Bias-corrected estimation of stable tail dependence function

Jan Beirlant[(1)], Mikael Escobar-Bach[(2)], Yuri Goegebeur[(2)] & Armelle Guillou[(3)]

[(1)] Department of Mathematics and LStat, KU Leuven, Celestijnenlaan 200B, 3001 Heverlee, Belgium

[(2)] Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

[(3)] Université de Strasbourg, CNRS, IRMA UMR 7501, F-67000 Strasbourg, France

Abstract

We consider the estimation of the stable tail dependence function. We propose a bias-corrected estimator and we establish its asymptotic behaviour under suitable assumptions. The finite sample performance of the proposed estimator is evaluated by means of an extensive simulation study where a comparison with alternatives from the recent literature is provided.

Keywords: Multivariate extreme value statistics, stable tail dependence function, bias correction.

AMS Subject classification: 62G05, 62G20, 62G32.

1 Introduction and notations

Many problems involving extreme events are inherently multivariate. For instance, de Haan and de Ronde (1998) estimate the probability that a storm will cause a sea wall near the town of Petten (the Netherlands) to collapse because of a dangerous combination of sea level and wave height. Other examples can be found in actuarial science, finance, environmental science and geology, to name but a few. A fundamental question that arises when studying more than one variable is that of extremal dependence. Similarly to classical statistics one can summarise extremal dependency in a number of well-chosen coefficients that give a representative picture of the dependency structure. Here, the prime example of such a dependency measure is the coefficient of tail dependence (Ledford and Tawn, 1997). Alternatively, a full characterization of the extremal dependence between variables can be obtained from functions like e.g. the stable tail dependence function, the spectral distribution function or the Pickands dependence function. We refer to Beirlant *et al.* (2004) and de Haan and Ferreira (2006), and the references therein, for more details. In this paper we will focus on bias-corrected estimation of the stable tail dependence function.

For any arbitrary dimension d, let $(X^{(1)}, ..., X^{(d)})$ be a multivariate vector with continuous marginal cumulative distribution functions (cdfs) $F_1, ..., F_d$. The stable tail dependence function is defined for each $x_i \in \mathbb{R}_+$, $i = 1, ..., d$, as

$$\lim_{t\to\infty} t\mathbb{P}\left(1 - F_1(X^{(1)}) \leqslant t^{-1}x_1 \quad \text{or} \quad ... \quad \text{or} \quad 1 - F_d(X^{(d)}) \leqslant t^{-1}x_d\right) = L(x_1, ..., x_d)$$

which can be rewritten as

$$\lim_{t\to\infty} t\left[1 - F\left(F_1^{-1}(1 - t^{-1}x_1), ..., F_d^{-1}(1 - t^{-1}x_d)\right)\right] = L(x_1, ..., x_d) \tag{1}$$

where F is the multivariate distribution function of the vector $(X^{(1)}, ..., X^{(d)})$.

Now, consider a sample of size n drawn from F and an intermediate sequence $k = k_n$, i.e. $k \to \infty$ as $n \to \infty$ with $k/n \to 0$. Let us denote $\mathbf{x} = (x_1, ..., x_d)$ a vector of the positive quadrant $\mathbb{R}_+^d$ and $X_{k,n}^{(j)}$ the $k-$th order statistic among n realisations of the margins $X^{(j)}$, $j = 1, ..., d$. The empirical estimator of L is then given by

$$\widehat{L}_k(\mathbf{x}) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}_{\{X_i^{(1)} \geqslant X_{n-[kx_1]+1,n}^{(1)} \quad \text{or} \quad ... \quad \text{or} \quad X_i^{(d)} \geqslant X_{n-[kx_d]+1,n}^{(d)}\}}.$$

The asymptotic behaviour of this estimator was first studied by Huang (1992); see also Drees and Huang (1998), and de Haan and Ferreira (2006). As is common in extreme value statistics, the empirical estimator $\widehat{L}_k(\mathbf{x})$ is affected by bias, which often complicates its application in practice. This bias-issue will be addressed in the present paper.

In the univariate framework there are numerous contributions to the bias-corrected estimation of the extreme value index and tail probabilities. Typically, the bias reduction of estimators for tail parameters is obtained by taking the second order structure of an extreme value model explicitly into account in the estimation stage. We refer here to Beirlant *et al.* (1999), Feuerverger and Hall (1999), Matthys and Beirlant (2003), and more recently, Gomes *et al.* (2008) and Caeiro *et al.* (2009) . In the bivariate framework some attention has been paid to bias-corrected estimation of the coefficient of tail dependence η. Goegebeur and Guillou (2013) obtained the bias correction by a properly weighted sum of two biased estimators, whereas Beirlant *et al.* (2011) fitted the extended Pareto distribution to properly transformed bivariate observations. Recently, a robust and bias-corrected estimator for η was introduced by Dutang *et al.* (2014). For what concerns the stable tail dependence function we are only aware of the estimator recently proposed by Fougères *et al.* (2015).

For the sequel, in order to study the behaviour of $\widehat{L}_k(\mathbf{x})$ or a function of it, we need to assume some conditions mentioned below and well-known in the extreme value framework:

First order condition: The limit in (1) exists and the convergence is uniform on $[0,T]^d$ for $T > 0$;

Second order condition: There exist a positive function α such that $\alpha(t) \to 0$ as $t \to \infty$ and a non null function M such that for all $\mathbf{x}$ with positive coordinates

$$\lim_{t\to\infty} \frac{1}{\alpha(t)} \left\{ t \left[1 - F\left(F_1^{-1}(1 - t^{-1}x_1), ..., F_d^{-1}(1 - t^{-1}x_d) \right) \right] - L(\mathbf{x}) \right\} = M(\mathbf{x}),$$

uniformly on $[0,T]^d$ for $T > 0$;

Third order condition: There exist a positive function β such that $\beta(t) \to 0$ as $t \to \infty$ and a non null function N such that for all $\mathbf{x}$ with positive coordinates

$$\lim_{t\to\infty} \frac{1}{\beta(t)} \left\{ \frac{t \left[1 - F\left(F_1^{-1}(1 - t^{-1}x_1), ..., F_d^{-1}(1 - t^{-1}x_d) \right) \right] - L(\mathbf{x})}{\alpha(t)} - M(\mathbf{x}) \right\} = N(\mathbf{x}), \qquad (2)$$

uniformly on $[0,T]^d$ for $T > 0$. This requires that N is not a multiple of M.

Note that these assumptions imply that the functions α and β are both regularly varying with indices ρ and ρ' respectively which are non positive. In the sequel we assume that both indices are negative. Remark also that the functions L, M and N have an homogeneity property, that is $L(a\mathbf{x}) = aL(\mathbf{x})$, $M(a\mathbf{x}) = a^{1-\rho}M(\mathbf{x})$ and $N(a\mathbf{x}) = a^{1-\rho-\rho'}N(\mathbf{x})$ for a positive scale parameter a.

The remainder of the paper is organised as follows. In the next section we introduce our estimators for $L(\mathbf{x})$, as well as for the second order quantities ρ and α, and study their asymptotic properties. The finite sample performance of the proposed bias-corrected estimator and of some estimators from the recent literature are evaluated by a simulation experiment in Section 3. The proofs of all results are given in the Appendix.

2 Estimators and asymptotic properties

Consider now the rescaled version

$$\widehat{L}_{k,a}(\mathbf{x}) := a^{-1}\widehat{L}_k(a\mathbf{x})$$

for a positive scale parameter a. Our first aim is to look at the behaviour of

$$\widetilde{L}_k(\mathbf{x}) := \frac{1}{k}\sum_{j=1}^{k} K(a_j)\widehat{L}_{k,a_j}(\mathbf{x})$$

where $a_j := \frac{j}{k+1}, j = 1, ..., k$, and K is a positive function on $[0,1]$ and such that $\int_0^1 K(u)du = 1$. This function is called a kernel function in the sequel. Let $\mathbf{e}_j$ be a $d-$vector with zeros, except for position j where it is one.

Theorem 1: *Let $\mathbf{X}_1, ..., \mathbf{X}_n$ be independent multivariate random vectors in $\mathbb{R}^d$ with common joint cdf F and continuous marginal cdfs F_j, $j = 1, ..., d$. Assume that the third order condition (2) holds with negative indices ρ and ρ' and that the first order partial derivatives of L, say $\partial_j L$, exist and that $\partial_j L$ is continuous on the set of points $\{\mathbf{x} \in \mathbb{R}_+^d : x_j > 0\}$. Suppose further that the function M is continuously differentiable and N continuous. Let K be a continuous kernel function on $[0,1]$ such that $\exists \varepsilon > 0$ with $\sup_{u\in[0,1]} K(u)u^{-\varepsilon} < +\infty$. Assuming that the* *intermediate sequence k satisfies $\sqrt{k}\alpha(n/k) \to \infty$ and $\sqrt{k}\alpha(n/k)\beta(n/k) \to 0$ as $n \to \infty$, we have*

$$\sqrt{k}\left\{\widetilde{L}_k(\mathbf{x}) - \frac{1}{k}\sum_{j=1}^{k} K(a_j)L(\mathbf{x}) - \alpha\left(\frac{n}{k}\right) M(\mathbf{x})\frac{1}{k}\sum_{j=1}^{k} K(a_j)a_j^{-\rho}\right\} \xrightarrow{d} \int_0^1 K(u)u^{-1}Z_L(u\mathbf{x})du$$

in $D([0,T]^d)$ for every $T > 0$ where $\int_0^1 K(u)u^{-1}Z_L(u\,\cdot)du$ is a centered Gaussian process and

$$Z_L(\mathbf{x}) := W_L(\mathbf{x}) - \sum_{j=1}^{d} W_L(x_j\mathbf{e}_j)\partial_j L(\mathbf{x})$$

with W_L a continuous centered Gaussian process with covariance structure

$$\mathbb{E}[W_L(\mathbf{x})W_L(\mathbf{y})] = \mu\{R(\mathbf{x}) \cap R(\mathbf{y})\}$$

where

$$R(\mathbf{x}) = \{\mathbf{u} \in \mathbb{R}_+^d : \text{there exists } j \text{ such that } 0 \leqslant u_j \leqslant x_j\}$$

and μ is the measure defined as

$$\mu\{R(\mathbf{x})\} := L(\mathbf{x}).$$

From our Theorem 1, we can easily deduce the following corollary which gives the asymptotic behaviour of an uncorrected estimator, $\frac{\widetilde{L}_k(\mathbf{x})}{\frac{1}{k}\sum_{j=1}^k K(a_j)}$, for L.

Corollary 1: *Under the assumptions of Theorem 1, we have*

$$\sqrt{k}\left\{\frac{\widetilde{L}_k(\mathbf{x})}{\frac{1}{k}\sum_{j=1}^k K(a_j)} - L(\mathbf{x}) - \alpha\left(\frac{n}{k}\right) M(\mathbf{x})\frac{\frac{1}{k}\sum_{j=1}^k K(a_j)a_j^{-\rho}}{\frac{1}{k}\sum_{j=1}^k K(a_j)}\right\} \xrightarrow{d} \int_0^1 K(u)u^{-1}Z_L(u\mathbf{x})du$$

in $D([0,T]^d)$ for every $T > 0$.

Now the idea is to remove from $\frac{\widetilde{L}_k(\mathbf{x})}{\frac{1}{k}\sum_{j=1}^k K(a_j)}$ the bias term by estimating the function $\alpha\left(\frac{n}{k}\right)M(\mathbf{x})$ as well as the second order parameter ρ. These quantities will be estimated externally with the same intermediate sequence $\overline{k} = \overline{k}_n$, which is such that $k = o(\overline{k})$. This idea was originally proposed in the univariate framework by Gomes and co-authors (see Caeiro *et al.*, 2009) and has the advantage that the variance of the reduced bias estimator is the same as that of the uncorrected estimator.

First we have to define an estimator for ρ. Similarly to Fougères *et al.* (2015), we propose the following estimator:

$$\widetilde{\rho}_k(\mathbf{x}^*) := \left(1 - \frac{1}{\ln r} \ln \left| \frac{\Delta_{k,a}(r\mathbf{x}^*)}{\Delta_{k,a}(\mathbf{x}^*)} \right| \right) \wedge 0 \tag{3}$$

at a fixed $d-$vector $\mathbf{x}^*$, where $r \in (0,1)$ and

$$\Delta_{k,a}(\mathbf{x}) := a^{-1}\widetilde{L}_k(a\mathbf{x}) - \widetilde{L}_k(\mathbf{x}).$$

Proposition 1. *For any fixed $d-$vector $\mathbf{x}^*$, under the assumptions of Theorem 1*

$$\sqrt{k}\alpha\left(\frac{n}{k}\right)(\widetilde{\rho}_k(\mathbf{x}^*) - \rho) \quad \xrightarrow{d} \quad Z_\rho(\mathbf{x}^*)$$

where

$$Z_\rho(\mathbf{x}^*) := \frac{\int_0^1 K(u)u^{-1}\left[a^{-1}Z_L(au\mathbf{x}^*) - Z_L(u\mathbf{x}^*) - r^{\rho-1}\left(a^{-1}Z_L(aru\mathbf{x}^*) - Z_L(ru\mathbf{x}^*)\right)\right]du}{\ln(r)M(\mathbf{x}^*)\int_0^1 K(u)u^{-\rho}du\left[a^{-\rho} - 1\right]}.$$

Secondly we study the estimation of $\breve{\alpha}_k(\mathbf{x}) := \alpha(n/k)M(\mathbf{x})$. To this aim consider the following regression model with the weighted sum of squared residuals

$$\sum_{j=1}^{k} \widetilde{K}(a_j)\left\{\widehat{L}_{k,a_j}(x) - L(x) - \breve{\alpha}_k(x)a_j^{-\rho}\right\}^2 \tag{4}$$

where $\widetilde{K}$ is another kernel function with similar properties as K.

Straightforward application of least squares estimation to (4), where ρ is fixed at $\widetilde{\rho}_k(\mathbf{x}^*)$ given in (3), leads to

$$\widetilde{\alpha}_k(\mathbf{x}) := \frac{\sum_{j=1}^k \sum_{\ell=1}^k \widetilde{K}(a_j)\widetilde{K}(a_\ell)\left(a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)} - a_\ell^{-\widetilde{\rho}_k(\mathbf{x}^*)}\right)\widehat{L}_{k,a_j}(\mathbf{x})}{\sum_{j=1}^k \sum_{\ell=1}^k \widetilde{K}(a_j)\widetilde{K}(a_\ell)\left(a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)} - a_\ell^{-\widetilde{\rho}_k(\mathbf{x}^*)}\right)a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)}}. \tag{5}$$

The aim of the next proposition is to establish the asymptotic behaviour of this estimator.

Proposition 2. *For any fixed $d-$vector $\mathbf{x}^*$, under the assumptions of Theorem 1, we have for a kernel function $\widetilde{K}$ satisfying the same assumptions as K*

$$\sqrt{k}\left(\widetilde{\alpha}_k(\mathbf{x}) - \alpha\left(\frac{n}{k}\right) M(\mathbf{x})\right) \xrightarrow{d} Z_{\alpha,\mathbf{x}^*}(\mathbf{x})$$

in $D([0,T]^d)$ for every $T > 0$, where

$$\begin{aligned} Z_{\alpha,\mathbf{x}^*}(\mathbf{x}) &:= Z_\rho(\mathbf{x}^*)M(\mathbf{x})\frac{\int_0^1 \widetilde{K}(u)u^{-2\rho}\ln(u)du - \int_0^1 \widetilde{K}(u)u^{-\rho}\ln(u)du \int_0^1 \widetilde{K}(u)u^{-\rho}du}{\int_0^1 \widetilde{K}(u)u^{-2\rho}du - \left(\int_0^1 \widetilde{K}(u)u^{-\rho}du\right)^2} \\ &+ \frac{\int_0^1 \widetilde{K}(u)u^{-1}\left[u^{-\rho} - \int_0^1 \widetilde{K}(v)v^{-\rho}dv\right] Z_L(u\mathbf{x})du}{\int_0^1 \widetilde{K}(u)u^{-2\rho}du - \left(\int_0^1 \widetilde{K}(u)u^{-\rho}du\right)^2}. \end{aligned}$$

We have now all the ingredients to study the behaviour of our bias-corrected estimator for L, namely

$$\overline{L}_{k,\overline{k}}(\mathbf{x}) := \frac{\widetilde{L}_k(\mathbf{x}) - \left(\frac{k}{\overline{k}}\right)^{\widetilde{\rho}_{\overline{k}}(\mathbf{x}^*)} \widetilde{\alpha}_{\overline{k}}(\mathbf{x})\frac{1}{k}\sum_{j=1}^k K(a_j)a_j^{-\widetilde{\rho}_{\overline{k}}(\mathbf{x}^*)}}{\frac{1}{k}\sum_{j=1}^k K(a_j)}.$$

This leads to our Theorem 2.

Theorem 2: *Under the assumptions of Proposition 2, satisfied for two intermediate sequences k and $\bar{k}$ such that $k = o(\bar{k})$ we have*

$$\sqrt{k}\left(\overline{L}_{k,\overline{k}}(\mathbf{x}) - L(\mathbf{x})\right) \xrightarrow{d} \int_0^1 K(u)u^{-1}Z_L(u\mathbf{x})du$$

in $D([0,T]^d)$ for every $T > 0$.

Note that the limiting process is independent of the value of ρ, and that the bias-corrected estimator has the same asymptotic variance as the uncorrected estimator of Corollary 1.

A problem of interest could be now to minimize the variance in our Theorem 2. To this aim, we illustrate in Corollary 2 that if we take a power kernel, that is a kernel of the form $K(t) = (\tau+1)\,t^\tau \mathbb{1}_{\{t\in[0,1]\}}$ with $\tau > 0$, the variance of our bias-corrected estimator $\overline{L}_{k,\overline{k}}(\mathbf{x})$ is converging as $\tau \to +\infty$, and it can reach the variance of the empirical estimator $\widehat{L}_k(\mathbf{x})$.

Corollary 2: *Under the assumptions of Theorem 2, for the power kernel* $K(t) = (\tau+1)\, t^{\tau} \mathbb{1}_{\{t\in[0,1]\}}$ *with* $\tau > 0$*, we have for any* $\mathbf{x} \in [0,T]^d$ *with* $T > 0$

$$Var\left(\int_0^1 K(u)u^{-1} Z_L(u\mathbf{x})du\right) \longrightarrow Var\left(Z_L(\mathbf{x})\right)$$

as $\tau \to +\infty$.

3 Simulation experiment

In order to evaluate the finite sample behaviour of our estimator $\overline{L}_{k,\overline{k}}(\mathbf{x})$, we perform a simulation study and we compare our estimator with the empirical one, $\widehat{L}_k(\mathbf{x})$, and two bias-corrected estimators recently proposed by Fougères *et al.* (2015), defined as follows

$$\begin{aligned} \dot{L}_{k,a,\overline{k}}(\mathbf{x}) &:= \widehat{L}_{k,a}(\mathbf{x}) - \widehat{\Delta}_{k,(a^{-\widehat{\rho}_{\overline{k}}(\mathbf{x}^*)}+1)^{-1/\widehat{\rho}_{\overline{k}}(\mathbf{x}^*)}}(\mathbf{x}) \\ \widetilde{L}_{k,a,\overline{k}}(\mathbf{x}) &:= \frac{\widehat{L}_k(\mathbf{x})\widehat{\Delta}_{\overline{k},a}(a\mathbf{x}) - \widehat{L}_k(a\mathbf{x})\widehat{\Delta}_{\overline{k},a}(\mathbf{x})}{\widehat{\Delta}_{\overline{k},a}(a\mathbf{x}) - a\widehat{\Delta}_{\overline{k},a}(\mathbf{x})} \end{aligned}$$

where $\widehat{\Delta}_{k,a}(\mathbf{x})$ is defined similarly as $\Delta_{k,a}(\mathbf{x})$ but based on the empirical estimator, that is

$$\widehat{\Delta}_{k,a}(\mathbf{x}) := a^{-1}\widehat{L}_k(a\mathbf{x}) - \widehat{L}_k(\mathbf{x})$$

and $\widehat{\rho}_k(\mathbf{x}^*)$ similarly as $\widetilde{\rho}_k(\mathbf{x}^*)$ but based on $\widehat{\Delta}_{k,a}(\mathbf{x})$:

$$\widehat{\rho}_k(\mathbf{x}^*) := \left(1 - \frac{1}{\ln r}\ln\left|\frac{\widehat{\Delta}_{k,a}(r\mathbf{x}^*)}{\widehat{\Delta}_{k,a}(\mathbf{x}^*)}\right|\right) \wedge 0. \tag{6}$$

For simplicity, we focus on $\mathbb{R}^2$ and using the homogeneity property, we consider only the estimation of $L(t, 1-t)$ for $t \in (0,1)$, corresponding to Pickands dependence function, and the same distributions as in Fougères *et al.* (2015), namely

- the Cauchy distribution, for which $L(x,y) = (x^2+y^2)^{1/2}$;
- the Student(ν) distribution, for which

$$L(x,y) = y\,F_{\nu+1}\left(\frac{(y/x)^{1/\nu}-\theta}{\sqrt{1-\theta^2}}\sqrt{\nu+1}\right) + x\,F_{\nu+1}\left(\frac{(x/y)^{1/\nu}-\theta}{\sqrt{1-\theta^2}}\sqrt{\nu+1}\right)$$

where $F_{\nu+1}$ is the cdf of the univariate Student distribution with $\nu+1$ degrees of freedom and θ is the Pearson correlation coefficient. We set $\theta = 0.5$ and $\nu = 2$;

• the bivariate Pareto of type II model, for which $L(x,y) = x + y - (x^{-p} + y^{-p})^{-1/p}$. We set $p = 3$ and we called this model BPII(3);
• the Symmetric logistic model, for which $L(x,y) = (x^{1/s} + y^{1/s})^s$. We set $s = 1/3$;
• the Archimax model with the logistic generator $L(x,y) = (x^2 + y^2)^{1/2}$ and with the mixed generator $L(x,y) = (x^2 + y^2 + xy)/(x + y)$.

For each distribution, we simulate 1000 samples of size 1000. As recommended by Fougères *et al.* (2015), we use the values $a = r = 0.4$, $\overline{k} = 990$, and we take $\mathbf{x}^* = \mathbf{x}$. We also need to choose the two kernels K and $\widetilde{K}$ such that $\sup_{u\in[0,1]} K(u)u^{-\varepsilon} < +\infty$ and $\sup_{u\in[0,1]} \widetilde{K}(u)u^{-\widetilde{\varepsilon}} < +\infty$ for some positive ε and $\widetilde{\varepsilon}$. In our simulations, we use the power kernel functions introduced in Corollary 2

$$K(u) := (1+\tau)u^{\tau}\mathbb{1}_{\{u\in[0,1]\}} \text{ and } \widetilde{K}(u) := (1+\widetilde{\tau})u^{\widetilde{\tau}}\mathbb{1}_{\{u\in[0,1]\}}.$$

Concerning the choice of $\widetilde{\tau}$, an extensive simulation study outlines that the closer $\widetilde{\tau}$ is to 0, the best are the results. Thus we use $\widetilde{\tau} = 10^{-6}$ overall in the paper. Since any stable tail dependence function satisfies $\max(t, 1-t) \leqslant L(t, 1-t) \leqslant 1$, all the estimators have been corrected so that they satisfy these bounds.

First, in Figure 1 we give the boxplots of $\widehat{\rho}_{\overline{k}}(0.5, 0.5)$ and $\widetilde{\rho}_{\overline{k}}(0.5, 0.5)$ for a power kernel with $\tau = 1, 5$ and 10 in case of a Student(2) distribution. It is clear from these boxplots that the estimators for ρ perform reasonably well with respect to the median but with some volatility. However, as we will see in the other figures, these uncertainties seem to have fortunately no impact on the performance of our estimators for L. Note also that our estimator $\widetilde{\rho}_{\overline{k}}(\mathbf{x}^*)$ is at least as competitive as the one proposed by Fougères *et al.* (2015), whatever the value of $\tau > 0$. Moreover, to avoid problems in the computation of $\dot{L}_{k,a,\overline{k}}(\mathbf{x})$ due to the fact that $\widehat{\rho}_{\overline{k}}(\mathbf{x}^*)$ can be too close to 0, we set $\widehat{\rho}_{\overline{k}}(\mathbf{x}^*) = -1$ if $\widehat{\rho}_{\overline{k}}(\mathbf{x}^*) \in [-0.1, 0]$ and its definition given in (6) otherwise. Similarly, for consistency, this modification has been used for $\widetilde{\rho}_{\overline{k}}(\mathbf{x}^*)$.

Next, we examine the performance of the estimators for $L(\mathbf{x})$ at specific points in $\mathbb{R}_+^2$. We also include the case $\tau = 0$ although it is not allowed by our assumptions on K to see the impact on the performances of our estimator. In Figures 2 and 3 we show the sample mean (left) and the empirical mean squared error (MSE, right) of $\overline{L}_{k,\overline{k}}(\mathbf{x})$ with $\tau = 0$ (dashed line), $\tau = 5$ (dotted line) and $\tau = 10$ (full line) as a function of k, for $\mathbf{x} = (0.5, 0.5)$ and $\mathbf{x} = (0.2, 0.8)$, respectively, on three of the above mentioned distributions for brevity. As is clear from these figures, our estimator performs quite well for $\tau > 0$, with a nice stable behaviour close to the true value of $L(\mathbf{x})$. This can be expected since our estimators are bias-corrected. On the contrary, for $\tau = 0$, the bias correction is more arduous with a loss of effectiveness in terms of MSE. In addition, it appears that the best performance is at both $\mathbf{x}$ positions obtained for $\tau = 10$ in terms of MSE, and therefore in subsequent comparisons of estimators we will focus on this choice for our estimator. It is noteworthy to observe that the MSE has a very low curvature, so the MSE-values are slightly larger than the minimum of the MSE for a very wide range of values for k. In Figures 4 and 5 we compare the performance of $\overline{L}_{k,\overline{k}}(\mathbf{x})$ (red full line) with the two estimators proposed by Fougères *et al.* (2015), $\dot{L}_{k,a,\overline{k}}(\mathbf{x})$ (blue dotted line) and $\widetilde{L}_{k,a,\overline{k}}(\mathbf{x})$ (purple dashed line), and the empirical one, $\widehat{L}_k(\mathbf{x})$ (black dash-dotted line), for $\mathbf{x} = (0.5, 0.5)$ and $\mathbf{x} = (0.2, 0.8)$, respectively. From the plots of the sample means we can clearly see the bias-correcting effect of $\overline{L}_{k,\overline{k}}(\mathbf{x})$ and $\widetilde{L}_{k,a,\overline{k}}(\mathbf{x})$: compared to the empirical estimator these estimators have a stable sample path that is close to the true value of $L(\mathbf{x})$ and this for a wide range of values of k. The estimator $\dot{L}_{k,a,\overline{k}}(\mathbf{x})$ has some bias-correcting effect, though the gain relative to the empirical estimator is quite distribution dependent. Unlike $\widetilde{L}_{k,a,\overline{k}}(\mathbf{x})$, the estimator $\overline{L}_{k,\overline{k}}(\mathbf{x})$ has a very smooth sample path, which can be expected as we take essentially a weighted sum of the empirical estimator calculated over different values of $\mathbf{x}$. In terms of MSE, $\overline{L}_{k,\overline{k}}(\mathbf{x})$ has a better performance than $\dot{L}_{k,a,\overline{k}}(\mathbf{x})$ and $\widetilde{L}_{k,a,\overline{k}}(\mathbf{x})$. Compared to the empirical estimator, one can see that except for the estimation of $L(0.5, 0.5)$ in case of the Student distribution, $\overline{L}_{k,\overline{k}}(\mathbf{x})$ has an MSE value that almost reaches the minimum MSE for $\widehat{L}_k(\mathbf{x})$, but it has this low MSE over a very wide range of k-values.

Finally, we compare the different estimators for L not at a single point but for the whole function, using the absolute bias and MSE, computed over $N = 1000$ replications as follows:

$$\begin{aligned}\text{Abias}(k) &:= \frac{1}{9}\sum_{t=1}^{9}\left|\frac{1}{N}\sum_{i=1}^{N}\widehat{L}_k^{(i)}(\mathbf{x}_t) - L(\mathbf{x}_t)\right| \\ \text{MSE}(k) &:= \frac{1}{9N}\sum_{t=1}^{9}\sum_{i=1}^{N}\left(\widehat{L}_k^{(i)}(\mathbf{x}_t) - L(\mathbf{x}_t)\right)^2\end{aligned}$$

where $\{\mathbf{x}_t := \left(\frac{t}{10}, 1 - \frac{t}{10}\right);\ t = 1, ..., 9\}$ and $\widehat{L}_k^{(i)}$ is the estimate based on the $i-$th sample. We naturally replace $\widehat{L}_k^{(i)}$ by $\overline{L}_{k,\overline{k}}^{(i)}, \dot{L}_{k,a,\overline{k}}^{(i)}$ and $\widetilde{L}_{k,a,\overline{k}}^{(i)}$. In Figure 6 we show the Abias(k) of the estimators under consideration as a function of k for the six distributions mentioned above. Figure 7 displays MSE(k). These global performance measures lead us to the same conclusions as the aforementioned study at specific points. Because of the bias correction, the estimators $\overline{L}_{k,\overline{k}}$ and $\widetilde{L}_{k,a,\overline{k}}$ keep the bias low for a wide range of k-values. Overall, the estimator $\overline{L}_{k,\overline{k}}$ has a more smooth sample path than $\dot{L}_{k,a,\overline{k}}$ and $\widetilde{L}_{k,a,\overline{k}}$, and therefore it also outperforms the estimators of Fougères *et al.* (2015) in terms of minimal MSE.

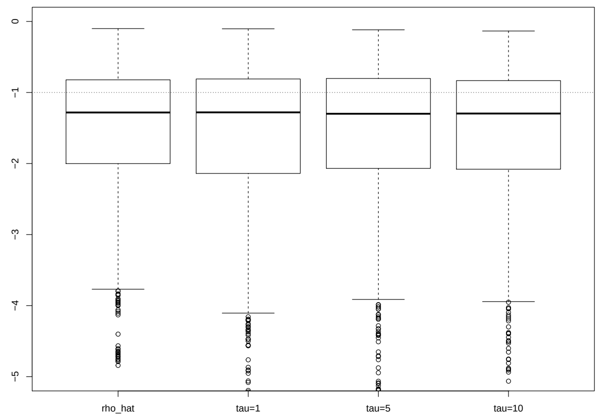

Figure 1: Boxplots of $\widehat{\rho}_{\overline{k}}(0.5, 0.5)$ and $\widetilde{\rho}_{\overline{k}}(0.5, 0.5)$ for a power kernel with $\tau = 1, 5$ and 10 in case of a Student(2) distribution.

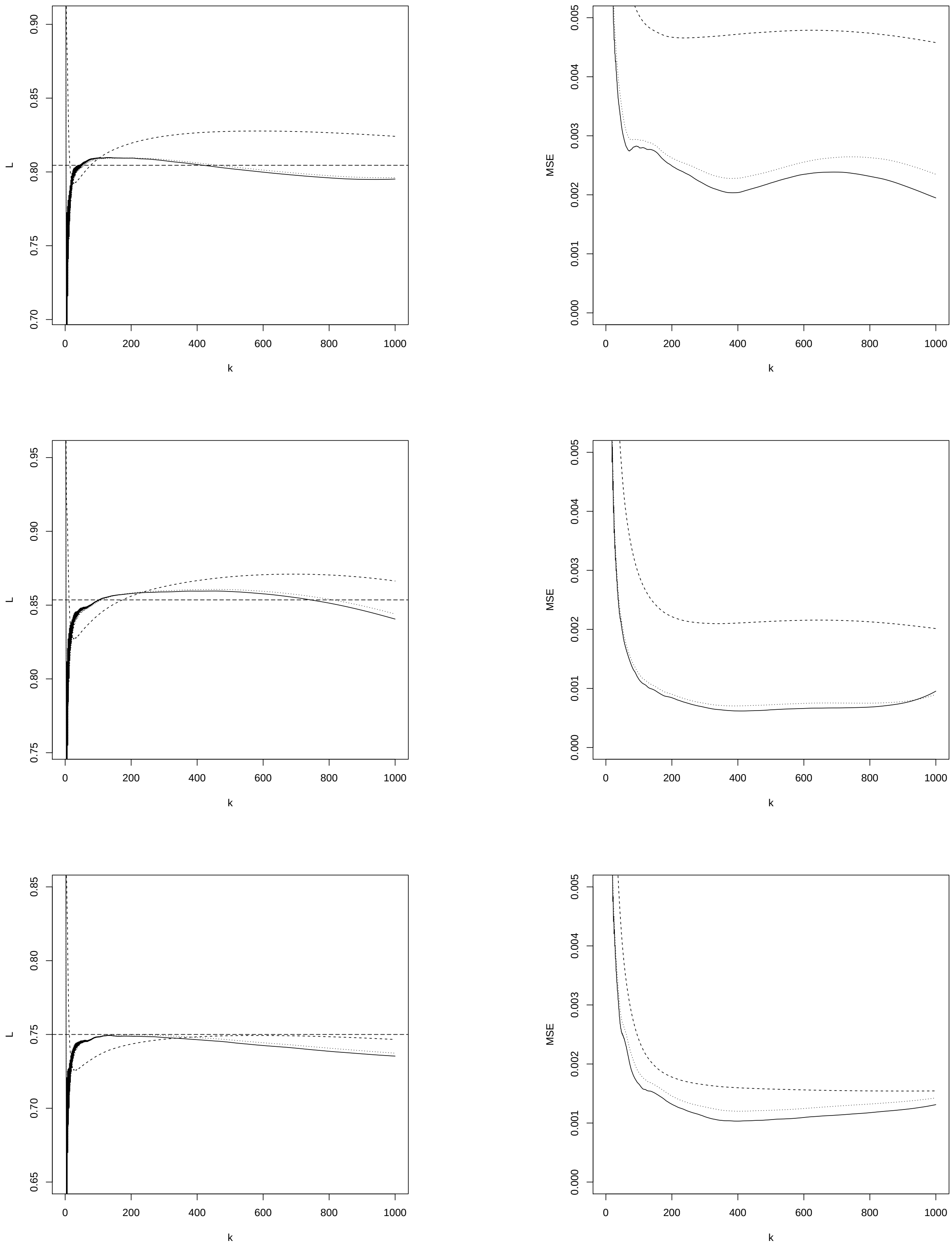

Figure 2: Mean (left) and MSE (right) of our estimator $\overline{L}_{k,\bar{k}}(0.5, 0.5)$ for different values of τ: $\tau = 0$ (dashed line), $\tau = 5$ (dotted line), $\tau = 10$ (full line). Three distributions have been considered: First line: Student(2); Second line: Cauchy; Third line: BPII(3).

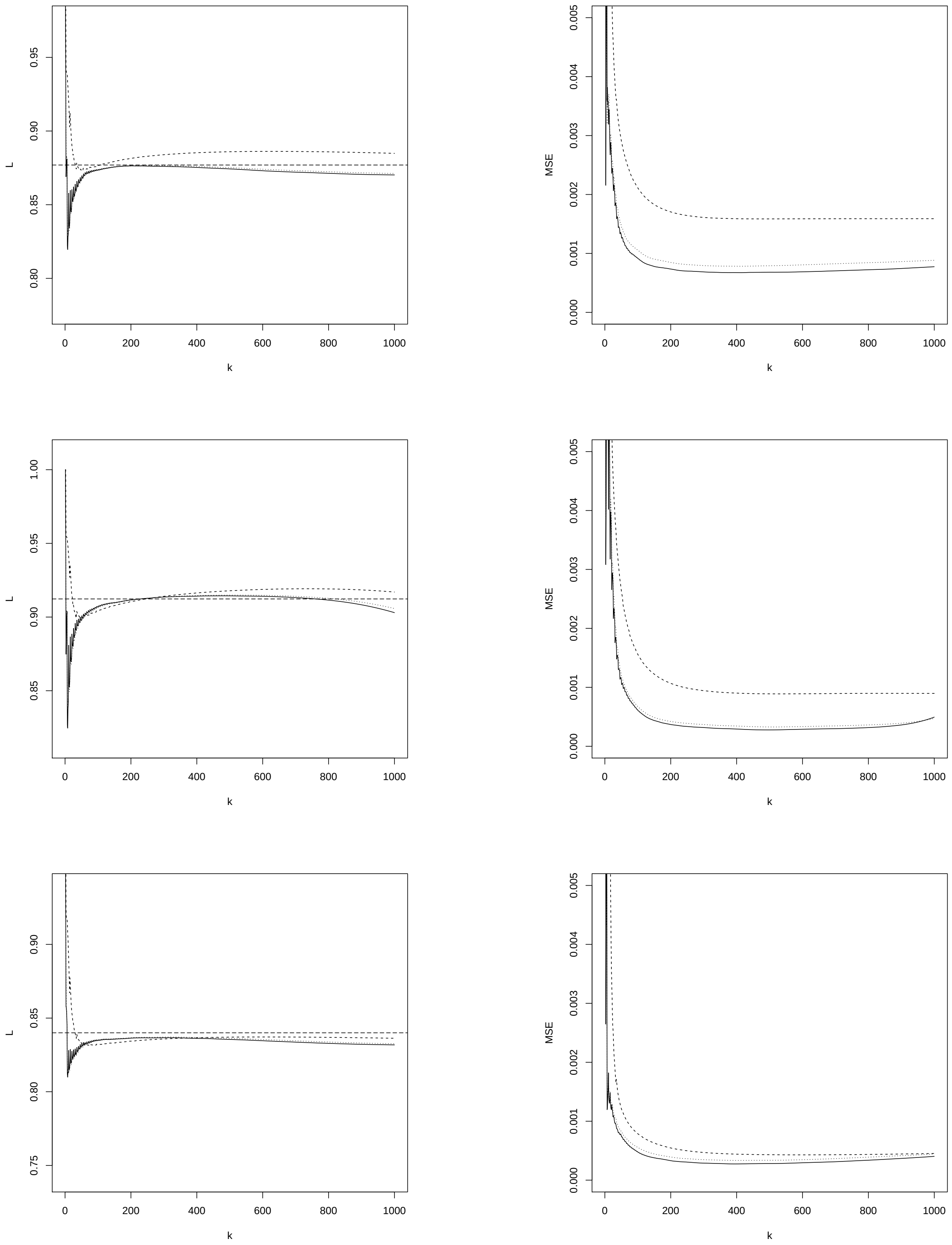

Figure 3: Mean (left) and MSE (right) of our estimator $\overline{L}_{k,\overline{k}}(0.2, 0.8)$ for different values of τ: $\tau = 0$ (dashed line), $\tau = 5$ (dotted line), $\tau = 10$ (full line). Three distributions have been considered: First line: Student(2); Second line: Cauchy; Third line: BPII(3).

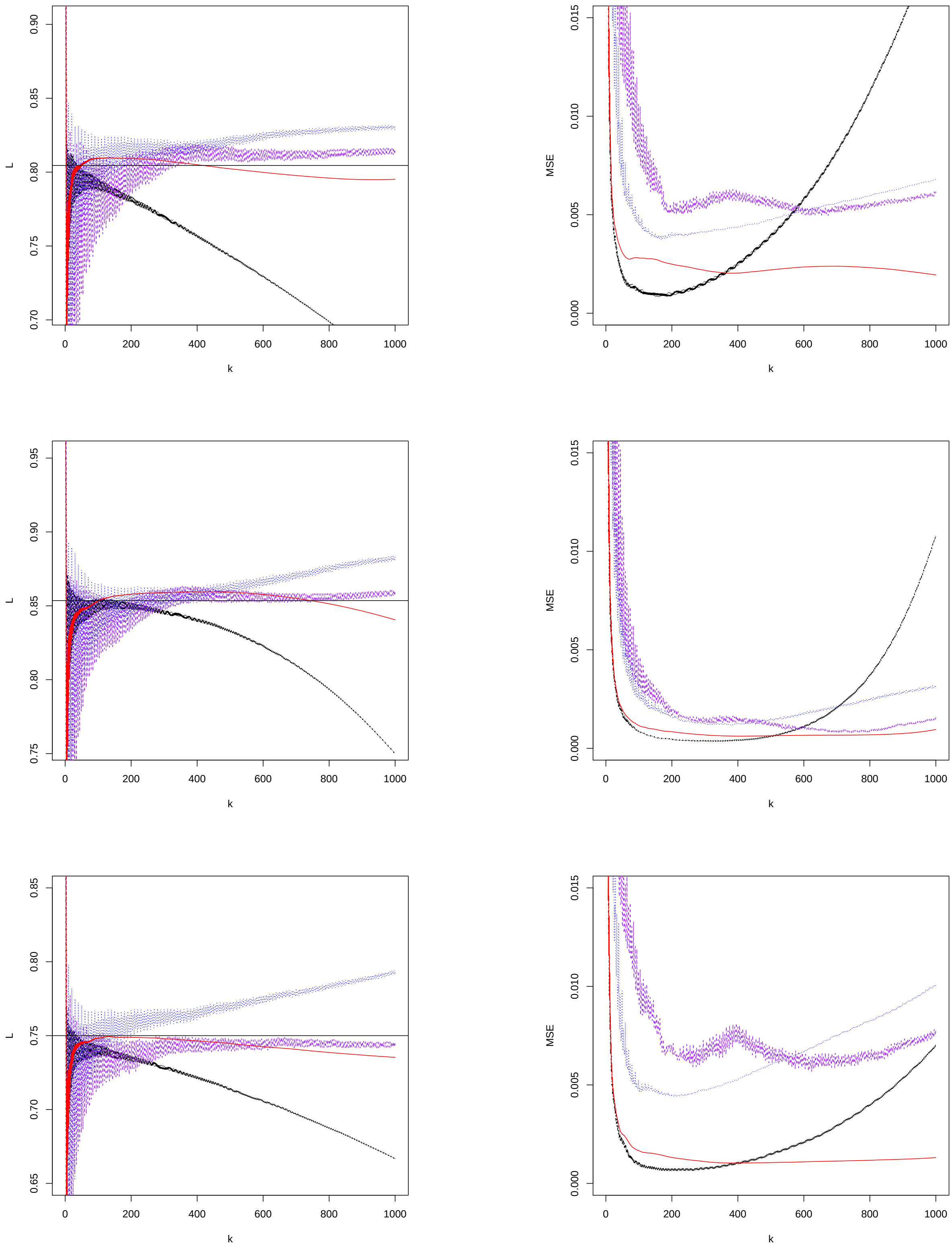

Figure 4: Mean (left) and MSE (right) of our estimator $\overline{L}_{k,\overline{k}}$ at $\mathbf{x} = (0.5, 0.5)$ with $\tau = 10$ (red full line) compared to the estimators of Fougères *et al.* (2015): $\dot{L}_{k,a,\overline{k}}$ (blue dotted line) and $\widetilde{L}_{k,a,\overline{k}}$ (purple dashed line) and the empirical estimator $\widehat{L}_k$ (black dash-dotted line). Three distributions have been considered: First line: Student(2); Second line: Cauchy; Third line: BPII(3).

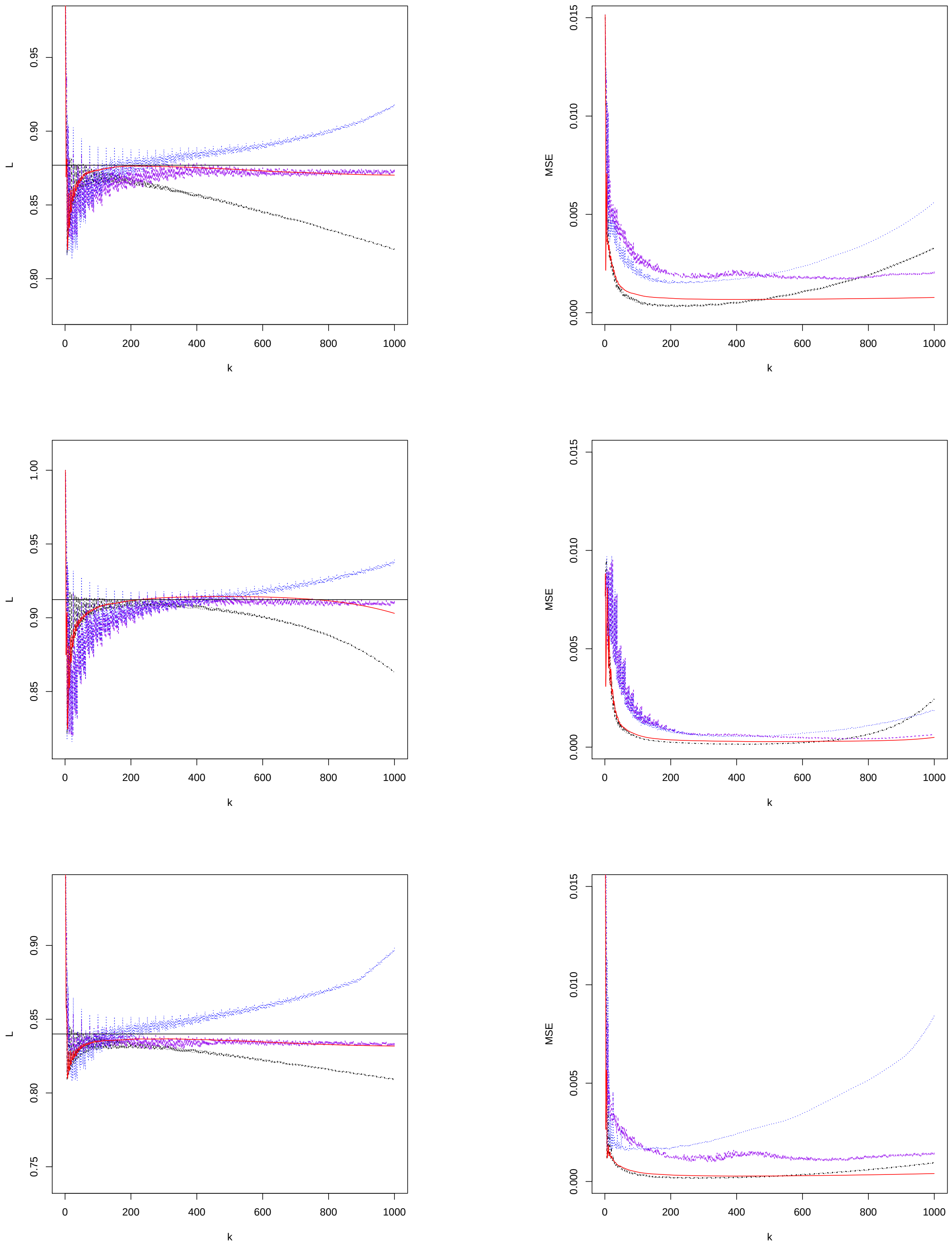

Figure 5: Mean (left) and MSE (right) of our estimator $\overline{L}_{k,\overline{k}}$ at $\mathbf{x} = (0.2, 0.8)$ with $\tau = 10$ (red full line) compared to the estimators of Fougères *et al.* (2015): $\dot{L}_{k,a,\overline{k}}$ (blue dotted line) and $\widetilde{L}_{k,a,\overline{k}}$ (purple dashed line) and the empirical estimator $\widehat{L}_k$ (black dash-dotted line). Three distributions have been considered: First line: Student(2); Second line: Cauchy; Third line: BPII(3).

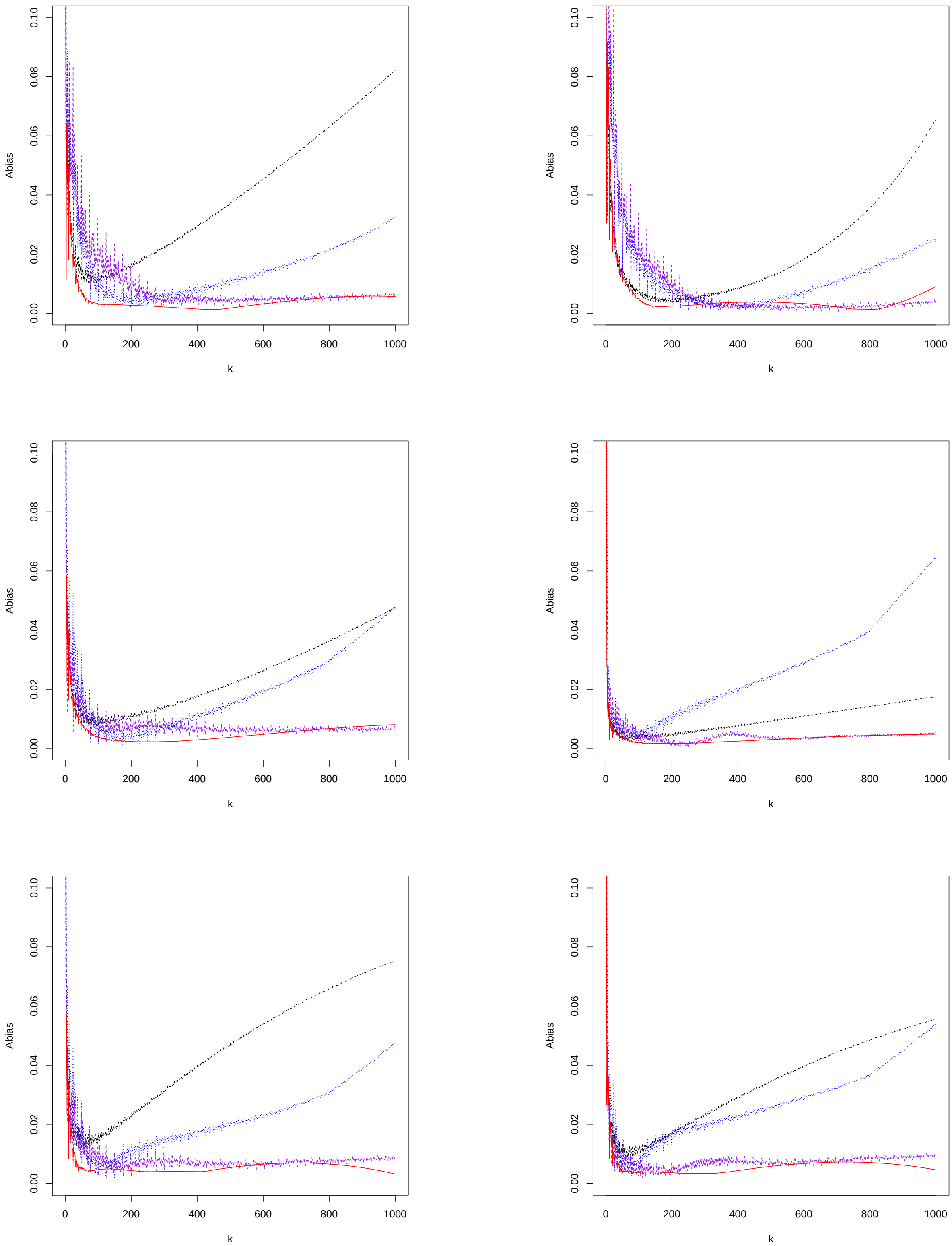

Figure 6: Absolute bias of our estimator $\overline{L}_{k,\overline{k}}$ with $\tau = 10$ (red full line) compared to the estimators of Fougères *et al.* (2015): $\dot{L}_{k,a,\overline{k}}$ (blue dotted line) and $\widetilde{L}_{k,a,\overline{k}}$ (purple dashed line) and the empirical estimator $\widehat{L}_k$ (black dash-dotted line). Six distributions have been considered: First line: Student(2) (left), Cauchy (right); Second line: BPII(3) (left), Symmetric logistic (right); Third line: Archimax mixed (left), Archimax logistic (right).

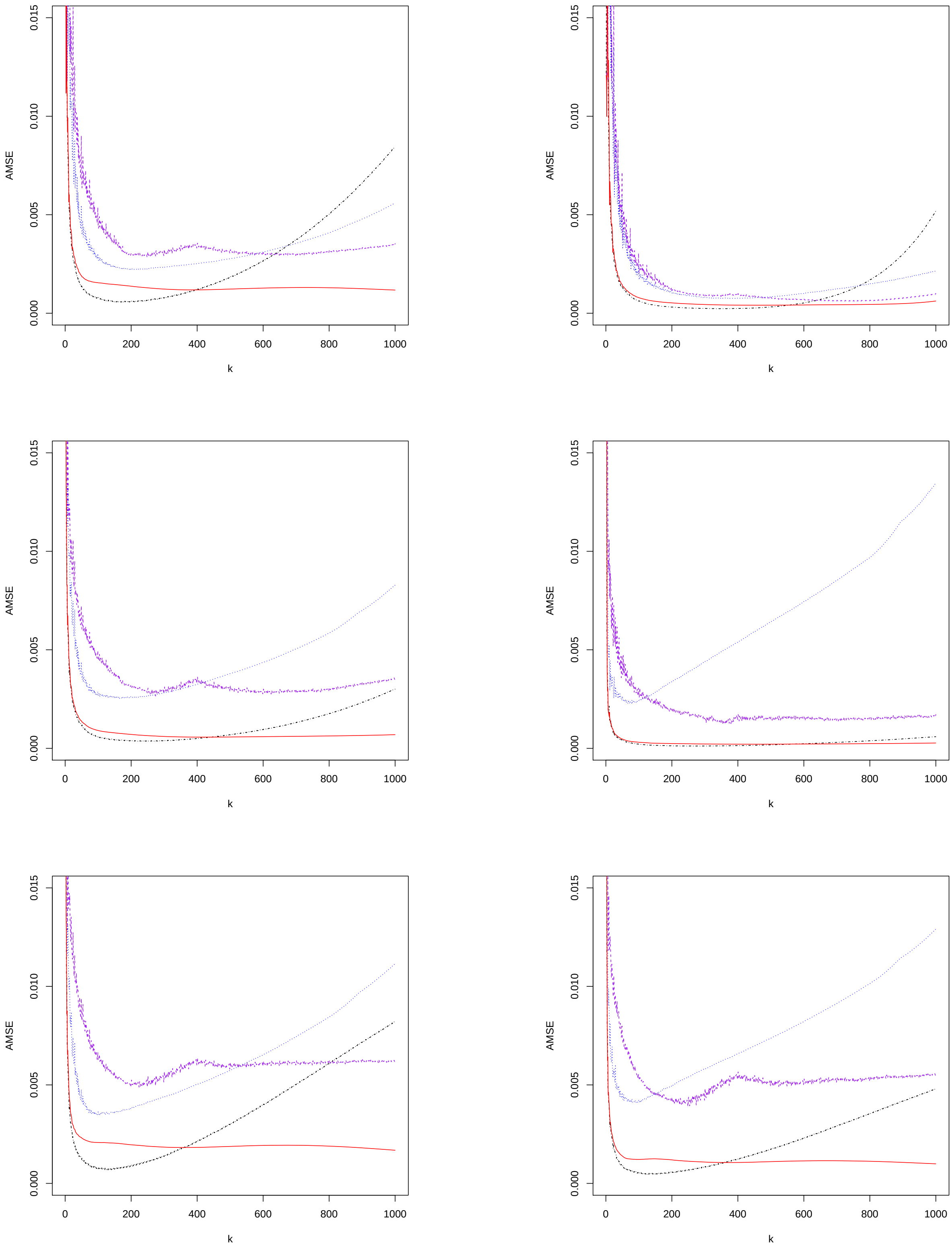

Figure 7: MSE of our estimator $\overline{L}_{k,\overline{k}}$ with $\tau = 10$ (red full line) compared to the estimators of Fougères *et al.* (2015): $\dot{L}_{k,a,\overline{k}}$ (blue dotted line) and $\widetilde{L}_{k,a,\overline{k}}$ (purple dashed line) and the empirical estimator $\widehat{L}_k$ (black dash-dotted line). Six distributions have been considered: First line: Student(2) (left), Cauchy (right); Second line: BPII(3) (left), Symmetric logistic (right); Third line: Archimax mixed (left), Archimax logistic (right).

4 Appendix: Proofs

4.1 Proof of Theorem 1

Using the Skorohod construction combined with the homogeneity properties of L and M, for any $j = 1, \ldots, k$ we have

$$\widehat{L}_{k,a_j}(\mathbf{x}) = L(\mathbf{x}) + \alpha\left(\frac{n}{k}\right) M(\mathbf{x}) a_j^{-\rho} + \frac{1}{\sqrt{k}}\left(a_j^{-1} Z_L(a_j\mathbf{x}) + a_j^{-1} o(1)\right)$$

where the error term is almost surely and uniform in j and $\mathbf{x}$. This leads to

$$\begin{aligned}\widetilde{L}_k(\mathbf{x}) &= L(x)\frac{1}{k}\sum_{j=1}^{k} K(a_j) + \alpha\left(\frac{n}{k}\right) M(\mathbf{x})\frac{1}{k}\sum_{j=1}^{k} K(a_j) a_j^{-\rho} \\ &\quad + \frac{1}{\sqrt{k}}\left(\frac{1}{k}\sum_{j=1}^{k} K(a_j) a_j^{-1} Z_L(a_j\mathbf{x}) + o(1)\frac{1}{k}\sum_{j=1}^{k} K(a_j) a_j^{-1}\right).\end{aligned}$$

Thus we have to show that

$$\sup_{\mathbf{x}\in[0,T]^d}\left|\frac{1}{k}\sum_{j=1}^{k} K(a_j) a_j^{-1} Z_L(a_j\mathbf{x}) - \int_0^1 K(u) u^{-1} Z_L(u\mathbf{x}) du\right| \to 0 \text{ a.s} \tag{7}$$

and

$$\frac{1}{k}\sum_{j=1}^{k} K(a_j) a_j^{-1} \to \int_0^1 K(u) u^{-1} du. \tag{8}$$

One can rewrite the weighted sum in (7) as

$$\frac{1}{k}\sum_{j=1}^{k} K(a_j) a_j^{-1} Z_L(a_j\mathbf{x}) = \int_0^1 K\left(\frac{\lceil ku\rceil}{k+1}\right)\left(\frac{\lceil ku\rceil}{k+1}\right)^{-1} Z_L\left(\frac{\lceil ku\rceil}{k+1}\mathbf{x}\right) du.$$

By continuity of Z_L, we have $C_Z := \sup_{\mathbf{x}\in[0,T]^d} |Z_L(\mathbf{x})| < +\infty$ almost surely. This leads to

$$\begin{aligned}&\left|\int_0^1 K\left(\frac{\lceil ku\rceil}{k+1}\right)\left(\frac{\lceil ku\rceil}{k+1}\right)^{-1} Z_L\left(\frac{\lceil ku\rceil}{k+1}\mathbf{x}\right) du - \int_0^1 K(u) u^{-1} Z_L(u\mathbf{x}) du\right| \\ \leqslant\ & C_Z \int_0^1 \left|K\left(\frac{\lceil ku\rceil}{k+1}\right)\left(\frac{\lceil ku\rceil}{k+1}\right)^{-1} - K(u) u^{-1}\right| du + \int_0^1 K(u) u^{-1}\left|Z_L\left(\frac{\lceil ku\rceil}{k+1}\mathbf{x}\right) - Z_L(u\mathbf{x})\right| du \text{ a.s} \\ =:\ & I_1 + I_2(\mathbf{x}).\end{aligned}$$

We start to study I_1. Note that $I_1 \to 0$ as $k \to +\infty$ means that (8) holds. To show this convergence, we will apply the dominated convergence theorem. Denoting $C_K := \sup_{u\in[0,1]} K(u) u^{-\varepsilon}$, we have

$$K\left(\frac{\lceil ku\rceil}{k+1}\right)\left(\frac{\lceil ku\rceil}{k+1}\right)^{-1} \leqslant C_K\left(\frac{\lceil ku\rceil}{k+1}\right)^{-1+\varepsilon}.$$

In case $\varepsilon \geqslant 1$, the right-hand side of the inequality is bounded whereas if $\varepsilon \in (0,1)$, one can find a universal constant $C > 0$ such that

$$\left(\frac{\lceil ku\rceil}{k+1}\right)^{-1+\varepsilon} \leqslant Cu^{-1+\epsilon}$$

which is integrable on $[0,1]$. Thus, by the dominated convergence theorem, we have $I_1 \to 0$ as $k \to +\infty$. Next, $\mathbf{x} \to Z_L(\mathbf{x})$ is almost surely uniformly continuous over $[0,T]^d$. In particular, this implies that

$$\sup_{u\in[0,1],\mathbf{x}\in[0,T]^d}\left|Z_L\left(\frac{\lceil ku\rceil}{k+1}\mathbf{x}\right) - Z_L(u\mathbf{x})\right| \to 0 \text{ a.s.}$$

Using again the assumptions on K, $\sup_{\mathbf{x}\in[0,T]^d} I_2(\mathbf{x}) \to 0$ almost surely and thus (7) is established.

It remains to show that $\int_0^1 K(u)u^{-1}Z_L(u\cdot)du$ is a Gaussian process, i.e that all finite dimensional distributions are Gaussian. To this aim, for any $\{\mathbf{x}_i\}_{1\leqslant i\leqslant \ell} \subset [0,T]^d$, according to (7) we have

$$\begin{aligned}\begin{pmatrix}\int_0^1 K(u)u^{-1}Z_L(u\mathbf{x}_1)du\\ \vdots \\ \int_0^1 K(u)u^{-1}Z_L(u\mathbf{x}_\ell)du\end{pmatrix} &= \begin{pmatrix}\frac{1}{k}\sum_{j=1}^k K(a_j)a_j^{-1}Z_L(a_j\mathbf{x}_1)\\ \vdots \\ \frac{1}{k}\sum_{j=1}^k K(a_j)a_j^{-1}Z_L(a_j\mathbf{x}_\ell)\end{pmatrix} + \underline{0}\\ &=: \quad V_k + \underline{0}\end{aligned}$$

where $\underline{0} \in \mathbb{R}^\ell$ with the $i-$th component converging to 0 uniformly for $\mathbf{x} \in [0,T]^d$. In order to show the limiting multivariate normality, we can make use of the Cramér-Wold device (see e.g. van der Vaart, 1998, p.16). Since Z_L is a centered Gaussian process, for all $\xi \in \mathbb{R}^\ell$ we have $\xi' V_k$ Gaussian with expectation 0 and variance $\xi'\Sigma_k\xi$, where ξ' denotes the transpose of the vector ξ and Σ_k the variance-covariance matrix of V_k. Hence, invoking the Lévy Continuity theorem and Slutsky's theorem, it remains to show that $\xi'\Sigma_k\xi$ admits a non null limit as $k \to +\infty$. This can be done by using the dominated convergence theorem and similar bounds as those used to show (7).

4.2 Proof of Proposition 1

The key elements in the proof are the homogeneity of the functions L and M. Thus from our Theorem 1, we deduce that

$$\begin{aligned}
&\Delta_{k,a}(\mathbf{x}^*) \\
= \ & \alpha\left(\frac{n}{k}\right) M(\mathbf{x}^*)\frac{1}{k}\sum_{j=1}^{k} K(a_j)a_j^{-\rho}[a^{-\rho}-1] + \frac{1}{\sqrt{k}}\int_0^1 K(u)u^{-1}\left[a^{-1}Z_L(au\mathbf{x}^*) - Z_L(u\mathbf{x}^*)\right]du + o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right).
\end{aligned}$$

A similar expression can be obtained for $\Delta_{k,a}(r\mathbf{x}^*)$ from which we can easily justify the expression of our estimator $\widetilde{\rho}_k(\mathbf{x}^*)$ given in (3), and deduce that

$$\widetilde{\rho}_k(\mathbf{x}^*) = \rho + \frac{1}{\sqrt{k}\alpha\left(\frac{n}{k}\right)}\frac{\int_0^1 K(u)u^{-\rho}du}{\frac{1}{k}\sum_{j=1}^{k} K(a_j)a_j^{-\rho}} Z_\rho(\mathbf{x}^*) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}\alpha\left(\frac{n}{k}\right)}\right),$$

from which we can easily deduce our Proposition 1.

4.3 Proof of Proposition 2

To prove the proposition, we use the Skorohod construction, meaning that we have to look at the difference

$$D := \left|\sqrt{k}\left(\widetilde{\alpha}_k(\mathbf{x}) - \alpha\left(\frac{n}{k}\right)M(\mathbf{x})\right) - Z_{\alpha,\mathbf{x}^*}(\mathbf{x})\right|$$

and to show that it tends almost surely uniformly to 0 as $n \to \infty$. According to Proposition 2 in Fougères *et al.* (2015) and using (5), we have under our assumptions that

$$\begin{aligned}
D \ \leqslant \ & \left|\sqrt{k}\alpha\left(\frac{n}{k}\right)M(\mathbf{x})\frac{\sum_{j=1}^{k}\sum_{\ell=1}^{k}\widetilde{K}(a_j)\widetilde{K}(a_\ell)[a_j^{-\rho} - a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)}][a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)} - a_\ell^{-\widetilde{\rho}_k(\mathbf{x}^*)}]}{\sum_{j=1}^{k}\sum_{\ell=1}^{k}\widetilde{K}(a_j)\widetilde{K}(a_\ell)a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)}[a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)} - a_\ell^{-\widetilde{\rho}_k(\mathbf{x}^*)}]}\right. \\
& \left. - \frac{\int_0^1 \widetilde{K}(u)u^{-2\rho}\ln(u)du - \int_0^1 \widetilde{K}(u)u^{-\rho}\ln(u)du\int_0^1 \widetilde{K}(u)u^{-\rho}du}{\int_0^1 \widetilde{K}(u)u^{-2\rho}du - \left(\int_0^1 \widetilde{K}(u)u^{-\rho}du\right)^2} Z_\rho(\mathbf{x}^*)M(\mathbf{x})\right| \\
& + \left|\frac{\sum_{j=1}^{k}\sum_{\ell=1}^{k}\widetilde{K}(a_j)\widetilde{K}(a_\ell)[a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)} - a_\ell^{-\widetilde{\rho}_k(\mathbf{x}^*)}]Z_L(a_j\mathbf{x})a_j^{-1}}{\sum_{j=1}^{k}\sum_{\ell=1}^{k}\widetilde{K}(a_j)\widetilde{K}(a_\ell)[a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)} - a_\ell^{-\widetilde{\rho}_k(\mathbf{x}^*)}]a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)}}\right. \\
& \left. - \frac{\int_0^1 \widetilde{K}(u)u^{-1}\left[u^{-\rho} - \int_0^1 \widetilde{K}(v)v^{-\rho}dv\right]Z_L(u\mathbf{x})du}{\int_0^1 \widetilde{K}(u)u^{-2\rho}du - \left(\int_0^1 \widetilde{K}(u)u^{-\rho}du\right)^2}\right| + o(1).
\end{aligned}$$

The main idea now is to replace everywhere the terms with $\widetilde{\rho}_k(\mathbf{x}^*)$ by the same terms with ρ and to study the difference by the mean value theorem. For instance, if we look at the term

$$\frac{1}{k^2}\sum_{j=1}^{k}\sum_{\ell=1}^{k}\widetilde{K}(a_j)\widetilde{K}(a_\ell)[a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)} - a_\ell^{-\widetilde{\rho}_k(\mathbf{x}^*)}]a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)}$$

appearing several times as a denominator in the bound of D, we can rewrite it as follows:

$$\begin{aligned}
&\frac{1}{k^2}\sum_{j=1}^{k}\sum_{\ell=1}^{k}\widetilde{K}(a_j)\widetilde{K}(a_\ell)[a_j^{-\rho} - a_\ell^{-\rho}]a_j^{-\rho}\\
&+\frac{1}{k^2}\sum_{j=1}^{k}\sum_{\ell=1}^{k}\widetilde{K}(a_j)\widetilde{K}(a_\ell)\left[a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)}(a_j^{-\widetilde{\rho}_k(\mathbf{x}^*)} - a_\ell^{-\widetilde{\rho}_k(\mathbf{x}^*)}) - a_j^{-\rho}(a_j^{-\rho} - a_\ell^{-\rho})\right]\\
=\;&\int_0^1 \widetilde{K}(u)u^{-2\rho}du - \left(\int_0^1 \widetilde{K}(u)u^{-\rho}du\right)^2 + o\,(1)\\
&+2(\widetilde{\rho}_k(\mathbf{x}^*) - \rho)\left[\left(\frac{1}{k}\sum_{j=1}^{k}\widetilde{K}(a_j)a_j^{-\breve{\rho}_k(\mathbf{x}^*)}\right)\left(\frac{1}{k}\sum_{j=1}^{k}\widetilde{K}(a_j)a_j^{-\breve{\rho}_k(\mathbf{x}^*)}\ln a_j\right)\right.\\
&\left.-\left(\frac{1}{k}\sum_{j=1}^{k}\widetilde{K}(a_j)\right)\left(\frac{1}{k}\sum_{j=1}^{k}\widetilde{K}(a_j)a_j^{-2\breve{\rho}_k(\mathbf{x}^*)}\ln a_j\right)\right],
\end{aligned}$$

by the mean value theorem where $\breve{\rho}_k(\mathbf{x}^*)$ is an intermediate value between $\widetilde{\rho}_k(\mathbf{x}^*)$ and ρ. Using the same approach for each term combining with Proposition 1 leads to Proposition 2.

4.4 Proof of Theorem 2

According to our Corollary 1, we clearly have the following decomposition

$$
\begin{aligned}
\sqrt{k}\left(\overline{L}_{k,\overline{k}}(\mathbf{x})-L(\mathbf{x})\right) &= \sqrt{k}\left[\alpha\left(\frac{n}{k}\right)M(\mathbf{x})-\left(\frac{\overline{k}}{k}\right)^{\widetilde{\rho}_{\overline{k}}(\mathbf{x}^*)}\widetilde{\alpha}_{\overline{k}}(\mathbf{x})\right]\frac{\frac{1}{k}\sum_{j=1}^{k}K(a_j)a_j^{-\rho}}{\frac{1}{k}\sum_{j=1}^{k}K(a_j)} \\
&+ \sqrt{k}\left(\frac{\overline{k}}{k}\right)^{\widetilde{\rho}_{\overline{k}}(\mathbf{x}^*)}\widetilde{\alpha}_{\overline{k}}(\mathbf{x})\frac{\frac{1}{k}\sum_{j=1}^{k}K(a_j)[a_j^{-\rho}-a_j^{-\widetilde{\rho}_{\overline{k}}(\mathbf{x}^*)}]}{\frac{1}{k}\sum_{j=1}^{k}K(a_j)} \\
&+ \int_0^1 K(u)u^{-1}Z_L(u\mathbf{x})du+o_{\mathbb{P}}(1) \\
&= \sqrt{k}\left[\alpha\left(\frac{n}{k}\right)M(\mathbf{x})-\left(\frac{\overline{k}}{k}\right)^{\widetilde{\rho}_{\overline{k}}(\mathbf{x}^*)}\alpha\left(\frac{n}{\overline{k}}\right)M(\mathbf{x})\right]\frac{\frac{1}{k}\sum_{j=1}^{k}K(a_j)a_j^{-\rho}}{\frac{1}{k}\sum_{j=1}^{k}K(a_j)} \\
&+ O_{\mathbb{P}}\left(\left(\frac{\overline{k}}{k}\right)^{\widetilde{\rho}_{\overline{k}}(\mathbf{x}^*)-\frac{1}{2}}\sqrt{\overline{k}}\alpha\left(\frac{n}{\overline{k}}\right)(\widetilde{\rho}_{\overline{k}}(\mathbf{x}^*)-\rho)\right) \\
&+ \int_0^1 K(u)u^{-1}Z_L(u\mathbf{x})du+o_{\mathbb{P}}(1)
\end{aligned}
$$

by applying our Proposition 2 and again the mean value theorem. Under the assumptions of our Theorem 2, using Proposition 1, we have

$$
\begin{aligned}
\sqrt{k}\left(\overline{L}_{k,\overline{k}}(\mathbf{x})-L(\mathbf{x})\right) &= \sqrt{k}\left[\alpha\left(\frac{n}{k}\right)M(\mathbf{x})-\left(\frac{\overline{k}}{k}\right)^{\widetilde{\rho}_{\overline{k}}(\mathbf{x}^*)}\alpha\left(\frac{n}{\overline{k}}\right)M(\mathbf{x})\right]\frac{\frac{1}{k}\sum_{j=1}^{k}K(a_j)a_j^{-\rho}}{\frac{1}{k}\sum_{j=1}^{k}K(a_j)} \\
&+ \int_0^1 K(u)u^{-1}Z_L(u\mathbf{x})du+o_{\mathbb{P}}(1).
\end{aligned}
$$

Recall now that the function α is regularly varying with index $\rho < 0$, that is $\alpha(t) = t^\rho \ell_\alpha(t)$ where ℓ_α is slowly varying at infinity. Thus

$$\begin{aligned}
\sqrt{k}\left(\overline{L}_{k,\overline{k}}(\mathbf{x}) - L(\mathbf{x})\right) &= \sqrt{\frac{k}{\overline{k}}}\sqrt{\overline{k}}\,\alpha\left(\frac{n}{\overline{k}}\right) M(\mathbf{x}) \left[\left(\frac{\overline{k}}{k}\right)^{\rho} - \left(\frac{\overline{k}}{k}\right)^{\widetilde{\rho}_{\overline{k}}(\mathbf{x}^*)}\right] \frac{\frac{1}{k}\sum_{j=1}^{k} K(a_j) a_j^{-\rho}}{\frac{1}{k}\sum_{j=1}^{k} K(a_j)} \\
&+ \sqrt{\frac{k}{\overline{k}}}\sqrt{\overline{k}}\,\alpha\left(\frac{n}{\overline{k}}\right)\left(\frac{\overline{k}}{k}\right)^{\rho} M(\mathbf{x}) \left[\frac{\ell_\alpha(\frac{n}{k})}{\ell_\alpha(\frac{n}{\overline{k}})} - 1\right] \frac{\frac{1}{k}\sum_{j=1}^{k} K(a_j) a_j^{-\rho}}{\frac{1}{k}\sum_{j=1}^{k} K(a_j)} \\
&+ \int_0^1 K(u) u^{-1} Z_L(u\mathbf{x}) du + o_{\mathbb{P}}(1) \\
&= \int_0^1 K(u) u^{-1} Z_L(u\mathbf{x}) du + o_{\mathbb{P}}(1) \\
&+ O_{\mathbb{P}}\left(\left(\frac{k}{\overline{k}}\right)^{\frac{1}{2}-\breve{\rho}_{\overline{k}}(\mathbf{x}^*)} \ln\left(\frac{k}{\overline{k}}\right)\sqrt{\overline{k}}\,\alpha\left(\frac{n}{\overline{k}}\right)\left(\widetilde{\rho}_{\overline{k}}(\mathbf{x}^*) - \rho\right)\right) \\
&+ \left(\frac{k}{\overline{k}}\right)^{\frac{1}{2}-\rho}\sqrt{\overline{k}}\,\alpha\left(\frac{n}{\overline{k}}\right)\beta\left(\frac{n}{\overline{k}}\right)\frac{1}{\beta\left(\frac{n}{\overline{k}}\right)}\left[\frac{\ell_\alpha(\frac{n}{k})}{\ell_\alpha(\frac{n}{\overline{k}})} - 1\right] O(1)
\end{aligned}$$

by the mean value theorem. Theorem 2 now follows under the assumptions since

$$\lim_{t\to\infty} \frac{1}{\beta(t)}\left(\frac{\ell_\alpha(tc)}{\ell_\alpha(t)} - 1\right) = O(1) \tag{9}$$

when $t \to \infty$ and $c = c(t) \to \infty$. To be convinced by (9), remark that for all $x, y > 0$ and $t > 0$ we have with $\beta_1(t) = \alpha(t)\beta(t)$ and $L_t(x\mathbf{1}) = t\left[1 - F\left(F_1^{-1}(1 - t^{-1}x), ..., F_d^{-1}(1 - t^{-1}x)\right)\right]$, and using the fact that $L_t(x\mathbf{1}) = xL_{t/x}(\mathbf{1})$, that

$$\begin{aligned}
&\frac{1}{x}\left\{\frac{L_t(xy\mathbf{1}) - L(xy\mathbf{1}) - \alpha(t)M(xy\mathbf{1})}{\beta_1(t)} - \frac{L_t(x\mathbf{1}) - L(x\mathbf{1}) - \alpha(t)M(x\mathbf{1})}{\beta_1(t)}\right\} \\
&\qquad = \left[\frac{L_{t/x}(y\mathbf{1}) - L(y\mathbf{1}) - \alpha(t/x)M(y\mathbf{1})}{\beta_1(t/x)} - \frac{L_{t/x}(\mathbf{1}) - L(\mathbf{1}) - \alpha(t/x)M(\mathbf{1})}{\beta_1(t/x)}\right]\frac{\beta_1(t/x)}{\beta_1(t)} \\
&\qquad + \left[\left\{\alpha(t/x) - \alpha(t)x^{-\rho}\right\}/\beta_1(t)\right]\left(M(y\mathbf{1}) - M(\mathbf{1})\right).
\end{aligned}$$

Taking limits for $t \to \infty$ we obtain that

$$\begin{aligned}
&\frac{1}{x}\left(N(xy\mathbf{1}) - N(x\mathbf{1})\right) \\
&\quad = \lim_{t\to\infty}\left\{\left(N(y\mathbf{1}) - N(\mathbf{1})\right)\frac{\beta_1(t/x)}{\beta_1(t)} + \frac{1}{\beta(t)}\left(\frac{\alpha(t/x)}{\alpha(t)} - x^{-\rho}\right)\left(M(y\mathbf{1}) - M(\mathbf{1})\right)\right\}.
\end{aligned} \tag{10}$$

Since the limit for $t \to \infty$ of $\beta_1(t/x)/\beta_1(t)$ exists, from (10) the limit for $t \to \infty$ of $\frac{1}{\beta(t)}\left(\frac{\alpha(t/x)}{\alpha(t)} - x^{-\rho}\right)$ exists such that, combining with Theorem B.2.1 in de Haan and Ferreira (2006):

$$\lim_{t\to\infty}\frac{1}{\beta(t)}\left(\frac{\alpha(t/x)}{\alpha(t)} - x^{-\rho}\right) = x^{-\rho}\lim_{t\to\infty}\frac{1}{\beta(t)}\left(\frac{\ell_\alpha(t/x)}{\ell_\alpha(t)} - 1\right) = x^{-\rho}\psi(1/x),$$

with $\psi(x) = a\,\frac{x^{\rho'}-1}{\rho'}$ and $a > 0$ (in case $a < 0$ one can redefine ℓ_α as $-\ell_\alpha$ and the result will follow with $a > 0$). Note that the positive constant a can in fact be incorporated in the function β, so from now on we take without loss of generality $a = 1$. We have thus, with $\widetilde{\beta} := \beta\ell_\alpha$,

$$\lim_{t\to\infty} \frac{\ell_\alpha(tx) - \ell_\alpha(t)}{\widetilde{\beta}(t)} = \psi(x).$$

Now write, for some function $\widetilde{\beta}^*$ where $\widetilde{\beta}^* \sim \widetilde{\beta}$,

$$\left|\frac{\ell_\alpha(tx) - \ell_\alpha(t)}{\widetilde{\beta}(t)}\right| \leqslant \frac{\widetilde{\beta}^*(t)}{\widetilde{\beta}(t)}\left[|\psi(x)| + \left|\frac{\ell_\alpha(tx) - \ell_\alpha(t)}{\widetilde{\beta}^*(t)} - \psi(x)\right|\right].$$

By Theorem B.2.18 in de Haan and Ferreira (2006) (see also Drees, 1998) we have for any $\varepsilon > 0$ there is a t_0 such that for $t, tx > t_0$,

$$\left|\frac{\ell_\alpha(tx) - \ell_\alpha(t)}{\widetilde{\beta}(t)}\right| \leqslant (1+\kappa)[|\psi(x)| + \varepsilon x^{\rho'+\delta}].$$

where $0 < \delta < -\rho'$ and $\kappa > 0$. Since both functions in the right-hand side are bounded for $x \in [x_0, \infty)$, $x_0 > 0$, result (9) follows.

4.5 Proof of Corollary 2

Let $u, v \in [0,1]$ and $\mathbf{x}, \mathbf{y} \in [0,T]^d$. By Proposition 2 in Fougères *et al.* (2015), using the property of a measure, we have

$$\mathbb{E}[W_L(\mathbf{x})W_L(\mathbf{y})] = L(\mathbf{x}) + L(\mathbf{y}) - L(\mathbf{x} \vee \mathbf{y})$$

where $\mathbf{x} \vee \mathbf{y} := (x_1 \vee y_1, \ldots, x_d \vee y_d)$. In particular, for any $i, j = 1, \ldots, d$, we have

$$\begin{aligned}
\mathbb{E}[W_L(u\mathbf{x})W_L(v\mathbf{x})] &= (u \wedge v)L(\mathbf{x}); \\
\mathbb{E}[W_L(u\mathbf{x})W_L(vx_j\mathbf{e}_j)] &= uL(\mathbf{x}) + vx_j - L(\{u\mathbf{x}\} \vee \{vx_j\mathbf{e}_j\}); \\
\mathbb{E}[W_L(ux_i\mathbf{e}_i)W_L(vx_j\mathbf{e}_j)] &= ux_i + vx_j - L(\{ux_i\mathbf{e}_i\} \vee \{vx_j\mathbf{e}_j\}).
\end{aligned}$$

Now, using the property that for any $i = 1, \ldots, d$, $\partial_i L$ is homogeneous of order 0, we have

$$\begin{aligned}
&\mathbb{E}\left[Z_L(u\mathbf{x})Z_L(v\mathbf{x})\right] \\
&= \mathbb{E}\left[W_L(u\mathbf{x})W_L(v\mathbf{x})\right] - \mathbb{E}\left[W_L(u\mathbf{x})\sum_{i=1}^{d} W_L(vx_i\mathbf{e}_i)\partial_i L(\mathbf{x})\right] \\
&\quad -\mathbb{E}\left[W_L(v\mathbf{x})\sum_{i=1}^{d} W_L(ux_i\mathbf{e}_i)\partial_i L(\mathbf{x})\right] + \mathbb{E}\left[\sum_{i=1}^{d}\sum_{j=1}^{d} W_L(ux_i\mathbf{e}_i)W_L(vx_j\mathbf{e}_j)\partial_i L(\mathbf{x})\partial_j L(\mathbf{x})\right] \\
&= (u \wedge v)L(\mathbf{x}) - \sum_{i=1}^{d}\partial_i L(\mathbf{x})\left[(v+u)\left(L(\mathbf{x})+x_i\right) - L(\{u\mathbf{x}\} \vee \{vx_i\mathbf{e}_i\}) - L(\{v\mathbf{x}\} \vee \{ux_i\mathbf{e}_i\})\right] \\
&\quad + \sum_{i=1}^{d}\sum_{j=1}^{d}\partial_i L(\mathbf{x})\partial_j L(\mathbf{x})\left[ux_i + vx_j - L(\{ux_i\mathbf{e}_i\} \vee \{vx_j\mathbf{e}_j\})\right].
\end{aligned}$$

We have in particular the variance of $Z_L(\mathbf{x})$ for $u = v = 1$, i.e,

$$\mathrm{Var}\left(Z_L(\mathbf{x})\right) = L(\mathbf{x}) - 2\sum_{i=1}^{d} x_i\partial_i L(\mathbf{x}) + \sum_{i=1}^{d}\sum_{j=1}^{d}\partial_i L(\mathbf{x})\partial_j L(\mathbf{x})\left[x_i + x_j - L(\{x_i\mathbf{e}_i\} \vee \{x_j\mathbf{e}_j\})\right].$$

Clearly, for $a \leqslant a'$ and $b \leqslant b'$, we have $R(\{a\mathbf{x}\} \vee \{b\mathbf{y}\}) \subset R(\{a'\mathbf{x}\} \vee \{b'\mathbf{y}\})$. Then since μ is a positive measure, this implies that $L(\{a\mathbf{x}\} \vee \{b\mathbf{y}\}) \leqslant L(\{a'\mathbf{x}\} \vee \{b'\mathbf{y}\})$. Consequently, using the homogeneity of L, we obtain the following inequalities

$$(u \wedge v)L(\mathbf{x}) = (u \wedge v)L(\mathbf{x} \vee \{x_i\mathbf{e}_i\}) \leqslant L(\{u\mathbf{x}\} \vee \{vx_i\mathbf{e}_i\}) \leqslant (u \vee v)L(\mathbf{x} \vee \{x_i\mathbf{e}_i\}) = (u \vee v)L(\mathbf{x})$$

$$(u \wedge v)L(\{x_i\mathbf{e}_i\} \vee \{x_j\mathbf{e}_j\}) \leqslant L(\{ux_i\mathbf{e}_i\} \vee \{vx_j\mathbf{e}_j\}) \leqslant (u \vee v)L(\{x_i\mathbf{e}_i\} \vee \{x_j\mathbf{e}_j\}).$$

Now, since $u \wedge v = u + v - u \vee v$ and

$$\lim_{\tau\to+\infty}\int_{[0,1]^2} K(u)u^{-1}K(v)v^{-1}u\,dudv = \lim_{\tau\to+\infty}\int_{[0,1]^2} K(u)u^{-1}K(v)v^{-1}(u \vee v)\,dudv = 1,$$

we can deduce from the previous inequalities that

$$\lim_{\tau\to+\infty}\int_{[0,1]^2} K(u)u^{-1}K(v)v^{-1}L(\{u\mathbf{x}\} \vee \{vx_i\mathbf{e}_i\})dudv = L(\mathbf{x})$$

and

$$\lim_{\tau\to+\infty}\int_{[0,1]^2} K(u)u^{-1}K(v)v^{-1}L(\{ux_i\mathbf{e}_i\} \vee \{vx_j\mathbf{e}_j\})dudv = L(\{x_i\mathbf{e}_i\} \vee \{x_j\mathbf{e}_j\}).$$

This achieves the proof of Corollary 2 since

$$\mathrm{Var}\left(\int_0^1 K(u)u^{-1}Z_L(u\mathbf{x})\,du\right) = \int_{[0,1]^2} K(u)u^{-1}K(v)v^{-1}\mathbb{E}\left[Z_L(u\mathbf{x})Z_L(v\mathbf{x})\right]\,dudv.$$

Acknowledgement

The authors kindly acknowledge Anne-Laure Fougères, Laurens de Haan and Cécile Mercadier for providing the program of their six distributions considered. This work was supported by a research grant (VKR023480) from VILLUM FONDEN and an international project for scientific cooperation (PICS-6416).

References

[1] Beirlant, J., Dierckx, G., Goegebeur, Y., Matthys, G., 1999. Tail index estimation and an exponential regression model. Extremes, 2, 177–200.

[2] Beirlant, J., Dierckx, G., Guillou, A., 2011. Bias-reduced estimators for bivariate tail modelling. Insurance: Mathematics and Economics, 49, 18–26.

[3] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. *Statistics of Extremes – Theory and Applications*. Wiley.

[4] Caeiro, F., Gomes, M.I., Rodrigues, L.H., 2009. Reduced-bias tail index estimators under a third order framework. Communications in Statistics - Theory and Methods, 38, 1019–1040.

[5] Drees, H., 1998. On smooth statistical tail functionals. Scandinavian Journal of Statistics, 25, 187–210.

[6] Drees, H., Huang, X., 1998. Best attainable rates of convergence for estimators of the stable tail dependence function. Journal of Multivariate Analysis, 64, 25–47.

[7] Dutang, C., Goegebeur, Y., Guillou, A., 2014. Robust and bias-corrected estimation of the coefficient of tail dependence. Insurance: Mathematics and Economics, 57, 46–57.

[8] Feuerverger, A., Hall, P., 1999. Estimating a tail exponent by modelling departure from a Pareto distribution. Annals of Statistics, 27, 760–781.

[9] Fougères, A.L., de Haan, L., Mercadier, C., 2015. Bias correction in multivariate extremes. Annals of Statistics, 43, 903–934.

[10] Goegebeur, Y., Guillou, A., 2013. Asymptotically unbiased estimation of the coefficient of tail dependence. Scandinavian Journal of Statistics, 40, 174–189.

[11] Gomes, M.I., de Haan, L., Rodrigues, L.H., 2008. Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. Journal of the Royal Statistical Society Series B, 70, 31–52.

[12] de Haan, L., Ferreira, A., 2006. *Extreme Value Theory - An Introduction*, Springer.

[13] de Haan, L., de Ronde, J., 1998. Sea and wind: multivariate extremes at work. Extremes, 1, 7–45.

[14] Huang, X., 1992. Statistics of bivariate extremes. PhD Thesis, Erasmus University Rotterdam, Tinbergen Institute Research series No. 22.

[15] Ledford, A.W., Tawn, J.A., 1997. Modelling dependence within joint tail regions. Journal of the Royal Statistical Society Series B, 59, 475–499.

[16] Matthys, G., Beirlant, J., 2003. Estimating the extreme value index and high quantiles with exponential regression models. Statistica Sinica 13, 853–880.

[17] van der Vaart, A.W., 1998. *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics.