CentER

Discussion Paper

No. 2004–91

NON-PARAMETRIC INFERENCE FOR BIVARIATE EXTREME-VALUE COPULAS

By J.J.J. Segers

September 2004

ISSN 0924-7815

TILBURG UNIVERSITY

Non-Parametric Inference for Bivariate Extreme-Value Copulas

JOHAN SEGERS*

Tilburg University & Catholic University Leuven

Abstract

Extreme-value copulas arise as the possible limits of copulas of component-wise maxima of independent, identically distributed samples. The use of bivariate extreme-value copulas is greatly facilitated by their representation in terms of Pickands dependence functions. The two main families of estimators of this dependence function are (variants of) the Pickands estimator and the Capéraà-Fougères-Genest estimator. In this paper, a unified treatment is given of these two families of estimators, and within these classes those estimators with the minimal asymptotic variance are determined. Main result is the explicit construction of an adaptive, minimum-variance estimator within a class of estimators that encompasses the Capéraà-Fougères-Genest estimator.

Key words: adaptive estimator; copula; extremes; Pickands dependence function
JEL: C13, C14
MSC 2000: 60G70, 62G32, 62H12

1 Introduction

The most general margin-free description of the dependence structure of a multivariate distribution is through its *copula* (Sklar 1959). Copulas have recently come to the attention in various sciences as a way to overcome the limitations of classical dependence measures as exemplified by the linear correlation, see for instance the monograph by Nelsen (1999).

A particular class of copulas are the *extreme-value copulas.* They arise as the possible limits of copulas of component-wise maxima of independent, identically distributed samples. For a bivariate sample (X_{i1}, X_{i2}), $i = 1, \ldots, n$, the vector of component-wise maxima is (M_{n1}, M_{n2}), with $M_{ij} = \bigvee_{i=1}^{n} X_{ij}$. If the pairs

*Tilburg University, Department of Econometrics and OR, P.O. Box 90153, NL-5000 LE Tilburg, the Netherlands (e-mail: jsegers@uvt.nl). The author is post-doctoral research fellow of the Fund for Scientific Research, Flanders. *September 2004*

(X_{i1}, X_{i2}) are independent and if they have a common bivariate distribution function F with continuous margins and copula C_F, then the joint distribution function of (M_{n1}, M_{n2}) is F^n with copula $C_{F^n}(u_1, u_2) = C_F^n(u_1^{1/n}, u_2^{1/n})$ for $(u_1, u_2) \in [0,1]^2$. If there exists a copula C such that $C_{F^n}(u_1, u_2) \to C(u_1, u_2)$ as $n \to \infty$, then C is by definition a bivariate extreme-value copula. The study of the limiting dependence structure of (M_{n1}, M_{n2}) dates back to Geffroy (1958/59), Gumbel (1960a), and Sibuya (1960), although not through the language of copulas.

In the above setting, if for instance the pair (X_{i1}, X_{i2}) records the heights of a certain river at day i at two different locations and if $n = 365$, then the pair (M_{n1}, M_{n2}) represents the maximal heights recorded during a given year. If (x_1, x_2) are the heights of the dikes at the two locations, then $F^n(x_1, x_2)$ is the probability that there will not be flood during a certain year. The copula C_{F^n} of F^n describes the dependence between the occurrences of floods at the two locations. As n is large, asymptotic theory suggests to model C_{F^n} by an extreme-value copula. Examples of this kind are already considered in Gumbel and Goldstein (1964). A more recent example in finance is Longin and Solnik (2001).

The use of bivariate extreme-value copulas is greatly facilitated by a representation discovered by Pickands (1981) and based on Balkema and Resnick (1977) and de Haan and Resnick (1977): a copula C is an extreme-value copula if and only if there exists a real-valued function A on the interval $[0,1]$ such that

$$C(u,v) = \exp\left\{\log(uv) A\left(\frac{\log(v)}{\log(uv)}\right)\right\},$$

for $0 < u < 1$ and $0 < v < 1$. The function A is called a *Pickands dependence function*. Necessary and sufficient requirements on a function A to make the function C in the preceding display a copula are the following: (i) $t \vee (1-t) \leq A(t) \leq 1$ for all $0 \leq t \leq 1$, and (ii) A is convex. The upper bound $A \equiv 1$ corresponds to independence, $C(u,v) = uv$, and the lower bound $A(t) = t \vee (1-t)$ to the comonotone copula, $C(u,v) = u \wedge v$. Dependence functions in higher dimensions are investigated in Obrenetov (1991).

Statistical inference on a bivariate extreme-value copula C may now be reduced to inference on its Pickands dependence function A. Since the requirements on A do not confine it to a parametric model, there are two approaches to inference. First, one may postulate a parametric model for A and estimate the unknown parameters by for instance the (pseudo) maximum-likelihood estimator in Genest, Ghoudi, and Rivest (1995). Second, one may construct non-parametric estimators of A in the full model.

A simple and popular parametric model is the Gumbel or logistic model (Gumbel 1960b), $A(t) = \{t^{1/\alpha} + (1-t)^{1/\alpha}\}^\alpha$, with parameter $0 \leq \alpha \leq 1$. The corresponding copula happens to be the only extreme-value copula that is also a Archimedean copula (Genest and Rivest 1989). More flexible models are the asymmetric logistic model and the mixed model (Tawn 1988). Overviews of the most common parametric models can be found in, for instance, Kotz and Nadarajah (2000) and Beirlant, Goegebeur, Segers, and Teugels (2004).

The focus of this paper, however, is on non-parametric inference on the Pickands dependence function A. There are two main families of estimators: first, the Pickands (1981) estimator and variants, and second, the estimator of Capéraà, Fougères, and Genest (1997) and variants. Next to these, there is also the non-parametric estimator by Tiago de Oliveira (1989), the convergence rate of which, however, is too slow to make it a serious candidate in practice, see Deheuvels and Tiago de Oliveira (1989).

The aims of this paper, then are the following: first, to give a unified treatment of the Pickands and Capéraà-Fougères-Genest estimators and variants; second, to find within these classes those estimators with the minimal asymptotic variances. Main result is the explicit construction of a minimum-variance estimator within a class of estimators that encompasses the Capéraà-Fougères-Genest estimators.

Two important issues remain open for further research. First, what are the asymptotic distributions of the estimators in case the marginal distributions are unknown? Second, what is the semi-parametric efficient lower bound for estimating the Pickands dependence function in the sense of Bickel, Klaassen, Ritov, and Wellner (1993)?

The structure of the paper is as follows. In Section 2, a tool is proposed to compute expected values with respect to extreme-value copulas, an interesting example being an explicit expression of the joint moment generating function. The estimators of Pickands and Capéraà-Fougères-Genest and variants are treated in a unified manner in Sections 3 and 4. Section 5 contains the construction of minimum-variance Pickands and Capéraà-Fougères-Genest estimators. All proofs are deferred to Section 6.

2 Bivariate extreme-value copulas

2.1 Conditional distributions

Let (U, V) be a random pair with standard uniform margins and joint distribution function equal to the bivariate extreme-value copula C with Pickands dependence function A. It is convenient to switch to standard exponential margins through $X = -\log U$ and $Y = -\log V$. The joint survivor function of (X, Y) is then given by

$$S(x, y) := \Pr[X > x, Y > y] = \exp\left\{-(x + y)A\left(\frac{y}{x + y}\right)\right\} \tag{1}$$

for $0 \leq x < \infty$ and $0 \leq y < \infty$ with $x + y > 0$.

For computing expectations of the form $\mathrm{E}[f(X, Y)]$, expression (1) is not directly useful. Better would be to have an expression for the conditional distribution of, say, Y given X. Since X is known to be a standard exponential random variable, the expectation $\mathrm{E}[f(X, Y)]$ could then be calculated as a double integral. However, as the pair (X, Y) in general does not possess a joint density with respect to two-dimensional Lebesgue measure, the conditional distribution of Y given X cannot be computed as a ratio of joint and marginal densities.

Let μ be the probability distribution of (X, Y). By Theorem 33.3 of Billingsley (1995), there exists a collection of probability measures $\{\mu_x : x \in (0, \infty)\}$ on $(0, \infty)$ such that μ can be desintegrated as $\mu(\mathrm{d}y\mathrm{d}x) = \mu_x(\mathrm{d}y)\mathrm{e}^{-x}\mathrm{d}x$. Formally, for a μ-integrable function f,

$$\int_{(0,\infty)^2} f(x,y)\mu(\mathrm{d}x\mathrm{d}y) = \int_0^\infty \left(\int_{(0,\infty)} f(x,y)\mu_x(\mathrm{d}y) \right) \mathrm{e}^{-x}\mathrm{d}x.$$

The measure μ_x is called the conditional distribution of Y given $X = x$. The expression in the above display states that $\mathrm{E}[f(X,Y)|X] = g(X)$ with $g(x) = \int_0^\infty f(x,y)\mu_x(\mathrm{d}y)$. This justifies the suggestive notation

$$\mathrm{E}[f(X,Y)|X = x] := \int_0^\infty f(x,y)\mu_x(\mathrm{d}y).$$

For indicator variables $f = \mathbf{1}_V$, we write

$$\Pr[(X,Y) \in V \mid X = x] := \int_0^\infty \mathbf{1}\{(x,y) \in V\}\mu_x(\mathrm{d}y).$$

The conditional distribution of Y given X can be expressed in terms of A and its right-hand derivative A'. Observe that A' always exists and is non-decreasing because A is convex.

Lemma 2.1 *For (X, Y) as in (1),*

$$\Pr[Y > y \mid X = x] = \mathrm{e}^x S(x,y) \left\{ A\left(\frac{y}{x+y}\right) - \frac{y}{x+y} A'\left(\frac{y}{x+y}\right) \right\}$$

for $0 < x < \infty$ and $0 \le y < \infty$.

Comments. From Lemma 2.1, it can be seen that the distribution of (X, Y) is, in the terminology of Lehmann (1966), monotone regression dependent, or, in modern terminology, stochastically increasing, that is, $\Pr[Y > y \mid X = x]$ is non-decreasing in x and $\Pr[X > x \mid Y = y]$ is non-decreasing in y (Garralda Guillem 2000). In particular, the distribution of (X, Y) is positively associated (Marshall and Olkin 1983). For stochastically increasing distributions, a conjecture by Hutchinson and Lai (1990) states that $-1 + \sqrt{1 + 3\tau} \le \rho_S \le \min\{(3/2)\tau, 2\tau - \tau^2\}$, where τ and ρ_S denote Kendall's tau and Spearman's rho, respectively. Using explicit expressions for τ and ρ_S in terms of A computed through Lemma 2.1, Hürlimann (2003) shows that the conjecture holds for the class of bivariate extreme-value copulas.

2.2 Joint moment generating function

Lemma 2.1 gives a handle to compute $\mathrm{E}[f(X,Y)]$ for various functions f. By way of illustration, here is an elegant expression for the joint moment generating function of (X,Y).

Lemma 2.2 *For (X,Y) as in (1),*

$$\mathrm{E}[\mathrm{e}^{aX+bY}] = \frac{1}{1-a} + \frac{1}{1-b} - 1 + ab \int_0^1 \frac{\mathrm{d}t}{\{A(t) - a(1-t) - bt\}^2}$$

with $(a,b) \in \mathbb{R}^2$ such that $\inf_{0 \le t \le 1}\{A(t) - a(1-t) - bt\} > 0$.

Since $A(t) \ge 1/2$ for all $0 \le t \le 1$, the range for (a,b) in Lemma 2.2 certainly includes $(-\infty, 1/2)^2$. Hence, joint moments $\mathrm{E}[X^r Y^s]$ for positive integer r and s can be found by differentiating the moment generating function at $(a,b) = (0,0)$. For instance,

$$\mathrm{E}[XY] = \frac{\partial^2}{\partial a\, \partial b} \mathrm{E}[\mathrm{e}^{aX+bY}] \bigg|_{a=b=0} = \int_0^1 \frac{\mathrm{d}t}{A^2(t)}$$

so that

$$\mathrm{Cov}(X,Y) = \int_0^1 \frac{\mathrm{d}t}{A^2(t)} - 1, \tag{2}$$

a result already stated in Tawn (1988).

We will have more opportunities to use Lemma 2.1 when computing the covariance function of certain Gaussian processes arising as the limit distribution of estimator processes for A.

3 Pickands estimator and variants

3.1 Estimators

Let the random pair (X,Y) be distributed as in (1). For $0 \le t \le 1$, define

$$\xi(t) = \frac{X}{1-t} \wedge \frac{Y}{t}, \tag{3}$$

with of course $\xi(0) = X$ and $\xi(1) = Y$. The random variable $\xi(t)$ is non-negative and its survivor function is

$$\Pr[\xi(t) > z] = \Pr[X > (1-t)z, Y > tz] = \exp\{-zA(t)\}, \tag{4}$$

for $0 \le z < \infty$. Hence $\xi(t)$ is exponentially distributed with expectation

$$\mathrm{E}[\xi(t)] = 1/A(t), \qquad 0 \le t \le 1. \tag{5}$$

Let (X_i, Y_i), $i = 1, \ldots, n$, be independent random pairs with the same distribution as (X, Y), and put

$$\xi_i(t) = \frac{X_i}{1-t} \wedge \frac{Y_i}{t}, \qquad 0 \le t \le 1.$$

Equation (5) motivates the Pickands (1981, 1989) estimator

$$1/\hat{A}_n^{\mathrm{p}}(t) = \bar{\xi}_n(t) = \frac{1}{n} \sum_{i=1}^n \xi_i(t), \qquad 0 \le t \le 1, \tag{6}$$

see also Smith, Tawn and Yuen (1990).

Unfortunately, the Pickands estimator $\hat{A}_n^{\mathrm{p}}$ is itself not a Pickands dependence function. For instance, the requirements $A(0) = 1$ and $A(1) = 1$ are not fulfilled. This was the reason for the following modification proposed by Deheuvels (1991): for $0 \le t \le 1$, set

$$1/\hat{A}_n^{\mathrm{d}}(t) = \bar{\xi}_n(t) - (1-t)\left(\bar{X}_n - 1\right) - t\left(\bar{Y}_n - 1\right) \tag{7}$$

where $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ and $\bar{Y}_n = n^{-1} \sum_{i=1}^n Y_i$. Indeed $\hat{A}_n^{\mathrm{d}}(0) = 1$ and $\hat{A}_n^{\mathrm{d}}(1) = 1$.

3.2 Asymptotics

It is convenient to formulate the estimators by Pickands and Deheuvels in an abstract way. Let $C[0,1]$ be the space of real-valued, continuous functions on the interval $[0,1]$, equipped with the topology of uniform convergence. For arbitrary $a, b \in C[0,1]$, define the operator $L_{a,b} : C[0,1] \to C[0,1]$ by

$$L_{a,b} f(t) = f(t) - a(t) f(0) - b(t) f(1) \tag{8}$$

for $0 \le t \le 1$. Define the *Pickands-type* estimator

$$\begin{aligned} 1/\hat{A}_n^{\mathrm{p}}(t; a, b) &= L_{a,b} \bar{\xi}_n(t) + a(t) + b(t) \\ &= \bar{\xi}_n(t) - a(t)\left(\bar{X}_n - 1\right) - b(t)\left(\bar{Y}_n - 1\right) \end{aligned} \tag{9}$$

for $0 \le t \le 1$. The Pickands estimator $\hat{A}_n^{\mathrm{p}}$ in (6) corresponds to the choice $a \equiv 0$ and $b \equiv 0$, while the Deheuvels estimator $\hat{A}_n^{\mathrm{d}}$ in (7) corresponds to $a(t) = 1 - t$ and $b(t) = t$ for $0 \le t \le 1$. Denote convergence in distribution by the arrow $\rightsquigarrow$.

Theorem 3.1 *Let (X_i, Y_i), $i = 1, \ldots, n$, be independent random pairs with distribution (1). Then*

$$n^{1/2}\left(1/\hat{A}_n^{\mathrm{p}} - 1/A\right) \rightsquigarrow \eta \tag{10}$$

in $C[0,1]$, where η is a centered Gaussian process in $C[0,1]$ with covariance function

$$\begin{aligned} &\operatorname{Cov}(\eta(s), \eta(t)) = \operatorname{Cov}(\xi(s), \xi(t)) \\ &= \frac{s}{t} \frac{1}{A^2(s)} + \frac{1-t}{1-s} \frac{1}{A^2(t)} + \frac{1}{(1-s)t} \int_s^t \frac{\mathrm{d}w}{A^2(w)} - \frac{1}{A(s)A(t)} \end{aligned} \tag{11}$$

for $0 \le s \le t \le 1$, with ξ as in (3).

Since the operator $L : C[0,1] \to C[0,1]$ is linear and bounded with respect to the topology of uniform convergence, the asymptotics for the Pickands-type estimator $\hat{A}_n^{\mathrm{p}}(\cdot; a, b)$ follow easily from those for $\bar{\xi}_n = 1/\hat{A}_n^{\mathrm{p}}$.

Corollary 3.2 *Under the assumptions and notations of Theorem 3.1, for $a, b \in C[0,1]$,*

$$n^{1/2}\left(\hat{A}_n^{\mathrm{p}}(\cdot; a, b) - A\right) \rightsquigarrow A^2 L_{a,b}\eta$$

in $C[0,1]$, with $L_{a,b}$ and $\hat{A}_n^{\mathrm{p}}(\cdot; a, b)$ as in (8) and (9), respectively.

Corollary 3.3 *Under the assumptions and notations of Theorem 3.1,*

$$\begin{aligned} n^{1/2}\left(\hat{A}_n^{\mathrm{p}} - A\right) &\rightsquigarrow A^2\eta, \\ n^{1/2}\left(\hat{A}_n^{\mathrm{d}} - A\right) &\rightsquigarrow A^2\{\eta - (1-I)\eta(0) - I\eta(1)\}, \end{aligned}$$

with $\hat{A}_n^{\mathrm{p}}$ and $\hat{A}_n^{\mathrm{d}}$ as in (6) and (7), respectively, and with $I(t) = t$ for $t \in [0,1]$.

Theorem 3.1 and Corollary 3.3 are essentially due to Deheuvels (1991), the novelty here being the explicit expression (11) for the covariance function.

Comments. There are of course other ways to modify the Pickands estimator so as to make it satisfy the boundary constraints at zero and one.

(i) Take for instance $L_{1-I,I}(\hat{A}_n^{\mathrm{p}}) + 1$. As $L_{1-I,I}(A) = A - 1$, Theorem 3.1 yields

$$n^{1/2}[\{L_{1-I,I}(\hat{A}_n^{\mathrm{p}}) + 1\} - A] \rightsquigarrow L_{1-I,I}(A^2\eta)$$

in $C[0,1]$. Observe the difference with the limit distribution $A^2 L_{1-I,I}\eta$ of $\hat{A}_n^{\mathrm{d}}$.

(ii) Hall and Tajvidi (2000) proposed

$$1/\hat{A}_n^{\mathrm{ht}}(t) = \frac{1}{n}\sum_{i=1}^{n} \xi_i^*(t)$$

where $\xi_i^*(t) = X_i^*/(1-t) \wedge Y_i^*/t$, with $X_i^* = X_i/\bar{X}_n$ and $Y_i^* = Y_i/\bar{Y}_n$. They proved that their estimator is $n^{1/2}$-consistent, although they did not report its asymptotic distribution. Simulations suggest that in fact it performs better than the Deheuvels estimator.

4 Capéraà-Fougères-Genest estimator and variants

4.1 Original definition

A different family of estimators for the Pickands dependence function A was obtained in Capéraà, Fougères, and Genest (1997). Let again (X, Y) be a random pair with standard exponential margins and joint survivor function given by (1). Define $W = Y/(X+Y)$. The distribution function of W can be found using the conditional distribution of Y given X as stated in Lemma 2.1:

$$\Pr[W \le w] = w + w(1-w)\frac{A'(w)}{A(w)}, \qquad 0 \le w \le 1, \tag{12}$$

see Ghoudi, Khoudraji, and Rivest (1998). Equation (12) implies

$$\frac{A'(w)}{A(w)} = \frac{\Pr[W \le w] - w}{w(1-w)}, \qquad 0 < w < 1.$$

Since $A(0) = 1$ and $A(1) = 1$, integrating both sides of the preceding display yields

$$\begin{aligned} \log A(t) &= \int_0^t \frac{\Pr[W \le w] - w}{w(1-w)} \mathrm{d}w \\ &= -\int_t^1 \frac{\Pr[W \le w] - w}{w(1-w)} \mathrm{d}w. \end{aligned}$$

Hence, for arbitrary $p(t)$,

$$\log A(t) = p(t) \int_0^t \frac{\Pr[W \le w] - w}{w(1-w)} \mathrm{d}w - (1-p(t)) \int_t^1 \frac{\Pr[W \le w] - w}{w(1-w)} \mathrm{d}w.$$

Now let (X_i, Y_i), $i = 1, \ldots, n$, be independent random pairs with the same distribution as (X, Y), and put $W_i = Y_i/(X_i + Y_i)$. Estimate the distribution function $\Pr[W \le w]$ by the empirical one $n^{-1} \sum_{i=1}^n \mathbf{1}(W_i \le w)$, and substitute the latter into the preceding display to get the Capéraà-Fougères-Genest (CFG) estimator

$$\begin{aligned} \log \hat{A}_n^{\mathrm{cfg}}(t; p) &= p(t) \int_0^t \frac{n^{-1} \sum_{i=1}^n \mathbf{1}(W_i \le w) - w}{w(1-w)} \mathrm{d}w \\ &\quad - (1-p(t)) \int_t^1 \frac{n^{-1} \sum_{i=1}^n \mathbf{1}(W_i \le w) - w}{w(1-w)} \mathrm{d}w. \end{aligned} \tag{13}$$

By taking for instance $p(t) = 1-t$, the aggregate $\hat{A}_n^{\mathrm{cfg}}(t; p)$ satisfies the boundary constraints $\hat{A}_n^{\mathrm{cfg}}(0; p) = 1$ and $\hat{A}_n^{\mathrm{cfg}}(1; p) = 1$.

A variant of the CFG estimator was proposed in Jiménez, Villa-Diharce and Flores (2001). Their estimator, however, is consistent only in case $\log(A)$ is convex.

4.2 Simplified definition

The original expression (13) for the CFG estimator can be simplified considerably. Recall $\xi_i(t) = X_i/(1-t) \wedge Y_i/t$ for $0 \le t \le 1$, with $\xi_i(0) = X_i$ and $\xi_i(1) = Y_i$.

Lemma 4.1 *For $0 \le t \le 1$, with $\hat{A}_n^{\mathrm{cfg}}(\cdot; p)$ as in (13),*

$$\log \hat{A}_n^{\mathrm{cfg}}(t;p) = -\frac{1}{n}\sum_{i=1}^{n} \log \xi_i(t) + p(t)\frac{1}{n}\sum_{i=1}^{n} \log(X_i) + (1-p(t))\frac{1}{n}\sum_{i=1}^{n} \log(Y_i).$$

Lemma 4.1 suggests a much simpler interpretation of the CFG estimator. Let

$$\gamma = -\int_0^\infty \log(x) \mathrm{e}^{-x} \mathrm{d}x = 0.577\ldots$$

be Euler's constant. Observe that $\gamma = -\mathrm{E}[\log(X)]$ with X a standard exponential random variable. Since the distribution of $\xi(t) = X/(1-t) \wedge Y/t$ is exponential with $\mathrm{E}[\xi(t)] = 1/A(t)$,

$$\mathrm{E}[\log \xi(t)] = -\log A(t) - \gamma, \qquad 0 \le t \le 1. \tag{14}$$

Based on (14), a naive estimator for $\log A$ would be

$$\log \hat{A}_n^{\mathrm{cfg}}(t) = -\frac{1}{n}\sum_{i=1}^{n} \log \xi_i(t) - \gamma. \tag{15}$$

However, this estimator does not fulfil the constraints $A(0) = A(1) = 1$. The CFG estimator then can be viewed as a modification of the naive estimator designed to accommodate for these constraints:

$$\log \hat{A}_n^{\mathrm{cfg}}(t;p) = \log \hat{A}_n^{\mathrm{cfg}}(t) - p(t) \log \hat{A}_n^{\mathrm{cfg}}(0) - (1-p(t)) \log \hat{A}_n^{\mathrm{cfg}}(1)$$

If p is chosen in such a way that $p(0) = 1$ and $p(1) = 0$, then indeed $\hat{A}_n^{\mathrm{cfg}}(0;p) = 1$ and $\hat{A}_n^{\mathrm{cfg}}(1;p) = 1$.

Capéraà *et al.* (1997) explicitly compute the function $p(t)$ that minimizes the asymptotic variance of the estimator. In practice, the simple choice $p(t) = 1-t$ turns out to work well too; see also Section 5.2.

4.3 Asymptotics

If $p \in C[0,1]$, then we can write in abstract notation

$$\log \hat{A}_n^{\mathrm{cfg}}(\cdot;p) = L_{p,1-p}(\log \hat{A}_n^{\mathrm{cfg}})$$

with $L_{a,b}$ as in (8). Since $A(0) = 1$ and $A(1) = 1$, we have $L_{a,b}(\log A) = \log A$ for arbitrary $a, b \in C[0,1]$. This suggests the definition of the more general *CFG-type* estimator

$$\begin{aligned} \log \hat{A}_n^{\mathrm{cfg}}(\cdot;a,b) &= L_{a,b}(\log \hat{A}_n^{\mathrm{cfg}}) \\ &= \log \hat{A}_n^{\mathrm{cfg}} - a \log \hat{A}_n^{\mathrm{cfg}}(0) - b \log \hat{A}_n^{\mathrm{cfg}}(1) \end{aligned} \tag{16}$$

The choice $a = p$ and $b = 1-p$ gives the original CFG estimator $\hat{A}_n^{\mathrm{cfg}}(\cdot;p) = \hat{A}_n^{\mathrm{cfg}}(\cdot;p,1-p)$. Since the operator $L_{a,b}$ is linear and bounded, asymptotic properties for $\hat{A}_n^{\mathrm{cfg}}(\cdot;a,b)$ follow easily from those of $\hat{A}_n^{\mathrm{cfg}}$.

Theorem 4.2 *Let* (X_i, Y_i), $i = 1, \ldots, n$, *be independent random pairs with distribution (1). Then, with* $\hat{A}_n^{\mathrm{cfg}}$ *as in (15),*

$$n^{1/2} \left(\log \hat{A}_n^{\mathrm{cfg}} - \log A \right) \rightsquigarrow \zeta \tag{17}$$

in $C[0,1]$, *where* ζ *is a centered Gaussian process in* $C[0,1]$ *with covariance function*

$$\begin{aligned} &\operatorname{Cov}(\zeta(s), \zeta(t)) = \operatorname{Cov}(\log \xi(s), \log \xi(t)) \\ = \quad & \frac{\pi^2}{6} - \log(t)\log(1-s) + \log(t)\log(1-t) + \int_s^t \log(w) \frac{\mathrm{d}w}{1-w} \\ & + \log\left(\frac{t}{s}\right) \log A(s) + \log\left(\frac{1-s}{1-t}\right) \log A(t) + \frac{1}{2}\left(\log \frac{A(s)}{A(t)}\right)^2 \\ & - \int_s^t \log A(w) \frac{\mathrm{d}w}{w(1-w)} \end{aligned} \tag{18}$$

for $0 \le s \le t \le 1$, *with* ξ *as in (3).*

Corollary 4.3 *Under the assumptions and notations of Theorem 4.2, for* $a, b \in C[0,1]$,

$$n^{1/2} \left(\hat{A}_n^{\mathrm{cfg}}(\cdot; a, b) - A \right) \rightsquigarrow A L_{a,b} \zeta$$

in $C[0,1]$, *with* $L_{a,b}$ *and* $\hat{A}_n^{\mathrm{cfg}}(\cdot; a, b)$ *as in (8) and (16), respectively.*

Corollary 4.4 *Under the assumptions and notations of Theorem 4.2, for* $p \in C[0,1]$,

$$n^{1/2} \left(\hat{A}_n^{\mathrm{cfg}}(\cdot; p) - A \right) \rightsquigarrow A\{\zeta - p\zeta(0) - (1-p)\zeta(1)\}$$

in $C[0,1]$, *with* $\hat{A}_n^{\mathrm{cfg}}(\cdot; p)$ *as in (13).*

Corollary 4.4 is essentially due to Capéraà *et al.* (1997) except for the explicit expression (18) of the limiting covariance function.

5 Minimizing the asymptotic variance

5.1 Minimal asymptotic variance

In Corollaries 3.2 and 4.3, it makes sense to look for those functions a and b such that the asymptotic variances $\operatorname{Var}(L_{a,b}\eta(t))$ and $\operatorname{Var}(L_{a,b}\zeta(t))$ are minimal for all $0 \le t \le 1$. Such functions indeed exist, and they can be found from the following simple result.

Lemma 5.1 *Let δ be a Gaussian process in $C[0,1]$ with covariance function $\sigma(s,t) = \mathrm{Cov}(\delta(s), \delta(t))$ for $(s,t) \in [0,1]^2$. Assume that the covariance matrix of $(\delta(0), \delta(1))$ is non-singular. Define $a_\delta, b_\delta \in C[0,1]$ by*

$$(a_\delta(t), b_\delta(t)) = (\sigma(t,0), \sigma(t,1)) \begin{pmatrix} \sigma(0,0) & \sigma(0,1) \\ \sigma(1,0) & \sigma(1,1) \end{pmatrix}^{-1} \tag{19}$$

for $0 \le t \le 1$. Then for $0 \le t \le 1$, with $L_{a,b}$ as in (8),

$$\begin{aligned} \mathrm{Var}\left(L_{a_\delta, b_\delta}\delta(t)\right) &= \sigma(t,t) - (a_\delta(t), b_\delta(t)) \begin{pmatrix} \sigma(t,0) \\ \sigma(t,1) \end{pmatrix} \\ &= \min_{a,b \in C[0,1]} \mathrm{Var}\left(L_{a,b}\delta(t)\right). \end{aligned}$$

Lemma 5.1 can be applied to the limiting processes η and ζ in Corollaries 3.2 and 4.3. The result is explicit expressions in terms of A for the functions $a, b \in C[0,1]$ that minimize the asymptotic variances of the estimators $\hat{A}_n^{\mathrm{p}}(\cdot; a, b)$ and $\hat{A}_n^{\mathrm{cfg}}(\cdot; a, b)$ in (9) and (16). The only provision is that $\xi(0) = X$ and $\xi(1) = Y$ are not completely dependent, that is, $A(1/2) > 1/2$. This is no great loss of generality, however, as in case of complete dependence $\hat{A}_n^{\mathrm{cfg}}(t; p) = t \vee (1-t) = A(t)$ with probability one for *any* function p.

5.2 Special case: independence

An interesting special case is that of independence, that is, $A \equiv 1$. In that case, the covariance functions σ_η and σ_ζ of the processes η and ζ in Theorems 3.1 and 4.2 satisfy

$$\begin{gathered} \sigma(0,1) = 0, \qquad \sigma(t,t) = \sigma(0,0), \qquad \sigma(s,t) = \sigma(1-t, 1-s), \\ \sigma_\eta(t) := \sigma_\eta(0,t) = 1 - t, \qquad \sigma_\zeta(t) := \sigma_\zeta(0,t) = L_2(1) - L_2(t), \end{gathered} \tag{20}$$

where L_2 denotes the *dilogarithm* function

$$L_2(t) = -\int_0^t \log(1-w) \frac{\mathrm{d}w}{w} = \sum_{k=1}^\infty \frac{t^k}{k^2}, \qquad -1 \le t \le 1,$$

with special value $L_2(1) = \pi^2/6$. Therefore, the variance-minimizing functions in equation (19) simplify to

$$a_\delta(t) = \frac{\sigma(t)}{\sigma(0)}, \qquad b_\delta(t) = \frac{\sigma(1-t)}{\sigma(0)}, \tag{21}$$

with corresponding minimal variance

$$\mathrm{Var}\left(L_{a_\delta, b_\delta}\delta(t)\right) = \sigma(0) - \frac{\sigma^2(t)}{\sigma(0)} - \frac{\sigma^2(1-t)}{\sigma(0)}.$$

Pickands-type estimators. For the Pickands-type estimators $\hat{A}_n(t;a,b)$ in (9), equations (20) and (21) lead to $a_\eta(t) = 1-t$ and $b_\eta(t) = t$. This shows that in case of independence, the Deheuvels estimator $\hat{A}_n^{\mathrm{d}}(t)$ in (7) has minimal asymptotic variance

$$\mathrm{Var}\{\eta(t) - (1-t)\eta(0) - t\eta(1)\} = 2t(1-t)$$

among all Pickands-type estimators $\hat{A}_n(t;a,b)$ as a and b range over $C[0,1]$.

CFG-type estimators. More interesting is the situation for the CFG-type estimators $\hat{A}_n^{\mathrm{cfg}}(\cdot;a,b)$ in (16). By (20) and (21), the variance-minimizing functions are

$$a_\zeta(t) = \frac{L_2(1) - L_2(t)}{L_2(1)}, \qquad b_\zeta(t) = \frac{L_2(1) - L_2(1-t)}{L_2(1)} \tag{22}$$

with corresponding minimal variance

$$\begin{aligned} &\mathrm{Var}\{\zeta(t) - a_\zeta(t)\zeta(0) - b_\zeta(t)\zeta(1)\} \\ &= \quad L_2(1) - \frac{\{L_2(1) - L_2(t)\}^2}{L_2(1)} - \frac{\{L_2(1) - L_2(1-t)\}^2}{L_2(1)}. \end{aligned}$$

Since $a_\zeta(t) + b_\zeta(t) \neq 1$ for $0 < t < 1$, the minimum-variance estimator $\hat{A}_n^{\mathrm{cfg}}(t;a_\zeta,b_\zeta)$ is *not* a CFG estimator $\hat{A}_n^{\mathrm{cfg}}(\cdot;p) = \hat{A}_n^{\mathrm{cfg}}(t;p,1-p)$ as in (13). Still, it is interesting to compute the function p_ζ for which the asymptotic variance $\mathrm{Var}(\{\zeta(t) - \zeta(1)\} - p(t)\{\zeta(0) - \zeta(1)\})$ of $\hat{A}_n^{\mathrm{cfg}}(t;p)$ is minimal as p ranges over $C[0,1]$. By Lemma 6.1 below, the variance-minimizing p is

$$p_\zeta(t) = \frac{L_2(1) - L_2(t) + L_2(1-t)}{2L_2(1)}, \tag{23}$$

with corresponding minimal variance

$$\begin{aligned} &\mathrm{Var}(\{\zeta(t) - \zeta(1)\} - p_\zeta(t)\{\zeta(0) - \zeta(1)\}) \\ &= \quad 2L_2(1-t) - \frac{\{L_2(1) - L_2(t) + L_2(1-t)\}^2}{2L_2(1)}. \end{aligned}$$

In contrast, the simple choice $p(t) = 1-t$ yields

$$\begin{aligned} &\mathrm{Var}(\{\zeta(t) - \zeta(1)\} - (1-t)\{\zeta(0) - \zeta(1)\}) \\ &= \quad 2tL_2(1-t) + 2(1-t)L_2(t) - 2t(1-t)L_2(1). \end{aligned}$$

Comparisons. Figure 1 shows the asymptotic relative efficiencies (the ratio of asymptotic variances) at independence ($A \equiv 1$) with respect to the optimal CFG-type estimator $\hat{A}_n^{\mathrm{cfg}}(t;a_\zeta,b_\zeta)$ as in (22) as function of $0 < t < 1$ of the following three estimators: the optimal CFG estimator $\hat{A}_n^{\mathrm{cfg}}(t;p_\zeta)$ as in (23); the CFG estimator $\hat{A}_n^{\mathrm{cfg}}(t;p)$ with $p(t) = 1-t$; the Deheuvels estimator $\hat{A}_n^{\mathrm{d}}(t)$.

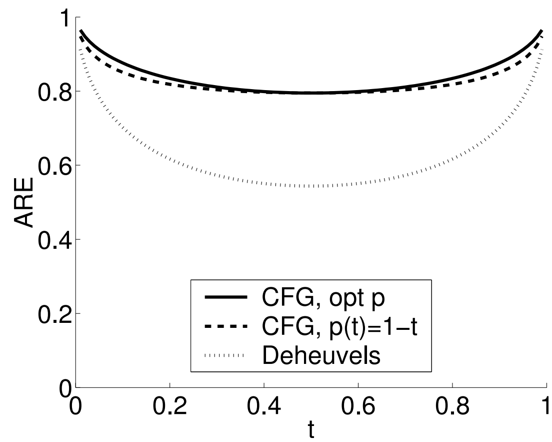

Figure 1: *Asymptotic relative efficiencies at independence with respect to CFG estimator with optimal a and b of: CFG estimator with optimal p (full), CFG estimator with* $p(t) = 1 - t$ *(dashed), Deheuvels estimator (dotted)*

The asymptotic variance of the Deheuvels estimator, which is the optimal Pickands-type estimator, is higher than the asymptotic variances of the three CFG-type estimators. This is quite remarkable: both Pickands and CFG-type estimators can be interpreted as estimators of the mean of an exponential distribution, the original Pickands estimators being the maximum-likelihood estimator.

Within the class of CFG estimators $\hat{A}_n^{\mathrm{cfg}}(t;p)$, the difference between the simple choice $p(t) = 1 - t$ and the optimal choice $p = p_\zeta$ of (23) is almost negligible. The asymptotic relative efficiency of the optimal CFG estimator $\hat{A}_n^{\mathrm{cfg}}(t;p_\zeta)$ with respect to the optimal CFG-type estimator $\hat{A}_n^{\mathrm{cfg}}(t;a_\zeta,b_\zeta)$ is at $t = 1/2$ approximately equal to 79.5%.

5.3 Adaptive estimators

The variance-minimizing functions a_δ and b_δ in Lemma 5.1 depend on the covariance structure of the process δ. In case δ represents one of the limiting processes η or ζ in Theorems 3.1 or 4.2, the corresponding covariance structures (11) and (18) in turn depend on the unknown Pickands dependence function A.

Replacing the optimal a_δ and b_δ by estimates leads to adaptive Pickands or CFG-type estimators. In case of CFG-type estimators, the procedure could go as follows:

1. Start with an initial estimator $\hat{A}_n$ for A (for instance $\hat{A}_n(t) = t \vee (1-t) \vee \{\hat{A}_n^{\mathrm{cfg}}(t;p) \wedge 1\}$ at $p(t) = 1 - t$).
2. In (18), replace A by $\hat{A}_n$ to find an estimate $\hat{\sigma}_n$ of the covariance function σ of the limiting process ζ in Theorem 4.2.

3. In (19), replace σ by $\hat{\sigma}_n$ to find estimates $(\hat{a}_n, \hat{b}_n)$ of the variance-minimizing functions (a, b).

4. Finally, estimate A by $\hat{A}_n^{\text{cfg}}(\cdot; \hat{a}_n, \hat{b}_n)$.

The following theorem states that the asymptotic distribution of the adaptive estimator $\hat{A}_n^{\text{cfg}}(\cdot; \hat{a}_n, \hat{b}_n)$ is under general conditions the same as when the true variance-minimizing functions of (19) would have been used instead.

Theorem 5.2 *Under the assumptions and notations of Theorem 4.2, if* $A(1/2) > 1/2$, *and if the initial estimator* $\hat{A}_n$ *satisfies* $t \vee (1-t) \leq \hat{A}_n(t) \leq 1$ *and* $\|\hat{A}_n - A\|_\infty = o_p(1)$, *then*

$$n^{1/2} \left(\hat{A}_n^{\text{cfg}}(\cdot; \hat{a}_n, \hat{b}_n) - A \right) \rightsquigarrow A L_{a_0, b_0} \zeta,$$

where $\hat{a}_n$ *and* $\hat{b}_n$ *are constructed according to the adaptive procedure above and with* (a_0, b_0) *the minimizers of* $\mathrm{Var}(L_{a,b}\zeta)$ *as* a *and* b *range over* $C[0,1]$.

A simple, alternative way to construct $\hat{a}_n$ and $\hat{b}_n$ is to estimate the unknown covariance $\sigma(s,t)$ in (19) not by the plug-in estimator based on (18) and an initial estimate $\hat{A}_n$ but by the sample covariance of the pairs $(\log \xi_i(s), \log \xi_i(t))$ for $i = 1, \ldots, n$. In fact, whatever estimator is used for the covariance function, as long as it is uniformly consistent, the asymptotic distribution of the adaptive estimator for A is as stated in Theorem 5.2.

A similar adaptive construction is feasible for Pickands-type estimators. The comparisons of asymptotic variances in case of independence in Section 5.2, however, raise the suspicion that the adaptive Pickands-type estimator will in general have a higher asymptotic variance than the adaptive CFG-type estimator.

Comment. The adaptive estimator with covariance function estimated by sample covariances can be cast within two familiar statistical procedures.

(i) It can be seen as the least-squares estimator in the linear regression model

$$\begin{aligned} &-\log \xi_i(t) - \gamma \\ &\quad = \quad \beta_0(t) \log A(t) + \beta_1(t)\{-\log(X_i) - \gamma\} + \beta_1(t)\{-\log(Y_i) - \gamma\} + \varepsilon_i(t) \end{aligned}$$

with regression coefficients $(\beta_0(t), \beta_1(t), \beta_2(t)) = (\log A(t), a_0(t), b_0(t))$.

(ii) It can also be seen as the constrained maximum empirical likelihood estimator in the sense of Owen (1991), section 3.1. In the latter approach, $\log A(t)$ would be estimated by $\sum_{i=1}^n p_{ni}(-\log \xi_i(t) - \gamma)$ with $p_n = (p_{n1}, \ldots, p_{nn})$ being the maximizer of the binomial likelihood $\prod_{i=1}^n p_{ni}$ subject to the constraints $p_{ni} \geq 0$, $\sum_{i=1}^n p_{ni} = 1$, $\sum_{i=1}^n p_{ni}(\log X_i + \gamma) = 0$ and $\sum_{i=1}^n p_{ni}(\log Y_i + \gamma) = 0$.

6 Proofs

6.1 Proofs for Section 2

Proof of Lemma 2.1

Fix $0 \leq y < \infty$. Since A is absolutely continuous, the function $x \mapsto S(x, y)$ is absolutely continuous as well with

$$S(x,y) = \int_x^\infty S(z,y) Q\left(\frac{y}{z+y}\right) \mathrm{d}z.$$

for $0 \leq x < \infty$, with $Q(r) = A(r) - rA'(r)$. Hence for $0 \leq x < \infty$,

$$\begin{aligned}
\mathrm{E}[\mathbf{1}(X > x) \Pr[Y > y \mid X]] &= S(x,y) \\
&= \mathrm{E}\left[\mathbf{1}(X > x) \mathrm{e}^X S(X,y) Q\left(\frac{y}{X+y}\right)\right].
\end{aligned}$$

Since this holds for all $0 \leq x < \infty$, in fact

$$\mathrm{E}[f(X) \Pr[Y > y \mid X]] = \mathrm{E}\left[f(X) \mathrm{e}^X S(X,y) Q\left(\frac{y}{X+y}\right)\right]$$

for all integrable f, whence the result. □

Proof of Lemma 2.2

First, fix $0 < x < \infty$ and $b < 0$. We have

$$\begin{aligned}
\mathrm{E}[\mathrm{e}^{bY} \mid X = x] &= \int_0^1 \Pr[\mathrm{e}^{bY} > u \mid X = x] \mathrm{d}u \\
&= \int_0^1 \Pr[Y < b^{-1} \log(u) \mid X = x] \mathrm{d}u \\
&= 1 - \int_0^1 \Pr[Y > b^{-1} \log(u) \mid X = x] \mathrm{d}u \\
&= 1 + b \int_0^\infty \Pr[Y > y \mid X = x] \mathrm{e}^{by} \mathrm{d}y.
\end{aligned}$$

For $0 \leq b < 1$, it is not hard to see that

$$\mathrm{E}[\mathrm{e}^{bY} \mid X = x] = 1 + b \int_0^\infty \Pr[Y > y \mid X = x] \mathrm{e}^{by} \mathrm{d}y$$

as well. Put $Q(t) = A(t) - tA'(t)$ for $0 \leq t \leq 1$. By Lemma 2.1, the integral on the right is equal to

$$\begin{aligned}
&\int_0^\infty \Pr[Y > y \mid X = x] \mathrm{e}^{by} \mathrm{d}y \\
&= \mathrm{e}^x \int_0^\infty \exp\left\{-(x+y) A\left(\frac{y}{x+y}\right)\right\} Q\left(\frac{y}{x+y}\right) \mathrm{e}^{by} \mathrm{d}y,
\end{aligned}$$

which by changing variables $t = y/(x+y)$ can be rewritten as

$$\begin{aligned} &\int_0^\infty \Pr[Y > y \mid X = x] \mathrm{e}^{by} \mathrm{d}y \\ &\quad = \int_0^1 x \exp\left\{ \left(1 - \frac{A(t) - bt}{1-t}\right) x \right\} Q(t) \frac{\mathrm{d}t}{(1-t)^2}. \end{aligned}$$

Therefore,

$$\begin{aligned} &\mathrm{E}[\mathrm{e}^{aX+bY}] = \mathrm{E}[\mathrm{e}^{aX} \mathrm{E}[\mathrm{e}^{bY} \mid X]] \\ &= \mathrm{E}[\mathrm{e}^{aX}] + b \int_0^1 \mathrm{E}\left[X \exp\left\{ \left(1 + a - \frac{A(t) - bt}{1-t}\right) X \right\}\right] Q(t) \frac{\mathrm{d}t}{(1-t)^2}. \end{aligned}$$

Since the distribution of X is standard exponential, $\mathrm{E}[\mathrm{e}^{\alpha X}] = (1-\alpha)^{-1}$ and $\mathrm{E}[X\mathrm{e}^{\alpha X}] = (1-\alpha)^{-2}$ for $\alpha < 1$, and infinity otherwise. Hence the inner expectation in the previous display is finite if and only if $A(t) - a(1-t) - bt > 0$ for all $0 \le t \le 1$. For such a and b,

$$\mathrm{E}\left[\mathrm{e}^{aX+bY}\right] = \frac{1}{1-a} + b \int_0^1 \frac{Q(t)}{\{A(t) - a(1-t) - bt\}^2} \mathrm{d}t.$$

Finally, the integral on the right-hand side can be manipulated as follows:

$$\begin{aligned} &\int_0^1 \frac{A(t) - tA'(t)}{\{A(t) - a(1-t) - bt\}^2} \mathrm{d}t \\ &= \int_0^1 \frac{\mathrm{d}t}{A(t) - a(1-t) - bt} \mathrm{d}t + \int_0^1 \frac{a(1-t) + bt - tA'(t)}{\{A(t) - a(1-t) - bt\}^2} \mathrm{d}t \\ &= \int_0^1 \frac{\mathrm{d}t}{A(t) - a(1-t) - bt} \mathrm{d}t + a \int_0^1 \frac{\mathrm{d}t}{\{A(t) - a(1-t) - bt\}^2} \\ &\quad - \int_0^1 t \frac{\{A(t) - a(1-t) - bt\}'}{\{A(t) - a(1-t) - bt\}^2} \mathrm{d}t \\ &= a \int_0^1 \frac{\mathrm{d}t}{\{A(t) - a(1-t) - bt\}^2} + \frac{1}{1-b}, \end{aligned}$$

the last equality following from partial integration. □

6.2 Proofs for Section 3

Proof of Theorem 3.1

Convergence of the finite-dimensional distributions follows immediately from the multivariate central limit theorem. The proof for the covariance function in (11) is given below. An alternative proof for tightness than the one by Deheuvels (1991) goes as follows.

Let P be the probability distribution of (X, Y) in (1), and let P_n be the empirical distribution of the sample (X_i, Y_i), $i = 1, \ldots, n$. For $0 \le t \le 1$, let

$$f_t(x, y) = \frac{x}{t} \wedge \frac{y}{1-t}, \qquad (x, y) \in (0, \infty). \tag{24}$$

Observe that $Pf_t = 1/A(t)$ and $\bar{\xi}_n(t) = P_n f_t$ for all positive integer n and all $0 \leq t \leq 1$.

Put $\mathcal{F} = \{f_t : 0 \leq t \leq 1\}$. According to Theorem 2.6.8 of van der Vaart and Wellner (1996), the class $\mathcal{F}$ is P-Donsker: First, it is has a P-square integrable envelope function because $0 \leq f_t(x, y) \leq 2(x \vee y)$ for all $0 \leq t \leq 1$ and all $(x, y) \in (0, \infty)^2$. Second, it is point-wise separable since $t \mapsto f_t(x, y)$ is continuous for all $(x, y) \in (0, \infty)^2$. Third, it is a Vapnik-Červonenkis subgraph (VC) class by repeated applications of Lemmas 2.6.15 and 2.6.18 of van der Vaart and Wellner (1996).

Let $\ell^\infty([0, 1])$ be the class of bounded, real-valued functions on $[0, 1]$. Equip $\ell^\infty([0, 1])$ with the topology of uniform convergence. The property that $\mathcal{F}$ is P-Donsker means weak convergence of the empirical processes

$$\eta_n = \left(n^{1/2}\left(P_n f_t - P f_t\right)\right)_{t \in [0,1]} \rightsquigarrow \eta \tag{25}$$

in $\ell^\infty([0, 1])$; here, we identified $\mathcal{F}$ with $[0, 1]$. The process η is a tight, centered Gaussian process on $[0, 1]$ with bounded sample paths and with covariance function (11). By Example 1.5.10 of van der Vaart and Wellner (1996), its sample paths are uniformly continuous with respect to the standard-deviation semimetric

$$\rho(s, t) = \left(\mathrm{E}[\{\eta(s) - \eta(t)\}^2]\right)^{1/2}$$

for $(s, t) \in [0, 1]^2$. Using (11), it is not difficult to see that there exists a positive constant C such that

$$\mathrm{E}[\{\eta(s) - \eta(t)\}^2] = \mathrm{Var}(\xi(s) - \xi(t)) \leq C|t - s|$$

for all $(s, t) \in [0, 1]^2$. Hence the sample paths of η are continuous with respect to the Euclidean distance on $[0, 1]$, so that η, like η_n, actually takes its values in the subspace $C[0, 1]$ of $\ell^\infty([0, 1])$. By Theorem 1.3.10 of van der Vaart and Wellner (1996), we can conclude that the weak convergence in (25) also holds true in $C[0, 1]$. □

Proof of equation (11)

Fix $0 < s \leq t < 1$. Since $(1 - s)t \geq s(1 - t)$,

$$\begin{aligned} \xi(s)\xi(t) &= \left(\frac{X}{1-s} \wedge \frac{Y}{s}\right)\left(\frac{X}{1-t} \wedge \frac{Y}{t}\right) \\ &= \frac{X^2}{(1-s)(1-t)} \wedge \frac{XY}{(1-s)t} \wedge \frac{Y^2}{st}. \end{aligned}$$

Hence

$$\begin{aligned} \mathrm{E}[\xi(s)\xi(t)] &= \int_0^\infty \Pr[\xi(s)\xi(t) > z]\,\mathrm{d}z \\ &= \int_0^\infty \Pr[X^2 > (1-s)(1-t)z, XY > (1-s)tz, Y^2 > stz]\,\mathrm{d}z. \end{aligned}$$

From Lemma 2.1, we have $\Pr[Y > y \mid X = x] = \mathrm{e}^x g(x,y)$ with $g(x,y) = S(x,y)Q(w)$, where $Q(w) = A(w) - wA'(w)$ and $w = y/(x+y)$. By the previous display,

$$\begin{aligned}
&\mathrm{E}[\xi(s)\xi(t)] \\
&= \int_0^\infty \mathrm{E}[\mathbf{1}\{X^2 > (1-s)(1-t)z\} \Pr[Y > (1-s)tz/X \vee (stz)^{1/2} \mid X]]\mathrm{d}z \\
&= \int_0^\infty \int_0^\infty \mathbf{1}\{x^2 > (1-s)(1-t)z\} g\left(x, (1-s)tz/x \vee (stz)^{1/2}\right) \mathrm{d}x\mathrm{d}z.
\end{aligned}$$

Change variables $(x,z) = (x, ux^2)$ and put $v(u) = (1-s)tu \vee (stu)^{1/2}$ to get

$$\mathrm{E}[\xi(s)\xi(t)] = \int_{u=0}^{\frac{1}{(1-s)(1-t)}} \int_{x=0}^\infty g\left(x, v(u)x\right) x^2 \mathrm{d}x\mathrm{d}u.$$

Observe that $g(x, vx) = \exp\{-(1+v)A(w)x\}Q(w)$ with $w = v/(1+v)$. Hence, denoting $w(u) = v(u)/(1+v(u))$,

$$\begin{aligned}
\mathrm{E}[\xi(s)\xi(t)] &= \int_{u=0}^{\frac{1}{(1-s)(1-t)}} Q(w(u)) \int_{x=0}^\infty x^2 \exp\{-(1+v(u))A(w(u))x\}\mathrm{d}x\mathrm{d}u \\
&= 2\int_0^{\frac{1}{(1-s)(1-t)}} \frac{Q(w(u))}{\{(1+v(u))A(w(u))\}^3}\mathrm{d}u.
\end{aligned}$$

Obviously, $(1-s)tu > (stu)^{1/2}$ if and only if $u > s(1-s)^{-2}t^{-1}$. Since $0 < s \le t < 1$ also implies $s(1-s)^{-2}t^{-1} \le (1-s)^{-1}(1-t)^{-1}$,

$$\begin{aligned}
\mathrm{E}[\xi(s)\xi(t)] &= 2\int_0^{\frac{s}{(1-s)^2 t}} \frac{Q\left(\frac{(stu)^{1/2}}{1+(stu)^{1/2}}\right)}{\left\{\left(1+(stu)^{1/2}\right) A\left(\frac{(stu)^{1/2}}{1+(stu)^{1/2}}\right)\right\}^3}\mathrm{d}u \\
&\quad + 2\int_{\frac{s}{(1-s)^2 t}}^{\frac{1}{(1-s)(1-t)}} \frac{Q\left(\frac{(1-s)tu}{1+(1-s)tu}\right)}{\left\{\left(1+(1-s)tu\right) A\left(\frac{(1-s)tu}{1+(1-s)tu}\right)\right\}^3}\mathrm{d}u.
\end{aligned}$$

In the first integral, put $w = (stu)^{1/2}/\left(1+(stu)^{1/2}\right)$; in the second integral, put $w = (1-s)tu/(1+(1-s)tu)$. The result is

$$\mathrm{E}[\xi(s)\xi(t)] = \frac{4}{st}\int_0^s w\frac{Q(w)}{A^3(w)}\mathrm{d}w + \frac{2}{(1-s)t}\int_s^t (1-w)\frac{Q(w)}{A^3(w)}\mathrm{d}w.$$

Since A is absolutely continuous with density A', the function $w \mapsto (w/A(w))^2$ is absolutely continuous with density $w \mapsto 2wQ(w)/A^3(w)$. Therefore, the first term on the right-hand side of the previous display simplifies to

$$\frac{4}{st}\int_0^s w\frac{Q(w)}{A^3(w)}\mathrm{d}w = \frac{2}{st}\left[\left(\frac{w}{A(w)}\right)^2\right]_0^s = 2\frac{s}{t}\frac{1}{A^2(s)},$$

while the second term can be manipulated by partial integration as

$$\begin{aligned}
&\frac{2}{(1-s)t}\int_s^t (1-w)\frac{Q(w)}{A^3(w)}\mathrm{d}w \\
&= \frac{1}{(1-s)t}\left[\frac{1-w}{w}\left(\frac{w}{A(w)}\right)^2\right]_s^t - \frac{1}{(1-s)t}\int_s^t \left(\frac{w}{A(w)}\right)^2 \mathrm{d}\frac{1-w}{w} \\
&= \frac{1-t}{1-s}\frac{1}{A^2(t)} - \frac{s}{t}\frac{1}{A^2(s)} + \frac{1}{(1-s)t}\int_s^t \frac{\mathrm{d}w}{A^2(w)}.
\end{aligned}$$

Collect the last three displays to get

$$\mathrm{E}[\xi(s)\xi(t)] = \frac{s}{t}\frac{1}{A^2(s)} + \frac{1-t}{1-s}\frac{1}{A^2(t)} + \frac{1}{(1-s)t}\int_s^t \frac{\mathrm{d}w}{A^2(w)}.$$

Finally, use $\mathrm{E}[\xi(s)] = 1/A(s)$ to obtain the result for the case $0 < s \le t < 1$.

The cases $s = 0$ and $t = 1$ follow from the case $0 < s \le t < 1$ by the continuity of the function $t \mapsto \log \xi(t)$ and the bounds $0 \le \xi(t) \le 2(X \vee Y)$. □

Proof of Corollary 3.2

Since the operator $L_{a,b}$ is linear and continuous, and since $A(0) = 1 = A(1)$ implies $L_{a,b}(1/A) = 1/A - a - b$,

$$\begin{aligned}
n^{1/2}\left(1/\hat{A}_n(\cdot;a,b) - 1/A\right) &= n^{1/2}\left(L_{a,b}(\bar{\xi}_n) + a + b - 1/A\right) \\
&= n^{1/2}\left(L_{a,b}(\bar{\xi}_n) - L_{a,b}(1/A)\right) \\
&= L\left(n^{1/2}(\bar{\xi}_n - 1/A)\right) \\
&\rightsquigarrow L_{a,b}\eta
\end{aligned}$$

in $C[0,1]$. In particular, $\|1/\hat{A}_n(\cdot;a,b) - 1/A\|_\infty = o_p(1)$. Since $1/2 \le A \le 1$,

$$\begin{aligned}
n^{1/2}\left(\hat{A}_n(\cdot;a,b) - A\right) &= \frac{n^{1/2}\left(1/\hat{A}_n(\cdot;a,b) - 1/A\right)}{1/\hat{A}_n(\cdot;a,b)\cdot 1/A} \\
&\rightsquigarrow \frac{L_{a,b}\eta}{(1/A)^2} = A^2 L_{a,b}\eta
\end{aligned}$$

in $C[0,1]$. □

6.3 Proofs for Section 4

Proof of Lemma 4.1

Observe that $\log \hat{A}_n^{\mathrm{cfg}}(t;p)$ is equal to

$$\frac{1}{n}\sum_{i=1}^n \int_0^t \frac{\mathbf{1}(W_i \le w) - w}{w(1-w)}\mathrm{d}w - (1-p(t))\frac{1}{n}\sum_{i=1}^n \int_0^1 \frac{\mathbf{1}(W_i \le w) - w}{w(1-w)}\mathrm{d}w. \quad (26)$$

The integrals can be computed explicitly:

$$\begin{aligned}\int_0^t \frac{\mathbf{1}(W_i \le w) - w}{w(1-w)} \mathrm{d}w &= -\int_0^{t\wedge W_i} \frac{\mathrm{d}w}{1-w} + \int_{t\wedge W_i}^t \frac{\mathrm{d}w}{w} \\ &= \log\{1-(t\wedge W_i)\} + \log(t) - \log\{t \wedge W_i\} \\ &= \log\left\{t\left(\frac{1-t}{t} \vee \frac{1-W_i}{W_i}\right)\right\}.\end{aligned}$$

Since $W_i = Y_i/(X_i+Y_i)$,

$$\begin{aligned}\int_0^t \frac{\mathbf{1}(W_i \le w) - w}{w(1-w)} \mathrm{d}w &= \log\left\{(1-t) \vee t\frac{X_i}{Y_i}\right\} \\ &= \log\{tX_i \vee (1-t)Y_i\} - \log(Y_i) \\ &= \log(X_i) - \log\{X_i/(1-t) \wedge Y_i/t\}. \qquad (27)\end{aligned}$$

In particular,

$$\int_0^1 \frac{\mathbf{1}(W_i \le w) - w}{w(1-w)} \mathrm{d}w = \log(X_i) - \log(Y_i). \qquad (28)$$

Combine (26), (27) and (28) to find the stated expression for $\log \hat{A}_n^{\mathrm{cfg}}(t;p)$. □

Proof of Theorem 4.2

The proof for the expression for the covariance function (18) is given below. For the rest, the proof is completely analogous to the one of Theorem 3.1, and we indicate only the appropriate changes. The appropriate class of functions is now $\mathcal{G} = \{g_t : 0 \le t \le 1\}$, with $g_t = \log f_t$ and f_t as in (24). That $\mathcal{G}$ has a P-square integrable envelope function follows from the bounds $x \wedge y \le f_t(x,y) \le 2(x \vee y)$. Equation (18) leads to the existence of a positive constant C such that $\mathrm{Var}(\log\xi(s) - \log\xi(t)) \le C|t-s|$ for all $(s,t) \in [0,1]^2$, and this implies that the sample paths of the limiting process ζ are continuous.

Proof of equation (18)

Fix $0 < s \le t < 1$. Put $u = s/(1-s)$ and $v = t/(1-t)$. Then

$$\begin{aligned}&\mathrm{E}[\log\xi(s)\log\xi(t)] \\ &\quad= \mathrm{E}[\{\log(uX\wedge Y) - \log(s)\}\{\log(vX \wedge Y) - \log(t)\}] \\ &\quad= \log(s)\log(t) - \log(s)\mathrm{E}[\log(vX\wedge Y)] - \log(t)\mathrm{E}[\log(uX\wedge Y)] \\ &\qquad + \mathrm{E}[\log(uX\wedge Y)\log(vX \wedge Y)]. \qquad (29)\end{aligned}$$

By (4),

$$\begin{aligned}\mathrm{E}[\log(uX\wedge Y)] &= \mathrm{E}[\log\xi(s)] + \log(s) \\ &= -\gamma - \log A(s) + \log(s), \qquad (30)\end{aligned}$$

and similarly for $\mathrm{E}[\log(vX \wedge Y)]$. Further,

$$\begin{aligned}
&\mathrm{E}[\log(uX \wedge Y)\log(vX \wedge Y)] \\
&\quad = \mathrm{E}[\{\log(Y) - \log(Y/(uX) \vee 1)\}\{\log(Y) - \log(Y/(vX) \vee 1)\}] \\
&\quad = \mathrm{E}[\{\log(Y)\}^2] - \mathrm{E}[\log(Y)\log(Y/(vX) \vee 1)] \\
&\qquad - \mathrm{E}[\log(Y)\log(Y/(uX) \vee 1)] + \mathrm{E}[\log(Y/(uX) \vee 1)\log(Y/(vX) \vee 1)].
\end{aligned} \tag{31}$$

As $\log(Y)$ is a Gumbel random variable, $\mathrm{E}[\{\log(Y)\}^2] = \pi^2/6 + \gamma^2$. Moreover,

$$\begin{aligned}
&\mathrm{E}[\log(Y)\log(Y/(uX) \vee 1)] \\
&\quad = \mathrm{E}[\{\log(Y/(uX)) + \log(uX)\}\log(Y/(uX) \vee 1)] \\
&\quad = \mathrm{E}[\{\log(Y/(uX) \vee 1)\}^2] + \mathrm{E}[\log(uX)\log(Y/(uX) \vee 1)],
\end{aligned} \tag{32}$$

and similarly for $\mathrm{E}[\log(Y)\log(Y/(vX) \vee 1)]$.

We separately treat the two terms on the right of the preceding display. First, as $\log(Y/(uX) \vee 1)^2 > z$ if and only if $Y > u\mathrm{e}^{\sqrt{z}}X$ for $0 < z < \infty$, by Lemma 2.1, denoting $Q(w) = A(w) - wA'(w)$,

$$\begin{aligned}
&\mathrm{E}[\{\log(Y/(uX) \vee 1)\}^2] \\
&\quad = \int_0^\infty \mathrm{E}\left[\Pr[Y > u\mathrm{e}^{\sqrt{z}}X \mid X]\right] \mathrm{d}z \\
&\quad = \int_0^\infty Q\left(\frac{u\mathrm{e}^{\sqrt{z}}}{1+u\mathrm{e}^{\sqrt{z}}}\right) \int_0^\infty \exp\left\{-\left(1+u\mathrm{e}^{\sqrt{z}}\right) A\left(\frac{u\mathrm{e}^{\sqrt{z}}}{1+u\mathrm{e}^{\sqrt{z}}}\right) x\right\} \mathrm{d}x\mathrm{d}z.
\end{aligned}$$

Calculate the inner integral to obtain

$$\mathrm{E}[\{\log(Y/(uX) \vee 1)\}^2] = \int_0^\infty \frac{Q\left(\frac{u\mathrm{e}^{\sqrt{z}}}{1+u\mathrm{e}^{\sqrt{z}}}\right)}{\left(1+u\mathrm{e}^{\sqrt{z}}\right) A\left(\frac{u\mathrm{e}^{\sqrt{z}}}{1+u\mathrm{e}^{\sqrt{z}}}\right)} \mathrm{d}z.$$

A change of variables $w = u\mathrm{e}^{\sqrt{z}}/(1+u\mathrm{e}^{\sqrt{z}})$ gives

$$\mathrm{E}[\{\log(Y/(uX) \vee 1)\}^2] = 2\int_s^1 \frac{Q(w)}{wA(w)} \log\left(\frac{1}{u}\frac{w}{1-w}\right) \mathrm{d}w$$

Since $Q(w)/(wA(w)) = 1/w - A'(w)/A(w)$, partial integration gives

$$\mathrm{E}[\{\log(Y/(uX) \vee 1)\}^2] = 2\int_s^1 \{\log A(w) - \log(w)\} \frac{\mathrm{d}w}{w(1-w)}. \tag{33}$$

Second, denoting again $\Pr[Y > y \mid X = x] = \mathrm{e}^x g(x,y)$ with g to be found from Lemma 2.1,

$$\begin{aligned}
\mathrm{E}[\log(Y/(uX) \vee 1) \mid X = x] &= \int_0^\infty \Pr[Y > u\mathrm{e}^z x \mid X = x]\mathrm{d}z \\
&= \mathrm{e}^x \int_0^\infty g(x, u\mathrm{e}^z x)\mathrm{d}z
\end{aligned}$$

and thus, with $Q(w) = A(w) - wA'(w)$,

$$
\begin{aligned}
&\mathrm{E}[\log(uX)\log(Y/(uX)\vee 1)] \\
&= \int_0^\infty \log(ux)\mathrm{e}^x \int_0^\infty g(x, u\mathrm{e}^z x)\mathrm{d}z \mathrm{e}^{-x}\mathrm{d}x \\
&= \frac{1}{u}\int_0^\infty \int_0^\infty \log(x) g(x/u, \mathrm{e}^z x)\mathrm{d}x\mathrm{d}z \\
&= \frac{1}{u}\int_0^\infty Q\left(\frac{u\mathrm{e}^z}{1+u\mathrm{e}^z}\right)\int_0^\infty \log(x)\exp\left\{-\frac{1+u\mathrm{e}^z}{u}A\left(\frac{u\mathrm{e}^z}{1+u\mathrm{e}^z}\right)x\right\}\mathrm{d}x\mathrm{d}z.
\end{aligned}
$$

The inner integral can be calculated explicitly, yielding

$$
\begin{aligned}
&\mathrm{E}[\log(uX)\log(Y/(uX)\vee 1)] \\
&= \int_0^\infty \frac{Q\left(\frac{u\mathrm{e}^z}{1+u\mathrm{e}^z}\right)}{(1+u\mathrm{e}^z)A\left(\frac{u\mathrm{e}^z}{1+u\mathrm{e}^z}\right)}\left[-\gamma - \log\left\{\frac{1+u\mathrm{e}^z}{u}A\left(\frac{u\mathrm{e}^z}{1+u\mathrm{e}^z}\right)\right\}\right]\mathrm{d}z.
\end{aligned}
$$

Change variables $w = u\mathrm{e}^z/(1+u\mathrm{e}^z)$ to obtain

$$
\begin{aligned}
&\mathrm{E}[\log(uX)\log(Y/(uX)\vee 1)] \\
&= \int_s^1 \frac{Q(w)}{wA(w)}\left[-\gamma - \log\left\{\frac{A(w)}{u(1-w)}\right\}\right]\mathrm{d}w \\
&= \int_s^1 \left(\frac{1}{w} - \frac{A(w)}{A'(w)}\right)\{\log(1-w) + \log(u) - \gamma\}\mathrm{d}w \\
&\quad - \int_s^1 \log A(w)\frac{\mathrm{d}w}{w} + \int_s^1 \frac{A'(w)}{A(w)}\log A(w)\mathrm{d}w.
\end{aligned}
$$

The first integral on the right can be rewritten using partial integration, while the last one can be calculated explicitly, leading to

$$
\begin{aligned}
&\mathrm{E}[\log(uX)\log(Y/(uX)\vee 1)] \\
&= \log(s)\{\log A(s) - \log(s)\} + \int_s^1 \log(w)\frac{\mathrm{d}w}{1-w} \\
&\quad - \int_s^1 \log A(w)\frac{\mathrm{d}w}{w(1-w)} - \frac{1}{2}(\log A(w))^2. \qquad (34)
\end{aligned}
$$

A formula for $\mathrm{E}[\log(Y)\log(Y/(uX)\vee 1)]$ now arises from the combination of equations (32), (33), and (34). Replace u by v and s by t to get a formula for $\mathrm{E}[\log(Y)\log(Y/(vX)\vee 1)]$ as well. This takes care of the two middle terms in (31). The last term in (31) can be dealt with as follows: as $0 < u \le v < \infty$,

$$
\begin{aligned}
&\mathrm{E}[\log(Y/(uX)\vee 1)\log(Y/(vX)\vee 1)] \\
&= \mathrm{E}[\{\log(v/u) + \log(Y/(vx)\vee 1)\}\log(Y/(vX)\vee 1)] \\
&= \log(v/u)\mathrm{E}[\log(Y/(vX)\vee 1)] + \mathrm{E}[\{\log(Y/(vX)\vee 1)\}^2]. \qquad (35)
\end{aligned}
$$

The first term in (35) is

$$\begin{aligned}\mathrm{E}[\log(Y/(vX)\vee 1)] &= \mathrm{E}[\log\{X/(1-t)\vee Y/t\}] - \mathrm{E}[\log\{X/(1-t)\}] \\ &= \mathrm{E}[\log\{Y/t\}] - \mathrm{E}[\log\xi(t)] \\ &= -\log(t) + \log A(t),\end{aligned}$$

while the second term in (35) can be found through (33) with u and s replaced by v and t, respectively.

All the terms in (31) have now been computed. In combination with (29) and (30), this results in a formula for $\mathrm{E}[\log\xi(s)\log\xi(t)]$. As we already knew that $\mathrm{E}[\log\xi(s)] = -\gamma - \log A(s)$, all is in place to compute $\mathrm{Cov}(\log\xi(s), \log\xi(t))$. Aggregating all the obtained expressions leads after some manipulation to the expression in the statement of the lemma.

The cases $s = 0$ and $t = 1$ follow from the case $0 < s \le t < 1$ by the continuity of the function $t \mapsto \log\xi(t)$ and the bounds $X \wedge Y \le \xi(t) \le 2(X \vee Y)$. □

Proof of Corollary 4.3

Analogous to the proof of Corollary 3.2. □

6.4 Proofs for Section 5

Proof of Lemma 5.1

This is a consequence of Lemma 6.1 below applied to $\boldsymbol{X} = (\delta(0), \delta(1))'$ and $\boldsymbol{Y} = \delta(t)$.

Lemma 6.1 *Let $\boldsymbol{Z} = (\boldsymbol{X}', \boldsymbol{Y}')'$ be a $(p+q)$-dimensional random vector with non-singular covariance matrix*

$$\mathrm{Var}(\boldsymbol{Z}) = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix} \in \mathbb{R}^{(p+q)\times(p+q)}.$$

Put $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1} \in \mathbb{R}^{q\times p}$. For arbitrary $\boldsymbol{\gamma} \in \mathbb{R}^{q\times p}$,

$$\mathrm{Var}(\boldsymbol{Y} - \boldsymbol{\gamma}\boldsymbol{X}) = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} + (\boldsymbol{\beta} - \boldsymbol{\gamma})\boldsymbol{\Sigma}_{xx}(\boldsymbol{\beta} - \boldsymbol{\gamma})'.$$

Proof. Assume without loss of generality that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are centered. Then $\mathrm{E}[(\boldsymbol{Y} - \boldsymbol{\beta}\boldsymbol{X})\boldsymbol{X}'] = \boldsymbol{\Sigma}_{yx} - \boldsymbol{\beta}\boldsymbol{\Sigma}_{xx} = \boldsymbol{0}$. Hence for $\boldsymbol{\gamma} \in \mathbb{R}^{q\times p}$,

$$\begin{aligned}\mathrm{Var}(\boldsymbol{Y} - \boldsymbol{\gamma}\boldsymbol{X}) &= \mathrm{E}[\{(\boldsymbol{Y} - \boldsymbol{\beta}\boldsymbol{X}) + (\boldsymbol{\beta} - \boldsymbol{\gamma})\boldsymbol{X}\}\{(\boldsymbol{Y} - \boldsymbol{\beta}\boldsymbol{X})' + \boldsymbol{X}'(\boldsymbol{\beta} - \boldsymbol{\gamma})'\}] \\ &= \mathrm{E}[(\boldsymbol{Y} - \boldsymbol{\beta}\boldsymbol{X})(\boldsymbol{Y} - \boldsymbol{\beta}\boldsymbol{X})'] + (\boldsymbol{\beta} - \boldsymbol{\gamma})\boldsymbol{\Sigma}_{xx}(\boldsymbol{\beta} - \boldsymbol{\gamma})'.\end{aligned}$$

Finally,

$$\begin{aligned}\mathrm{E}[(\boldsymbol{Y} - \boldsymbol{\beta}\boldsymbol{X})(\boldsymbol{Y} - \boldsymbol{\beta}\boldsymbol{X})'] &= \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\beta}' - \boldsymbol{\beta}\boldsymbol{\Sigma}_{xy} + \boldsymbol{\beta}\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta}' \\ &= \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}.\end{aligned}$$

□

Proof of Theorem 5.2

Note that the assumption $A(1/2) > 1/2$ is needed since otherwise the variance-minimizing functions (a_0, b_0) in (19) are not well-defined. We can write

$$\begin{aligned} & n^{1/2}\left(\log \hat{A}_n^{\text{cfg}}(t; \hat{a}_n, \hat{b}_n) - \log \hat{A}_n^{\text{cfg}}(t; a_0, b_0)\right) \\ &= \{a_0(t) - \hat{a}_n(t)\} n^{1/2} \log \hat{A}_n^{\text{cfg}}(0) + \{b_0(t) - \hat{b}_n(t)\} n^{1/2} \log \hat{A}_n^{\text{cfg}}(1). \end{aligned}$$

As $n^{1/2} \log \hat{A}_n^{\text{cfg}}(0) = O_p(1)$ and $n^{1/2} \log \hat{A}_n^{\text{cfg}}(1) = O_p(1)$ by the central limit theorem, it is therefore sufficient to show that $|\hat{a}_n(t) - a_0(t)| = o_p(1)$ and $|\hat{b}_n(t) - b_0(t)| = o_p(1)$ uniformly over $t \in [0,1]$. But since $\hat{a}_n$ and $\hat{b}_n$ are constructed through (19) with σ replaced by $\hat{\sigma}_n$, it is sufficient that $|\hat{\sigma}_n(s,t) - \sigma(s,t)| = o_p(1)$ uniformly over $(s,t) \in [0,1]^2$. If $\hat{\sigma}_n$ arises by replacing A by $\hat{A}_n$ in (18), the result now follows from the stated conditions on $\hat{A}_n$. □

References

Balkema, A.A. and Resnick, S.I. (1977) "Max-infinite divisibility", *Journal of Applied Probability* **14**, 309–319.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004) *Statistics of Extremes: Theory and Applications.* Wiley.

Bickel, P.J., Klaassen, C.A., Ritov, Y., and Wellner, J.A. (1993) *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press.

Billingsley, P. (1995) *Probability and Measure.* Wiley.

Capéraà, P., Fougères, A.L., and Genest, C. (1997) "A nonparametric estimation procedure for bivariate extreme value copulas", *Biometrika* **84**, 567–577.

Deheuvels, P. (1991) "On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions", *Statistics & Probability Letters* **12**, 429–439.

Deheuvels, P. and Tiago de Oliveira, J. (1989) "On the non-parametric estimation of the bivariate extreme-value distributions", *Statistics & Probability Letters* **8**, 315–323.

Garralda Guillem, A.I. (2000) "Structure de dépendance des lois de valeurs extrêmes bivariées", *Comptes Rendues de l'Académie des Sciences de Paris* **330**, 593–596.

Geffroy, J. (1958/59) "Contribution à la théorie des valuers extrêmes", *Publications de l'Institut de Statistique de l'Université de Paris* **7**, 37–121, and **8**, 123–184.

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

Genest, C. and Rivest, L.P. (1989) "A characterization of Gumbel's family of extreme value distributions", *Statistics & Probability Letters* **8**, 207–211.

Ghoudi, K., Khoudraji, A., and Rivest, L.P. (1998) "Propriétés statistiques des copules de valeurs extrêmes bidimensionelles", *Canadian Journal of Statistics* **14**, 145–159.

Gumbel, E.J. (1960a) "Distributions des valeurs extrêmes en plusieurs dimensions", *Publications de l'Institute de Statistique de l'Université de Paris* **9**, 171–173.

Gumbel, E.J. (1960b) "Bivariate exponential distributions", *Journal of the American Statistical Association* **55**, 698–707.

Gumbel, E.J. and Goldstein, N. (1964) "Analysis of empirical bivariate extremal distributions", *Journal of the American Statistical Association* **59**, 794–816.

de Haan, L. and Resnick, S.I. (1977) "Limit theory for multivariate sample extremes", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **40**, 317–337.

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

Hürlimann, W. (2003) "Hutchinson-Lai's conjecture for bivariate extreme value copulas", *Statistics & Probability Letters* **61**, 191–198.

Hutchinson, T.P. and Lai, C.D. (1990) *Continuous bivariate distributions, emphasizing applications.* Rumsby Scientific, Adelaide.

Jiménez, J.R., Villa-Diharce, E. and Flores, M. (2001) "Nonparametric estimation of the dependence function in bivariate extreme value distributions", *Journal of Multivariate Analysis* **76**, 159–191.

Kotz, S. and Nadarajah, S. (2000) *Extreme Value Distributions.* Imperial College Press.

Lehmann, E.L. (1966) "Some concepts of dependence", *The Annals of Mathematical Statistics* **37**, 1137–1153.

Longin, F. and Solnik, B. (2001) "Extreme correlation of international equity markets", *The Journal of Finance* **56**, 649–676

Marshall, A.W. and Olkin, I. (1983) "Domains of attraction of multivariate extreme value distributions", *Annals of Probability* **11**, 168–177.

Obrenetov, A. (1991) "On the dependence function of Sibuya in multivariate extreme value theory", *Journal of Multivariate Analysis* **36**, 35–43.

Owen, A. (1991) "Empirical likelihood for linear models", *Annals of Statistics* **4**, 1725–1747.

Pickands, J. (1981) "Multivariate extreme value distributions", in *Bulletin of the International Statistical Institute, Proceedings of the 43rd Session, Buenos Aires 1981*, pp. 859–878.

Pickands, J. (1989) "Multivariate negative exponential and extreme value distributions", in *Extreme Value Theory: Proceedings, Oberwolfach 1987* (eds. J. Hüsler and R.D. Reiss), pp. 262–274. Lecture Notes in Statistics **51**.

Resnick, S.I. (1987) *Extreme Values, Regular Variation and Point Processes.* Springer-Verlag.

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

Smith, R.L., Tawn, J.A., and Yuen, H.K. (1990) "Statistics of multivariate extremes", *International Statistical Review* **58**, 47–58.

Tawn, J.A. (1988) "Bivariate extreme value theory: models and estimation", *Biometrika* **75**, 397–415.

Tiago de Oliveira, J. (1989) "Intrinsic estimation of the dependence structure for bivariate extremes", *Statistics & Probability Letters* **8**, 213–218.