_1} = \{f \in l^\infty(T_i) : \inf_{x \in T_i} |f(x)| > 0\} \subset l^\infty(T_i)$. By Lemma 12.2 in Kosorok (2008), it follows that Φ_1 is Hadamard-differentiable at C, tangentially to $l^\infty(T_i)$, with derivative $\Phi'_{1,C}(f) = f/C$. Since $\tilde{C}_n \geq n^{-\gamma}$ and $C \geq \Pi$ we have $\tilde{C}_n, C \in \mathbb{D}_{\Phi_1}$ and the functional delta method [see Theorem 2.8 in Kosorok (2008)] yields

$$\sqrt{n}(\log \tilde{C}_n - \log C) \overset{w}{\rightsquigarrow} \mathbb{G}_C / C$$

in $l^\infty(T_i)$. Next, we consider the operator

$$\Phi_2 : \begin{cases} l^\infty(T_i) \to l^\infty([1/i, 1] \times [0, 1]), \\ f \mapsto f \circ \varphi, \end{cases}$$

where the mapping $\varphi : [1/i, 1] \times [0, 1] \to T_i$ is defined by $\varphi(y, t) = (y^{1-t}, y^t)$. Observing

$$\sup_{(y,t) \in [1/i,1] \times [0,1]} |f \circ \varphi(y, t) - g \circ \varphi(y, t)| \leq \sup_{\mathbf{x} \in T_i} |f(\mathbf{x}) - g(\mathbf{x})|$$

we can conclude that Φ_2 is Lipschitz-continuous. By the continuous mapping theorem [see, e.g., Theorem 7.7 in Kosorok (2008)] and conditions (3.9) and (3.10), we immediately obtain

$$\sqrt{n}\big(\log \tilde{C}_n(y^{1-t}, y^t) - \log C(y^{1-t}, y^t)\big) w(y, t) \overset{w}{\rightsquigarrow} \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} w(y, t)$$

in $l^\infty([1/i, 1] \times [0, 1])$. The assertion in (i) now follows by continuity of integration with respect to the variable y.

For the proof of assertion (ii), we simply note that $\mathbb{G}_C$ is bounded on $[0, 1]^2$ and that

$$K(y, t) = \frac{w(y, t)}{C(y^{1-t}, y^t)}$$

is uniformly bounded with respect to $t \in [0, 1]$ by the integrable function $\bar{K}(y) = \bar{w}(y) y^{-1}$.

For the proof of assertion (iii), choose some $\alpha \in (0, 1/2)$ such that $\lambda\alpha > \gamma$ and consider the decomposition

$$\begin{aligned} &W_n(t) - W_{i,n}(t) \\ &\quad = \int_0^{1/i} \sqrt{n}\big(\log \tilde{C}_n(y^{1-t}, y^t) - \log C(y^{1-t}, y^t)\big) w(y, t)\, dy \qquad \text{(A.2)} \\ &\quad = B_i^{(1)}(t) + B_i^{(2)}(t), \end{aligned}$$

where

$$B_i^{(j)}(t) = \int_{I_{B_i^{(j)}(t)}} \sqrt{n} \log \frac{\tilde{C}_n}{C}(y^{1-t}, y^t) w(y,t)\, dy, \qquad j = 1, 2, \tag{A.3}$$

and

$$\begin{aligned} I_{B_i^{(1)}(t)} &= \{0 < y < 1/i \,|\, C(y^{1-t}, y^t) > n^{-\alpha}\}, \\ I_{B_i^{(2)}(t)} &= (0,1) \setminus I_{B_i^{(1)}(t)}. \end{aligned} \tag{A.4}$$

The usual estimate

$$\begin{aligned} &\mathbb{P}^*\Big(\sup_{t\in[0,1]} |W_{i,n}(t) - W_n(t)| > \varepsilon\Big) \\ &\quad \le \mathbb{P}^*\Big(\sup_{t\in[0,1]} \big|B_i^{(1)}(t)\big| > \varepsilon/2\Big) + \mathbb{P}^*\Big(\sup_{t\in[0,1]} \big|B_i^{(2)}(t)\big| > \varepsilon/2\Big) \end{aligned} \tag{A.5}$$

allows for individual investigation of both expressions, and we begin with the term $\sup_{t\in[0,1]}|B_i^{(1)}(t)|$. By the mean value theorem applied to the logarithm, we have

$$\begin{aligned} \log \frac{\tilde{C}_n}{C}(y^{1-t}, y^t) &= \log \tilde{C}_n(y^{1-t}, y^t) - \log C(y^{1-t}, y^t) \\ &= (\tilde{C}_n - C)(y^{1-t}, y^t) \frac{1}{C^*(y,t)}, \end{aligned} \tag{A.6}$$

where $C^*(y,t)$ is some intermediate point satisfying $|C^*(y,t) - C(y^{1-t}, y^t)| \le |\tilde{C}_n(y^{1-t}, y^t) - C(y^{1-t}, y^t)|$. Especially, observing $C \ge \Pi$ we have

$$C^*(y,t) \ge (C \wedge \tilde{C}_n)(y^{1-t}, y^t) \ge y \wedge \left(y \frac{\tilde{C}_n}{C}(y^{1-t}, y^t)\right) \tag{A.7}$$

and therefore

$$\begin{aligned} \sup_{t\in[0,1]} \big|B_i^{(1)}(t)\big| &\le \sup_{t\in[0,1]} \int_{I_{B_i^{(1)}(t)}} \sqrt{n} |(\tilde{C}_n - C)(y^{1-t}, y^t)| \\ &\qquad \times \left| 1 \vee \frac{C}{\tilde{C}_n}(y^{1-t}, y^t)\right| w(y,t) y^{-1}\, dy \\ &\le \sup_{\mathbf{x}\in[0,1]^2} \sqrt{n} |\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})| \\ &\qquad \times \left(1 \vee \sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x}) > n^{-\alpha}} \left| \frac{C}{\tilde{C}_n}(\mathbf{x})\right|\right) \times \psi(i) \end{aligned}$$

with $\psi(i) = \int_0^{1/i} \bar{w}(y) y^{-1}\, dy = o(1)$ for $i \to \infty$. This yields for the first term on the right-hand side of (A.5)

$$\mathbb{P}^*\Bigl(\sup_{t\in[0,1]} |B_i^{(1)}(t)| > \varepsilon\Bigr) \le \mathbb{P}^*\Biggl(\sup_{\mathbf{x}\in[0,1]^2} \sqrt{n}|\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})| > \sqrt{\frac{\varepsilon}{\psi(i)}}\Biggr) + \mathbb{P}^*\Biggl(1 \vee \sup_{C(\mathbf{x})>n^{-\alpha}} \left|\frac{C}{\tilde{C}_n}(\mathbf{x})\right| > \sqrt{\frac{\varepsilon}{\psi(i)}}\Biggr). \tag{A.8}$$

Since $\sup_{\mathbf{x}\in[0,1]^2} \sqrt{n}|\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})|$ is asymptotically tight, we immediately obtain

$$\lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*\Biggl(\sup_{\mathbf{x}\in[0,1]^2} \sqrt{n}|\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})| > \sqrt{\frac{\varepsilon}{\psi(i)}}\Biggr) = 0. \tag{A.9}$$

For the estimation of the second term in (A.8), we note that

$$\sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})>n^{-\alpha}} \left|\frac{\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})}{C(\mathbf{x})}\right| < n^{\alpha} \sup_{\mathbf{x}\in[0,1]^2} |\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})| \xrightarrow{\mathbb{P}^*} 0, \tag{A.10}$$

which in turn implies

$$\begin{aligned}
\sup_{C(\mathbf{x})>n^{-\alpha}} \left|\frac{C}{\tilde{C}_n}(\mathbf{x})\right| &= \sup_{C(\mathbf{x})>n^{-\alpha}} \left|1 + \frac{\tilde{C}_n - C}{C}(\mathbf{x})\right|^{-1} \\
&\le \Biggl(1 - \sup_{C(\mathbf{x})>n^{-\alpha}} \left|\frac{\tilde{C}_n - C}{C}(\mathbf{x})\right|\Biggr)^{-1} \mathbb{I}_{A_n} \\
&\quad + \Biggl(\sup_{C(\mathbf{x})>n^{-\alpha}} \left|1 + \frac{\tilde{C}_n - C}{C}(\mathbf{x})\right|^{-1}\Biggr) \mathbb{I}_{\Omega\setminus A_n} \\
&\xrightarrow{\mathbb{P}^*} 1,
\end{aligned} \tag{A.11}$$

where $A_n = \{\sup_{C(\mathbf{x})>n^{-\alpha}} |\frac{\tilde{C}_n - C}{C}(\mathbf{x})| < 1/2\}$. This implies that the function $\max\{1, \sup_{C(\mathbf{x})>n^{-\alpha}} |\frac{C}{\tilde{C}_n}(\mathbf{x})|\}$ can be bounded by a function that converges to one in outer probability, and thus

$$\lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*\Biggl(1 \vee \sup_{C(\mathbf{x})>n^{-\alpha}} \left|\frac{C}{\tilde{C}_n}(\mathbf{x})\right| > \sqrt{\frac{\varepsilon}{\psi(i)}}\Biggr) = 0.$$

Observing (A.8) and (A.9) it remains to estimate the second term on the right-hand side of (A.5). We make use of the mean value theorem again [see (A.6)] but use the estimate

$$C^*(y,t) \ge (C \wedge \tilde{C}_n)(y^{1-t}, y^t) \ge y^{\lambda} \wedge y^{\lambda} \frac{\tilde{C}_n}{C^{\lambda}}(y^{1-t}, y^t) \tag{A.12}$$

[recall that $\lambda > 1$ by assumption (3.11)]. This yields

$$\sup_{t\in[0,1]} \left|B_i^{(2)}(t)\right| \le \sup_{t\in[0,1]} \int_{I_{B_i^{(2)}(t)}} \sqrt{n}|(\tilde{C}_n - C)(y^{1-t}, y^t)| \times \left|1 \vee \frac{C^\lambda}{\tilde{C}_n}(y^{1-t}, y^t)\right| w(y,t) y^{-\lambda}\, dy$$

$$\le \sup_{\mathbf{x}\in[0,1]^2} \sqrt{n}|\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})| \times \left(1 \vee \sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})\le n^{-\alpha}} \left|\frac{C^\lambda}{\tilde{C}_n}(\mathbf{x})\right|\right) \times \phi(i),$$

where $\phi(i) = \int_0^{1/i} \bar{w}(y) y^{-\lambda}\, dy = o(1)$ for $i \to \infty$ by condition (3.11). Using analogous arguments as for the estimation of $\sup_{t\in[0,1]}|B_i^{(1)}(t)|$ the assertion follows from

$$\sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})\le n^{-\alpha}} \left|\frac{C^\lambda}{\tilde{C}_n}(\mathbf{x})\right| \le \sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})\le n^{-\alpha}} |n^\gamma C^\lambda(\mathbf{x})| \le n^{\gamma - \lambda\alpha} = o(1)$$

due to the choice of γ and α. $\square$

PROOF OF THEOREM 3.4. The proof will also be based on Lemma B.1 in Appendix B verifying conditions (i)–(iii) in (A.1). A careful inspection of the previous proof shows that the verification of condition (i) in (A.1) remains valid. Regarding condition (ii), we have to show that the process $\frac{\mathbb{G}_C}{C}(y^{1-t}, y^t)$ is integrable on the interval $(0, 1)$. For this purpose, we write

$$\mathbb{G}_C(\mathbf{x}) = \mathbb{B}_C(\mathbf{x}) - \partial_1 C(\mathbf{x})\mathbb{B}_C(x_1, 1) - \partial_2 C(\mathbf{x})\mathbb{B}_C(1, x_2)$$

and consider each term separately. From Theorem G.1 in Genest and Segers (2009), we know that for any $\omega \in (0, 1/2)$ the process

$$\tilde{\mathbb{B}}_C(\mathbf{x}) = \begin{cases} \dfrac{\mathbb{B}_C(\mathbf{x})}{(x_1 \wedge x_2)^\omega (1 - x_1 \wedge x_2)^\omega}, & \text{if } x_1 \wedge x_2 \in (0,1), \\ 0, & \text{if } x_1 = 0 \text{ or } x_2 = 0 \text{ or } \mathbf{x} = (1,1), \end{cases}$$

has continuous sample paths on $[0, 1]^2$. Considering $C(y^{1-t}, y^t) \ge y$ and using the notation

$$K_1(y,t) = q_\omega(y^{1-t} \wedge y^t) y^{-1}, \tag{A.13}$$

$$K_2(y,t) = \partial_1 C(y^{1-t}, y^t) q_\omega(y^{1-t}) y^{-1}, \tag{A.14}$$

$$K_3(y,t) = \partial_2 C(y^{1-t}, y^t) q_\omega(y^t) y^{-1} \tag{A.15}$$

with $q_\omega(t) = t^\omega(1 - t)^\omega$ it remains to show that there exist integrable functions $K_j^*(y)$ with $K_j(y,t) \le K_j^*(y)$ for all $t \in [0, 1]$ $(j = 1, 2, 3)$. For K_1 this is immediate because $K_1(y,t) \le (y^{1-t} \wedge y^t)^\omega y^{-1} \le y^{\omega/2-1}$. For K_2, note that $\partial_1 C(y^{1-t}, y^t) = \mu(t) y^{A(t)-(1-t)}$, with $\mu(t) = A(t) - tA'(t)$. Therefore,

$$K_2(y,t) \le \mu(t) y^{A(t)-(1-\omega)(1-t)-1} \le \mu(t) y^{\omega/2-1} \le 2y^{\omega/2-1}, \tag{A.16}$$

where the second estimate follows from the inequality $t \vee (1-t) \le A(t) \le 1$ and holds for $\omega \in (0,2)$. A similar argument works for the term K_3.

For the verification of condition (iii), we proceed along similar lines as in the previous proof. We begin by choosing some $\beta \in (1, 9/8)$, $\omega \in (1/4, 1/2)$ and some $\alpha \in (4/9, \gamma \wedge (2-\omega)^{-1})$ in such a way that $\gamma < \beta\alpha$. First, note that $y \le 1/(n+2)^2$ implies $\tilde{C}_n(y^{1-t}, y^t) = n^{-\gamma}$ for all $t \in [0,1]$. This yields

$$\int_0^{(n+2)^{-2}} \sqrt{n}(\log \tilde{C}_n - \log C)(y^{1-t}, y^t)\, dy = O\left(\frac{\log n}{n^{3/2}}\right)$$

uniformly with respect to $t \in [0,1]$, and therefore it is sufficient to consider the decomposition in (A.2) with the sets

$$I_{B_i^{(1)}(t)} = \{1/(n+2)^2 < y < 1/i \,|\, C(y^{1-t}, y^t) > n^{-\alpha}\},$$
$$I_{B_i^{(2)}(t)} = \bigl(1/(n+2)^2, 1/i\bigr) \setminus I_{B_i^{(1)}(t)}.$$

We can estimate the term $B_i^{(1)}(t)$ analogously to the previous proof by

$$|B_i^{(1)}(t)| \le \int_{I_{B_i^{(1)}(t)}} \sqrt{n}|(\tilde{C}_n - C)(y^{1-t}, y^t)| \times \left|1 \vee \frac{C}{\tilde{C}_n}(y^{1-t}, y^t)\right| y^{-1}\, dy.$$

Let H_n denote the empirical distribution function of the standardized sample $(F(X_1), G(Y_1)), \dots, (F(X_n), G(Y_n))$. By the results in Segers [(2010), Section 5] we can decompose $\sqrt{n}(\tilde{C}_n - C) = \sqrt{n}(C_n \vee n^{-\gamma} - C)$ as follows:

$$\begin{aligned} \sqrt{n}(\tilde{C}_n - C)(\mathbf{x}) &= \sqrt{n}(C_n - C)(\mathbf{x}) + \sqrt{n}(\tilde{C}_n - C_n)(\mathbf{x}) \\ &= \alpha_n(\mathbf{x}) - \partial_1 C(\mathbf{x})\alpha_n(x_1, 1) - \partial_2 C(\mathbf{x})\alpha_n(1, x_2) \\ &\quad + \tilde{R}_n(\mathbf{x}), \end{aligned} \tag{A.17}$$

where $\alpha_n(\mathbf{x}) = \sqrt{n}(H_n - C)(\mathbf{x})$ and the remainder satisfies

$$\sup_{\mathbf{x}\in[0,1]^2} |\tilde{R}_n(\mathbf{x})| = O\bigl(n^{1/2-\gamma} + n^{-1/4}(\log n)^{1/2}(\log\log n)^{3/4}\bigr) \qquad \text{a.s.} \tag{A.18}$$

Note that the estimate of (A.18) requires validity of condition 5.1 in Segers (2010). This condition is satisfied provided that the function A is assumed to be twice continuously differentiable; see Example 6.3 in Segers (2010). With (A.17), we can estimate the term $|B_i^{(1)}(t)|$ analogously to decomposition (A.2) by $B_{i,1}^{(1)}(t) +$

$\cdots + B_{i,4}^{(1)}(t)$, where

$$B_{i,1}^{(1)}(t) = \int_{I_{B_i^{(1)}(t)}} |\alpha_n(y^{1-t}, y^t)| \left| 1 \vee \frac{C}{\tilde{C}_n}(y^{1-t}, y^t) \right| y^{-1}\, dy,$$

$$B_{i,2}^{(1)}(t) = \int_{I_{B_i^{(1)}(t)}} \partial_1 C(y^{1-t}, y^t) |\alpha_n(y^{1-t}, 1)| \left| 1 \vee \frac{C}{\tilde{C}_n}(y^{1-t}, y^t) \right| y^{-1}\, dy,$$

$$B_{i,3}^{(1)}(t) = \int_{I_{B_i^{(1)}(t)}} \partial_2 C(y^{1-t}, y^t) |\alpha_n(1, y^t)| \left| 1 \vee \frac{C}{\tilde{C}_n}(y^{1-t}, y^t) \right| y^{-1}\, dy,$$

$$B_{i,4}^{(1)}(t) = \int_{I_{B_i^{(1)}(t)}} |\tilde{R}_n(y^{1-t}, y^t)| \left| 1 \vee \frac{C}{\tilde{C}_n}(y^{1-t}, y^t) \right| y^{-1}\, dy.$$

The decomposition in (A.17), Theorem G.1 in Genest and Segers (2009) and the inequality $\alpha < \gamma \wedge (2-\omega)^{-1}$ may be used to conclude

$$\sup_{(y,t)\,:\,C(y^{1-t},y^t)>n^{-\alpha}} \left| \frac{\tilde{C}_n - C}{C}(y^{1-t}, y^t) \right| = o_{\mathbb{P}^*}(1),$$

which in turn implies

$$1 \vee \sup_{(y,t)\,:\,C(y^{1-t},y^t)>n^{-\alpha}} \left| \frac{C}{\tilde{C}_n}(y^{1-t}, y^t) \right| = O_{\mathbb{P}^*}(1) \tag{A.19}$$

analogously to (A.11). Together with (A.18) and observing the inequality $\int_{(n+2)^{-2}}^{1/i} y^{-1}\, dy \le 2\log(n+2)$, we obtain, for $n \to \infty$

$$\sup_{t\in[0,1]} B_{i,4}^{(1)}(t) = O_{\mathbb{P}^*}\big(n^{1/2-\gamma}\log n + n^{-1/4}(\log n)^{3/2}(\log\log n)^{1/4}\big) = o_{\mathbb{P}^*}(1),$$

which implies

$$\lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*\Big(\sup_{t\in[0,1]} B_{i,4}^{(1)}(t) > \varepsilon/4 \Big) = 0. \tag{A.20}$$

Observing that $q_\omega(y^{1-t} \wedge y^t) \le y^{\omega/2}$ the first term $B_{i,1}^{(1)}(t)$ can be estimated by

$$\begin{aligned}\sup_{t\in[0,1]} B_{i,1}^{(1)}(t) \le{}& \sup_{\mathbf{x}\in[0,1]^2} \frac{|\alpha_n(\mathbf{x})|}{q_\omega(x_1 \wedge x_2)} \\ &\times \left(1 \vee \sup_{(y,t)\,:\,C(y^{1-t},y^t)>n^{-\alpha}} \left| \frac{C}{\tilde{C}_n}(y^{1-t}, y^t) \right| \right) \times \psi(i),\end{aligned}$$

where $\psi(i) = \int_0^{1/i} y^{-1+\omega/2}\, dy = o(1)$ for $i \to \infty$. Using analogous arguments as in the previous proof we can conclude, using of (A.19) and Theorem G.1 in Genest and Segers (2009), that $\lim_{i\to\infty}\limsup_{n\to\infty}\mathbb{P}^*(\sup_{t\in[0,1]}B_{i,1}^{(1)}(t)>\varepsilon/4)=0$. For the second summand, we note that

$$\sup_{t\in[0,1]}B_{i,2}^{(1)}(t)\le\sup_{x_1\in[0,1]}\frac{|\alpha_n(x_1,1)|}{q_\omega(x_1)}\times\biggl(1\vee\sup_{(y,t):C(y^{1-t},y^t)>n^{-\alpha}}\biggl|\frac{C}{\tilde{C}_n}(y^{1-t},y^t)\biggr|\biggr)$$
$$\times\sup_{t\in[0,1]}\int_0^{1/i}K_2(y,t)\,dy,$$

where $K_2(y,t)$ is defined in (A.14). Observing the estimate in (A.16), we easily obtain $\lim_{i\to\infty}\sup_{t\in[0,1]}\int_0^{1/i}K_2(y,t)\,dy=0$. Again, under consideration of (A.19) and Theorem G.1 in Genest and Segers (2009), we have $\lim_{i\to\infty}\limsup_{n\to\infty}\mathbb{P}^*(\sup_{t\in[0,1]}B_{i,2}^{(1)}(t)>\varepsilon/4)=0$. A similar argument works for $B_{i,3}^{(1)}$ and from the estimates for the different terms the assertion

$$\lim_{i\to\infty}\limsup_{n\to\infty}\mathbb{P}^*\Bigl(\sup_{t\in[0,1]}\bigl|B_i^{(1)}(t)\bigr|>\varepsilon\Bigr)=0$$

follows. Considering the term $\sup_{t\in[0,1]}|B_i^{(2)}(t)|$, we proceed along similar lines as in the proof of Theorem 3.1. For the sake of brevity, we only state the important differences: in estimation (A.12) replace λ by β, then make use of decomposition (A.17), calculations similar to (A.16), and Theorem G.1 in Genest and Segers (2009) again and for the estimation of the remainder note that $\int_{1/(n+2)^2}^{1/i}y^{-\beta}\,dy=O(n^{2(\beta-1)})$. $\square$

PROOF OF THEOREM 3.8. Let η denote a probability measure minimizing the functional V defined in (3.19). Note that V is convex and define for $\alpha\in[0,1]$ and a further probability measure ξ on $[0,1]$ the function

$$g(\alpha)=V\bigl(\alpha\xi+(1-\alpha)\eta\bigr).$$

Because V is convex it follows that η is optimal if and only if the directional derivative of η in the direction $\xi-\eta$ satisfies

$$\begin{aligned}0\le g'(0+)&=\lim_{\alpha\to0+}\frac{g(\alpha)-g(0)}{\alpha}\\&=2\int_0^1\int_0^1k_t(x,y)\,d\xi(x)\,d\eta(y)\\&\quad-2\int_0^1\int_0^1k_t(x,y)\,d\eta(x)\,d\eta(y)\end{aligned}$$

for all probability measures ξ. Using Dirac measures for ξ yields that this inequality is equivalent to (3.20), which proves Theorem 3.8. $\square$

PROOF OF THEOREM 4.1. Since the integration mapping is continuous, it suffices to establish the weak convergence $W_n(t) \overset{w}{\rightsquigarrow} W(t)$ in $l^\infty[0,1]$ where we define

$$W_n(t) = \int_0^1 n\left(\log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right)^2 \bar{w}(y)\,dy - nB_h\big(\hat{A}_{n,h}(t) - A^*(t)\big)^2,$$

$$W(t) = \int_0^1 \left(\frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right)^2 \bar{w}(y)\,dy - B_h \mathbb{A}_{C,h}^2(t).$$

We prove this assertion along similar lines as in the proof of Theorem 3.1. For $i \geq 2$, we recall the notation $\bar{w}(y) = h^*(y)/(\log y)^2$ and consider the following random functions in $l^\infty[0,1]$:

$$W_{i,n}(t) = \int_{1/i}^1 n\left(\log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right)^2 \bar{w}(y)\,dy$$
$$- B_h^{-1}\left(\int_{1/i}^1 \sqrt{n}\left(\log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right) \frac{h^*(y)}{\log y}\,dy\right)^2,$$

$$W_i(t) = \int_{1/i}^1 \left(\frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right)^2 \bar{w}(y)\,dy$$
$$- B_h^{-1}\left(\int_{1/i}^1 \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} \frac{h^*(y)}{\log y}\,dy\right)^2.$$

By an application of Lemma B.1 in Appendix B, it suffices to show the conditions listed in (A.1). By arguments similar to those in the proof of Theorem 3.1, we obtain

$$\sqrt{n} \log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} \overset{w}{\rightsquigarrow} \frac{\mathbb{G}_C(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} \tag{A.21}$$

in $l^\infty([1/i, 1] \times [0,1])$. Assertion (i) now follows immediately by the boundedness of the functions $\bar{w}(y)$ and $h^*(y)(-\log y)^{-1}$ on $[1/i, 1]$ [see conditions (3.9), (3.10) and (3.13)] and the continuous mapping theorem.

For the proof of assertion (ii), we simply note that $\mathbb{G}_C^2$ and $\mathbb{G}_C$ are bounded on $[0,1]^2$ and $K_1(y,t) = \frac{\bar{w}(y)}{C^2(y^{1-t}, y^t)}$ and $K_2(y,t) = \frac{h^*(y)}{C(y^{1-t}, y^t)}$ are bounded uniformly with respect to $t \in [0,1]$ by the integrable functions $\bar{K}_1(y) = \bar{w}(y) y^{-2}$ and $\bar{K}_2(y) = h^*(y)(-\log y)^{-1} y^{-1}$.

For the proof of assertion (iii), we fix some $\alpha \in (0, 1/2)$ such that $\lambda\alpha > 2\gamma$ and consider the decomposition

$$W_n(t) - W_{i,n}(t) = B_i^{(1)}(t) + B_i^{(2)}(t) + B_i^{(3)}(t), \tag{A.22}$$

where

$$B_i^{(1)}(t) = \int_{I_{B_i^{(1)}(t)}} n\left(\log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right)^2 \bar{w}(y)\,dy, \tag{A.23}$$

$$B_i^{(2)}(t) = \int_{I_{B_i^{(2)}(t)}} n\left(\log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right)^2 \bar{w}(y)\,dy, \tag{A.24}$$

$$B_i^{(3)}(t) = -B_h^{-1} I(t, 1/i)\big(2I(t,1) - I(t, 1/i)\big), \tag{A.25}$$

$I_{B_i^{(1)}}(t)$ and $I_{B_i^{(2)}}(t)$ are defined in (A.4) and

$$I(t,a) = \sqrt{n}\int_0^a \left(\log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right)\frac{h^*(y)}{\log y}\,dy.$$

By the same arguments as in the proof of Theorem 3.1, we have for every $\varepsilon > 0$

$$\lim_{i\to\infty}\limsup_{n\to\infty}\mathbb{P}^*\Big(\sup_{t\in[0,1]}|I(t,1/i)| > \varepsilon\Big) = 0,$$

and $\sup_{t\in[0,1]}|I(t,1)| = O_{\mathbb{P}^*}(1)$, which yields the asymptotic negligibility $\lim_{i\to\infty}\limsup_{n\to\infty}\mathbb{P}^*(\sup_{t\in[0,1]}|B_i^{(3)}(t)| > \varepsilon) = 0$. For $B_i^{(1)}(t)$, we obtain the estimate

$$\begin{aligned}
&\sup_{t\in[0,1]}\big|B_i^{(1)}(t)\big| \\
&\quad\le \sup_{t\in[0,1]}\int_{I_{B_i^{(1)}(t)}} n|(\tilde{C}_n - C)(y^{1-t}, y^t)|^2\left|1 \vee \frac{C^2}{\tilde{C}_n^2}(y^{1-t}, y^t)\right|\bar{w}(y)y^{-2}\,dy \\
&\quad\le \sup_{\mathbf{x}\in[0,1]^2} n|\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})|^2 \times \left(1 \vee \sup_{\mathbf{x}\in[0,1]^2:\,C(\mathbf{x})>n^{-\alpha}}\left|\frac{C^2}{\tilde{C}_n^2}(\mathbf{x})\right|\right) \times \psi(i),
\end{aligned}$$

where $\psi(i) := \int_0^{1/i}\bar{w}(y)y^{-2}\,dy$, which can be handled by the same arguments as in the proof of Theorem 3.1. Finally, the term $B_i^{(2)}(t)$ can be estimated by

$$\begin{aligned}
&\sup_{t\in[0,1]}\big|B_i^{(2)}(t)\big| \\
&\quad\le \sup_{t\in[0,1]}\int_{I_{B_i^{(2)}(t)}} n|(\tilde{C}_n - C)(y^{1-t}, y^t)|^2\left|1 \vee \frac{C^\lambda}{\tilde{C}_n^2}(y^{1-t}, y^t)\right|\bar{w}(y)y^{-\lambda}\,dy \\
&\quad\le \sup_{\mathbf{x}\in[0,1]^2} n|\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})|^2 \times \left(1 \vee \sup_{\mathbf{x}\in[0,1]^2:\,C(\mathbf{x})\le n^{-\alpha}}\left|\frac{C^\lambda}{\tilde{C}_n^2}(\mathbf{x})\right|\right) \times \phi(i),
\end{aligned}$$

where $\phi(i) = \int_0^{1/i}\bar{w}(y)y^{-\lambda}\,dy = o(1)$ for $i\to\infty$ by condition (3.11). Mimicking the arguments from the proof of Theorem 3.1 completes the proof. $\square$

PROOF OF THEOREM 4.2. Recall the decomposition $M_h(\tilde{C}_n, \hat{A}_{n,h}) - M_h(C, A^*) = S_1 + S_2 + S_3$ where S_1, S_2 and S_3 are defined in (4.1). With the notation $\bar{v}(y) := 2h^*(y)/(-\log y)$ it follows that $|v(y,t)| \le \bar{v}(y)$ and the assumptions on h yield the validity of (3.9)–(3.11) for $v(y,t)$. This allows for an application of Theorem 3.1 and together with the continuous mapping theorem we obtain $\sqrt{n} S_1 \stackrel{w}{\rightsquigarrow} Z_1$, where Z_1 is the limiting process defined in (4.2). Thus, it remains to verify the negligibility of $S_2 + S_3$. For S_3, we note that by Theorem 3.2 and the continuous mapping theorem we have $S_3 = O_{\mathbb{P}^*}(1/n)$ and it remains to consider S_2. To this end, we fix some $\alpha \in (0, 1/2)$ such that $(1 + (\lambda - 1)/2)\alpha > \gamma$ and consider the decomposition

$$\begin{aligned}
\int_0^1 \log^2 \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} \frac{h^*(y)}{(\log y)^2}\, dy
&= \int_{I_{B_1^{(1)}(t)}} \log^2 \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} \frac{h^*(y)}{(\log y)^2}\, dy \\
&\quad + \int_{I_{B_1^{(2)}(t)}} \log^2 \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} \frac{h^*(y)}{(\log y)^2}\, dy \\
&=: T_1(t,n) + T_2(t,n),
\end{aligned}$$

where the sets $I_{B_1^{(j)}(t)}$, $j = 1, 2$ are defined in (A.4). On the set $I_{B_1^{(1)}(t)}$, we use the estimate

$$\begin{aligned}
\log^2 \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)} &\le \frac{|\tilde{C}_n - C|^2}{(C^*)^2}(y^{1-t}, y^t) \le n^\alpha \frac{|\tilde{C}_n - C|^2}{C^*} \frac{1}{1 \wedge \tilde{C}_n/C}(y^{1-t}, y^t) \\
&\le n^\alpha \frac{|\tilde{C}_n - C|^2}{C^*}(y^{1-t}, y^t)\left(1 \vee \sup_{\mathbf{x} \in [0,1]^2 : C(\mathbf{x}) > n^{-\alpha}} \frac{C(\mathbf{x})}{\tilde{C}_n(\mathbf{x})}\right),
\end{aligned}$$

where $|C^*(y,t) - C(y^{1-t}, y^t)| \le |\tilde{C}_n(y^{1-t}, y^t) - C(y^{1-t}, y^t)|$. By arguments similar to those used in the proof of Theorem 3.1, it is now easy to see that

$$\begin{aligned}
\sqrt{n} \sup_t |T_1(t,n)| &\le \sup_{\mathbf{x} \in [0,1]^2} n^{\alpha + 1/2} |\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})|^2 \\
&\qquad \times \left(1 \vee \sup_{\mathbf{x} \in [0,1]^2 : C(\mathbf{x}) > n^{-\alpha}} \left| \frac{C}{\tilde{C}_n}(\mathbf{x}) \right| \right)^2 \times K \\
&= o_{\mathbb{P}^*}(1),
\end{aligned}$$

where $K := \int_0^1 \bar{w}(y) y^{-1}\, dy < \infty$ denotes a finite constant [see condition (3.11)]. Now set $\beta := (\lambda - 1)/2 > 0$. From the estimate

$$C^*(y,t) \ge y^{1+\beta}\left(1 \wedge \frac{\tilde{C}_n}{C^{1+\beta}}(y^{1-t}, y^t)\right) = y^{-\beta} y^\lambda \left(1 \wedge \frac{\tilde{C}_n}{C^{1+\beta}}(y^{1-t}, y^t)\right)$$

we obtain by similar arguments as in the proof of the negligibility of $|B_i^{(2)}(t)|$ in the proof of Theorem 3.1 (note that on $I_{B_1^{(2)}(t)}$ we have $y \le C(y^{1-t}, y^t) \le n^{-\alpha}$)

$$\sup_{t\in[0,1]} |T_2(t,n)| \le \log(n) n^{-\beta\alpha} \sup_{\mathbf{x}\in[0,1]^2} \sqrt{n}|\tilde{C}_n(\mathbf{x}) - C(\mathbf{x})|$$

$$\times \left(1 \vee \sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})\le n^{-\alpha}} \left|\frac{C^{1+\beta}}{\tilde{C}_n}(\mathbf{x})\right|\right) \times \tilde{K},$$

where $\tilde{K} := \gamma \int_0^1 (1 - \log y) \frac{h^*(y)}{(\log y)^2} y^{-\lambda}\, dy$ denotes a finite constant [see conditions (3.11) and (3.14)] and we used the estimate

$$\left|\log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right|^2 \le (\gamma \log n - \log y)\left|\log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right|$$

$$\le \gamma \log(n)(1 - \log y)\left|\log \frac{\tilde{C}_n(y^{1-t}, y^t)}{C(y^{1-t}, y^t)}\right|,$$

which holds for sufficiently large n. Finally, we observe that

$$\sup_{\mathbf{x}\in[0,1]^2 : C(\mathbf{x})\le n^{-\alpha}} \left|\frac{C^{1+\beta}}{\tilde{C}_n}(\mathbf{x})\right| \le \sup_{\mathbf{x} : C(\mathbf{x})\le n^{-\alpha}} |n^{\gamma} C^{1+\beta}(\mathbf{x})| \le n^{\gamma-(1+\beta)\alpha} = o(1).$$

Now the proof is complete. $\square$

PROOF OF THEOREM 4.4. The conditions on the weight function imply that all integrals in the definition of Z_0 are proper and therefore the mapping $(\mathbb{G}_C, C) \mapsto Z_0(\mathbb{G}_C, C)$ is continuous. Hence, the result follows by the continuous mapping theorem for the bootstrap [see, e.g., Theorem 10.8 in Kosorok (2008)] provided the conditional weak convergence in (4.4) holds under the nonrestrictive smoothness assumption (3.6). To see this, proceed similar as in Bücher and Dette (2010) and show Hadamard-differentiability of the mapping $H \mapsto H(H_1^-, H_2^-)$, which is defined for some distribution function H on the unit square whose marginals $H_1 = H(\cdot, 1)$ and $H_2 = H(1, \cdot)$ satisfy $H_1(0) = H_2(0) = 0$. This can be done by similar arguments as in Segers (2010) and the details are omitted for the sake of brevity. $\square$

APPENDIX B: AN AUXILIARY RESULT

LEMMA B.1. *Let $X_n, X_{i,n} : \Omega \to \mathbb{D}$ for $i, n \in \mathbb{N}$ be arbitrary maps with values in the metric space $(\mathbb{D}, d)$ and $X_i, X : \Omega \to \mathbb{D}$ be Borel-measurable. Suppose that*:

(i) *For every $i \in \mathbb{N}$* $\quad X_{i,n} \overset{w}{\rightsquigarrow} X_i \quad$ *for $n \to \infty$*,

(ii) $X_i \overset{w}{\rightsquigarrow} X \quad$ *for $i \to \infty$*,

(iii) *For every $\varepsilon > 0$* $\quad \lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*\big(d(X_{i,n}, X_n) > \varepsilon\big) = 0.$

Then $X_n \overset{w}{\rightsquigarrow} X$ *for* $n \to \infty$.

PROOF. Let $F \subset \mathbb{D}$ be closed and fix $\varepsilon > 0$. If $F^{\varepsilon} = \{x \in \mathbb{D} : d(x, F) \leq \varepsilon)$ denotes the ε-enlargement of F we obtain

$$\mathbb{P}^*(X_n \in F) \leq \mathbb{P}^*(X_{i,n} \in F^{\varepsilon}) + \mathbb{P}^*\bigl(d(X_{i,n}, X_n) > \varepsilon\bigr).$$

By hypothesis (i) and the Portmanteau theorem [see Van der Vaart and Wellner (1996)]

$$\limsup_{n\to\infty} \mathbb{P}^*(X_n \in F) \leq \mathbb{P}(X_i \in F^{\varepsilon}) + \limsup_{n\to\infty} \mathbb{P}^*\bigl(d(X_{i,n}, X_n) > \varepsilon\bigr).$$

By conditions (ii) and (iii) $\limsup_{n\to\infty} \mathbb{P}^*(X_n \in F) \leq P(X \in F^{\varepsilon})$ and since $F^{\varepsilon} \downarrow F$ for $\varepsilon \downarrow 0$ and closed F the result follows by the Portmanteau theorem. $\square$

Acknowledgments. The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. The authors would also like to thank Christian Genest for pointing out important references and Johan Segers for many fruitful discussions on the subject. We are also grateful to two unknown referees and an Associate Editor for their constructive comments on an earlier version of this manuscript, which led to a substantial improvement of the paper.

REFERENCES

BEN GHORBAL, N., GENEST, C. and NEŠLEHOVÁ, J. (2009). On the Ghoudi, Khoudraji, and Rivest test for extreme-value dependence. *Canad. J. Statist.* **37** 534–552. MR2588948

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

BÜCHER, A. and DETTE, H. (2010). A note on bootstrap approximations for the empirical copula process. *Statist. Probab. Lett.* **80** 1925–1932.

BÜCHER, A., DETTE, H. and VOLGUSHEV, S. (2010). New estimators of the Pickands dependence function and a test for extreme-value dependence. Available at http://www.ruhr-uni-bochum.de/mathematik3/research/index.html.

CAPÉRAÀ, P., FOUGÈRES, A. L. and GENEST, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84** 567–577. MR1603985

CEBRIAN, A., DENUIT, M. and LAMBERT, P. (2003). Analysis of bivariate tail dependence using extreme values copulas: An application to the SOA medical large claims database. *Belgian Actuarial Journal* **3** 33–41.

COLES, S., HEFFERNAN, J. and TAWN, J. (1999). Dependence measures for extreme value analyses. *Extremes* **2** 339–365.

DEHEUVELS, P. (1984). Probabilistic aspects of multivariate extremes. In *Statistical Extremes and Applications* (J. T. de Oliveira, ed.) 117–130. Reidel, Dordrecht. MR0784817

DEHEUVELS, P. (1991). On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statist. Probab. Lett.* **12** 429–439. MR1142097

FERMANIAN, J.-D., RADULOVIĆ, D. and WEGKAMP, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10** 847–860. MR2093613

FILS-VILLETARD, A., GUILLOU, A. and SEGERS, J. (2008). Projection estimators of Pickands dependence functions. *Canad. J. Statist.* **36** 369–382. MR2456011

GENEST, C. and SEGERS, J. (2009). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist.* **37** 2990–3022. MR2541453

GHOUDI, K., KHOUDRAJI, A. and RIVEST, L. (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Canad. J. Statist.* **26** 187–197. MR1624413

GUMBEL, É. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris* **9** 171–173. MR0119279

HALL, P. and TAJVIDI, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli* **6** 835–844. MR1791904

HSING, T. (1989). Extreme value theory for multivariate stationary sequences. *J. Multivariate Anal.* **29** 274–291. MR1004339

HÜSLER, J. and REISS, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statist. Probab. Lett.* **7** 283–286. MR0980699

JIMÉNEZ, J. R., VILLA-DIHARCE, E. and FLORES, M. (2001). Nonparametric estimation of the dependence function in bivariate extreme value distributions. *J. Multivariate Anal.* **76** 159–191. MR1821817

JOE, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist. Probab. Lett.* **9** 75–81. MR1035994

KARATZOGLOU, A., SMOLA, A., HORNIK, K. and ZEILEIS, A. (2004). Kernlab—an S4 package for Kernel methods in R. *Journal of Statistical Software* **11** 1–20.

KOJADINOVIC, I. and YAN, J. (2010). Nonparametric rank-based tests of bivariate extreme-value dependence. *J. Multivariate Anal.* **101** 2234–2249. MR2671214

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York. MR2724368

MARSHALL, A. W. (1970). Discussion of Barlow and van Zwet's papers. In *Nonparametric Techniques in Statistical Inference* (M. L. Pudi, ed.) 175–176. Cambridge Univ. Press, London. MR0273755

NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. Springer, New York. MR2197664

NOCEDAL, J. and WRIGHT, S. J. (2006). *Numerical Optimization*, 2nd ed. Springer, New York. MR2244940

PICKANDS, J. (1981). Multivariate extreme value distributions (with a discussion). In *Proceedings of the* 43*rd Session of the International Statistical Institute. Bull. Inst. Internat. Statist.* **49** 859–878, 894–902. MR0820979

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1996). *Ordered Restricted Statistical Inference*. Wiley, New York.

RÜSCHENDORF, L. (1976). Asymptotic distributions of multivariate rank order statistics. *Ann. Statist.* **4** 912–923. MR0420794

SCAILLET, O. (2005). A Kolmogorov–Smirnov type test for positive quadrant dependence. *Canad. J. Statist.* **33** 415–427. MR2193983

SEGERS, J. (2007). Nonparametric inference for bivariate extreme-value copulas. In *Topics in Extreme Values* (M. Ahsanullah and S. N. U. A. Kirmani, eds.) 185–207. Nova Science Publishers, New York.

SEGERS, J. (2010). Weak convergence of empirical copula processes under nonrestrictive smoothness assumptions. Available at ArXiv:1012.2133v1.

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

STRIMMER, K. (2009). fdrtool: Estimation and control of (local) false discovery rates. R package Version 1.2.6.

TAWN, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika* **75** 397–415. MR0967580

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

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

WANG, J. L. (1986). Asymptotically minimax estimators for distributions with increasing failure rate. *Ann*. *Statist*. **14** 1113–1131. MR0856809

ZHANG, D., WELLS, M. T. and PENG, L. (2008). Nonparametric estimation of the dependence function for a multivariate extreme value distribution. *J*. *Multivariate Anal*. **99** 577–588. MR2406072

FAKULTÄT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
UNIVERSITÄTSSTRASSE 150
44780 BOCHUM
GERMANY
E-MAIL: axel.buecher@ruhr-uni-bochum.de
holger.dette@ruhr-uni-bochum.de
stanislav.volgushev@ruhr-uni-bochum.de