Bias correction in conditional multivariate extremes

Mikael Escobar-Bach, Yuri Goegebeur, Armelle Guillou

Bias correction in conditional multivariate extremes

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

[(1)] LAREMA, UMR CNRS 6093, Université d'Angers, 2, Bd Lavoisier Angers Cedex 01, 49045, France
[(2)] Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark
[(3)] Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France

Abstract

We consider bias-corrected estimation of the stable tail dependence function in the regression context. To this aim, we first estimate the bias of a smoothed estimator of the stable tail dependence function, and then we subtract it from the estimator. The weak convergence, as a stochastic process, of the resulting asymptotically unbiased estimator of the conditional stable tail dependence function, correctly normalized, is established under mild assumptions, the covariate argument being fixed. The finite sample behaviour of our asymptotically unbiased estimator is then illustrated on a simulation study and compared to two alternatives, which are not bias corrected. Finally, our methodology is applied to a dataset of air pollution measurements.
Keywords: Bias correction, conditional stable tail dependence function, stochastic convergence.

1 Introduction

Most of the practical problems involving extreme events are inherently multivariate. Consequently, being able to estimate the extremal dependence between random variables is useful. To this aim, we can either use some extremal coefficients, that give a representative picture of the full dependency structure (see, e.g., Ledford and Tawn, 1997), or functions, such as the spectral distribution function or the stable tail dependence function, that provide a full characterization of the extreme dependence between variables. 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 the stable tail dependence function, which can be defined as follows.

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

$$\lim_{t\to\infty} t\mathbb{P}\left(1 - F_1(Y^{(1)}) \leqslant t^{-1}y_1 \text{ or } \ldots \text{ or } 1 - F_d(Y^{(d)}) \leqslant t^{-1}y_d\right) = L(y_1, \ldots, y_d), \tag{1}$$

provided that this limit exists. We refer to Huang (1992), and de Haan and Ferreira (2006) for more details.

Several estimators for L have been proposed in the literature, see, e.g., Huang (1992), Drees and Huang (1998), Fils-Villetard *et al.* (2008), Bücher *et al.* (2014), but as usual in the extreme value framework, the classical estimators are affected by bias, which often complicates their practical application. To solve this issue, Fougères *et al.* (2015) and Beirlant *et al.* (2016) have introduced bias-corrected estimators and they have established the main properties of their estimators as stochastic processes.

Taking care of the bias is important, but in practical applications, we are also often faced with the presence of covariates in addition to the random vector $(Y^{(1)}, \ldots, Y^{(d)})$. It is thus important to be able to estimate the stable tail dependence function when random covariates X are present, i.e., to consider the regression situation with a multivariate response. In that case, we want to describe the extremal dependence between the variables $(Y^{(1)}, \ldots, Y^{(d)})$ given some observed value x for the covariate $X \in \mathbb{R}^p$. Thus, the notion of conditional stable tail dependence function $L(\cdot|x)$ can be introduced and the classical framework (1) can be extended into

$$\lim_{t\to\infty} t\mathbb{P}\left(1 - F_1(Y^{(1)}|X) \leqslant t^{-1}y_1 \text{ or } \ldots \text{ or } 1 - F_d(Y^{(d)}|X) \leqslant t^{-1}y_d | X = x\right) = L(y_1, \ldots, y_d|x), \quad (2)$$

where $F_j(\cdot|x), j = 1, \ldots, d$, denote the continuous conditional distribution function of $Y^{(j)}$ given $X = x$. To the best of our knowledge, the estimation of the conditional stable tail dependence function has only been studied very recently by Escobar-Bach *et al.* (2018b), where a local estimator was proposed and its weak convergence as a stochastic process was established. In related work, Gardes and Girard (2015) introduced an estimator for the conditional tail copula and studied its finite dimensional convergence. However, being in the regression context, of course does not solve the bias problem of the estimator of $L(\cdot|x)$. Thus, combining bias-correction and regression will be the subject of this paper. As far as we know, this topic is completely new in the literature.

The remainder of the paper is organized as follows. In Section 2, we introduce our bias-corrected estimator of the conditional stable tail dependence function and we establish its weak convergence as a stochastic process, the covariate being fixed. Then in Section 3, we illustrate the performance of this estimator on a small simulation study where we compare it with two alternatives, that are not asymptotically unbiased. Section 4 is devoted to a data analysis of air pollution measurements. All the proofs are postponed to Section 5.

2 Estimators and convergence results

Denote $(Y, X) := (Y^{(1)}, \ldots, Y^{(d)}, X)$, a random vector satisfying (2), and let $(Y_1, X_1), \ldots, (Y_n, X_n)$, be independent copies of (Y, X), where X has density function f. We introduce a local estimator for L, based on an empirical version of the left-hand side of (2), for large values of t. As is usual in the extreme value context, we consider an intermediate sequence $k = k_n$, i.e., $k \to \infty$ as

$n \to \infty$ with $k/n \to 0$. Since the margins $F_j(\cdot|x)$ appearing in (2) are unknown in practice, we have to replace them by estimators such as the empirical kernel estimator

$$\widehat{F}_{n,j}(y|x) := \frac{\sum_{i=1}^n K_c(x - X_i)\mathbb{1}_{\{Y_i^{(j)} \leqslant y\}}}{\sum_{i=1}^n K_c(x - X_i)},\ j = 1, \ldots, d,$$

where $K_c(\cdot) := K(\cdot/c)/c^p$ with K a density function on $\mathbb{R}^p$, and $c := c_n$ is a positive non-random sequence satisfying $c_n \to 0$ as $n \to \infty$. Denote with $y := (y_1, \ldots, y_d)$ a vector of the positive quadrant $\mathbb{R}_+^d$. According to Escobar-Bach *et al.* (2018b)

$$\widehat{L}_k(y|x) := \frac{\frac{1}{k}\sum_{i=1}^n K_h\left(x - X_i\right)\mathbb{1}_{\left\{1-\widehat{F}_{n,1}(Y_i^{(1)}|X_i) \leqslant \frac{k}{n}\, y_1 \text{ or } \ldots \text{ or } 1-\widehat{F}_{n,d}(Y_i^{(d)}|X_i) \leqslant \frac{k}{n}\, y_d\right\}}}{\frac{1}{n}\sum_{i=1}^n K_h(x - X_i)}$$

with $h := h_n$ a positive non-random sequence satisfying $h_n \to 0$ as $n \to \infty$, is an estimator of the conditional stable tail dependence function $L(y|x)$. Note that in $\widehat{F}_{n,j}(y|x)$ and $\widehat{L}_k(y|x)$ the same kernel function K has been used, but they can of course be taken different. As in Escobar-Bach *et al.* (2018b), the bandwidths for $\widehat{F}_{n,j}$ and $\widehat{L}_k$ are though different.

The aim of the paper is to propose an asymptotically unbiased estimator for $L(\cdot|x)$. To the best of our knowledge this topic has not been considered previously in the literature in the regression context, on the contrary to the classical framework without covariates where we can mention the contributions of Fougères *et al.* (2015) and Beirlant *et al.* (2016).

The main results of the paper will be derived as stochastic weak convergence results for processes in $y \in [\varepsilon, T]^d$, for any $\varepsilon > 0$ and $T > \varepsilon$, but with the covariate argument fixed, meaning that we will focus our study only around one reference position $x_0 \in Int(S_X)$, the interior of the support S_X of f, assumed to be non-empty. To this aim, we need to introduce some conditions mentioned below and well-known in the extreme value framework. Let $\|.\|$ be some norm on $\mathbb{R}^p$, and denote by $B_x(\tau)$ the closed ball with respect to $\|.\|$, centered at x and radius $\tau > 0$. The event $A_{t,y}$ is defined for any $t \geqslant 0$ and $y \in \mathbb{R}_+^d$ as

$$A_{t,y} := \left\{1 - F_1(Y^{(1)}\,|X\,) \leqslant t^{-1}y_1 \text{ or } \ldots \text{ or } 1 - F_d(Y^{(d)}\,|X\,) \leqslant t^{-1}y_d\right\}.$$

First order condition: *The limit in (2) exists for all $x \in S_X$ and $y \in \mathbb{R}_+^d$, and the convergence is uniform on $[0,T]^d \times B_{x_0}(\tau)$ for any $T > 0$ and a $\tau > 0$.*

Second order condition: *For any $x \in S_X$ there exist a positive function $\alpha(\cdot|x)$ such that $\alpha(t|x) \to 0$ as $t \to \infty$ and a non null function $M(\cdot|x)$ such that for all $y \in \mathbb{R}_+^d$*

$$\lim_{t\to\infty} \frac{1}{\alpha(t|x)}\left\{t\mathbb{P}\left(A_{t,y}\,|X = x\,\right) - L(y|x)\right\} = M(y|x),$$

uniformly on $[0,T]^d \times B_{x_0}(\tau)$ for any $T > 0$ and a $\tau > 0$.

Third order condition: *For any $x \in S_X$ there exist a positive function $\beta(\cdot|x)$ such that $\beta(t|x) \to 0$ as $t \to \infty$ and a non null function $N(\cdot|x)$ such that for all $y \in \mathbb{R}_+^d$*

$$\lim_{t\to\infty} \frac{1}{\beta(t|x)} \left\{ \frac{t\mathbb{P}\left(A_{t,y}\,|X = x\right) - L(y|x)}{\alpha(t|x)} - M(y|x) \right\} = N(y|x),$$

uniformly on $[0,T]^d \times B_{x_0}(\tau)$ for any $T > 0$ and a $\tau > 0$, and where N is not a multiple of M.

Note that these assumptions imply that the functions $\alpha(\cdot|x)$ and $\beta(\cdot|x)$ 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 $L(\cdot|x)$, $M(\cdot|x)$ and $N(\cdot|x)$ have a homogeneity property, that is $L(ay|x) = aL(y|x)$, $M(ay|x) = a^{1-\rho}M(y|x)$ and $N(ay|x) = a^{1-\rho-\rho'}N(y|x)$ for all $a > 0$ and all $y \in \mathbb{R}_+^d$.

Due to the regression context, we need some Hölder-type conditions.

Assumption $(\mathcal{F}_m)$**.** *There exist $M_{F_j} > 0$ and $\eta_{F_j} > 0$ such that $|F_j(y|x) - F_j(y|z)| \leqslant M_{F_j}\|x - z\|^{\eta_{F_j}}$, for all $y \in \mathbb{R}$, all $(x,z) \in S_X \times S_X$ and $j = 1,\ldots,d$.*

Assumption $(\mathcal{D})$**.** *There exist $M_f > 0$ and $\eta_f > 0$ such that $|f(x) - f(z)| \leqslant M_f\|x - z\|^{\eta_f}$, for all $(x,z) \in S_X \times S_X$.*

Assumption $(\mathcal{L})$**.** *There exist $M_L > 0$ and $\eta_L > 0$ such that $|L(y|x) - L(y|z)| \leqslant M_L\|x - z\|^{\eta_L}$, for all $(x,z) \in B_{x_0}(\tau) \times B_{x_0}(\tau)$, $\tau > 0$, and $y \in [0,T]^d, T > 0$.*

Assumption $(\mathcal{A})$**.** *There exist $M_\alpha > 0$ and $\eta_\alpha > 0$ such that $|\alpha(t|x) - \alpha(t|z)| \leqslant M_\alpha\|x - z\|^{\eta_\alpha}$, for all $(x,z) \in S_X \times S_X$ and $t \geqslant 0$.*

Assumption $(\mathcal{B})$**.** *There exist $M_\beta > 0$ and $\eta_\beta > 0$ such that $|\beta(t|x) - \beta(t|z)| \leqslant M_\beta\|x - z\|^{\eta_\beta}$, for all $(x,z) \in S_X \times S_X$ and $t \geqslant 0$.*

Assumption $(\mathcal{M})$**.** *There exist $M > 0$ and $\eta_M > 0$ such that $|M(y|x) - M(y|z)| \leqslant M\|x - z\|^{\eta_M}$, for all $(x,z) \in B_{x_0}(\tau) \times B_{x_0}(\tau)$, $\tau > 0$, and $y \in [0,T]^d, T > 0$.*

Also a condition is required on the kernel function K.

Assumption $(\mathcal{K})$**.** *K is a bounded density function on $\mathbb{R}^p$ with support S_K included in the unit ball of $\mathbb{R}^p$ with respect to the norm $\|.\|$. Moreover, we assume that there exists $\delta, m > 0$ such that $B_0(\delta) \subset S_K$ and $K(u) \geqslant m$ for all $u \in B_0(\delta)$, and K belongs to the linear span (the set of finite linear combinations) of functions $k \geqslant 0$ satisfying the following property: the subgraph of k, $\{(s,u) : k(s) \geqslant u\}$, can be represented as a finite number of Boolean operations among sets of the form $\{(s,u) : q(s,u) \geqslant \varphi(u)\}$, where q is a polynomial on $\mathbb{R}^p \times \mathbb{R}$ and φ is an arbitrary real function.*

The latter assumption is common, and used already in, e.g., Giné and Guillou (2002) and Escobar-Bach *et al.* (2018a, 2018b), and allows to measure the discrepancy between the conditional distribution function $F_j(\cdot|x)$ and its empirical kernel version $\widehat{F}_{n,j}(\cdot|x)$.

2.1 Asymptotic result for $\widehat{L}_k(\cdot|x_0)$ under a third order condition

Escobar-Bach *et al.* (2018b) have established the weak convergence of $\widehat{L}_k(y|x_0)$ as a stochastic process in $y \in [0,T]^d$ and for a fixed covariate position $x_0 \in \mathbb{R}^p$, under the second order condition. In order to construct an asymptotically unbiased estimator for $L(\cdot|x_0)$, the third order condition is required and thus we need to know if under this new condition, a similar convergence result can be stated.

Theorem 2.1 *Assume the third order condition, $(x,y) \to N(y|x)$ continuous on $B_{x_0}(\tau) \times [0,T]^d$, with $B_{x_0}(\tau) \subset S_X$, and that there exists $b > 0$ with $f(x) \geqslant b$, $\forall x \in S_X \subset \mathbb{R}^p$ and f bounded. Under $(\mathcal{F}_m)$, $(\mathcal{D})$, $(\mathcal{L})$, $(\mathcal{A})$, $(\mathcal{B})$, $(\mathcal{M})$, $(\mathcal{K})$, and assuming that there exists an $\varepsilon > 0$ such that for n sufficiently large*

$$\inf_{x \in S_X} \lambda(\{u \in B_0(1) : x - cu \in S_X\}) > \varepsilon,$$

where λ denotes the Lebesgue measure, consider sequences $k \to \infty$, $h \to 0$ and $c \to 0$ as $n \to \infty$ such that $k/n \to 0$ with

$$\begin{aligned}
\sqrt{kh^p}\, h^{\min(\eta_f,\eta_L,\eta_\alpha)} &\longrightarrow 0,\\
\sqrt{kh^p}\, \alpha(n/k|x_0)\, h^{\min(\eta_M,\eta_\beta)} &\longrightarrow 0,\\
\sqrt{kh^p}\, \alpha(n/k|x_0) &\longrightarrow \infty,\\
\sqrt{kh^p}\, \alpha(n/k|x_0)\, \beta(n/k|x_0) &\longrightarrow \mu_1(x_0) \in \mathbb{R}_+,
\end{aligned}$$

and for some $q > 1$ and $0 < \eta < \min(\eta_{F_1}, ..., \eta_{F_d})$

$$n\sqrt{\frac{h^p}{k}} \max\left(\sqrt{\frac{|\log c|^q}{nc^p}}, c^\eta\right) \longrightarrow 0. \tag{3}$$

Then the process

$$\left\{\sqrt{kh^p}\left(\widehat{L}_k(y|x_0) - L(y|x_0) - \alpha\left(\frac{n}{k}\middle|x_0\right) M(y|x_0) - \alpha\left(\frac{n}{k}\middle|x_0\right)\beta\left(\frac{n}{k}\middle|x_0\right) N(y|x_0)\right), y \in [0,T]^d\right\}$$

weakly converges in $D([0,T]^d)$ towards a tight centered Gaussian process $\{B(y), y \in [0,T]^d\}$, for any $T > 0$, with covariance structure given by

$$Cov\big(B(y), B(y')\big) = \frac{\|K\|_2^2}{f(x_0)}\left(L(y|x_0) + L(y'|x_0) - L(y \vee y'|x_0)\right),$$

where $y, y' \in [0,T]^d$ and $\|K\|_2 := \sqrt{\int_{S_K} K^2(u)du}$.

2.2 Smoothed estimator for $L(\cdot|x_0)$

Inspired by the homogeneity of $L(\cdot|x_0)$, consider now the rescaled statistic

$$\widehat{L}_{k,a}(y|x_0) := \frac{1}{a}\widehat{L}_k(ay|x_0)$$

for a positive scale parameter a. Our uncorrected (in terms of bias) estimator for $L(\cdot|x_0)$ will be the following weighted version of the rescaled statistic, defined for any $r > 1$ and $\xi > 0$ as

$$\breve{L}_k(y|x_0) := \left(\frac{\int_1^r \widehat{L}_{k,a}^{\xi}(y|x_0)da}{r-1} \right)^{1/\xi}.$$

The weak convergence of this new estimator as a stochastic process is established in the following theorem.

Theorem 2.2 *Under the assumptions of Theorem 2.1 together with* $\sqrt{kh^p}\alpha^2(n/k|x_0) \to \mu_2(x_0) \in \mathbb{R}_+$, *for any* $r > 1$ *and* $\xi > 0$, *we have*

$$\sqrt{kh^p}\Big\{\breve{L}_k(y|x_0) - L(y|x_0) - \alpha\left(\tfrac{n}{k}\Big|x_0\right) M(y|x_0)c(r;\rho) - \alpha\left(\tfrac{n}{k}\Big|x_0\right)\beta\left(\tfrac{n}{k}\Big|x_0\right) N(y|x_0)c(r;\rho+\rho') - \tfrac{\alpha^2(n/k|x_0)}{2}\tfrac{M^2(y|x_0)}{L(y|x_0)}d(r,\xi;\rho)\Big\} \xrightarrow{d} \frac{1}{r-1}\int_1^r \frac{B(ay)}{a}da,$$

in $D([\varepsilon,T]^d)$, *for every* $\varepsilon > 0$ *and* $T > \varepsilon$, *where* B *is defined in Theorem 2.1 and*

$$\begin{aligned} c(r;\rho) &:= \frac{r^{1-\rho}-1}{(r-1)(1-\rho)}, \\ d(r,\xi;\rho) &:= [c(r;2\rho) - c^2(r;\rho)](\xi-1). \end{aligned}$$

Based on this result, in order to construct an asymptotically unbiased estimator for $L(\cdot|x_0)$, we need now to estimate ρ and $\alpha_k(y|x_0) := \alpha(n/k|x_0)M(y|x_0)$. This is the aim of the next section.

2.3 Estimation of ρ and $\alpha(n/k|x_0)M(y|x_0)$

Let $(\xi_1,\xi_2,\xi_3,\xi_4) \in \mathbb{R}_+^4$, $r_1 \neq r_2 > 1$ and $s > 0$. We propose to estimate ρ by

$$\breve{\rho}_k := 1 - \frac{1}{\log s}\log\left(\frac{\left(\frac{\int_1^{r_1}\widehat{L}_{k,a}^{\xi_1}(sy|x_0)da}{r_1-1}\right)^{1/\xi_1} - \left(\frac{\int_1^{r_2}\widehat{L}_{k,a}^{\xi_2}(sy|x_0)da}{r_2-1}\right)^{1/\xi_2}}{\left(\frac{\int_1^{r_1}\widehat{L}_{k,a}^{\xi_3}(y|x_0)da}{r_1-1}\right)^{1/\xi_3} - \left(\frac{\int_1^{r_2}\widehat{L}_{k,a}^{\xi_4}(y|x_0)da}{r_2-1}\right)^{1/\xi_4}} \right). \tag{4}$$

Theorem 2.3 *Under the assumptions of Theorem 2.1, and additionally assuming that* M *never vanishes except on the axes and that* $\sqrt{kh^p}\alpha^2(n/k|x_0) \to \mu_2(x_0) \in \mathbb{R}_+$, *for any* $(\xi_1,\xi_2,\xi_3,\xi_4) \in$

$\mathbb{R}_+^4$, $r_1 \neq r_2 > 1$ *and* $s > 0$, *we have*

$$\sqrt{kh^p}\alpha\left(\frac{n}{k}\Big|x_0\right)\left\{\breve{\rho}_k - \rho + \frac{\alpha(n/k|x_0)}{2\log s}\frac{M(y|x_0)}{L(y|x_0)}\left[s^{-\rho}\frac{d(r_1,\xi_1;\rho)-d(r_2,\xi_2;\rho)}{c(r_1;\rho)-c(r_2;\rho)} - \frac{d(r_1,\xi_3;\rho)-d(r_2,\xi_4;\rho)}{c(r_1;\rho)-c(r_2;\rho)}\right]\right.$$
$$\left. +\beta\left(\frac{n}{k}\Big|x_0\right)\frac{N(y|x_0)}{M(y|x_0)}\frac{s^{-\rho'}-1}{\log s}\frac{c(r_1;\rho+\rho')-c(r_2;\rho+\rho')}{c(r_1;\rho)-c(r_2;\rho)}\right\}$$
$$\xrightarrow{d} -\frac{\frac{s^{\rho-1}}{\log s}\left\{\frac{1}{r_1-1}\int_1^{r_1}\frac{B(asy)}{a}da - \frac{1}{r_2-1}\int_1^{r_2}\frac{B(asy)}{a}da\right\} - \frac{1}{\log s}\left\{\frac{1}{r_1-1}\int_1^{r_1}\frac{B(ay)}{a}da - \frac{1}{r_2-1}\int_1^{r_2}\frac{B(ay)}{a}da\right\}}{M(y|x_0)\left[c(r_1;\rho)-c(r_2;\rho)\right]},$$

in $D([\varepsilon,T]^d)$, *for every* $\varepsilon > 0$ *and* $T > \varepsilon$, *where* B *is defined in Theorem 2.1.*

Let $(\xi_5,\xi_6) \in \mathbb{R}_+^2$ and $r_3 \neq r_4 > 1$. To estimate $\alpha_k(y|x_0)$, we propose

$$\breve{\alpha}_k(y|x_0) := \frac{\left(\frac{\int_1^{r_3}\widehat{L}_{k,a}^{\xi_5}(y|x_0)da}{r_3-1}\right)^{1/\xi_5} - \left(\frac{\int_1^{r_4}\widehat{L}_{k,a}^{\xi_6}(y|x_0)da}{r_4-1}\right)^{1/\xi_6}}{c(r_3;\breve{\rho}_k)-c(r_4;\breve{\rho}_k)}. \tag{5}$$

In the sequel, we denote by $c'(r;\rho)$ the derivative of $c(r;\rho)$ with respect to ρ.

Theorem 2.4 *Under the assumptions of Theorem 2.3, we have*

$$\sqrt{kh^p}\alpha\left(\frac{n}{k}\Big|x_0\right)\left\{\frac{\breve{\alpha}_k(y|x_0)}{\alpha(n/k|x_0)M(y|x_0)} - 1 - \alpha\left(\frac{n}{k}\Big|x_0\right)\frac{M(y|x_0)}{L(y|x_0)}\frac{1}{2[c(r_3;\rho)-c(r_4;\rho)]}\right.$$
$$\times\left(d(r_3,\xi_5;\rho) - d(r_4,\xi_6;\rho) + \frac{c'(r_3;\rho)-c'(r_4;\rho)}{c(r_1;\rho)-c(r_2;\rho)}\left[\frac{s^{-\rho}}{\log s}\left(d(r_1,\xi_1;\rho)-d(r_2,\xi_2;\rho)\right)\right.\right.$$
$$\left.\left.-\frac{1}{\log s}\left(d(r_1,\xi_3;\rho)-d(r_2,\xi_4;\rho)\right)\right]\right)$$
$$-\beta\left(\frac{n}{k}\Big|x_0\right)\frac{N(y|x_0)}{M(y|x_0)}\frac{1}{c(r_3;\rho)-c(r_4;\rho)}$$
$$\left.\times\left(c(r_3;\rho+\rho')-c(r_4;\rho+\rho') + \frac{[c'(r_3;\rho)-c'(r_4;\rho)][c(r_1;\rho+\rho')-c(r_2;\rho+\rho')]}{c(r_1;\rho)-c(r_2;\rho)}\frac{s^{\rho'}-1}{\log s}\right)\right\}$$
$$\xrightarrow{d} \frac{1}{c(r_3;\rho)-c(r_4;\rho)}\frac{1}{M(y|x_0)}\left\{\frac{1}{r_3-1}\int_1^{r_3}\frac{B(ay)}{a}da - \frac{1}{r_4-1}\int_1^{r_4}\frac{B(ay)}{a}da\right.$$
$$-\frac{c'(r_4;\rho)-c'(r_3;\rho)}{c(r_1;\rho)-c(r_2;\rho)}\left[\frac{s^{\rho-1}}{\log s}\left(\frac{1}{r_1-1}\int_1^{r_1}\frac{B(asy)}{a}da - \frac{1}{r_2-1}\int_1^{r_2}\frac{B(asy)}{a}da\right)\right.$$
$$\left.\left.-\frac{1}{\log s}\left(\frac{1}{r_1-1}\int_1^{r_1}\frac{B(ay)}{a}da - \frac{1}{r_2-1}\int_1^{r_2}\frac{B(ay)}{a}da\right)\right]\right\}$$

in $D([\varepsilon,T]^d)$, *for every* $\varepsilon > 0$ *and* $T > \varepsilon$, *where* B *is defined in Theorem 2.1.*

2.4 Bias correction of $\breve{L}_k(y|x_0)$

Now we have all the ingredients to construct an asymptotically unbiased estimator for $L(\cdot|x_0)$ by removing from $\breve{L}_k(y|x_0)$ the bias term where $\alpha(n/k|x_0)M(y|x_0)$ together with the second order rate parameter ρ have been estimated externally, using the same intermediate sequence $\overline{k} = \overline{k}_n$, which is such that $k = o(\overline{k})$. This idea has been originally proposed by Gomes and co-authors (see, e.g., Gomes *et al.*, 2008, Caeiro *et al.*, 2009) in the univariate framework and has the advantage that the variance of the bias-corrected estimator and the uncorrected one is the same. Thus, we propose the following bias-corrected estimator for $L(\cdot|x_0)$

$$\overline{L}_{k,\overline{k}}(y|x_0) := \breve{L}_k(y|x_0) - \breve{\alpha}_{\overline{k}}(y|x_0)c(r;\breve{\rho}_{\overline{k}})\left(\frac{\overline{k}}{k}\right)^{\breve{\rho}_{\overline{k}}}. \tag{6}$$

Theorem 2.5 *Assume the third order condition, M never vanishes except on the axes, $(x,y) \to N(y|x)$ continuous on $B_{x_0}(\tau) \times [0,T]^d$, with $B_{x_0}(\tau) \subset S_X$, and that there exists $b > 0$ with $f(x) \geqslant b$, $\forall x \in S_X \subset \mathbb{R}^p$ and f bounded. Under $(\mathcal{F}_m)$, $(\mathcal{D})$, $(\mathcal{L})$, $(\mathcal{A})$, $(\mathcal{B})$, $(\mathcal{M})$, $(\mathcal{K})$, and assuming that there exists an $\varepsilon > 0$ such that for n sufficiently large*

$$\inf_{x \in S_X} \lambda(\{u \in B_0(1) : x - cu \in S_X\}) > \varepsilon,$$

consider sequences $k \to \infty$, $h \to 0$, $c \to 0$ as $n \to \infty$ and $\overline{k}$ such that $k = o(\overline{k})$, $\overline{k}/n \to 0$, and with

$$\begin{aligned}
\sqrt{\overline{k}h^p}\, h^{\min(\eta_f,\eta_L,\eta_\alpha)} &\longrightarrow 0,\\
\sqrt{\overline{k}h^p}\, \alpha(n/\overline{k}|x_0)\, h^{\min(\eta_M,\eta_\beta)} &\longrightarrow 0,\\
\sqrt{kh^p}\, \alpha(n/k|x_0) &\longrightarrow \infty,\\
\sqrt{\overline{k}h^p}\, \alpha(n/\overline{k}|x_0)\, \beta(n/\overline{k}|x_0) &\longrightarrow \mu_1(x_0) \in \mathbb{R}_+\\
\sqrt{\overline{k}h^p}\alpha^2(n/\overline{k}|x_0) &\longrightarrow \mu_2(x_0) \in \mathbb{R}_+
\end{aligned}$$

and for some $q > 1$ and $0 < \eta < \min(\eta_{F_1}, ..., \eta_{F_d})$

$$n\sqrt{\frac{h^p}{k}} \max\left(\sqrt{\frac{|\log c|^q}{nc^p}}, c^\eta\right) \longrightarrow 0.$$

Then we have

$$\begin{aligned}
\sqrt{kh^p}\left\{\overline{L}_{k,\overline{k}}(y|x_0) - L(y|x_0) - \alpha\left(\tfrac{n}{k}\middle|x_0\right)\beta\left(\tfrac{n}{k}\middle|x_0\right)N(y|x_0)c(r;\rho+\rho') - \tfrac{\alpha^2(n/k|x_0)}{2}\tfrac{M^2(y|x_0)}{L(y|x_0)}d(r,\xi;\rho)\right\}\\
\xrightarrow{d} \frac{1}{r-1}\int_1^r \frac{B(ay)}{a}da
\end{aligned}$$

in $D([\varepsilon,T]^d)$, for every $\varepsilon > 0$ and $T > \varepsilon$, where B is defined in Theorem 2.1.

Note that this bias-corrected estimator $\overline{L}_{k,\overline{k}}(\cdot|x_0)$ has the same asymptotic variance as the uncorrected estimator $\breve{L}_k(\cdot|x_0)$ (see Theorem 2.2).

3 Simulation study

Our aim in this section is to illustrate the bias-correcting effect in the estimation of $L(\cdot|x_0)$. We focus on dimensions $d = 2$ and $p = 1$. We consider the two models studied in Escobar-Bach *et al.* (2018b), which both satisfy our third order condition, together with Assumptions $(\mathcal{D})$, $(\mathcal{L})$, $(\mathcal{A})$, $(\mathcal{B})$, $(\mathcal{M})$ and $(\mathcal{F}_m)$. In particular, these models are the following:

- Model 1: The bivariate Student distribution with density function

$$f_{(Y^{(1)},Y^{(2)})}(y_1, y_2) = \frac{\sqrt{1-\theta^2}}{2\pi}\left(1 + \frac{y_1^2 - 2\theta y_1 y_2 + y_2^2}{\nu}\right)^{-\frac{\nu+2}{2}}, \quad (y_1, y_2) \in \mathbb{R}^2,$$

where θ is the Pearson correlation coefficient. The stable tail dependence function can be described as

$$L(y_1, y_2|\theta) = y_1 F_{\nu+1}\left(\frac{(y_1/y_2)^{1/\nu} - \theta}{\sqrt{1-\theta^2}}\sqrt{\nu+1}\right) + y_2 F_{\nu+1}\left(\frac{(y_2/y_1)^{1/\nu} - \theta}{\sqrt{1-\theta^2}}\sqrt{\nu+1}\right),$$

where $F_{\nu+1}$ is the distribution function of the univariate Student distribution with $(\nu+1)$ degrees of freedom. Also

$$\begin{aligned}
M(y_1, y_2|\theta) &= C_1\left[y_1^{2/\nu+1} F_{\nu+3}\left(\frac{(y_1/y_2)^{1/\nu} - \theta}{\sqrt{1-\theta^2}}\sqrt{\nu+3}\right) + y_2^{2/\nu+1} F_{\nu+3}\left(\frac{(y_2/y_1)^{1/\nu} - \theta}{\sqrt{1-\theta^2}}\sqrt{\nu+3}\right)\right],\\
N(y_1, y_2|\theta) &= C_2\left[y_1^{4/\nu+1} F_{\nu+5}\left(\frac{(y_1/y_2)^{1/\nu} - \theta}{\sqrt{1-\theta^2}}\sqrt{\nu+5}\right) + y_2^{4/\nu+1} F_{\nu+5}\left(\frac{(y_2/y_1)^{1/\nu} - \theta}{\sqrt{1-\theta^2}}\sqrt{\nu+5}\right)\right],\\
C_1 &:= -\frac{\nu^{2/\nu+1}\pi^{1/\nu}(\nu+1)}{2(\nu+2)}\left(\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right)^{2/\nu},\\
C_2 &:= \frac{\nu^{4/\nu+1}\pi^{2/\nu}(\nu+1)(\nu+3)}{8(\nu+4)}\left(\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right)^{4/\nu},\\
\alpha(t|\theta) &= t^{-2/\nu},\\
\beta(t|\theta) &= t^{-2/\nu}.
\end{aligned}$$

We set $\theta = X$, where X is uniformly distributed on $[0, 1]$. In the simulations, we use $\nu = 4$, which corresponds to $\rho = \rho' = -1/2$;

- Model 2: a particular case of the Archimax bivariate copulas introduced in Capéraà *et al.* (2000) and also mentioned in Fougères *et al.* (2015), namely:

$$C(y_1, y_2|x) = \left\{1 + L(y_1^{-1} - 1, y_2^{-1} - 1|x)\right\}^{-1},$$

where we use for L the asymmetric logistic stable tail dependence function defined by

$$L(y_1, y_2|x) = (1-t_1)y_1 + (1-t_2)y_2 + \left[(t_1 y_1)^{\theta_x} + (t_2 y_2)^{\theta_x}\right]^{1/\theta_x},$$

where $0 \leqslant t_1, t_2 \leqslant 1$, and $\theta_x := \min(1/x, 100)$, with the covariate X uniformly distributed on $[0,1]$. The marginal distributions are taken to be unit Fréchet. For this model

$$\begin{aligned}
M(y_1,y_2|x) &= y_1^2\partial_1 L(y_1,y_2|x) + y_2^2\partial_2 L(y_1,y_2|x) - L^2(y_1,y_2|x),\\
N(y_1,y_2|x) &= \partial_1 L(y_1,y_2|x)(y_1^3 - 2L(y_1,y_2|x)y_1^2) + \partial_2 L(y_1,y_2|x)(y_2^3 - 2L(y_1,y_2|x)y_2^2)\\
&\quad + \frac{1}{2}\partial_{11} L(y_1,y_2|x)y_1^4 + \frac{1}{2}\partial_{22} L(y_1,y_2|x)y_2^4 + \partial_{12} L(y_1,y_2|x)y_1^2 y_2^2 + L^3(y_1,y_2|x),\\
\alpha(t|x) &= t^{-1},\\
\beta(t|x) &= t^{-1}.
\end{aligned}$$

Hence $\rho = \rho' = -1$. In the simulations, different values for the pair (t_1, t_2) have been tried but the results seem to be not too much influenced by them, thus we exhibit only the results in case $(t_1, t_2) = (0.4, 0.6)$ which corresponds to an asymmetric tail dependence function.

For each model, we simulate 500 samples of size 1000, and we compare three estimators of $L(\cdot|x_0)$: the two uncorrected estimators, $\widehat{L}_k(\cdot|x_0)$ and its smoothed version $\breve{L}_k(\cdot|x_0)$, and our bias-corrected estimator $\overline{L}_{k,\overline{k}}(\cdot|x_0)$, for three positions $x_0 = 0.3, 0.5$ and 0.7. Concerning the kernel, we always use the bi-quadratic function

$$K(u) := \frac{15}{16}(1-u^2)^2 \mathbb{1}_{\{u\in[-1,1]\}}.$$

Each estimator requires the selection of some tuning parameters. This will be done as follows.

For the uncorrected estimator $\widehat{L}_k(\cdot|x_0)$ of Escobar-Bach *et al.* (2018b), we follow their approach, i.e., we use their cross-validation criterion for both bandwidth parameters c_1 and c_2, corresponding to the marginals approximation, and for the sequence h, we use

$$h = \frac{\min(c_1,c_2)}{|\log(\min(c_1,c_2))|^{1.1}}\,\frac{k}{n},$$

coming from condition (3), as described in their paper.

For the uncorrected smoothed estimator $\breve{L}_k(\cdot|x_0)$, the pair (r,ξ) is selected in a data-driven way using the homogeneity of the function $L(y|x_0)$, namely, for all y and k

$$(r^*,\xi^*) := \underset{(r,\xi)\in R\times E}{\operatorname{argmin}} \sum_{t\in T}\left(\breve{L}_k(ty|x_0) - t\breve{L}_k(y|x_0)\right)^2,$$

where $R := \{1.1, 1.2, \ldots, 2\}$, $E := \{1,2,3\}$ and $T := \{1/3, 2/3, 1, 4/3, 5/3\}$. The grids of values are selected after an extensive simulation study.

For the bias-corrected estimator $\overline{L}_{k,\overline{k}}(\cdot|x_0)$, also a data-driven method has been used for all the parameters involved. More precisely, $\overline{L}_{k,\overline{k}}(\cdot|x_0)$ defined in (6) is based on the uncorrected smoothed estimator $\breve{L}_k(\cdot|x_0)$ computed with (r^*,ξ^*) from which we remove the bias, based on

estimates $\breve{\alpha}_{\overline{k}}(\cdot|x_0)$ and $\breve{\rho}_{\overline{k}}$, derived according to the following algorithm:
Step 1. Let $y^* = (0.5, 0.5)$, $s = 0.4$ and $\overline{k} = \lfloor 0.99n \rfloor$ as suggested by Fougères *et al.* (2015);
Step 2. Note that $\breve{\rho}_{\overline{k}}$ is an estimate of ρ, and as such is independent of y. Define $\mathcal{R} := \{(r_1, r_2) \in R^2 : r_1 \neq r_2\}$, $\Xi := \{(\xi_1, \xi_2, \xi_3, \xi_4) \in E^4 : \xi_1 = \xi_3, \xi_2 = \xi_4\}$ and denote $\breve{\rho}_{\overline{k}}(y^*, r_1, r_2, \xi_1, \xi_2, \xi_3, \xi_4) := \breve{\rho}_{\overline{k}}$ as in (4) for all $(r_1, r_2) \in \mathcal{R}$, $(\xi_1, \xi_2, \xi_3, \xi_4) \in \Xi$ and $y = y^*$. Then, find $(r_1^*, r_2^*, \xi_1^*, \xi_2^*)$ the values of $(\breve{r}_1, \breve{r}_2, \breve{\xi}_1, \breve{\xi}_2) \in \mathcal{R} \times E^2$ minimizing the criterion

$$\sum_{(r_1,r_2,\xi_1,\xi_2)\in\mathcal{R}\times E^2} \left(\breve{\rho}_{\overline{k}}\left(y^*, \breve{r}_1, \breve{r}_2, \breve{\xi}_1, \breve{\xi}_2, \breve{\xi}_1, \breve{\xi}_2\right) - \breve{\rho}_{\overline{k}}\left(y^*, r_1, r_2, \xi_1, \xi_2, \xi_1, \xi_2\right)\right)^2.$$

The estimate $\breve{\rho}_{\overline{k}}$ in (6) is finally computed as $\breve{\rho}_{\overline{k}}(y^*, r_1^*, r_2^*, \xi_1^*, \xi_2^*, \xi_1^*, \xi_2^*)$;
Step 3. Let $\breve{\alpha}_{\overline{k}}(\cdot, r_3, r_4, \xi_5, \xi_6|x_0) := \breve{\alpha}_{\overline{k}}(\cdot|x_0)$ as defined in (5). We use the homogeneity of $M(\cdot|x_0)$ in order to select the parameters (r_3, r_4, ξ_5, ξ_6). More precisely, $\breve{\alpha}_{\overline{k}}(\cdot|x_0)$ in (6) is computed as $\breve{\alpha}_{\overline{k}}(\cdot, r_3^*, r_4^*, \xi_5^*, \xi_5^*|x_0)$ where

$$(r_3^*, r_4^*, \xi_5^*) := \underset{(r_3,r_4,\xi_5)\in\mathcal{R}\times E}{\operatorname{argmin}} \sum_{t\in T} \left(\breve{\alpha}_{\overline{k}}(ty^*, r_3, r_4, \xi_5, \xi_5|x_0) - t^{1-\breve{\rho}_{\overline{k}}}\, \breve{\alpha}_{\overline{k}}(y^*, r_3, r_4, \xi_5, \xi_5|x_0)\right)^2.$$

In the latter, $\breve{\rho}_{\overline{k}}$ is the value obtained in Step 2.

In Figure 1, we show the sample mean (left) and the empirical mean squared error (MSE, right) of $\widehat{L}_k(y|x_0)$ (dotted line), $\breve{L}_k(y|x_0)$ (dashed line) and $\overline{L}_{k,\overline{k}}(y|x_0)$ (full line) as a function of k in case of Model 1 with $x_0 = 0.3$ and four possible values of y, corresponding to the different rows: from the top to the bottom, $y = (0.2, 0.8), (0.4, 0.6), (0.6, 0.4)$ and $(0.8, 0.2)$, respectively. The horizontal line on the left panel represents the true value of $L(y|x_0)$. Figures 2 and 3 are constructed similarly, but for $x_0 = 0.5$ and 0.7, respectively, whereas Figures 4 to 6 concern Model 2 and the same values of x_0 and y. Based on these simulations, we can draw the following conclusions:

- Our estimator $\overline{L}_{k,\overline{k}}(y|x_0)$ clearly outperforms the two alternatives. In terms of bias, the sample means show very stable paths as a function of k, close to the true value. In terms of MSE, it is still competitive, almost always better than $\widehat{L}_k(y|x_0)$ and $\breve{L}_k(y|x_0)$, or otherwise at least similar, and again very stable as a function of k. Those are very nice features since in our case, the selection of k is not very crucial, while it is for $\widehat{L}_k$ and $\breve{L}_k$.
- For Model 2, the estimation is more difficult for y far away from the diagonal, whereas for Model 1, it does not depend on y. Also, the performance of our bias-corrected estimator $\overline{L}_{k,\overline{k}}(y|x_0)$ does not seem to depend on the position in the covariate space.

4 Application to air pollution data

In this section, we illustrate the practical applicability of our bias-corrected estimator on a dataset of air pollution measurements. We consider the data collected by the United States Environmental Protection Agency (EPA), publicly available at https://aqsdr1.epa.gov/aqsweb/aqstmp/airdata

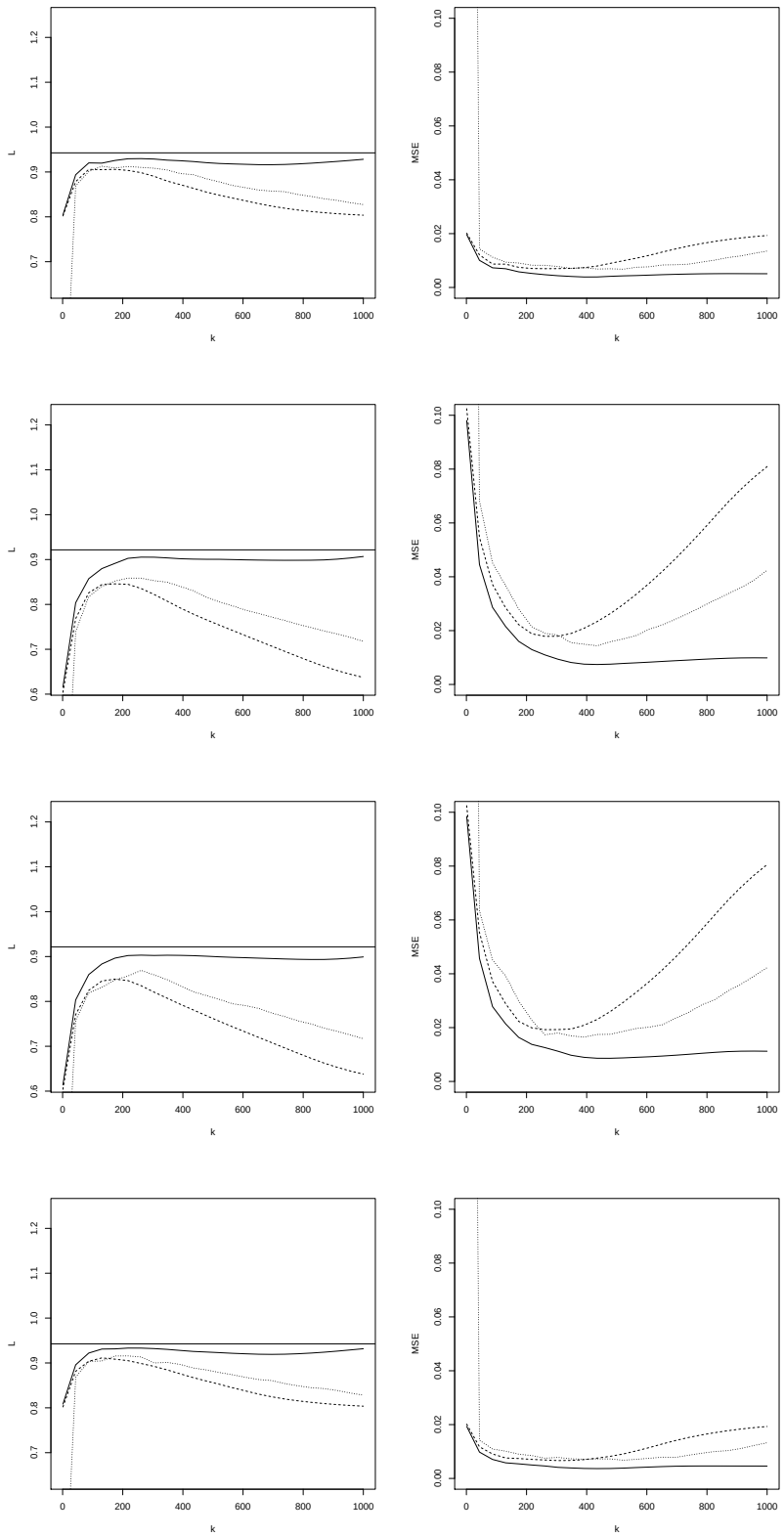

Figure 1: Model 1: Mean (left) and MSE (right) of three estimators of $L(y|0.3)$: $\widehat{L}_k(y|0.3)$ (dotted line), $\breve{L}_k(y|0.3)$ (dashed line), $\overline{L}_{k,\overline{k}}(y|0.3)$ (full line) as a function of k for different values of y corresponding to each row: $y = (0.2, 0.8), (0.4, 0.6), (0.6, 0.4), (0.8, 0.2)$. The horizontal line on the left panel corresponds to the true value of $L(y|0.3)$.

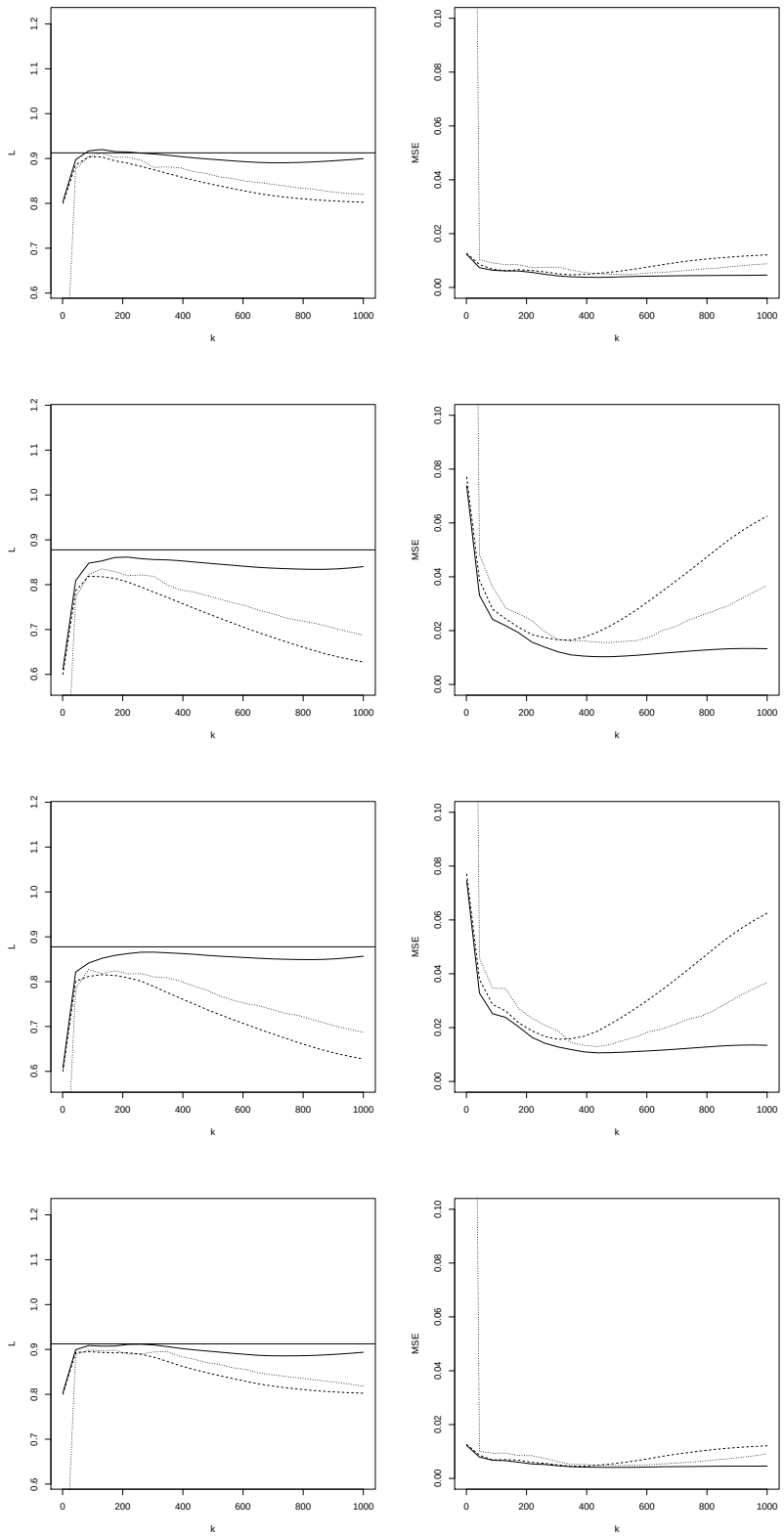

Figure 2: Model 1: Mean (left) and MSE (right) of three estimators of $L(y|0.5)$: $\widehat{L}_k(y|0.5)$ (dotted line), $\breve{L}_k(y|0.5)$ (dashed line), $\overline{L}_{k,\overline{k}}(y|0.5)$ (full line) as a function of k for different values of y corresponding to each row: $y = (0.2, 0.8), (0.4, 0.6), (0.6, 0.4), (0.8, 0.2)$. The horizontal line on the left panel corresponds to the true value of $L(y|0.5)$.

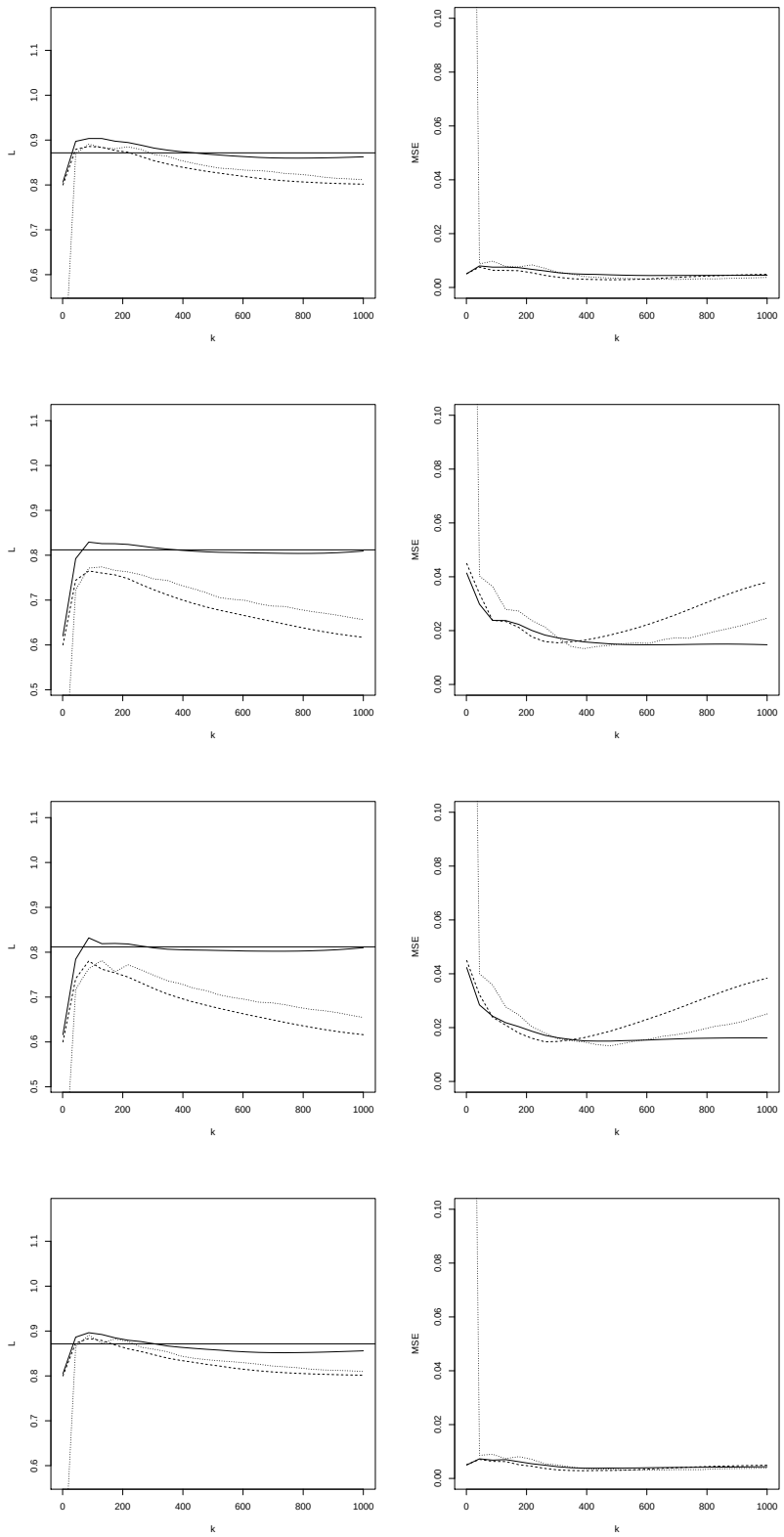

Figure 3: Model 1: Mean (left) and MSE (right) of three estimators of $L(y|0.7)$: $\widehat{L}_k(y|0.7)$ (dotted line), $\breve{L}_k(y|0.7)$ (dashed line), $\overline{L}_{k,\overline{k}}(y|0.7)$ (full line) as a function of k for different values of y corresponding to each row: $y = (0.2, 0.8), (0.4, 0.6), (0.6, 0.4), (0.8, 0.2)$. The horizontal line on the left panel corresponds to the true value of $L(y|0.7)$.

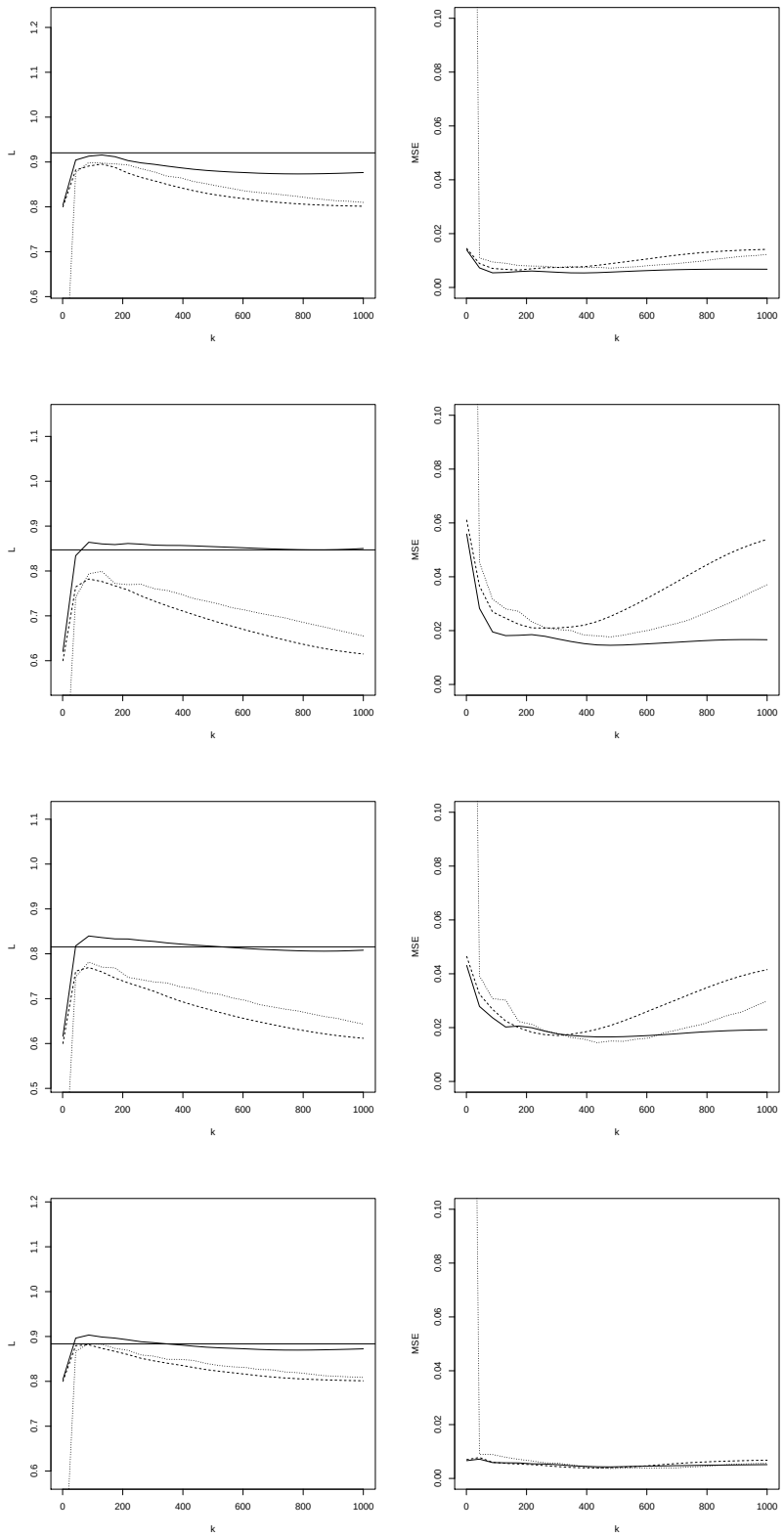

Figure 4: Model 2: Mean (left) and MSE (right) of three estimators of $L(y|0.3)$: $\widehat{L}_k(y|0.3)$ (dotted line), $\breve{L}_k(y|0.3)$ (dashed line), $\overline{L}_{k,\overline{k}}(y|0.3)$ (full line) as a function of k for different values of y corresponding to each row: $y = (0.2, 0.8), (0.4, 0.6), (0.6, 0.4), (0.8, 0.2)$. The horizontal line on the left panel corresponds to the true value of $L(y|0.3)$.

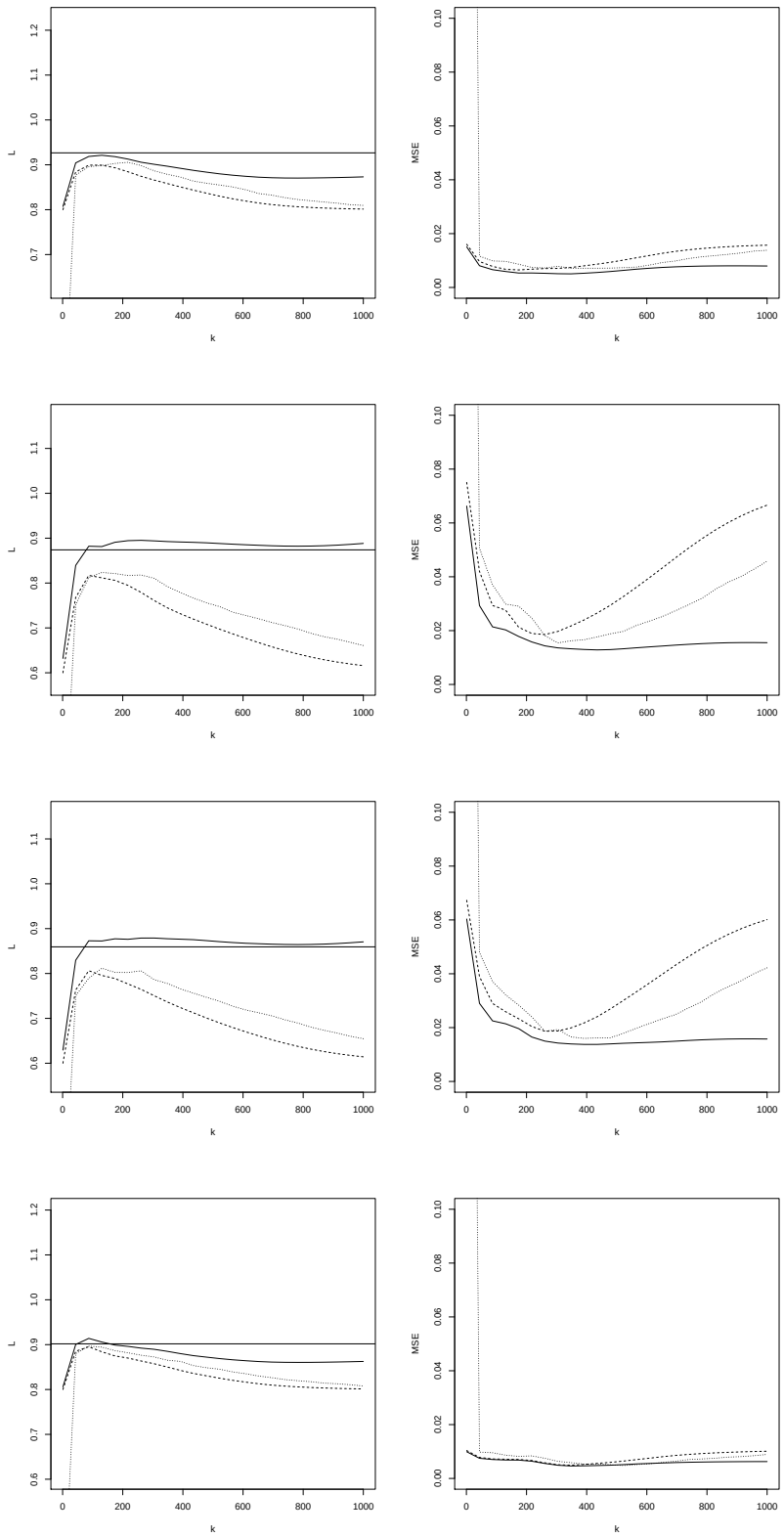

Figure 5: Model 2: Mean (left) and MSE (right) of three estimators of $L(y|0.5)$: $\widehat{L}_k(y|0.5)$ (dotted line), $\breve{L}_k(y|0.5)$ (dashed line), $\overline{L}_{k,\overline{k}}(y|0.5)$ (full line) as a function of k for different values of y corresponding to each row: $y = (0.2, 0.8), (0.4, 0.6), (0.6, 0.4), (0.8, 0.2)$. The horizontal line on the left panel corresponds to the true value of $L(y|0.5)$.

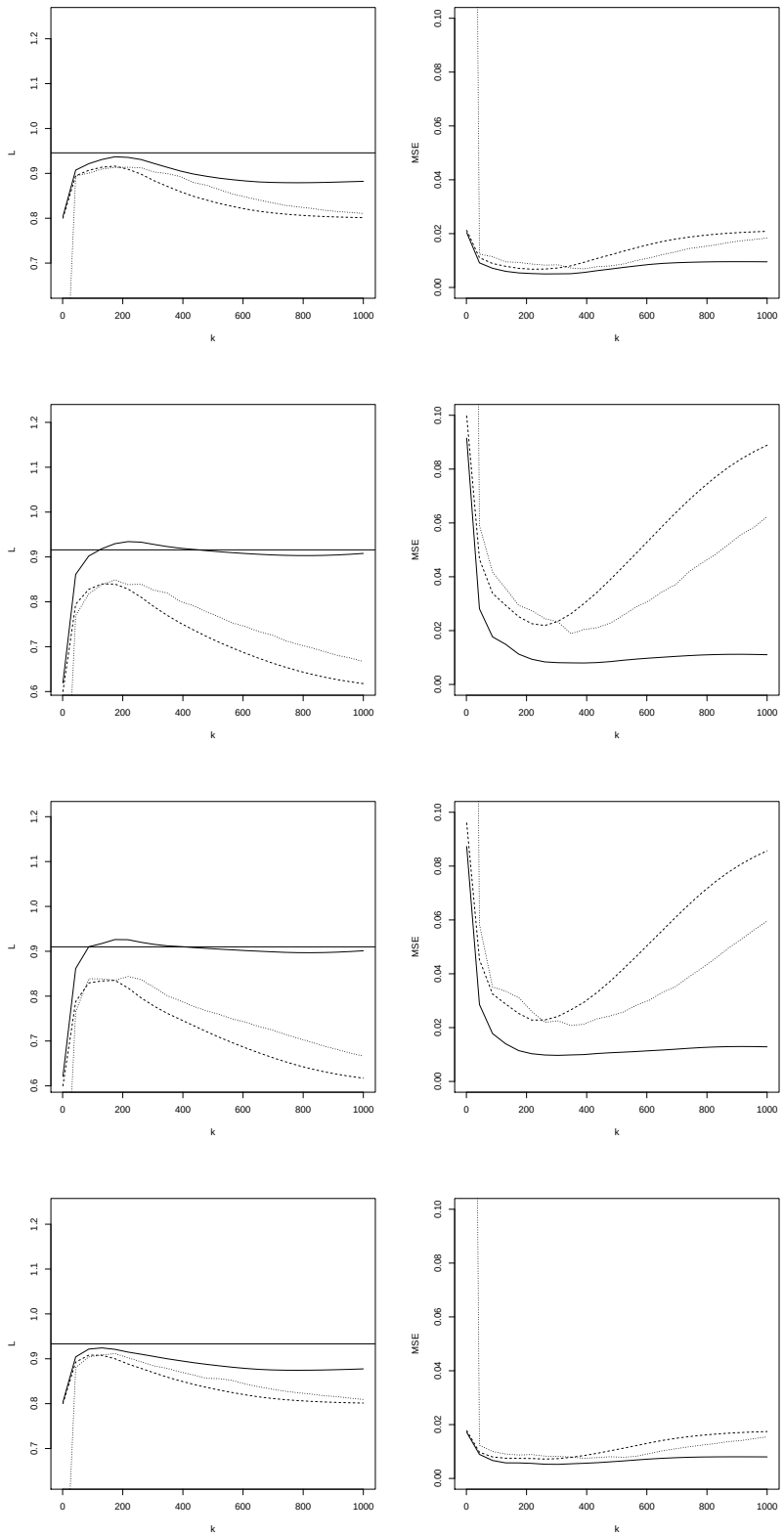

Figure 6: Model 2: Mean (left) and MSE (right) of three estimators of $L(y|0.7)$: $\widehat{L}_k(y|0.7)$ (dotted line), $\breve{L}_k(y|0.7)$ (dashed line), $\overline{L}_{k,\overline{k}}(y|0.7)$ (full line) as a function of k for different values of y corresponding to each row: $y = (0.2, 0.8), (0.4, 0.6), (0.6, 0.4), (0.8, 0.2)$. The horizontal line on the left panel corresponds to the true value of $L(y|0.7)$.

/download_files.html. The dataset contains daily measurements of, among others, maximum temperature, ground-level ozone, carbon monoxide and particulate matter concentrations, for the period 1999 to 2013, and this for stations spread over the U.S. Monitoring levels of these pollutants is of crucial importance, as extreme temperature and high levels of pollutants like ground-level ozone and particulate matter pose a major threat to human health. We estimate the stable tail dependence function for the variables temperature and ozone concentration, conditional on time and location, where the latter is expressed by latitude and longitude. In the estimation, the covariates are standardised to the interval $[0, 1]$, and the tuning parameters are selected with the algorithm described in Section 3. In order to keep the computational time requirements under control, the tuning parameters selected at steps 1. to 3. of the algorithm are computed with a random sampling of size $\lceil 0.1n \rceil$ where $n = 127328$ refers to the initial sample size. As kernel function K^*, we use the following generalisation of the bi-quadratic kernel K:

$$K^*(x_1, x_2, x_3) := \prod_{i=1}^{3} K(x_i),$$

where x_1, x_2, x_3, refer to the covariates time, latitude and longitude, respectively, in standardised form. Note that K^* has as support the unit ball with respect to the max-norm on $\mathbb{R}^3$.

We report here only the results at two different time points, January 15, 2007 and June 15, 2007, and for two locations, Fresno and Los Angeles (both in California). In Figure 7, we show the estimates median$\{\widetilde{L}_k(t, 1-t|x), k = n/4, \cdots, n/2\}$, with a range of $k-$values based on 25 equally spaced integers, where $\widetilde{L}_k$ is either $\overline{L}_{k,\overline{k}}$ (full line), $\widehat{L}_k$ (dotted line) or $\breve{L}_k$ (dashed line), for the cities Fresno (top row) and Los Angeles (bottom row) on January 15, 2007 (first column) and June 15, 2007 (second column). For both stations, the bias-corrected estimate for the stable tail dependence function indicates a stronger extreme dependence between temperature and ozone concentration in winter than in summer. In winter the extreme dependence in Fresno is stronger than in Los Angeles. The results obtained with the uncorrected estimators $\widehat{L}_k$ and $\breve{L}_k$ are typically similar to each other, and correspond more or less with the analysis reported in Escobar-Bach *et al.* (2018b). The estimate $\overline{L}_{k,\overline{k}}$ differs considerably from $\widehat{L}_k$ and $\breve{L}_k$ for Fresno, winter and Los Angeles, summer. Note that in these cases, the bias-corrected estimate tends to be higher than the uncorrected estimates, indicating a weaker extremal dependence. This was also observed in the simulation experiment, where the bias-corrected estimator tends to be larger (and closer to the true value) than the uncorrected estimators. The observed discrepancy indicates that estimation of tail dependence between temperature and ozone concentration can suffer from bias, and therefore it is recommended to use the bias-corrected estimator in order to get a better estimate of the stable tail dependence function.

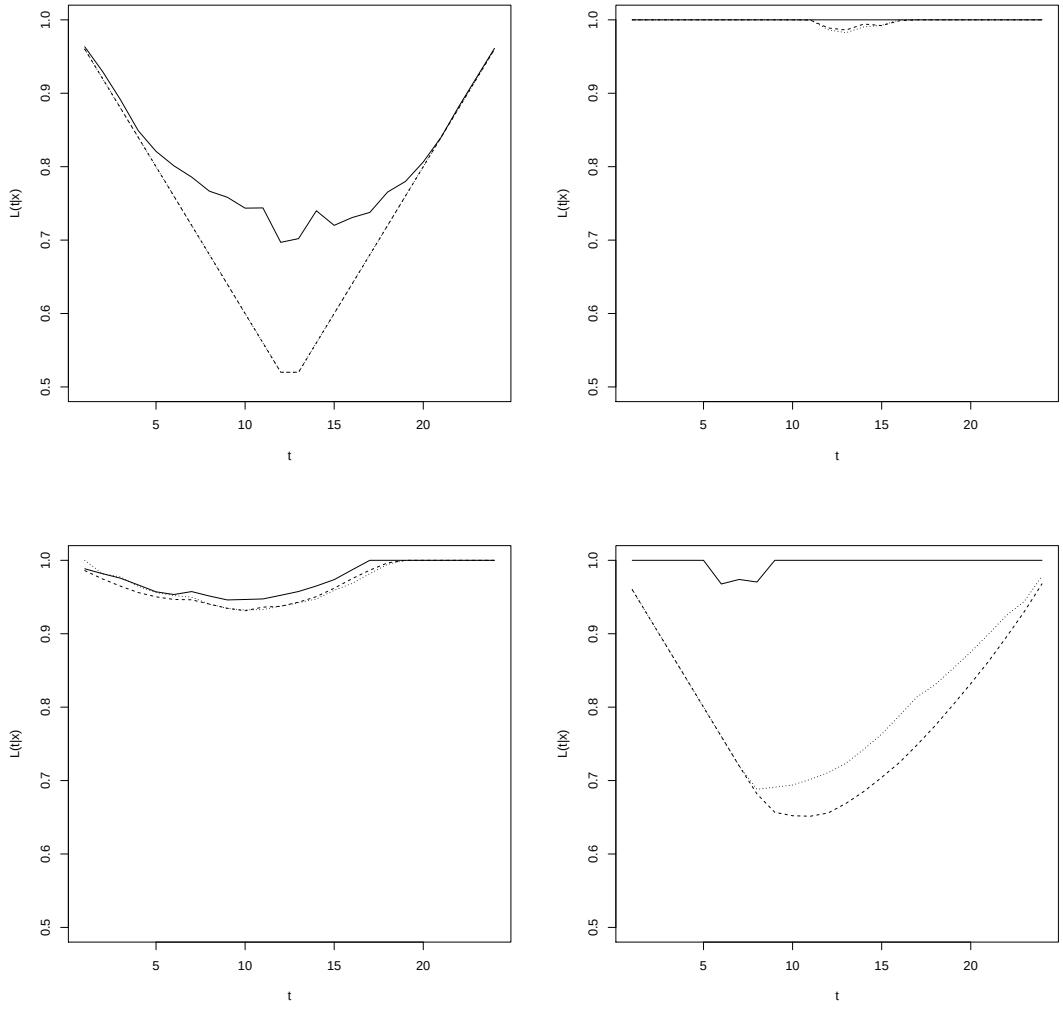

Figure 7: Air pollution data : Estimates of median$\{\widetilde{L}_k(t, 1-t|x), k = n/4, \cdots n/2\}$, with a range of $k-$values based on 25 equally spaced integers, for Fresno (top) and Los Angeles (bottom) on January 15, 2007 (first column) and June 15, 2007 (second column).

5 Proofs

5.1 Proof of Theorem 2.1

We follow the lines of proof of Theorems 2.1 and 2.3 in Escobar-Bach *et al.* (2018b). The only difference lies in the fact that, under the third order condition, we have

$$\begin{aligned}
&\frac{n}{k}\mathbb{P}\left(A_{n/k,y}|X=x_0-hu\right)-L(y|x_0-hu)\\
&\quad=\quad \alpha\left(\frac{n}{k}\Big|x_0-hu\right)M(y|x_0-hu)+\alpha\left(\frac{n}{k}\Big|x_0-hu\right)\beta\left(\frac{n}{k}\Big|x_0-hu\right)N(y|x_0-hu)\\
&+\alpha\left(\frac{n}{k}\Big|x_0-hu\right)\beta\left(\frac{n}{k}|x_0-hu\right)\left\{\frac{\dfrac{\frac{n}{k}\mathbb{P}\left(A_{n/k,y}\,|X=x_0-hu\right)-L(y|x_0-hu)}{\alpha(n/k|x_0-hu)}-M(y|x_0-hu)}{\beta\left(n/k|x_0-hu\right)}-N(y|x_0-hu)\right\},
\end{aligned}$$

where

$$\begin{aligned}
&\sup_{y\in[0,T]^d}\left|\frac{\dfrac{\frac{n}{k}\mathbb{P}\left(A_{n/k,y}\,|X=x_0-hu\right)-L(y|x_0-hu)}{\alpha(n/k|x_0-hu)}-M(y|x_0-hu)}{\beta\left(n/k|x_0-hu\right)}-N(y|x_0-hu)\right|\\
&\leqslant \sup_{y\in[0,T]^d,x\in B_{x_0}(\tau)}\left|\frac{\dfrac{\frac{n}{k}\mathbb{P}\left(A_{n/k,y}\,|X=x\right)-L(y|x)}{\alpha(n/k|x)}-M(y|x)}{\beta\left(n/k|x\right)}-N(y|x)\right|\to 0,
\end{aligned}$$

since for n large enough $x_0-hu\in B_{x_0}(\tau)$. This leads to

$$\begin{aligned}
&\frac{n}{k}\mathbb{P}\left(A_{n/k,y}|X=x_0-hu\right)-L(y|x_0-hu)\\
&\quad=\quad \alpha\left(\frac{n}{k}\Big|x_0\right)M(y|x_0-hu)+\alpha\left(\frac{n}{k}\Big|x_0\right)\beta\left(\frac{n}{k}\Big|x_0\right)\left[N(y|x_0-hu)+o(1)\right]\\
&\qquad+\left[\alpha\left(\frac{n}{k}\Big|x_0-hu\right)-\alpha\left(\frac{n}{k}\Big|x_0\right)\right]M(y|x_0-hu)\\
&\qquad+\left[\alpha\left(\frac{n}{k}\Big|x_0-hu\right)\beta\left(\frac{n}{k}\Big|x_0-hu\right)-\alpha\left(\frac{n}{k}\Big|x_0\right)\beta\left(\frac{n}{k}\Big|x_0\right)\right]\left[N(y|x_0-hu)+o(1)\right]\\
&\quad=\quad \alpha\left(\frac{n}{k}\Big|x_0\right)M(y|x_0)+\alpha\left(\frac{n}{k}\Big|x_0\right)\beta\left(\frac{n}{k}\Big|x_0\right)N(y|x_0)\\
&\qquad+O\left(\alpha\left(\frac{n}{k}\Big|x_0\right)h^{\eta_M\wedge\eta_\beta}\right)+O\left(h^{\eta_\alpha}\right)+o\left(\alpha\left(\frac{n}{k}\Big|x_0\right)\beta\left(\frac{n}{k}\Big|x_0\right)\right),
\end{aligned}$$

where the error terms are all independent from y.

5.2 Proof of Theorem 2.2

Using Theorem 2.1, the homogeneity properties of the functions $L(\cdot|x_0)$, $M(\cdot|x_0)$, $N(\cdot|x_0)$, and the Skorohod representation (but keeping the same notation), we have

$$\begin{aligned}\widehat{L}_{k,a}(y|x_0) \stackrel{d}{=} L(y|x_0) &+ \alpha\left(\frac{n}{k}\Big|x_0\right) a^{-\rho} M(y|x_0) + \alpha\left(\frac{n}{k}\Big|x_0\right)\beta\left(\frac{n}{k}\Big|x_0\right) a^{-\rho-\rho'} N(y|x_0) \\ &+ \frac{1}{\sqrt{kh^p}}\frac{B(ay)}{a} + \frac{1}{a} o\left(\frac{1}{\sqrt{kh^p}}\right),\end{aligned}$$

where the $o-$term is almost surely and uniform in a and y. This implies that for any $r > 1$ and $\xi > 0$, under the assumptions of Theorem 2.2, we have by straightforward application of Taylor's theorem that

$$\begin{aligned}\frac{\int_1^r \widehat{L}^{\xi}_{k,a}(y|x_0)da}{r-1} \stackrel{d}{=} L^{\xi}(y|x_0) &+ \alpha\left(\frac{n}{k}\Big|x_0\right)\xi L^{\xi-1}(y|x_0)M(y|x_0)c(r;\rho) \\ &+ \alpha\left(\frac{n}{k}\Big|x_0\right)\beta\left(\frac{n}{k}\Big|x_0\right)\xi L^{\xi-1}(y|x_0)N(y|x_0)c(r;\rho+\rho') \\ &+ \alpha^2\left(\frac{n}{k}\Big|x_0\right)\frac{\xi(\xi-1)}{2}L^{\xi-2}(y|x_0)M^2(y|x_0)c(r;2\rho) \\ &+ \frac{1}{\sqrt{kh^p}}\frac{\xi}{r-1}L^{\xi-1}(y|x_0)\int_1^r \frac{B(ay)}{a}da + o\left(\frac{1}{\sqrt{kh^p}}\right).\end{aligned}$$

Theorem 2.2 follows then from another application of Taylor's theorem.

5.3 Proof of Theorem 2.3

According to Theorem 2.2, using the homogeneity properties of the functions $L(\cdot|x_0), M(\cdot|x_0)$, $N(\cdot|x_0)$, and the Skorohod representation, we have

$$\begin{aligned}\left(\frac{\int_1^r \widehat{L}^{\xi}_{k,a}(sy|x_0)da}{r-1}\right)^{1/\xi} \stackrel{d}{=}\ & s\,L(y|x_0) + \alpha\left(\frac{n}{k}\Big|x_0\right)s^{1-\rho}M(y|x_0)c(r;\rho) \\ &+\alpha\left(\frac{n}{k}\Big|x_0\right)\beta\left(\frac{n}{k}\Big|x_0\right)s^{1-\rho-\rho'}N(y|x_0)c(r;\rho+\rho') \\ &+\alpha^2\left(\frac{n}{k}\Big|x_0\right)\frac{s^{1-2\rho}}{2}\frac{M^2(y|x_0)}{L(y|x_0)}d(r,\xi;\rho) \\ &+\frac{1}{\sqrt{kh^p}}\frac{1}{r-1}\int_1^r \frac{B(asy)}{a}da + o\left(\frac{1}{\sqrt{kh^p}}\right).\end{aligned}$$

Several Taylor series expansions allow us to achieve the proof of Theorem 2.3.

5.4 Proof of Theorem 2.4

Consider the decomposition

$$\breve{\alpha}_k(y|x_0) = \frac{\left(\frac{\int_1^{r_3}\hat{L}_{k,a}^{\xi_5}(y|x_0)da}{r_3-1}\right)^{1/\xi_5} - \left(\frac{\int_1^{r_4}\hat{L}_{k,a}^{\xi_6}(y|x_0)da}{r_4-1}\right)^{1/\xi_6}}{c(r_3;\rho)-c(r_4;\rho)} \left\{1 + \frac{[c(r_3;\rho)-c(r_3;\breve{\rho}_k)]-[c(r_4;\rho)-c(r_4;\breve{\rho}_k)]}{c(r_3;\breve{\rho}_k)-c(r_4;\breve{\rho}_k)}\right\}$$
$$=: T_1\ \{1+T_2\}.$$

From Theorem 2.2

$$\begin{aligned}\frac{T_1}{\alpha(n/k|x_0)M(y|x_0)} &\stackrel{d}{=} 1+\beta\left(\frac{n}{k}\Big|x_0\right)\frac{N(y|x_0)}{M(y|x_0)}\frac{c(r_3;\rho+\rho')-c(r_4;\rho+\rho')}{c(r_3;\rho)-c(r_4;\rho)}\\ &+ \alpha\left(\frac{n}{k}\Big|x_0\right)\frac{M(y|x_0)}{L(y|x_0)}\frac{d(r_3,\xi_5;\rho)-d(r_4,\xi_6;\rho)}{2[c(r_3;\rho)-c(r_4;\rho)]}\\ &+ \frac{1}{\sqrt{kh^p}\alpha(n/k|x_0)M(y|x_0)}\frac{\frac{1}{r_3-1}\int_1^{r_3}\frac{B(ay)}{a}da-\frac{1}{r_4-1}\int_1^{r_4}\frac{B(ay)}{a}da}{c(r_3;\rho)-c(r_4;\rho)}\\ &+ o\left(\frac{1}{\sqrt{kh^p}\alpha(n/k|x_0)}\right)\end{aligned}$$

and by a Taylor series expansion

$$T_2 \stackrel{d}{=} \frac{[c'(r_4;\rho)-c'(r_3;\rho)](\breve{\rho}_k-\rho)}{c(r_3;\rho)-c(r_4;\rho)}(1+o(1)).$$

Combining these results leads to Theorem 2.4.

5.5 Proof of Theorem 2.5

From Theorem 2.2, we have

$$\begin{aligned}\overline{L}_{k,\overline{k}}(y|x_0) &\stackrel{d}{=} L(y|x_0)+\left(\alpha\left(\frac{n}{k}\Big|x_0\right)M(y|x_0)-\breve{\alpha}_{\overline{k}}(y|x_0)\left(\frac{\overline{k}}{k}\right)^{\breve{\rho}_{\overline{k}}}\right)c(r;\rho)\\ &-\breve{\alpha}_{\overline{k}}(y|x_0)\left(\frac{\overline{k}}{k}\right)^{\breve{\rho}_{\overline{k}}}\left(c(r;\breve{\rho}_{\overline{k}})-c(r;\rho)\right)+\alpha\left(\frac{n}{k}\Big|x_0\right)\beta\left(\frac{n}{k}\Big|x_0\right)N(y|x_0)c(r;\rho+\rho')\\ &+\frac{\alpha^2(n/k|x_0)}{2}\frac{M^2(y|x_0)}{L(y|x_0)}d(r,\xi;\rho)+\frac{1}{\sqrt{kh^p}}\frac{1}{r-1}\int_1^r\frac{B(ay)}{a}da+o\left(\frac{1}{\sqrt{kh^p}}\right).\end{aligned}$$

Now remark that, by the mean value theorem, for $\widetilde{\rho}_{\overline{k}}$ an intermediate value between $\breve{\rho}_{\overline{k}}$ and ρ

$$\begin{aligned}&\alpha\left(\frac{n}{k}\Big|x_0\right)M(y|x_0)-\breve{\alpha}_{\overline{k}}(y|x_0)\left(\frac{\overline{k}}{k}\right)^{\breve{\rho}_{\overline{k}}}\\ &\quad=\left\{\alpha\left(\frac{n}{k}\Big|x_0\right)M(y|x_0)-\breve{\alpha}_{\overline{k}}(y|x_0)\left(\frac{\overline{k}}{k}\right)^{\rho}\right\}-\breve{\alpha}_{\overline{k}}(y|x_0)\left(\frac{\overline{k}}{k}\right)^{\widetilde{\rho}_{\overline{k}}}\log\frac{\overline{k}}{k}\left(\breve{\rho}_{\overline{k}}-\rho\right).\end{aligned}$$

Using Theorems 2.3 and 2.4, we can thus deduce that

$$\begin{aligned}
&\alpha\left(\frac{n}{k}\Big|x_0\right) M(y|x_0) - \breve{\alpha}_{\overline{k}}(y|x_0)\left(\frac{\overline{k}}{k}\right)^{\breve{\rho}_{\overline{k}}} \\
&\quad \stackrel{d}{=} \left\{\alpha\left(\frac{n}{k}\Big|x_0\right) - \alpha\left(\frac{n}{\overline{k}}\Big|x_0\right)\left(\frac{\overline{k}}{k}\right)^{\rho}\right\} M(y|x_0) - \breve{\alpha}_{\overline{k}}(y|x_0)\left(\frac{\overline{k}}{k}\right)^{\widetilde{\rho}_{\overline{k}}} \log\frac{\overline{k}}{k}\left(\breve{\rho}_{\overline{k}} - \rho\right) + o\left(\frac{1}{\sqrt{kh^p}}\right) \\
&\quad \stackrel{d}{=} \left\{\alpha\left(\frac{n}{k}\Big|x_0\right) - \alpha\left(\frac{n}{\overline{k}}\Big|x_0\right)\left(\frac{\overline{k}}{k}\right)^{\rho}\right\} M(y|x_0) + o\left(\frac{1}{\sqrt{kh^p}}\right).
\end{aligned}$$

Recall that $\alpha(.|x_0)$ is regularly varying with index $\rho < 0$, which means that $\alpha(y|x_0) = y^{\rho}\ell_{\alpha}(y|x_0)$ where $\ell_{\alpha}(\cdot|x_0)$ is a slowly varying function at infinity. Following the lines of proof of Theorem 2 in Beirlant *et al.* (2016), this implies that

$$\begin{aligned}
\alpha\left(\frac{n}{k}\Big|x_0\right) M(y|x_0) - \breve{\alpha}_{\overline{k}}(y|x_0)\left(\frac{\overline{k}}{k}\right)^{\breve{\rho}_{\overline{k}}} &\stackrel{d}{=} \left(\frac{\overline{k}}{k}\right)^{\rho}\alpha\left(\frac{n}{\overline{k}}\Big|x_0\right)\beta\left(\frac{n}{\overline{k}}\Big|x_0\right)\left\{\frac{\frac{\ell_{\alpha}(n/k|x_0)}{\ell_{\alpha}(n/\overline{k}|x_0)} - 1}{\beta(n/\overline{k}|x_0)}\right\} M(y|x_0) + o\left(\frac{1}{\sqrt{kh^p}}\right) \\
&= O\left(\left(\frac{\overline{k}}{k}\right)^{\rho}\alpha\left(\frac{n}{\overline{k}}\Big|x_0\right)\beta\left(\frac{n}{\overline{k}}\Big|x_0\right)\right) + o\left(\frac{1}{\sqrt{kh^p}}\right) \\
&= o\left(\frac{1}{\sqrt{kh^p}}\right).
\end{aligned}$$

Consequently, since we have also by Theorem 2.3

$$\breve{\alpha}_{\overline{k}}(y|x_0)\left(\frac{\overline{k}}{k}\right)^{\breve{\rho}_{\overline{k}}}\left(c(r;\breve{\rho}_{\overline{k}}) - c(r;\rho)\right) = O\left(\left(\frac{\overline{k}}{k}\right)^{\breve{\rho}_{\overline{k}}}\frac{1}{\sqrt{\overline{k}h^p}}\right) = o\left(\frac{1}{\sqrt{kh^p}}\right),$$

and thus Theorem 2.5 follows.

Acknowledgement

This work was supported by a research grant (VKR023480) from VILLUM FONDEN. Computation/simulation for the work described in this paper was supported by the DeIC National HPC Centre, SDU.

References

Beirlant, J., Escobar-Bach, M., Goegebeur, Y. and Guillou, A. (2016). Bias-corrected estimation of stable tail dependence function. *Journal of Multivariate Analysis*, **143**, 453–466.

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

Bücher, A., Segers, J. and Volgushev, S. (2014). When uniform weak convergence fails: empirical processes for dependence functions and residuals via epi- and hypographs. *Annals of Statistics*, **42**, 1598–1634.

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

Capéraà, P., Fougères, A.L. and Genest, C. (2000). Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis*, **72**, 30–49.

Drees, H. and Huang, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis*, **64**, 25–47.

Escobar-Bach, M., Goegebeur, Y. and Guillou, A. (2018a). Local robust estimation of the Pickands dependence function. *Annals of Statistics*, **46**, 2806–2843.

Escobar-Bach, M., Goegebeur, Y. and Guillou, A. (2018b). Local estimation of the conditional stable tail dependence function. *Scandinavian Journal of Statistics*, **45**, 590–617.

Fils-Villetard, A., Guillou, A. and Segers, J. (2008). Projection estimators of Pickands dependence functions. *Canadian Journal of Statistics*, **36**, 369–382.

Fougères, A.L., de Haan, L. and Mercadier, C. (2015). Bias correction in multivariate extremes. *Annals of Statistics*, **43**, 903–934.

Gardes, L. and Girard, S. (2015). Nonparametric estimation of the conditional tail copula. *Journal of Multivariate Analysis*, **137**, 1–16.

Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques*, **38**, 907–921.

Gomes, M.I., de Haan, L. and 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.

de Haan, L. and Ferreira, A. (2006). *Extreme value theory - An introduction*, Springer.

Huang, X. (1992). *Statistics of bivariate extremes.* PhD Thesis, Erasmus University Rotterdam, Tinbergen Institute Research series No. 22.

Ledford, A.W. and Tawn, J.A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society Series B*, **59**, 475–499.