

# On kernel estimation of the second order rate parameter in multivariate extreme value statistics

Yuri Goegebeur \*

Armelle Guillou †

Jing Qin ‡

**Abstract.** We introduce a flexible class of kernel type estimators of a second order parameter appearing in the multivariate extreme value framework. Such an estimator is crucial in order to construct asymptotically unbiased estimators of dependence measures, as e.g. the stable tail dependence function. We establish the asymptotic properties of this class of estimators under suitable assumptions. The behaviour of some examples of kernel estimators is illustrated by a simulation study in which they are also compared with a benchmark estimator of a second order parameter recently introduced in the literature.

**Keywords:** Multivariate extreme value statistics, second order parameter, stable tail dependence function.

## 1 Introduction

Measuring the strength of the dependence in the extremes is a challenging topic which has received considerable attention in the recent multivariate extreme value literature. Several tools can be used, either some coefficients of tail dependence or some functions, such as the Pickands

---

\*Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: yuri.goegebeur@imada.sdu.dk).

†Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France (email: armelle.guillou@math.unistra.fr).

‡Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: qin@imada.sdu.dk).

dependence function or the spectral distribution function. In this paper, we focus on the stable tail dependence function, denoted  $L$ , originally introduced by Huang (1992), which can be defined as follows. For any arbitrary dimension  $d$ , let  $(X^{(1)}, \dots, X^{(d)})$  be a random vector with continuous marginal distribution functions (dfs)  $F_1, \dots, F_d$ . The stable tail dependence function is defined for each  $x_i \in \mathbb{R}_+$ ,  $i = 1, \dots, d$ , as

$$\lim_{t \rightarrow \infty} t [1 - F(F_1^{\leftarrow}(1 - t^{-1}x_1), \dots, F_d^{\leftarrow}(1 - t^{-1}x_d))] = L(x_1, \dots, x_d) \quad (1)$$

provided that this limit exists, where  $F$  is the distribution function of the vector  $(X^{(1)}, \dots, X^{(d)})$ , and  $F_i^{\leftarrow}$  is the generalised inverse of  $F_i$ , i.e.  $F_i^{\leftarrow}(p) := \inf\{x : F_i(x) \geq p\}$ ,  $p \in (0, 1)$ ,  $i = 1, \dots, d$ . Let  $\mathbf{x} := (x_1, \dots, x_d)$ . Note that  $L$  is homogeneous of order one, that is  $L(a\mathbf{x}) = aL(\mathbf{x})$ , for all  $\mathbf{x} \in \mathbb{R}_+^d$  and all  $a > 0$ .

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

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

Recently the poor performance of this empirical estimator in terms of bias has been emphasized in the literature. This bias-issue is common in extreme value statistics, e.g. it is also present in the univariate and the regression contexts, and generally it complicates the practical application of extreme value methods. All the contributions dealing with bias reduction in the multivariate framework (see Fougères et al., 2015, Beirlant et al., 2016, Escobar-Bach et al., 2016) require the following second or third order conditions, depending on the type of asymptotic properties that one wants to establish.

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

$$\lim_{t \rightarrow \infty} \frac{1}{\alpha(t)} \{t [1 - F(F_1^{\leftarrow}(1 - t^{-1}x_1), \dots, F_d^{\leftarrow}(1 - t^{-1}x_d))] - L(\mathbf{x})\} = M(\mathbf{x}), \quad (2)$$

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

The second order condition implies that the function  $\alpha$  is regularly varying at infinity of index

$\rho \leq 0$ , i.e.  $\alpha(ty)/\alpha(t) \rightarrow y^\rho$  as  $t \rightarrow \infty$  for all  $y > 0$ , and  $M(\mathbf{x})$  is homogeneous of order  $1 - \rho$ .

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

$$\lim_{t \rightarrow \infty} \frac{1}{\beta(t)} \left\{ \frac{t [1 - F(F_1^{\leftarrow}(1 - t^{-1}x_1), \dots, F_d^{\leftarrow}(1 - t^{-1}x_d))] - L(\mathbf{x})}{\alpha(t)} - M(\mathbf{x}) \right\} = N(\mathbf{x}), \quad (3)$$

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

It can be shown that the third order condition implies that  $\beta$  is regularly varying at infinity of order  $\rho' \leq 0$  and that  $N$  is homogeneous of order  $1 - \rho - \rho'$ .

The asymptotically unbiased estimators of the stable tail dependence function  $L$  proposed in the abovementioned recent papers depend on the second order parameter  $\rho$ , which has to be estimated from the sample. The problem of estimating second order parameters is also present in the univariate framework, but in that context, several estimators have already been proposed in the literature and they perform reasonably well in practice, although not very stable as a function of the intermediate sequence  $k$  used to compute them. We refer for instance to Fraga Alves et al. (2003), Ciuperca and Mercadier (2010) and Goegebeur et al. (2010), among others. On the contrary, in the multivariate context this topic is still in its infancy. We are only aware of the estimators proposed by Fougères et al. (2015) and Beirlant et al. (2016). However, these papers are mainly focused on bias-corrected estimation of  $L$ , and the estimation of  $\rho$  is only an obstacle to overcome in order to obtain the bias-correction. Thus, although these papers introduce estimators for  $\rho$ , the performance of these estimators has not been studied in detail. Until now, due to the difficulty of estimating this parameter  $\rho$ , it is often suggested in practice to replace it by a canonical value like e.g. -1 (see Escobar-Bach et al., 2016). However, a mis-specification of this parameter implies that from a theoretical perspective one loses the bias correction. Thus, similarly to the univariate context, it is an important challenge to be able to estimate  $\rho$  in the multivariate context. To reach this goal, we introduce in Section 2 a flexible class of kernel estimators for  $\rho$ , for which we derive the asymptotic normality under suitable assumptions, in particular the third order condition mentioned above. Our estimator depends on some tuning parameters that we have to select. For some specific values of them, our estimator encompasses the one proposed in Beirlant et al. (2016). In Section 3, the performance of our

estimator is compared with a benchmark from the recent literature. All the proofs of the results are postponed to the appendix.

## 2 Kernel estimator and asymptotic properties

Motivated by the homogeneity property of the function  $L$  we introduce the following scaled estimator

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

for a positive parameter  $a$ . The basic building block for the  $\rho$  estimator will be the following kernel statistic

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

where  $a_j := \frac{j}{k+1}$ ,  $j = 1, \dots, k$ ,  $\xi \in \mathbb{N}$ , and  $K$  is a function defined on  $[0, 1]$  which is positive and such that  $\int_0^1 K(u) du = 1$ . This function is called a kernel function in the sequel.

As a first step in the development of the estimator we derive the weak limit of  $\widehat{L}_k(\mathbf{x})$ , when properly normalised. Let  $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$  be the canonical basis vectors of  $\mathbb{R}^d$ .

**Theorem 1.** *Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be independent random vectors in  $\mathbb{R}^d$  with common joint df  $F$  and continuous marginal dfs  $F_j$ ,  $j = 1, \dots, d$ . Assume that the third order condition (3) holds with negative indices  $\rho$  and  $\rho'$  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. Assuming that the intermediate sequence  $k$  satisfies  $\sqrt{k}\alpha(n/k) \rightarrow \infty$  and  $\sqrt{k}\alpha(n/k)\beta(n/k) \rightarrow \lambda_1 \in \mathbb{R}$ , we have*

$$\sqrt{k} \left\{ \widehat{L}_k(\mathbf{x}) - L(\mathbf{x}) - \alpha \left( \frac{n}{k} \right) M(\mathbf{x}) - \alpha \left( \frac{n}{k} \right) \beta \left( \frac{n}{k} \right) N(\mathbf{x}) \right\} \xrightarrow{d} Z_L(\mathbf{x})$$

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

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

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

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

Note that we assume  $\sqrt{k}\alpha(n/k)\beta(n/k) \rightarrow \lambda_1 \in \mathbb{R}$ , while Fougères et al. (2015) took  $\lambda_1 = 0$ . Our assumption will allow us to study the estimation of  $\rho$  in a more general context. In particular we can make the bias term explicit, unlike Fougères et al. (2015), where no information about the bias of  $\rho$  is available.

Based on the result of Theorem 1 we can now establish the weak limit of the kernel statistic  $\tilde{L}_k(\mathbf{x}; K, \xi)$ . Let  $I(K, r) := \int_0^1 K(u)u^{-r}du$ .

**Theorem 2.** *Let  $K$  be a continuous kernel function satisfying*

$$\left\{ \begin{array}{l} I(K, (\xi + 1)/2) < \infty, \\ \frac{1}{k} \sum_{j=1}^k K(a_j) = 1 + o\left(\frac{1}{\sqrt{k}}\right), \\ \frac{1}{k} \sum_{j=1}^k K(a_j)a_j^{-\rho} = I(K, \rho) + o\left(\frac{1}{\sqrt{k}}\right). \end{array} \right.$$

Suppose that the conditions of Theorem 1 hold together with  $\sqrt{k}\alpha^2\left(\frac{n}{k}\right) \rightarrow \lambda_2 \in \mathbb{R}$ . Then, we have

$$\begin{aligned} \sqrt{k} \left\{ \tilde{L}_k(\mathbf{x}; K, \xi) - L^\xi(\mathbf{x}) - \alpha\left(\frac{n}{k}\right) \xi L^{\xi-1}(\mathbf{x})M(\mathbf{x})I(K, \rho) - \alpha\left(\frac{n}{k}\right) \beta\left(\frac{n}{k}\right) \xi L^{\xi-1}(\mathbf{x})N(\mathbf{x})I(K, \rho + \rho') \right. \\ \left. - \alpha^2\left(\frac{n}{k}\right) \frac{\xi(\xi-1)}{2} L^{\xi-2}(\mathbf{x})M^2(\mathbf{x})I(K, 2\rho) \right\} \xrightarrow{d} \xi L^{\xi-1}(\mathbf{x}) \int_0^1 K(u) \frac{1}{u} Z_L(u\mathbf{x}) du \end{aligned}$$

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

**Definition** For  $(\xi_1, \xi_2) \in \mathbb{N} \times \mathbb{N}$ ,  $a, r \in (0, 1)$ , let

$$\tilde{\rho}_k(\mathbf{x}; K, \xi_1, \xi_2) := \left( 1 - \frac{1}{\log r} \log \left| \frac{\tilde{\Delta}_{k,a}(r\mathbf{x}; K, \xi_1, \xi_2)}{\tilde{\Delta}_{k,a}(\mathbf{x}; K, \xi_1, \xi_2)} \right| \right) \wedge 0$$

where

$$\tilde{\Delta}_{k,a}(\mathbf{x}; K, \xi_1, \xi_2) := \left[ \frac{1}{a^{\xi_1}} \tilde{L}_k(a\mathbf{x}; K, \xi_1) \right]^{\frac{1}{\xi_1}} - \left[ \tilde{L}_k(\mathbf{x}; K, \xi_2) \right]^{\frac{1}{\xi_2}}.$$

The weak convergence of this estimator for  $\rho$  is established in the next theorem.

**Theorem 3.** *Assume the conditions of Theorem 2 hold, that the function  $M$  never vanishes*

except on the axes, and consider real numbers  $a$  and  $r$  both in  $(0, 1)$ . Then we have

$$\begin{aligned} & \sqrt{k}\alpha\left(\frac{n}{k}\right) \left\{ \tilde{\rho}_k(\mathbf{x}; K, \xi_1, \xi_2) - \rho + \frac{I(K, \rho + \rho')}{I(K, \rho)} \frac{N(\mathbf{x})}{M(\mathbf{x})} \frac{a^{-\rho-\rho'} - 1}{a^{-\rho} - 1} \frac{r^{-\rho'} - 1}{\log r} \beta\left(\frac{n}{k}\right) \right. \\ & \left. + \left( \frac{I(K, 2\rho)}{I(K, \rho)} - I(K, \rho) \right) \frac{\frac{\xi_1-1}{2}a^{-2\rho} - \frac{\xi_2-1}{2}}{a^{-\rho} - 1} \frac{M(\mathbf{x})}{L(\mathbf{x})} \frac{r^{-\rho} - 1}{\log r} \alpha\left(\frac{n}{k}\right) \right\} \\ & \xrightarrow{d} -\frac{1}{(a^{-\rho} - 1)M(\mathbf{x})I(K, \rho)\log r} \left( \frac{\frac{1}{a} \int_0^1 K(u) \frac{Z_L(uarx)}{u} du - \int_0^1 K(u) \frac{Z_L(urx)}{u} du}{r^{1-\rho}} \right. \\ & \left. - \frac{1}{a} \int_0^1 K(u) \frac{Z_L(uax)}{u} du + \int_0^1 K(u) \frac{Z_L(ux)}{u} du \right), \end{aligned}$$

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

Note that in the particular case where  $\xi_1 = \xi_2 = 1$ , we recover the estimator proposed by Beirlant et al. (2016) but we provide its asymptotic normality in Theorem 3 under more general assumptions which allow us to exhibit the bias of our estimator. Also, we provide here the result in terms of weak convergence of a stochastic process in  $\mathbf{x}$ , whereas Beirlant et al. (2016) only provide a pointwise convergence result. Remark that our asymptotic limit does not depend on the values of  $\xi_1$  and  $\xi_2$ .

### 3 Simulation study

To assess the performance of our class of estimators in practice, we simulate  $B = 1000$  samples of size  $n = 1000$  from several distributions originally proposed by Fougères et al. (2015) and used in Beirlant et al. (2016). To keep the length of the paper reasonable, we only include the following ones:

- the logistic model, for which  $L(x, y) = (x^{1/s} + y^{1/s})^s$  and  $\rho = -1$ . We set  $s = 1/3$ ;
- the bivariate Pareto of type II model, called BP $II(p)$ , for which  $L(x, y) = x + y - (x^{-p} + y^{-p})^{-1/p}$ .

We set  $p = 4$  which corresponds to  $\rho = -0.5$ ;

- the bivariate Cauchy distribution, for which  $L(x, y) = (x^2 + y^2)^{1/2}$  and  $\rho = -2$ .

In this simulation study, we compare our new class of kernel estimators for  $\rho$  with the benchmark estimator already proposed in the literature by Fougères et al. (2015) and defined as

$$\hat{\rho}_k(\mathbf{x}) := \left( 1 - \frac{1}{\log r} \log \left| \frac{\hat{\Delta}_{k,a}(r\mathbf{x})}{\hat{\Delta}_{k,a}(\mathbf{x})} \right| \right) \wedge 0$$

where  $\widehat{\Delta}_{k,a}(\mathbf{x}) := a^{-1}\widehat{L}_k(a\mathbf{x}) - \widehat{L}_k(\mathbf{x})$ , with advocated values  $r = a = 0.4$  for practical use.

Our class of estimators depends on several tuning parameters that we have to select in practice.

An extensive simulation study leads us to the following conclusions:

- the performance of our estimator, whatever the distribution, is almost the same for all the pairs of  $\xi_i$ ,  $i = 1, 2$ , used. To keep the length of the paper under control, we only report the results when  $(\xi_1, \xi_2) = (1, 1)$  (which coincides with the estimator introduced in Beirlant et al., 2016) and  $(4, 4)$ ;
- concerning the kernel, we tried also different types of families, all satisfying the conditions of our Theorem 2, but again we show only the results for the power kernel,  $\mathbb{K}_\tau(t) := (\tau + 1)t^\tau \mathbb{1}_{\{t \in [0,1]\}}$ , where  $I(\mathbb{K}_\tau, r) = \frac{1+\tau}{1+\tau-r}$  and we set  $\tau = 2$  and  $10$ ;
- concerning  $a$  and  $r$ , as in Fougères et al. (2015), the value  $0.4$  for both parameters seems overall best, although in some particular cases, a larger value, say  $0.5$ , is better.

Figure 1 displays the results for the logistic and BPII(4) distributions, in case of the particular choices  $a = r = 0.4$ ,  $(\xi_1, \xi_2) = (1, 1)$  (first and third rows) or  $(4, 4)$  (second and fourth rows), whereas Figure 2 shows similar plots for the bivariate Cauchy distribution in case  $a = r = 0.4$  and  $0.5$  and the same pairs for  $(\xi_1, \xi_2)$ . The estimates for  $\rho$  were calculated at  $\mathbf{x} = (0.5, 0.5)$ . Also other values of  $\mathbf{x}$  have been tried, but it turns out that the estimation results are not very sensitive with respect to the choice of this position. Each time, our class of estimators is computed with the power kernel with  $\tau = 2$  and  $10$ . The left panel of the figures represents the mean of the estimators over the  $B$  samples as a function of  $k$ , whereas the right panel is the mean squared error (MSE), also as a function of  $k$ . On the left panel, the horizontal reference line (solid black line) indicates the true value of  $\rho$ . Due to the high volatility of the estimator  $\widehat{\rho}_k$ , we follow the advices of Fougères et al. (2015) by drawing a horizontal line (solid grey line) corresponding to the mean value computed at  $k = 990$  to represent the performance of this estimator. Since this mean value is sometimes far away from the true one, the corresponding MSE may not appear on the right-hand side of the figures in view of the scale used. The values  $\tau = 2$  or  $10$  seem to give always good results with a slightly better performance to  $\tau = 10$  for some distributions. When  $\rho = -1$ , our estimator has almost no bias and is very stable as a function of  $k$ . When  $\rho = -1/2$  some bias appears, but this is expected since it is well-known in

the extreme value literature, that bias occurs in case  $\rho \in (-1, 0]$ , making estimation practically difficult. Now if  $\rho < -1$ , a bias also appears, whatever the value of  $\tau$ , though for large values of  $k$ , which is the range to consider, the estimator gets close to the true value. This bias is not usual but due to the value of the pair  $(a, r)$  which is not appropriate to that framework. Indeed, as illustrated in Figure 2 for this distribution,  $a = r = 0.5$  leads to less bias than in case  $a = r = 0.4$ , but an MSE more or less at the same level. On the contrary, for distributions with  $\rho = -1$  the choice  $a = r = 0.5$  leads to more bias and a larger MSE compared to  $a = r = 0.4$ . This motivates why we essentially illustrate the performance of our estimator in case  $a = r = 0.4$  which is overall the best choice.

To conclude, similarly to the univariate framework, no particular values of the tuning parameters are best for the whole parameter space. This has been already observed by Fraga Alves et al. (2003) and Goegebeur et al. (2010). However, on the contrary to the univariate context, our class of estimators for  $\rho$  with the specific values of the parameters considered in this section seems to have nice bias properties with rather stable sample paths as a function of  $k$ , which alleviates the choice of  $k$  to some extent, and it outperforms the benchmark estimator,  $\hat{\rho}_k$ . Indeed, due to the large variability of  $\hat{\rho}_k$  one has that  $\tilde{\rho}_k$  is typically better in terms of MSE.

## 4 Appendix: Proofs of the results

**Proof of Theorem 1.** Let  $U_i^{(j)}$ ,  $j = 1, \dots, d$ , be uniform random variables defined as  $U_i^{(j)} := 1 - F_j(X_i^{(j)})$  for  $j = 1, \dots, d$ . Introduce now

$$V_k(\mathbf{x}) := \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{U_i^{(1)} \leq kx_1/n \text{ or } \dots \text{ or } U_i^{(d)} \leq kx_d/n\}}$$

Then, we can use the following decomposition



$$\begin{aligned}
& \widehat{L}_k(\mathbf{x}) - L(\mathbf{x}) - \alpha \binom{n}{k} M(\mathbf{x}) - \alpha \binom{n}{k} \beta \binom{n}{k} N(\mathbf{x}) \\
&= V_k \left( \frac{n}{k} U_{[kx_1],n}^{(1)}, \dots, \frac{n}{k} U_{[kx_d],n}^{(d)} \right) - \frac{n}{k} \left[ 1 - F \left\{ F_1^{\leftarrow} \left( 1 - U_{[kx_1],n}^{(1)} \right), \dots, F_d^{\leftarrow} \left( 1 - U_{[kx_d],n}^{(d)} \right) \right\} \right] \\
&\quad + \left( \frac{n}{k} \left[ 1 - F \left\{ F_1^{\leftarrow} \left( 1 - U_{[kx_1],n}^{(1)} \right), \dots, F_d^{\leftarrow} \left( 1 - U_{[kx_d],n}^{(d)} \right) \right\} \right] - L \left( \frac{n}{k} U_{[kx_1],n}^{(1)}, \dots, \frac{n}{k} U_{[kx_d],n}^{(d)} \right) \right. \\
&\quad \quad \left. - \alpha \binom{n}{k} M \left( \frac{n}{k} U_{[kx_1],n}^{(1)}, \dots, \frac{n}{k} U_{[kx_d],n}^{(d)} \right) - \alpha \binom{n}{k} \beta \binom{n}{k} N \left( \frac{n}{k} U_{[kx_1],n}^{(1)}, \dots, \frac{n}{k} U_{[kx_d],n}^{(d)} \right) \right) \\
&\quad + L \left( \frac{n}{k} U_{[kx_1],n}^{(1)}, \dots, \frac{n}{k} U_{[kx_d],n}^{(d)} \right) - L(\mathbf{x}) \\
&\quad + \alpha \binom{n}{k} \left[ M \left( \frac{n}{k} U_{[kx_1],n}^{(1)}, \dots, \frac{n}{k} U_{[kx_d],n}^{(d)} \right) - M(\mathbf{x}) \right] \\
&\quad + \alpha \binom{n}{k} \beta \binom{n}{k} \left[ N \left( \frac{n}{k} U_{[kx_1],n}^{(1)}, \dots, \frac{n}{k} U_{[kx_d],n}^{(d)} \right) - N(\mathbf{x}) \right] \\
&=: A_{1,k}(\mathbf{x}) + A_{2,k}(\mathbf{x}) + A_{3,k}(\mathbf{x}) + A_{4,k}(\mathbf{x}) + A_{5,k}(\mathbf{x}).
\end{aligned}$$

We have to study each term separately. According to the proof of Proposition 1 in Fougères et al. (2015), we have

$$\sqrt{k} A_{1,k}(\mathbf{x}) \xrightarrow{d} W_L(\mathbf{x}) \quad (4)$$

$$\sqrt{k} A_{3,k}(\mathbf{x}) \xrightarrow{d} - \sum_{j=1}^d W_L(x_j \mathbf{e}_j) \partial_j L(\mathbf{x}) \quad (5)$$

in  $D([0, T]^d)$  for every  $T > 0$ . Now according to the proof of Theorem 7.2.2 in de Haan and Ferreira (2006), we have

$$\sup_{x \in [0, T]} \left| \sqrt{k} \left( \frac{n}{k} U_{[kx],n}^{(j)} - x \right) + W_L(x \mathbf{e}_j) \right| \rightarrow 0 \quad \text{a.s.} \quad (6)$$

Under the assumptions on the function  $M$ , this implies that

$$\sup_{0 \leq x_1, \dots, x_d \leq T} \left| \sqrt{k} \left\{ M \left( \frac{n}{k} U_{[kx_1],n}^{(1)}, \dots, \frac{n}{k} U_{[kx_d],n}^{(d)} \right) - M(\mathbf{x}) \right\} + \sum_{j=1}^d W_L(x_j \mathbf{e}_j) \partial_j M(\mathbf{x}) \right| \rightarrow 0 \quad \text{a.s.}$$

Consequently

$$\sup_{0 \leq x_1, \dots, x_d \leq T} \left| \sqrt{k} A_{4,k}(\mathbf{x}) \right| \xrightarrow{\mathbb{P}} 0. \quad (7)$$

Similarly using the continuity of the function  $N$  and (6), we have

$$\sup_{0 \leq x_1, \dots, x_d \leq T} \left| N(\mathbf{x}) - N \left( \frac{n}{k} U_{[kx_1],n}^{(1)}, \dots, \frac{n}{k} U_{[kx_d],n}^{(d)} \right) \right| \rightarrow 0 \quad \text{a.s.}$$

which entails that

$$\sup_{0 \leq x_1, \dots, x_d \leq T} \left| \sqrt{k} A_{5,k}(\mathbf{x}) \right| \longrightarrow 0 \quad \text{a.s.} \quad (8)$$

under the assumption that  $\sqrt{k} \alpha \left( \frac{n}{k} \right) \beta \left( \frac{n}{k} \right) \rightarrow \lambda_1$ . The last term that we have to consider is  $A_{2,k}(\mathbf{x})$ . According to the third order condition and (6), we have

$$\sqrt{k} A_{2,k}(\mathbf{x}) = o \left( \sqrt{k} \alpha \left( \frac{n}{k} \right) \beta \left( \frac{n}{k} \right) \right) = o(1) \quad \text{a.s.} \quad (9)$$

uniformly on  $[0, T]^d$ . Combining (4), (5), (7), (8) and (9), Theorem 1 follows.

**Proof of Theorem 2.** From Theorem 1, and by using the Skorohod construction, we have

$$\widehat{L}_k(\mathbf{x}) = L(\mathbf{x}) + \alpha \left( \frac{n}{k} \right) M(\mathbf{x}) + \alpha \left( \frac{n}{k} \right) \beta \left( \frac{n}{k} \right) N(\mathbf{x}) + \frac{Z_L(\mathbf{x})}{\sqrt{k}} + o \left( \frac{1}{\sqrt{k}} \right),$$

where the  $o$ -term is a.s. uniform on  $[0, T]^d$ . Using again the homogeneity of the functions  $L, M, N$  we deduce that

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

Now, by a Taylor series expansion, we have

$$\begin{aligned} \frac{1}{k} \sum_{j=1}^k K(a_j) \widehat{L}_{k,a_j}^\xi(\mathbf{x}) &= L^\xi(\mathbf{x}) \frac{1}{k} \sum_{j=1}^k K(a_j) + \alpha \left( \frac{n}{k} \right) \xi L^{\xi-1}(\mathbf{x}) M(\mathbf{x}) \frac{1}{k} \sum_{j=1}^k K(a_j) a_j^{-\rho} \\ &\quad + \alpha \left( \frac{n}{k} \right) \beta \left( \frac{n}{k} \right) \xi L^{\xi-1}(\mathbf{x}) N(\mathbf{x}) \frac{1}{k} \sum_{j=1}^k K(a_j) a_j^{-\rho-\rho'} \\ &\quad + \frac{\xi L^{\xi-1}(\mathbf{x})}{\sqrt{k}} \frac{1}{k} \sum_{j=1}^k K(a_j) \frac{1}{a_j} Z_L(a_j \mathbf{x}) + R_n^{(1)}(\mathbf{x}) + R_n^{(2)}(\mathbf{x}) + o \left( \frac{1}{\sqrt{k}} \right), \end{aligned}$$

where

$$\begin{aligned} R_n^{(1)}(\mathbf{x}) &:= \frac{\xi(\xi-1)}{2} L^{\xi-2}(\mathbf{x}) \frac{1}{k} \sum_{j=1}^k K(a_j) y_j^2, \\ R_n^{(2)}(\mathbf{x}) &:= \frac{\xi(\xi-1)(\xi-2)}{6} \frac{1}{k} \sum_{i=1}^k K(a_j) \{L(\mathbf{x}) + \theta y_j\}^{\xi-3} y_j^3, \quad \text{with } \theta \in (0, 1), \\ y_j &:= \alpha \left( \frac{n}{k} \right) a_j^{-\rho} M(\mathbf{x}) + \alpha \left( \frac{n}{k} \right) \beta \left( \frac{n}{k} \right) a_j^{-\rho-\rho'} N(\mathbf{x}) + \frac{1}{\sqrt{k} a_j} Z_L(a_j \mathbf{x}) + \frac{1}{a_j} o \left( \frac{1}{\sqrt{k}} \right). \end{aligned}$$

First we establish the convergence to zero of

$$\Delta_k := \left| \frac{1}{k} \sum_{j=1}^k K(a_j) \frac{Z_L(a_j \mathbf{x})}{a_j} - \int_0^1 K(u) \frac{Z_L(u \mathbf{x})}{u} du \right|.$$

To this aim it is instructive to realise that  $\{W_L(ax)/\sqrt{L(\mathbf{x})}; 0 \leq a \leq 1\}$  is a Wiener process. Using the fact that the sample paths of a Wiener process are a.s. Hölder continuous of order  $\gamma < 1/2$ , and by Lévy's global modulus of continuity, one has for any  $\varepsilon > 0$ ,  $T > \varepsilon$  and  $\delta \in (0, 1/2)$  that

$$\sup_{\mathbf{x} \in [\varepsilon, T]^d, (a,b) \in [0,1]^2} \frac{\left| \frac{W_L(ax)}{\sqrt{L(\mathbf{x})}} - \frac{W_L(b\mathbf{x})}{\sqrt{L(\mathbf{x})}} \right|}{|a-b|^{1/2-\delta}} \leq C \quad \text{a.s.}, \quad (10)$$

where  $C$  is a positive constant.

We have, with  $\tilde{a}_j$  being a value between  $a_j$  and  $a_{j+1}$ ,  $j = 1, \dots, k$ , that

$$\begin{aligned} \Delta_k &\leq \frac{1}{k+1} \sum_{j=1}^k \left| K(a_j) \frac{Z_L(a_j \mathbf{x})}{a_j} - K(\tilde{a}_j) \frac{Z_L(\tilde{a}_j \mathbf{x})}{\tilde{a}_j} \right| + \left| \int_0^{1/(k+1)} K(u) \frac{Z_L(u \mathbf{x})}{u} du \right| \\ &\quad + \frac{1}{k+1} \left| \frac{1}{k} \sum_{j=1}^k K(a_j) \frac{Z_L(a_j \mathbf{x})}{a_j} \right| \\ &=: T_1 + T_2 + T_3. \end{aligned}$$

Concerning  $T_1$ , we easily establish

$$\begin{aligned} T_1 &\leq \frac{1}{k+1} \sum_{j=1}^k K(a_j) \frac{1}{a_j} |Z_L(a_j \mathbf{x}) - Z_L(\tilde{a}_j \mathbf{x})| + \frac{1}{k+1} \sum_{j=1}^k \left| K(a_j) \frac{1}{a_j} - K(\tilde{a}_j) \frac{1}{\tilde{a}_j} \right| |Z_L(\tilde{a}_j \mathbf{x})| \\ &=: T_{1,1} + T_{1,2}. \end{aligned}$$

Now, by the definition of  $Z_L(\mathbf{x})$  and using the fact that  $\partial_i L(\mathbf{x})$  is homogeneous of order zero, we obtain the inequality

$$\begin{aligned} T_{1,1} &\leq \frac{1}{k+1} \sum_{j=1}^k K(a_j) \frac{1}{a_j} |W_L(a_j \mathbf{x}) - W_L(\tilde{a}_j \mathbf{x})| + \sum_{i=1}^d \frac{\partial_i L(\mathbf{x})}{k+1} \sum_{j=1}^k K(a_j) \frac{1}{a_j} |W_L(a_j x_i \mathbf{e}_i) - W_L(\tilde{a}_j x_i \mathbf{e}_i)| \\ &=: T_{1,1,1} + T_{1,1,2}. \end{aligned}$$

Using (10), we have for some small  $\iota > 0$ ,

$$T_{1,1,1} \leq \frac{\tilde{C}}{(k+1)^\iota} \frac{1}{k+1} \sum_{j=1}^k K(a_j) a_j^{-1/2-\delta-\iota},$$

where  $\tilde{C} > 0$  is a constant, and hence we conclude  $T_{1,1,1} = o(1)$  a.s. uniformly in  $\mathbf{x} \in [\varepsilon, T]^d$ . Using the continuity of  $\partial_i L(\mathbf{x})$ , we obtain in the same way  $T_{1,1,2} = o(1)$  a.s. uniformly in  $\mathbf{x} \in [\varepsilon, T]^d$ .

For  $T_{1,2}$  we write

$$\begin{aligned} T_{1,2} &\leq \frac{1}{k+1} \sum_{j=1}^k K(a_j) \left| \frac{1}{a_j} - \frac{1}{\tilde{a}_j} \right| |Z_L(\tilde{a}_j \mathbf{x})| + \frac{1}{k+1} \sum_{j=1}^k |K(a_j) - K(\tilde{a}_j)| \frac{1}{\tilde{a}_j} |Z_L(\tilde{a}_j \mathbf{x})| \\ &=: T_{1,2,1} + T_{1,2,2}. \end{aligned}$$

Again using (10), and with arguments similar to those used for  $T_{1,1}$ , we obtain  $T_{1,2,1} = o(1)$  a.s. uniformly in  $\mathbf{x} \in [\varepsilon, T]^d$ . Concerning  $T_{1,2,2}$ , use the fact that  $K$  is uniformly continuous over  $[0, 1]$  and obtain, for  $k$  large enough

$$T_{1,2,2} \leq \frac{\omega}{k+1} \sum_{j=1}^k \frac{1}{\tilde{a}_j} |Z_L(\tilde{a}_j \mathbf{x})|,$$

where  $\omega > 0$  can be chosen arbitrary small, and hence, again using (10) we have  $T_{1,2,2} = o(1)$  a.s. uniformly in  $\mathbf{x} \in [\varepsilon, T]^d$ . Also  $T_2$  and  $T_3$  can be analysed in a similar way, and allow us to conclude  $\Delta_k = o(1)$  a.s. uniformly in  $\mathbf{x} \in [\varepsilon, T]^d$  as  $k \rightarrow \infty$ .

After tedious calculations one can establish, provided  $\xi \geq 2$ ,

$$R_n^{(1)}(\mathbf{x}) = \alpha^2 \binom{n}{k} \frac{\xi(\xi-1)}{2} L^{\xi-2}(\mathbf{x}) M^2(\mathbf{x}) I(K, 2\rho) + o_{\mathbb{P}} \left( \frac{1}{\sqrt{k}} \right),$$

where the  $o_{\mathbb{P}}$  term is uniform in  $\mathbf{x} \in [\varepsilon, T]^d$ . Concerning  $R_n^{(2)}(\mathbf{x})$ , the cases  $\xi = 1$  and  $\xi = 2$ , are trivial since  $R_n^{(2)}(\mathbf{x})$  is exactly zero. In case  $\xi = 3$ , some straightforward but tedious computations lead to  $R_n^{(2)}(\mathbf{x}) = o_{\mathbb{P}} \left( \frac{1}{\sqrt{k}} \right)$ , uniformly in  $\mathbf{x} \in [\varepsilon, T]^d$ . Finally, if  $\xi \in \mathbb{N}$  such that  $\xi > 3$ , then, for  $n$  sufficiently large, we can use the following bound,

$$\begin{aligned} |L(\mathbf{x}) + \theta y_j|^{\xi-3} &\leq (L(\mathbf{x}) + |y_j|)^{\xi-3} \\ &\leq a_j^{(3-\xi)/2} \left( L(\mathbf{x}) + \alpha \binom{n}{k} |M(\mathbf{x})| + \alpha \binom{n}{k} \beta \binom{n}{k} |N(\mathbf{x})| + 2|Z_L(\mathbf{x})| + o(1) \right)^{\xi-3} \\ &=: \widehat{C} a_j^{(3-\xi)/2}, \end{aligned}$$

where  $\widehat{C}$  is independent from  $j$  and  $\mathbf{x}$ , which allows us to deduce that

$$R_n^{(2)}(\mathbf{x}) \leq \widehat{C} \frac{\xi(\xi-1)(\xi-2)}{6} \frac{1}{k} \sum_{j=1}^k K(a_j) a_j^{(3-\xi)/2} |y_j|^3 = o_{\mathbb{P}} \left( \frac{1}{\sqrt{k}} \right)$$

uniformly in  $\mathbf{x} \in [\varepsilon, T]^d$ . This achieves the proof of Theorem 2.

**Proof of Theorem 3.** The result can be established by using Theorem 2, the Skorohod representation, the homogeneity of the functions  $L, M, N$ , (10), and several Taylor series expansions.

## Acknowledgement

The authors thank the editor and the reviewers for their helpful comments which led to improvements of our first version of the paper. 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., Escobar-Bach, M., Goegebeur, Y., Guillou, A., 2016. Bias-corrected estimation of stable tail dependence function. *J. Multivariate Anal.* 143, 453–466.
- [2] Ciuperca, G., Mercadier, C., 2010. Semi-parametric estimation for heavy tailed distributions. *Extremes* 13, 55–87.
- [3] Escobar-Bach, M., Goegebeur, Y., Guillou, A., You, A., 2016. Bias-corrected and robust estimation of the bivariate stable tail dependence function. Accepted for publication in *Test*.
- [4] Fougères, A.L., de Haan, L., Mercadier, C., 2015. Bias correction in multivariate extremes. *Ann. Statist.* 43, 903–934.
- [5] Fraga Alves, M.I., Gomes, M.I., de Haan, L., 2003. A new class of semi-parametric estimators of the second order parameter. *Port. Math.* 60, 193–213.
- [6] Goegebeur, Y., Beirlant, J., de Wet, T., 2010. Kernel estimators of the second order parameter in extreme value statistics. *J. Statist. Plann. Inference* 140, 2632–2652.
- [7] de Haan, L., Ferreira, A., 2006. *Extreme Value Theory - An Introduction*, Springer, New York.
- [8] Huang, X., 1992. *Statistics of bivariate extremes*. PhD Thesis, Erasmus University Rotterdam, Tinbergen Institute Research series No. 22.

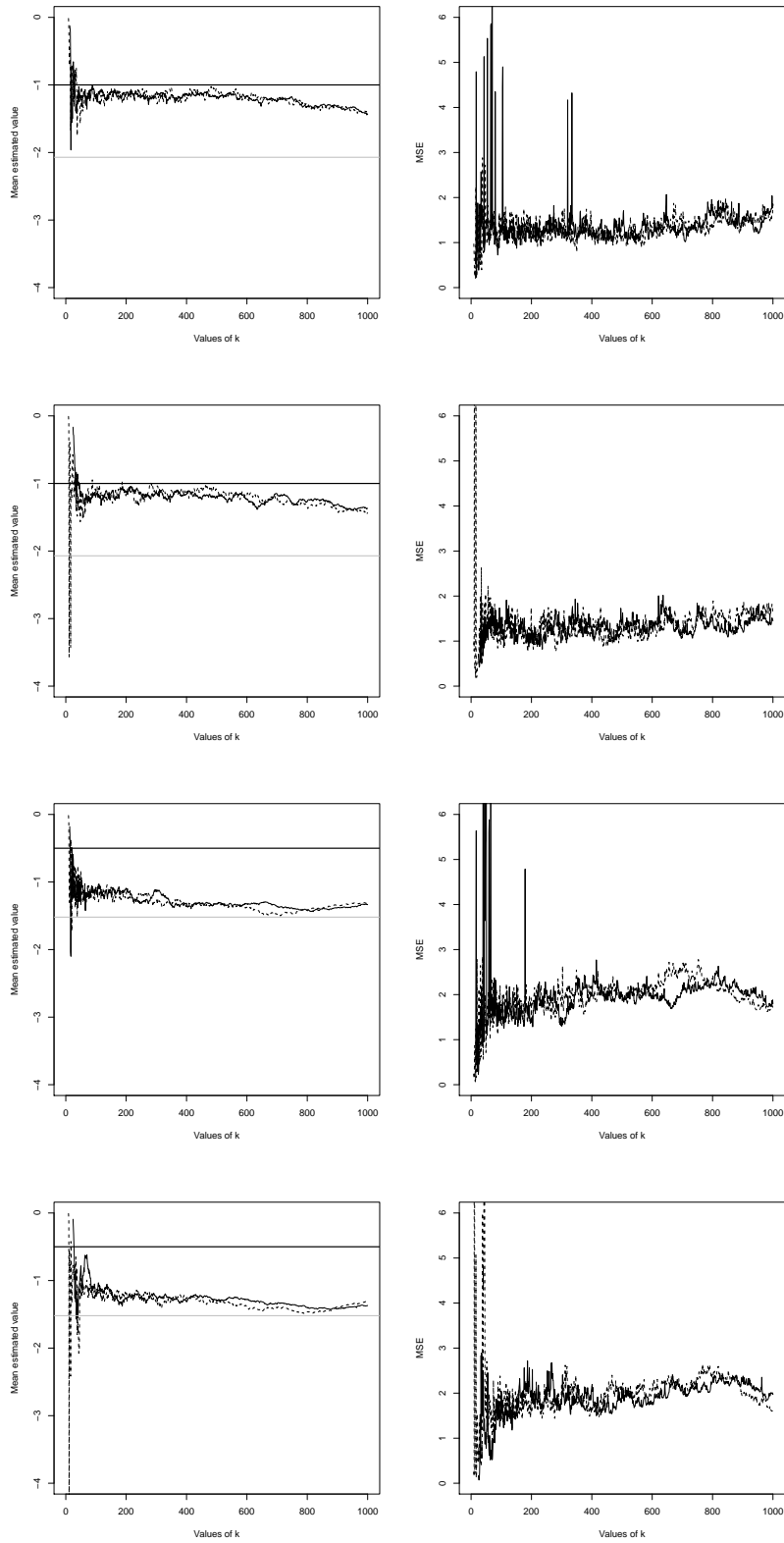


Figure 1: Comparison between  $\hat{\rho}_k$  at  $k = 990$  (horizontal grey solid line) and our power kernel estimator with  $a = r = 0.4$  in case of the logistic distribution (two first rows) and the BPII(4) distribution (two last rows) with  $(\xi_1, \xi_2) = (1, 1)$  (first and third rows) and  $(4, 4)$  (second and fourth rows), and  $\tau = 2$  (black solid line) and  $\tau = 10$  (black dashed line). All estimators are computed at  $\mathbf{x} = (0.5, 0.5)$ .

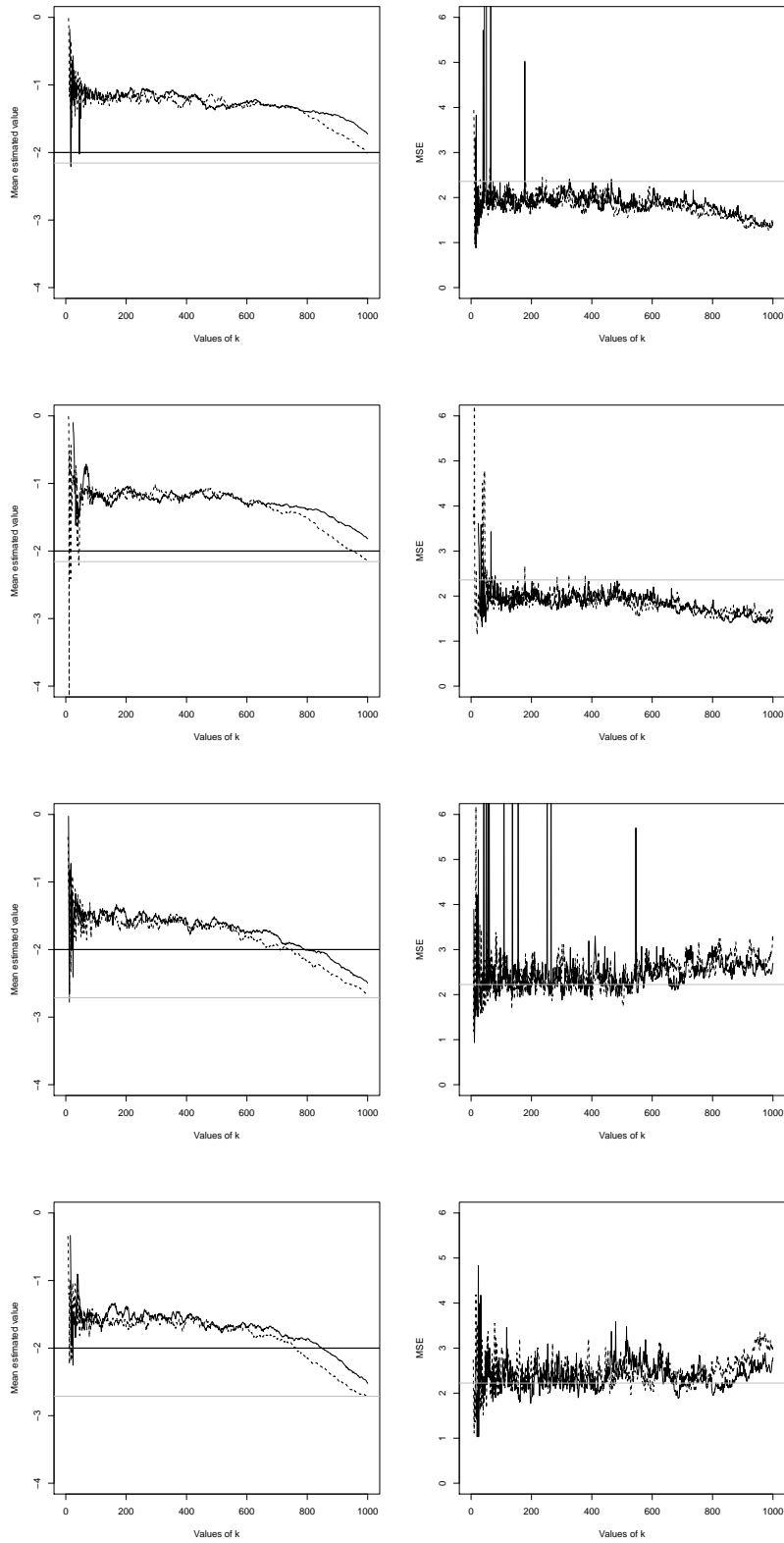


Figure 2: Bivariate Cauchy:  $\hat{\rho}_k$  at  $k = 990$  (horizontal grey solid line) and our power kernel estimator with  $(\xi_1, \xi_2) = (1, 1)$  (first row),  $(4, 4)$  (second row) for  $a = r = 0.4$ , whereas  $(\xi_1, \xi_2) = (1, 1)$  (third row),  $(4, 4)$  (bottom row) for  $a = r = 0.5$ . For all the plots,  $\tau = 2$  (black solid line) and  $\tau = 10$  (black dashed line). All estimators are computed at  $\mathbf{x} = (0.5, 0.5)$ .