

Local estimation of the conditional stable tail dependence function

MIKAEL ESCOBAR-BACH

Department of Mathematics and Computer Science, University of Southern Denmark

YURI GOEGBEUR

Department of Mathematics and Computer Science, University of Southern Denmark

ARMELLE GUILLOU

Institut Recherche Mathématique Avancée, Université de Strasbourg et CNRS

November 8, 2017

Abstract

We consider the local estimation of the stable tail dependence function when a random covariate is observed together with the variables of main interest. Our estimator is a weighted version of the empirical estimator adapted to the covariate framework. We provide the main asymptotic properties of our estimator, when properly normalized, in particular the convergence of the empirical process towards a tight centered Gaussian process. The finite sample performance of our estimator is illustrated on a small simulation study and on a dataset of air pollution measurements.

Keywords: conditional stable tail dependence function, empirical process, stochastic convergence.

Running headline: Local estimation of tail dependence

1 Introduction

A central topic in multivariate extreme value statistics is the estimation of the extremal dependence between two or more random variables. Ledford & Tawn (1997) introduced the coefficient

of tail dependence as a summary measure of extremal dependence, and also proposed an estimator for this parameter. See also Peng (1999), Beirlant & Vandewalle (2002), Beirlant *et al.* (2011), Goegebeur & Guillou (2013), Dutang *et al.* (2014) for alternative estimators of this parameter. Other examples of summary dependence measures for extremes can be found in Coles *et al.* (1999). As an alternative to these summary measures, one can work with functions that give a complete characterisation of the extremal dependence, like the spectral distribution function (Einmahl *et al.*, 1997), the Pickands dependence function (Pickands, 1981) or the stable tail dependence function (Huang, 1992). These functions can be seen as the analogues of copulas in classical statistics. In the present paper we focus on the stable tail dependence function.

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

$$\lim_{t \rightarrow \infty} t \mathbb{P} \left(1 - F_1(Y^{(1)}) \leq t^{-1}y_1 \quad \text{or} \quad \dots \quad \text{or} \quad 1 - F_d(Y^{(d)}) \leq t^{-1}y_d \right) = L(y_1, \dots, y_d),$$

provided that this limit exists, which can be rewritten as

$$\lim_{t \rightarrow \infty} t \left[1 - F \left(F_1^{-1}(1 - t^{-1}y_1), \dots, F_d^{-1}(1 - t^{-1}y_d) \right) \right] = L(y_1, \dots, y_d),$$

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

Now, consider a random 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 $y = (y_1, \dots, y_d)$ a vector of the positive quadrant \mathbb{R}_+^d and $Y_{k,n}^{(j)}$ the k -th order statistic among n realisations of the margins $Y^{(j)}$, $j = 1, \dots, d$. The empirical estimator of L is then given by

$$\hat{L}_k(y) = \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{Y_i^{(1)} \geq Y_{n-[ky_1]+1,n}^{(1)} \quad \text{or} \quad \dots \quad \text{or} \quad Y_i^{(d)} \geq Y_{n-[ky_d]+1,n}^{(d)}\}}.$$

The asymptotic behaviour of this estimator was first studied by Huang (1992); see also Drees & Huang (1998), de Haan & Ferreira (2006) and Bücher *et al.* (2014). We also refer to Peng (2010), Fougères *et al.* (2015), Beirlant *et al.* (2016) and Escobar-Bach *et al.* (2017b) where alternative estimators for L were introduced. In the present paper we extend the empirical estimator to the

situation where we observe a random covariate X together with the variables of main interest $(Y^{(1)}, \dots, Y^{(d)})$. We consider thus a regression problem where we want to describe the extremal dependence between the random variables $(Y^{(1)}, \dots, Y^{(d)})$ given some observed value x for the covariate X . Our approach is nonparametric and based on local estimation in the covariate space.

In the univariate context there is a quite extensive literature on estimation of tail parameters in presence of random covariates. In the framework of heavy-tailed distributions, nonparametric kernel methods were introduced by Daouia *et al.* (2011), who used a fixed number of extreme conditional quantile estimators to estimate the conditional extreme value index, for instance using the Hill (Hill, 1975) and Pickands (Pickands, 1975) estimators, whereas Goegebeur *et al.* (2014b) developed a nonparametric and asymptotically unbiased estimator based on weighted exceedances over a high threshold. The extension of this regression estimation of tail parameters to the full max-domain of attraction, has been considered in Daouia *et al.* (2013), who generalized Daouia *et al.* (2011), and also by Stupfler (2013) and Goegebeur *et al.* (2014a) where an adjustment of the moment estimator, originally proposed by Dekkers *et al.* (1989), to this setting of local estimation has been proposed. On the contrary, the development of extreme value methodology for regression problems with a multivariate response vector is still in its infancy. In de Carvalho & Davison (2014), a procedure was introduced to infer about extremal dependence in the presence of qualitative independent variables, that is, an ANOVA-type setting. Portier & Segers (2017) considered the estimation of a bivariate extreme value distribution under the simplifying assumption that the dependence between $Y^{(1)}$ and $Y^{(2)}$ does not depend on the value taken by the covariate, so that the dependence of the model on the covariates is only through the marginal distributions. Escobar-Bach *et al.* (2017a) studied the robust estimation of the conditional Pickands dependence function using the minimum density power divergence criterion, adapted to the context of local estimation. However, in that paper it is assumed that a random sample from a conditional bivariate extreme value distribution is available. In the present paper we relax this assumption and introduce a local estimator for the conditional stable tail dependence function assuming only that we have data available from a distribution with a dependence structure converging to that of an extreme value distribution.

Thus, we extend the above framework to the case where the vector $(Y^{(1)}, \dots, Y^{(d)})$ is recorded along with a random covariate $X \in \mathbb{R}^p$. In that context, the stable tail dependence function together with the marginal distribution functions depend on the covariate X . In the sequel, for $j = 1, \dots, d$, we denote by $F_j(\cdot|x)$, the continuous conditional distribution function of $Y^{(j)}$ given $X = x$ and $L(\cdot|x)$ the conditional stable tail dependence function defined as

$$\lim_{t \rightarrow \infty} t\mathbb{P} \left(1 - F_1(Y^{(1)}|X) \leq t^{-1}y_1 \text{ or } \dots \text{ or } 1 - F_d(Y^{(d)}|X) \leq t^{-1}y_d | X = x \right) = L(y_1, \dots, y_d|x). \quad (1)$$

We establish the weak convergence of the empirical process of the properly normalized estimator using Donsker results for changing function classes and arguments based on the theory of Vapnik-Červonenkis classes (*VC*-classes). To the best of our knowledge this type of regression problem has not been considered in the multivariate extreme value literature.

The remainder of the paper is organised as follows. In the next section we introduce the local estimator for the conditional stable tail dependence function and study its asymptotic properties. In first instance we assume that the marginal conditional distribution functions are known, whereafter this assumption is removed and the unknown marginal conditional distribution functions are estimated locally using a kernel method. Finally, in Section 3, we illustrate the finite sample behaviour of our estimator with a small simulation study and on a dataset of air pollution measurements. All the proofs of the results are collected in the Appendix.

2 Estimator and asymptotic properties

Denote $(Y, X) := (Y^{(1)}, \dots, Y^{(d)}, X)$, a random vector satisfying (1), and let $(Y_1, X_1), \dots, (Y_n, X_n)$, be independent copies of (Y, X) , where $X \in \mathbb{R}^p$ has density function f . As is usual in the extreme value context, we consider an intermediate sequence $k = k_n$, i.e. $k \rightarrow \infty$ as $n \rightarrow \infty$ with $k/n \rightarrow 0$. Let us denote $y := (y_1, \dots, y_d)$ a vector of the positive quadrant \mathbb{R}_+^d . The event $A_{t,y}$ is defined for any $t \geq 0$ and $y \in \mathbb{R}_+^d$ as

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

and $A_{t,y}^{(i)}$ denotes its analogue for observation (Y_i, X_i) , $i = 1, \dots, n$. The conditional empirical estimator is then given for any $x \in \mathbb{R}^p$ by

$$\begin{aligned}\widehat{T}_k(y|x) &:= \frac{1}{k} \sum_{i=1}^n K_h(x - X_i) \mathbb{1}_{\left\{1 - F_1(Y_i^{(1)}|X_i) \leq \frac{k}{n} y_1 \text{ or } \dots \text{ or } 1 - F_d(Y_i^{(d)}|X_i) \leq \frac{k}{n} y_d\right\}} \\ &= \frac{1}{k} \sum_{i=1}^n K_h(x - X_i) \mathbb{1}_{A_{n/k,y}^{(i)}},\end{aligned}\tag{2}$$

where $K_h(\cdot) := K(\cdot/h)/h^p$ with K a joint density function and $h = h_n$ is a positive non-random sequence satisfying $h_n \rightarrow 0$ as $n \rightarrow \infty$.

The aim of the paper is to derive stochastic convergence results for empirical processes based on (2), with $y \in [0, T]^d$, $T > 0$, but with the covariate argument fixed, meaning that we will focus our study only around one reference position $x_0 \in \text{Int}(S_X)$, the interior of the support S_X of f . In order to derive the asymptotic behaviour of $\widehat{T}_k(y|x_0)$, we need to introduce some conditions mentioned below and well-known in the extreme value framework. Let $\|\cdot\|$ be some norm on \mathbb{R}^p , and denote by $B_x(r)$ the closed ball with respect to $\|\cdot\|$ centered at x and radius $r > 0$.

First order condition: *The limit in (1) exists for all $x \in S_X$ and $y \in \mathbb{R}_+^d$, and the convergence is uniform on $[0, T]^d \times B_{x_0}(r)$ for any $T > 0$ and a $r > 0$.*

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

$$\lim_{t \rightarrow \infty} \frac{1}{\alpha_x(t)} \{t\mathbb{P}(A_{t,y} | X = x) - L(y|x)\} = M_x(y),$$

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

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

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

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

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

Also a usual condition is assumed on the kernel function K .

Assumption (\mathcal{K}_1). *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 $\|\cdot\|$.*

2.1 Marginal conditional distributions known

In this section, we restrict our interest to the case where the marginal conditional distribution functions $F_j(\cdot|x)$, $j = 1, \dots, d$, are known. We start by showing the convergence in probability of our main statistic under some weak assumptions.

Lemma 2.1 *Let $y \in \mathbb{R}_+^d$. Assume the first order condition, (\mathcal{K}_1) and that the functions f and $x \rightarrow L(y|x)$ are continuous at $x_0 \in \text{Int}(S_X)$ non-empty. If for $n \rightarrow \infty$ we have $k \rightarrow \infty$ and $h \rightarrow 0$ in such a way that $k/n \rightarrow 0$ and $kh^p \rightarrow \infty$, then for x_0 such that $f(x_0) > 0$, we have*

$$\widehat{T}_k(y|x_0) \xrightarrow{\mathbb{P}} f(x_0)L(y|x_0).$$

This result indicates that in order to estimate $L(y|x_0)$, the statistic $\widehat{T}_k(y|x_0)$ will need to be divided by an estimator for $f(x_0)$. Our main objective in this section is to show the weak convergence of the stochastic process

$$\left\{ \sqrt{kh^p} \left(\frac{\widehat{T}_k(y|x_0)}{\widehat{f}_n(x_0)} - L(y|x_0) - \alpha_{x_0} \left(\frac{n}{k} \right) M_{x_0}(y) \right), y \in [0, T]^d \right\}, \quad (3)$$

for any $T > 0$, where \widehat{f}_n is the usual kernel density function estimator for f :

$$\widehat{f}_n(x) := \frac{1}{n} \sum_{i=1}^n K_h(x - X_i),$$

see e.g. Parzen (1962). Note that for convenience we use here for \widehat{f}_n the same kernel function K and bandwidth parameter h as for $\widehat{T}_k(y|x_0)$.

As a preliminary step we deduce the covariance structure of the limiting process (apart from the scaling by $1/\widehat{f}_n(x_0)$).

Lemma 2.2 Under the assumptions of Lemma 2.1, we have for any $y, y' \in \mathbb{R}_+^d$

$$kh^p \text{Cov} \left(\widehat{T}_k(y|x_0), \widehat{T}_k(y'|x_0) \right) \rightarrow f(x_0) \|K\|_2^2 \left(L(y|x_0) + L(y'|x_0) - L(y \vee y'|x_0) \right), \text{ as } n \rightarrow \infty.$$

Here, $y \vee y' := (y_1 \vee y'_1, y_2 \vee y'_2, \dots, y_d \vee y'_d)$ and $\|K\|_2 := \sqrt{\int_{S_K} K^2(u) du}$.

We derive now the weak convergence of (3) using Donsker type results for empirical processes with changing function classes and arguments based on the theory of VC-classes, as formulated in van der Vaart & Wellner (1996). These allow us to obtain weak convergence results by mainly focusing on the class of functions involved in our estimator. It should be mentioned that our main weak convergence results are derived in the usual Skorohod space, here $\mathcal{D}([0, T]^d)$ equipped with the sup norm $\|\cdot\|_\infty$.

Theorem 2.1 Assume the second order condition, (\mathcal{D}) , (\mathcal{L}) , (\mathcal{A}) , (\mathcal{K}_1) , and $(x, y) \rightarrow M_x(y)$ being continuous on $B_{x_0}(r) \times [0, T]^d$, with $B_{x_0}(r) \subset S_X$. Consider sequences $k \rightarrow \infty$ and $h \rightarrow 0$ as $n \rightarrow \infty$, in such a way that $k/n \rightarrow 0$, $kh^p \rightarrow \infty$ and

$$\sqrt{kh^p} h^{\min(\eta_f, \eta_L, \eta_\alpha)} \rightarrow 0 \text{ and } \sqrt{kh^p} \alpha_{x_0}(n/k) \rightarrow \lambda_{x_0} \in \mathbb{R}_+.$$

Then, for x_0 such that $f(x_0) > 0$, the process

$$\left\{ \sqrt{kh^p} \left(\frac{\widehat{T}_k(y|x_0)}{\widehat{f}_n(x_0)} - L(y|x_0) - \alpha_{x_0} \left(\frac{n}{k} \right) M_{x_0}(y) \right), y \in [0, T]^d \right\},$$

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

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

where $y, y' \in [0, T]^d$.

2.2 Marginal conditional distributions unknown

In this section, we consider the general framework where all $F_j(\cdot|x)$, $j = 1, \dots, d$, are unknown conditional distribution functions. We want to mimic what has been done in the previous section

in case where these conditional distributions are assumed to be known. To this aim, we consider the random vectors

$$\left(\widehat{F}_{n,1}(Y_i^{(1)}|X_i), \widehat{F}_{n,2}(Y_i^{(2)}|X_i), \dots, \widehat{F}_{n,d}(Y_i^{(d)}|X_i), X_i \right), \quad i = 1, \dots, n,$$

for suitable estimators $\widehat{F}_{n,j}$ of F_j , $j = 1, \dots, d$. Then, similarly as in Section 2.1, we study the statistic

$$\check{T}_k(y|x_0) := \frac{1}{k} \sum_{i=1}^n K_h(x_0 - X_i) \mathbb{1}_{\{1 - \widehat{F}_{n,1}(Y_i^{(1)}|X_i) \leq \frac{k}{n} y_1 \text{ or } \dots \text{ or } 1 - \widehat{F}_{n,d}(Y_i^{(d)}|X_i) \leq \frac{k}{n} y_d\}}.$$

Our final goal is still the same, that is the weak convergence of the stochastic process

$$\left\{ \sqrt{kh^p} \left(\frac{\check{T}_k(y|x_0)}{\widehat{f}_n(x_0)} - L(y|x_0) - \alpha_{x_0} \left(\frac{n}{k} \right) M_{x_0}(y) \right), y \in [0, T]^d \right\}.$$

The idea will be to decompose the process $\left\{ \sqrt{kh^p} \left(\check{T}_k - \mathbb{E}[\check{T}_k] \right) (y|x_0), y \in [0, T]^d \right\}$ into the two terms

$$\left\{ \sqrt{kh^p} \left(\widehat{T}_k - \mathbb{E}[\widehat{T}_k] \right) (y|x_0) + \sqrt{kh^p} \left([\check{T}_k - \widehat{T}_k] - \mathbb{E}[\check{T}_k - \widehat{T}_k] \right) (y|x_0), y \in [0, T]^d \right\}. \quad (4)$$

The first term in the above display can be dealt with using the results of Section 2.1 whereas we have to show that the second term is uniformly negligible. To achieve this objective, let us introduce the following empirical kernel estimator of the unknown conditional distribution functions

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

where $c := c_n$ is a positive non-random sequence satisfying $c_n \rightarrow 0$ as $n \rightarrow \infty$. Here we kept the same kernel K as for $\check{T}_k(y|x_0)$, but of course any other kernel function can be used.

We need to impose again some assumptions, in particular a Hölder-type condition on each marginal conditional distribution function F_j similar to the one imposed on the conditional stable tail dependence function.

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

Concerning the kernel K a stronger assumption than (\mathcal{K}_1) is needed.

Assumption (\mathcal{K}_2) . K satisfies Assumption (\mathcal{K}_1) , there exists $\delta, m > 0$ such that $B_0(\delta) \subset S_K$ and $K(u) \geq m$ for all $u \in B_0(\delta)$, and K belongs to the linear span (the set of finite linear combinations) of functions $k \geq 0$ satisfying the following property: the subgraph of k , $\{(s, u) : k(s) \geq u\}$, can be represented as a finite number of Boolean operations among sets of the form $\{(s, u) : q(s, u) \geq \varphi(u)\}$, where q is a polynomial on $\mathbb{R}^p \times \mathbb{R}$ and φ is an arbitrary real function.

This assumption has already been used in Giné & Guillou (2002) and Escobar-Bach *et al.* (2017a). In particular, we refer to the latter to enunciate the following lemma that measures the discrepancy between the conditional distribution function F_j and its empirical kernel version $\widehat{F}_{n,j}$.

Lemma 2.3 *Assume that there exists $b > 0$ such that $f(x) \geq b, \forall x \in S_X \subset \mathbb{R}^p$, f is bounded, and (\mathcal{K}_2) and (\mathcal{F}_m) hold. Consider a sequence c tending to 0 as $n \rightarrow \infty$ such that for some $q > 1$*

$$\frac{|\log c|^q}{nc^p} \rightarrow 0.$$

Also assume that there exists an $\varepsilon > 0$ such that for n sufficiently large

$$\inf_{x \in S_X} \lambda(\{u \in B_0(1) : x - cu \in S_X\}) > \varepsilon, \quad (5)$$

where λ denotes the Lebesgue measure. Then for any $0 < \eta < \min(\eta_{F_1}, \dots, \eta_{F_d})$, we have

$$\sup_{(y,x) \in \mathbb{R} \times S_X} \left| \widehat{F}_{n,j}(y|x) - F_j(y|x) \right| = o_{\mathbb{P}} \left(\max \left(\sqrt{\frac{|\log c|^q}{nc^p}}, c^\eta \right) \right), \text{ for } j = 1, \dots, d.$$

This rate of convergence allows us to study the second term in (4) and to show that it is uniformly negligible.

Theorem 2.2 *Assume that there exists $b > 0$ such that $f(x) \geq b, \forall x \in S_X \subset \mathbb{R}^p$, f is bounded, (\mathcal{K}_2) , (\mathcal{F}_m) , (\mathcal{D}) , the first order condition and condition (5), and also for any $y \in [0, T]^d$ that $x \rightarrow L(y|x)$ continuous at $x_0 \in \text{Int}(S_X)$ non-empty. Consider sequences $k \rightarrow \infty$, $h \rightarrow 0$ and $c \rightarrow 0$ as $n \rightarrow \infty$, such that $k/n \rightarrow 0$, $kh^p \rightarrow \infty$, and with for some $q > 1$ and $0 < \eta < \min(\eta_{F_1}, \dots, \eta_{F_d})$*

$$n\sqrt{\frac{h^p}{k}} r_n := n\sqrt{\frac{h^p}{k}} \max \left(\sqrt{\frac{|\log c|^q}{nc^p}}, c^\eta \right) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Then

$$\sup_{y \in [0, T]^d} \sqrt{kh^p} \left| \check{T}_k - \hat{T}_k - \mathbb{E} \left[\check{T}_k - \hat{T}_k \right] \right| (y|x_0) = o_{\mathbb{P}}(1).$$

Finally, the decomposition (4) combined with Theorem 2.2 and the results from Section 2.1, yield the desired final result of this paper.

Theorem 2.3 *Assume the second order condition, $(x, y) \rightarrow M_x(y)$ continuous on $B_{x_0}(r) \times [0, T]^d$, with $B_{x_0}(r) \subset S_X$, and that there exists $b > 0$ with $f(x) \geq b, \forall x \in S_X \subset \mathbb{R}^p$, f bounded. Under (\mathcal{D}) , (\mathcal{L}) , (\mathcal{A}) , (\mathcal{F}_m) , (\mathcal{K}_2) and condition (5), consider sequences $k \rightarrow \infty$, $h \rightarrow 0$ and $c \rightarrow 0$ as $n \rightarrow \infty$, such that $k/n \rightarrow 0$, $kh^p \rightarrow \infty$ with*

$$\sqrt{kh^p} h^{\min(\eta_f, \eta_L, \eta_\alpha)} \rightarrow 0, \quad \sqrt{kh^p} \alpha_{x_0}(n/k) \rightarrow \lambda_{x_0} \in \mathbb{R}_+,$$

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

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

Then, the process

$$\left\{ \sqrt{kh^p} \left(\frac{\check{T}_k(y|x_0)}{\hat{f}_n(x_0)} - L(y|x_0) - \alpha_{x_0} \left(\frac{n}{k} \right) M_{x_0}(y) \right), y \in [0, T]^d \right\},$$

weakly converges in $\mathcal{D}([0, T]^d)$ towards a tight centered Gaussian process $\{B_y, y \in [0, T]^d\}$, for any $T > 0$, with covariance structure given in Theorem 2.1.

3 Simulation and real data analysis

3.1 A small simulation study

Our aim in this section is to illustrate the finite sample behaviour of our estimator

$$\bar{L}_k(y|x) := \frac{\check{T}_k(y|x)}{\hat{f}_n(x)}$$

with a small simulation study. We focus on dimension $d = 2$ and we consider the two following models:

- Model 1: We consider the bivariate Student distribution with density function

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{\sqrt{1 - \theta^2}}{2\pi} \left(1 + \frac{y_1^2 - 2\theta y_1 y_2 + y_2^2}{\nu} \right)^{-\frac{\nu+2}{2}}, \quad (y_1, y_2) \in \mathbb{R}^2,$$

and θ being the Pearson correlation coefficient. The stable tail dependence function can be described as

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

where $F_{\nu+1}$ is the distribution function of the univariate Student distribution with $\nu+1$ degrees of freedom. Asymptotic independence can be reached for $\theta = -1$ and complete positive dependence for $\theta = 1$. We set $\theta = X$, where X is uniformly distributed on $[0, 1]$. This model satisfies the second order condition with

$$\begin{aligned} M_x(y_1, y_2) &= C \left[y_2^{2/\nu+1} F_{\nu+3} \left(\frac{(y_2/y_1)^{1/\nu} - \theta}{\sqrt{1 - \theta^2}} \sqrt{\nu + 3} \right) + y_1^{2/\nu+1} F_{\nu+3} \left(\frac{(y_1/y_2)^{1/\nu} - \theta}{\sqrt{1 - \theta^2}} \sqrt{\nu + 3} \right) \right], \\ C &:= -\frac{\nu^{2/\nu+1} \pi^{1/\nu} (\nu + 1)}{2(\nu + 2)} \left(\frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})} \right)^{2/\nu}, \\ \alpha_x(t) &= t^{-2/\nu}. \end{aligned}$$

Moreover, one could check that the uniform property in the first and second order conditions are verified since we have continuity of the involved functions. The model satisfies also conditions (\mathcal{D}) , (\mathcal{L}) , (\mathcal{A}) and (\mathcal{F}_m) . In the simulations we set $\nu = 1$.

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

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

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 $0 \leq t_1, t_2 \leq 1$, 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. This model satisfies our second

order condition with

$$\begin{aligned} M_x(y_1, y_2) &= y_1^2 \partial_1 L(y_1, y_2 | x) + y_2^2 \partial_2 L(y_1, y_2 | x) - L^2(y_1, y_2 | x), \\ \alpha_x(t) &= t^{-1}, \end{aligned}$$

and also satisfies (\mathcal{D}) , (\mathcal{L}) , (\mathcal{A}) and (\mathcal{F}_m) . 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.

To compute our estimator \bar{L}_k , two sequences h and c have to be chosen. Concerning c , we can use the following cross validation criterion introduced by Yao (1999), implemented by Gannoun *et al.* (2002), and already used in an extreme value context by Daouia *et al.* (2011, 2013) or Goegebeur *et al.* (2015):

$$c_j := \arg \min_{c \in \mathcal{C}_g} \sum_{i=1}^n \sum_{k=1}^n \left[\mathbb{1}_{\{Y_i^{(j)} \leq Y_k^{(j)}\}} - \tilde{F}_{n,-i,j}(Y_k^{(j)} | X_i) \right]^2, \quad j = 1, 2,$$

where \mathcal{C}_g is a grid of values of c and $\tilde{F}_{n,-i,j}(y|x) := \frac{\sum_{k=1, k \neq i}^n K_c(x - X_k) \mathbb{1}_{\{Y_k^{(j)} \leq y\}}}{\sum_{k=1, k \neq i}^n K_c(x - X_k)}$. We select the sequence h from the condition

$$n \sqrt{\frac{h^p}{k}} \sqrt{\frac{|\log c|^q}{nc^p}} \rightarrow 0$$

by taking $h = c(k/n)^{1/p} |\log c|^{-\xi}$, where $\xi p > q$ and $c := \min(c_1, c_2)$.

For each distribution, we simulate $N = 500$ samples of size $n = 1000$, and we consider several positions $\{y_t := (t/10, 1 - t/10); t = 1, \dots, 9\}$. Since all stable tail dependence functions satisfy $\max(t, 1 - t) \leq L(t, 1 - t) \leq 1$, all the estimators have been corrected so that they satisfy these bounds. However, the estimators have not been forced to be convex although this could have been done for instance by using a constrained spline smoothing method (Hall & Tajvidi, 2000), or a projection technique (Fils-Villetard *et al.*, 2008). In all the settings, $\mathcal{C}_g = \{0.06, 0.12, 0.18, 0.24, 0.3\}$ and $\xi = 1.1$ are used as chosen in Escobar-Bach *et al.* (2017a). An extensive simulation study has also indicated that these choices seem to give always reason-

able results. Concerning the kernel, each time we use the bi-quadratic function

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

As a qualitative measure of the efficiency over the different positions $\{y_t, t = 1, \dots, 9\}$, we define the absolute bias and the mean squared error (MSE) respectively as follows

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

Figures 1-3 (respectively Figures 4-6) represent the sample means in case of Model 1 (respectively Model 2), based on N samples of size n , of our estimator $\bar{L}_k(y|x)$ as a function of k . Each of these figures shows the behaviour of our estimator at the positions $y \in \{y_t, t = 1, \dots, 9\}$ for a given value of the covariate ($x = 0.2, 0.5$ and 0.8 , respectively). Based on these simulations, we can conclude that in general, our estimator behaves well for not too large values of k with a good proximity to the true value, while some bias appears for k large, which can be expected from our theoretical results, since for k large $\alpha_x(n/k)$ is not necessary negligible. The estimates obtained for the asymmetric logistic model show more bias than those for the bivariate Student distribution, since $\alpha_x(t)$ converges faster to zero as $t \rightarrow \infty$ for the latter. Indeed, for the bivariate Student distribution with $\nu = 1$ we have $\alpha_x(t) = t^{-2}$ while $\alpha_x(t) = t^{-1}$ for the asymmetric logistic distribution. From the figures it also seems that the estimation is more difficult for y close to the diagonal.

In Figures 7 and 8 we show the summary performance measures Abias and MSE for Model 1 and Model 2, respectively, as a function of k for each of the covariate positions. As is clear from these figures, the performance measures do not critically depend on the position in the covariate space for k not too large. The results also seem to indicate that the estimator performs better for stronger dependence than for weaker dependence.

Insert Figures 1-8 here

3.2 Application to air pollution data

In this section, the proposed methodology is applied to a dataset of air pollution measurements. Being able to analyse the dependence between temperature and ozone concentration is of primary importance in order to identify population health effects of high ozone concentration and extreme temperature. The dataset contains daily measurements on, among others, maximum temperature and ground level ozone concentration, for the time period 1999 to 2013, collected at stations spread over the U.S. by the United States Environmental Protection Agency (EPA). It is publicly available at https://aqhdr1.epa.gov/aqswb/aqstmp/airdata/download_files.html. We estimate the stable tail dependence function conditional on time and location, where the latter is expressed by latitude and longitude. The estimation method is the same as the one described in Section 3.1 apart from the dimension of the covariate space, namely here $p = 3$ which implies $\xi = 1.1/3$. We use the same grid values \mathcal{C}_g for the cross-validation, after standardising the covariates to the interval $[0, 1]$. 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 in California. California has one of the largest economies in the world and as such there is a high emission of air pollutants. First, Figure 9 represents the stations in California as markers with different colors corresponding to the value of the estimates $\text{median}\{\bar{L}_k(0.5, 0.5|x), k = n/4, \dots, n/2\}$ of $L(0.5, 0.5|x)$. The range over which the median is computed can be motivated from the simulations, see e.g. Figures 1 till 6. Clearly, the extremal dependence between daily maximum temperature and ground level ozone concentration varies a lot across measurement stations. This could be explained by the fact that the climate of California varies widely, from hot desert to subarctic, depending on the location. As is also clear from Figure 9, the extremal dependence also varies over time.

Insert Figure 9 here

In order to get a better idea of the extremal dependence between temperature and ozone, we show in Figure 10 the time plot of the estimates of the conditional extremal coefficient $\eta(x) := 2 \times L(0.5, 0.5|x) \in [1, 2]$ at two specific stations, Fresno and Los Angeles. This coefficient is often used in the literature as a summary measure of the extremal dependence, with perfect dependence corresponding to the value 1 and independence to the value 2. These two cities exhibit a different extremal dependence throughout the year. Indeed, for Fresno the extremal dependence is strong in the winter months but becomes weaker in summer, while the opposite holds for Los Angeles. Note that these time plots exhibit quite some variability in the estimate of the extremal coefficient. This volatile pattern could be smoothed out by allowing e.g. a different bandwidth parameter for each covariate, which would result in a more flexible estimation procedure. To get a more detailed picture of the extremal dependence we show in Figure 11 the estimate $\text{median}\{\bar{L}_k(t, 1-t|x), k = n/4, \dots, n/2\}$ of the Pickands dependence function for the cities Fresno (top row) and Los Angeles (bottom row) on January 15, 2007 (first column) and June 15, 2007 (second column). In Los Angeles the extremal dependence is stronger in summer than in spring and winter, which corresponds to the typical pattern (see e.g. Mahmud *et al.*, 2008). Fresno deviates from this typical pattern, and the two variables are close to asymptotic independence during summer. This could be explained by the fact that ozone formation seems to be suppressed at extremely high temperatures, say above 312 Kelvin, due to different chemical and biophysical feedback mechanisms, and such temperature conditions are not unusual for the Central Valley of California; see Steiner *et al.* (2010).

Insert Figures 10 & 11 here

Acknowledgement

This work was supported by a research grant (VKR023480) from VILLUM FONDEN and an international project for scientific cooperation (PICS-6416). The authors sincerely thank the editor, associate editor and the referees for their helpful comments and suggestions that led to substantial improvement of the paper.

References

- [1] Beirlant, J., Dierckx, G. & Guillou, A. (2011). Bias-reduced estimators for bivariate tail modelling. *Insurance Math. Econom.* **49**, 18–26.
- [2] Beirlant, J., Escobar-Bach, M., Goegebeur, Y. & Guillou, A. (2016). Bias-corrected estimation of stable tail dependence function. *J. Multivariate Anal.* **143**, 453–466.
- [3] Beirlant, J. & Vandewalle, B. (2002). Some comments on the estimation of a dependence index in bivariate extreme value in statistics. *Statist. Probab. Lett.* **60**, 265–278.
- [4] Bücher, A., Segers, J. & Volgushev, S. (2014). When uniform weak convergence fails: empirical processes for dependence functions and residuals via epi- and hypographs. *Ann. Statist.* **42**, 1598–1634.
- [5] Capéraà, P., Fougères, A. L. & Genest, C. (2000). Bivariate distributions with given extreme value attractor. *J. Multivariate Anal.* **72**, 30–49.
- [6] de Carvalho, M. & Davison, A. C. (2014). Spectral density ratio models for multivariate extremes. *J. Amer. Statist. Assoc.* **109**, 764–776.
- [7] Coles, S. G., Heffernan, J. E. & Tawn, J. A. (1999). Dependence measures for extreme value analyses. *Extremes* **2**, 339–365.
- [8] Daouia, A., Gardes, L. & Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli* **19**, 2557–2589.
- [9] Daouia, A., Gardes, L., Girard, S. & Lekina, A. (2011). Kernel estimators of extreme level curves. *TEST* **20**, 311–333.
- [10] Dekkers, A. L. M., Einmahl, J. H. J. & de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17**, 1833–1855.
- [11] Drees, H. & Huang, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivariate Anal.* **64**, 25–47.

- [12] Dutang, C., Goegebeur, Y. & Guillou, A. (2014). Robust and bias-corrected estimation of the coefficient of tail dependence. *Insurance Math. Econom.* **57**, 46–57.
- [13] Einmahl, J. H. J., de Haan, L. & Sinha, A. K. (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Process. Appl.* **70**, 143–171.
- [14] Escobar-Bach, M., Goegebeur, Y. & Guillou, A. (2017a). Local robust estimation of the Pickands dependence function. *To appear in Ann. Statist.*.
- [15] Escobar-Bach, M., Goegebeur, Y., Guillou, A. & You, A. (2017b). Bias-corrected and robust estimation of the bivariate stable tail dependence function. *TEST* **26**, 284–307.
- [16] Fils-Villetard, A., Guillou, A. & Segers, J. (2008). Projection estimators of Pickands dependence functions. *Canad. J. Statist.* **36**, 369–382.
- [17] Fougères, A. L., de Haan, L. & Mercadier, C. (2015). Bias correction in multivariate extremes. *Ann. Statist.* **43**, 903–934.
- [18] Gannoun, A., Girard, S., Guinot, C. & Saracco, J. (2002). Reference ranges based on nonparametric quantile regression. *Stat. Med.* **21**, 3119–3135.
- [19] Giné, E. & Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **38**, 907–921.
- [20] Goegebeur, Y. & Guillou, A. (2013). Asymptotically unbiased estimation of the coefficient of tail dependence. *Scand. J. Stat.* **40**, 174–189.
- [21] Goegebeur, Y., Guillou, A. & Osmann, M. (2014a). A local moment type estimator for the extreme value index in regression with random covariates. *Canad. J. Statist.* **42**, 487–507.
- [22] Goegebeur, Y., Guillou, A. & Rietsch, T. (2015). Robust conditional Weibull-type estimation. *Ann. Inst. Statist. Math.* **67**, 479–514.
- [23] Goegebeur, Y., Guillou, A. & Schorgen, A. (2014b). Nonparametric regression estimation of conditional tails: the random covariate case. *Statistics* **48**, 732–755.

- [24] de Haan, L. & Ferreira, A. (2006). *Extreme value theory - An introduction*, Springer.
- [25] Hall, P. & Tajvidi, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli* **6**, 835–844.
- [26] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163–1174.
- [27] Huang, X. (1992). *Statistics of bivariate extremes*. PhD Thesis, Erasmus University Rotterdam, Tinbergen Institute Research series No. 22.
- [28] Ledford, A. W. & Tawn, J. A. (1997). Modelling dependence within joint tail regions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **59**, 475–499.
- [29] Mahmud, A., Tyree, M., Cayan, D., Motallebi, N. & Kleeman, M. J. (2008). Statistical downscaling of climate change impacts on ozone concentrations in California. *Journal of Geophysical Research* **113**, doi:10.1029/2007JD009534.
- [30] Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics* **33**, 1065–1076.
- [31] Peng, L. (1999). Estimation of the coefficient of tail dependence in bivariate extremes. *Statist. Probab. Lett.* **43**, 399–409.
- [32] Peng, L. (2010). A practical way for estimating tail dependence functions. *Statist. Sinica* **20**, 365–378.
- [33] Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131.
- [34] Pickands, J. (1981). Multivariate extreme value distributions. *Bulletin of the International Statistical Institute* **49**, 859–878.
- [35] Portier, F. & Segers, J. (2017). On the weak convergence of the empirical conditional copula under a simplifying assumption. *ArXiv:1511.06544*.

- [36] Steiner, A. L., Davis, A. J., Sillman, S., Owen, R. C., Michalak, A. M. & Fiore, A. M. (2010). Observed suppression of ozone formation at extremely high temperatures due to chemical and biophysical feedbacks. *Proceedings of the National Academy of Sciences* **107**, 19685–19690.
- [37] Stupfler, G. (2013). A moment estimator for the conditional extreme-value index. *Electron. J. Stat.* **7**, 2298–2353.
- [38] van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.
- [39] van der Vaart, A. W. & Wellner, J. A. (1996). *Weak convergence and empirical processes, with applications to statistics*. Springer Series in Statistics, Springer-Verlag, New York.
- [40] van der Vaart, A. W. & Wellner, J. A. (2007). Empirical processes indexed by estimated functions. Asymptotics: Particles, Processes and Inverse Problems, *IMS Lecture Notes Monograph Series* **55**, 234–252.
- [41] Yao, Q. (1999). *Conditional predictive regions for stochastic processes*. Technical report, University of Kent at Canterbury.

Yuri Goegebeur, Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark.

E-mail: Yuri.Goegebeur@imada.sdu.dk

A Proofs

A.1 Proof of Lemma 2.1

In order to prove Lemma 2.1, we only need to verify that

$$\mathbb{E} \left[\widehat{T}_k(y|x_0) \right] \rightarrow f(x_0)L(y|x_0) \quad \text{and} \quad \text{Var} \left(\widehat{T}_k(y|x_0) \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

We have

$$\begin{aligned}
\mathbb{E} \left[\widehat{T}_k(y|x_0) \right] &= \int_{S_K} K(u) \frac{n}{k} \mathbb{P}(A_{n/k,y} | X = x_0 - hu) f(x_0 - hu) du \\
&= \int_{S_K} K(u) L(y|x_0 - hu) f(x_0 - hu) du \\
&+ \int_{S_K} K(u) \left(\frac{n}{k} \mathbb{P}(A_{n/k,y} | X = x_0 - hu) - L(y|x_0 - hu) \right) f(x_0 - hu) du. \quad (\text{A.1})
\end{aligned}$$

Since $u \in S_K$, for n large enough, using the continuity of f and L at $x_0 \in \text{Int}(S_X)$ non-empty, we have boundedness in a neighborhood of x_0 and thus

$$\sup_{u \in S_K} L(y|x_0 - hu) < +\infty \text{ and } \sup_{u \in S_K} f(x_0 - hu) < +\infty,$$

and by the first order condition

$$\sup_{u \in S_K} \left| \frac{n}{k} \mathbb{P}(A_{n/k,y} | X = x_0 - hu) - L(y|x_0 - hu) \right| \rightarrow 0,$$

as $n \rightarrow \infty$.

Note that $\int K = 1$, and hence by Lebesgue's dominated convergence theorem, we obtain, for $n \rightarrow \infty$,

$$\int_{S_K} K(u) L(y|x_0 - hu) f(x_0 - hu) du \rightarrow f(x_0) L(y|x_0),$$

and

$$\int_{S_K} K(u) \left(\frac{n}{k} \mathbb{P}(A_{n/k,y} | X = x_0 - hu) - L(y|x_0 - hu) \right) f(x_0 - hu) du \rightarrow 0,$$

implying the first statement.

Next,

$$\begin{aligned}
&\text{Var} \left(\widehat{T}_k(y|x_0) \right) \\
&= \frac{1}{k} \left\{ h^{-p} \int_{S_K} K^2(u) \frac{n}{k} \mathbb{P}(A_{n/k,y} | X = x_0 - hu) f(x_0 - hu) du \right\} - \frac{1}{n} (f(x_0) L(y|x_0) + o(1))^2 \\
&= \frac{1}{kh^p} (\|K\|_2^2 f(x_0) L(y|x_0) + o(1)) - \frac{1}{n} (f(x_0) L(y|x_0) + o(1))^2,
\end{aligned}$$

which goes to zero since $kh^p \rightarrow \infty$.

A.2 Proof of Lemma 2.2

Clearly, we have

$$\begin{aligned}
kh^p \text{Cov} \left(\widehat{T}_k(y|x_0), \widehat{T}_k(y'|x_0) \right) &= \int_{S_K} K^2(u) \frac{n}{k} \mathbb{P} \left(A_{n/k,y} \cap A_{n/k,y'} \mid X = x_0 - hu \right) f(x_0 - hu) du \\
&- h^p \frac{k}{n} \left(f(x_0)^2 L(y|x_0) L(y'|x_0) + o(1) \right) \\
&= \int_{S_K} K^2(u) \frac{n}{k} \mathbb{P} \left(A_{n/k,y} \cap A_{n/k,y'} \mid X = x_0 - hu \right) f(x_0 - hu) du + o(1).
\end{aligned}$$

Then, we easily deduce that

$$\begin{aligned}
\mathbb{P} \left(A_{n/k,y} \cap A_{n/k,y'} \mid X = x_0 - hu \right) &= \mathbb{P} \left(A_{n/k,y} \mid X = x_0 - hu \right) + \mathbb{P} \left(A_{n/k,y'} \mid X = x_0 - hu \right) \\
&- \mathbb{P} \left(A_{n/k,y} \cup A_{n/k,y'} \mid X = x_0 - hu \right).
\end{aligned}$$

Naturally we can describe the sets $A_{n/k,y}$ and $A_{n/k,y'}$ as a finite union like for any $y \in \mathbb{R}_+^d$

$$A_{n/k,y} = \bigcup_{j=1}^d \left\{ 1 - F_j(Y^{(j)}|X) \leq \frac{k}{n} y_j \right\} =: \bigcup_{j=1}^d A_{n/k,y_j,j}.$$

Thus, we have

$$\begin{aligned}
A_{n/k,y} \cup A_{n/k,y'} &= \bigcup_{j=1}^d \left\{ A_{n/k,y_j,j} \cup A_{n/k,y'_j,j} \right\} \\
&= \bigcup_{j=1}^d \left\{ 1 - F_j(Y^{(j)}|X) \leq \frac{k}{n} (y_j \vee y'_j) \right\} \\
&= \bigcup_{j=1}^d A_{n/k,y_j \vee y'_j,j} = A_{n/k,y \vee y'},
\end{aligned}$$

which implies that

$$\begin{aligned}
\mathbb{P} \left(A_{n/k,y} \cap A_{n/k,y'} \mid X = x_0 - hu \right) &= \mathbb{P} \left(A_{n/k,y} \mid X = x_0 - hu \right) + \mathbb{P} \left(A_{n/k,y'} \mid X = x_0 - hu \right) \\
&- \mathbb{P} \left(A_{n/k,y \vee y'} \mid X = x_0 - hu \right),
\end{aligned}$$

and using the same arguments as in the proof of Lemma 2.1, the result follows.

A.3 Proof of Theorem 2.1

As a first step we consider the process

$$\left\{ \sqrt{kh^p} \left(\widehat{T}_k(y|x_0) - f(x_0) \left[L(y|x_0) + \alpha_{x_0} \left(\frac{n}{k} \right) M_{x_0}(y) \right] \right), y \in [0, T]^d \right\},$$

and study its weak convergence. Based on the decomposition

$$\begin{aligned} & \sqrt{kh^p} \left(\widehat{T}_k(y|x_0) - f(x_0) \left[L(y|x_0) + \alpha_{x_0} \left(\frac{n}{k} \right) M_{x_0}(y) \right] \right) \\ &= \sqrt{kh^p} \left(\widehat{T}_k(y|x_0) - \mathbb{E} \left[\widehat{T}_k(y|x_0) \right] \right) \\ & \quad + \sqrt{kh^p} \left(\mathbb{E} \left[\widehat{T}_k(y|x_0) \right] - f(x_0) \left[L(y|x_0) + \alpha_{x_0} \left(\frac{n}{k} \right) M_{x_0}(y) \right] \right), \end{aligned}$$

we have that the main task is to study the weak convergence of the process

$$\left\{ \sqrt{kh^p} \left(\widehat{T}_k(y|x_0) - \mathbb{E} \left[\widehat{T}_k(y|x_0) \right] \right), y \in [0, T]^d \right\}, \quad (\text{A.2})$$

since

$$\lim_{n \rightarrow \infty} \sup_{y \in [0, T]^d} \sqrt{kh^p} \left| \mathbb{E} \left[\widehat{T}_k(y|x_0) \right] - f(x_0) \left[L(y|x_0) + \alpha_{x_0} \left(\frac{n}{k} \right) M_{x_0}(y) \right] \right| = 0.$$

Indeed, if we look at (A.1) in the proof of Lemma 2.1

$$\begin{aligned} \mathbb{E} \left[\widehat{T}_k(y|x_0) \right] &= f(x_0) L(y|x_0) + O(h^{\eta_f \wedge \eta_L}) \\ & \quad + \int_{S_K} K(u) \left(\frac{n}{k} \mathbb{P}(A_{n/k, y} | X = x_0 - hu) - L(y|x_0 - hu) \right) f(x_0 - hu) du, \end{aligned}$$

where the big O term is independent from y . Then

$$\begin{aligned} & \frac{n}{k} \mathbb{P}(A_{n/k, y} | X = x_0 - hu) - L(y|x_0 - hu) \\ &= \alpha_{x_0 - hu} \left(\frac{n}{k} \right) \left\{ M_{x_0 - hu}(y) + \left[\frac{\frac{n}{k} \mathbb{P}(A_{n/k, y} | X = x_0 - hu) - L(y|x_0 - hu)}{\alpha_{x_0 - hu}(n/k)} - M_{x_0 - hu}(y) \right] \right\}, \end{aligned}$$

where

$$\begin{aligned} & \sup_{y \in [0, T]^d} \left| \frac{\frac{n}{k} \mathbb{P}(A_{n/k, y} | X = x_0 - hu) - L(y|x_0 - hu)}{\alpha_{x_0 - hu}(n/k)} - M_{x_0 - hu}(y) \right| \\ & \leq \sup_{y \in [0, T]^d, x \in B_{x_0}(r)} \left| \frac{\frac{n}{k} \mathbb{P}(A_{n/k, y} | X = x) - L(y|x)}{\alpha_x(n/k)} - M_x(y) \right| \rightarrow 0, \end{aligned}$$

combining the second order condition with the fact that for n large enough $x_0 - hu \in B_{x_0}(r)$.

This leads to

$$\begin{aligned} & \frac{n}{k} \mathbb{P} \left(A_{n/k,y} | X = x_0 - hu \right) - L(y|x_0 - hu) \\ &= \alpha_{x_0-hu} \left(\frac{n}{k} \right) (M_{x_0-hu}(y) + o(1)) \\ &= \alpha_{x_0} \left(\frac{n}{k} \right) (M_{x_0-hu}(y) + o(1)) + \left(\alpha_{x_0-hu} \left(\frac{n}{k} \right) - \alpha_{x_0} \left(\frac{n}{k} \right) \right) (M_{x_0-hu}(y) + o(1)), \end{aligned}$$

where the little o component doesn't depend on y . Now,

- according to the Hölder condition on α and $\|u\| \leq 1$

$$\sup_{t \geq 0} |\alpha_{x_0-hu}(t) - \alpha_{x_0}(t)| \leq M_\alpha \|hu\|^{\eta_\alpha} = O(h^{\eta_\alpha}),$$

- by uniform continuity of $(x, y) \rightarrow M_x(y)$ over $B_{x_0}(r) \times [0, T]^d$

$$\sup_{y \in [0, T]^d} |M_{x_0-hu}(y) - M_{x_0}(y)| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Hence, we can deduce that

$$\frac{n}{k} \mathbb{P} \left(A_{n/k,y} | X = x_0 - hu \right) - L(y|x_0 - hu) = \alpha_{x_0} \left(\frac{n}{k} \right) M_{x_0}(y) + \alpha_{x_0} \left(\frac{n}{k} \right) o(1) + O(h^{\eta_\alpha}),$$

which implies that

$$\begin{aligned} & \sqrt{kh^p} \left| \mathbb{E} \left[\hat{T}_k(y|x_0) \right] - f(x_0) \left[L(y|x_0) + \alpha_{x_0} \left(\frac{n}{k} \right) M_{x_0}(y) \right] \right| \\ &= O \left(\sqrt{kh^p} h^{\min(\eta_f, \eta_L, \eta_\alpha)} \right) + \sqrt{kh^p} \alpha_{x_0} \left(\frac{n}{k} \right) o(1) \rightarrow 0. \end{aligned}$$

Define now the covering number $N(\mathcal{F}, L_2(Q), \tau)$ as the minimal number of $L_2(Q)$ -balls of radius τ needed to cover the class of functions \mathcal{F} and the uniform entropy integral as

$$J(\delta, \mathcal{F}, L_2) := \int_0^\delta \sqrt{\log \sup_{Q \in \mathcal{Q}} N(\mathcal{F}, L_2(Q), \tau \|F\|_{Q,2})} d\tau,$$

where \mathcal{Q} is the set of all probability measures Q for which $0 < \|F\|_{Q,2}^2 := \int F^2 dQ < \infty$ and F is an envelope function for the class \mathcal{F} .

Let P be the distribution measure of (Y, X) , and denote the expected value under P , the empirical version and empirical process as follows

$$Pf := \int f dP, \quad \mathbb{P}_n f := \frac{1}{n} \sum_{i=1}^n f(Y_i, X_i), \quad \mathbb{G}_n f := \sqrt{n}(\mathbb{P}_n - P)f,$$

for any real-valued measurable function $f : \mathbb{R}^d \times \mathbb{R}^p \rightarrow \mathbb{R}$.

We introduce our sequence of classes \mathcal{F}_n on $\mathbb{R}^d \times \mathbb{R}^p$ as

$$\mathcal{F}_n := \left\{ (u, z) \rightarrow f_{n,y}(u, z), \quad y \in [0, T]^d \right\}$$

where

$$f_{n,y}(u, z) := \sqrt{\frac{n}{k}} h^p K_h(x_0 - z) \mathbb{1}_{\{1 - F_1(u_1|z) \leq k/n y_1 \text{ or } \dots \text{ or } 1 - F_d(u_d|z) \leq k/n y_d\}}.$$

Denote also by F_n an envelope function of the class \mathcal{F}_n . Now, according to Theorem 19.28 in van der Vaart (1998), the weak convergence of the stochastic process (A.2) follows from the following four conditions. Let ρ_{x_0} be a semimetric, possibly depending on x_0 , making $[0, T]^d$ totally bounded. We have to prove that

$$\sup_{\rho_{x_0}(y, y') \leq \delta_n} P(f_{n,y} - f_{n,y'})^2 \longrightarrow 0 \text{ for every } \delta_n \searrow 0, \quad (\text{A.3})$$

$$PF_n^2 = O(1), \quad (\text{A.4})$$

$$PF_n^2 \mathbb{1}_{\{F_n > \varepsilon \sqrt{n}\}} \longrightarrow 0 \text{ for every } \varepsilon > 0, \quad (\text{A.5})$$

$$J(\delta_n, \mathcal{F}_n, L_2) \longrightarrow 0 \text{ for every } \delta_n \searrow 0. \quad (\text{A.6})$$

We start with proving (A.3). We have

$$P(f_{n,y} - f_{n,y'})^2 = \int_{S_K} K(u)^2 \frac{n}{k} \mathbb{E} \left[\left(\mathbb{1}_{A_{n/k,y}} - \mathbb{1}_{A_{n/k,y'}} \right)^2 \mid X = x_0 - hu \right] f(x_0 - hu) du.$$

But, for any $x' \in S_X$

$$\begin{aligned} \mathbb{E} \left[\left(\mathbb{1}_{A_{n/k,y}} - \mathbb{1}_{A_{n/k,y'}} \right)^2 \mid X = x' \right] &= \mathbb{P}(A_{n/k,y} \cup A_{n/k,y'} \mid X = x') - \mathbb{P}(A_{n/k,y} \cap A_{n/k,y'} \mid X = x') \\ &= \mathbb{P}(\{A_{n/k,y} \cup A_{n/k,y'}\} \setminus \{A_{n/k,y} \cap A_{n/k,y'}\} \mid X = x'). \end{aligned} \quad (\text{A.7})$$

Using the same notation as in the proof of Lemma 2.2, we have

$$\begin{aligned}
A_{n/k,y} \cap A_{n/k,y'} &= \left\{ \bigcup_{j=1}^d A_{n/k,y_j,j} \right\} \cap \left\{ \bigcup_{j=1}^d A_{n/k,y'_j,j} \right\} \\
&= \bigcup_{j=1}^d \bigcup_{i=1}^d \left\{ A_{n/k,y_j,j} \cap A_{n/k,y'_i,i} \right\} \\
&\supseteq \bigcup_{j=1}^d \left\{ A_{n/k,y_j,j} \cap A_{n/k,y'_j,j} \right\} \\
&= \bigcup_{j=1}^d A_{n/k,y_j \wedge y'_j,j}.
\end{aligned}$$

Then, with A^c denoting the complement of any set A , it follows that

$$\begin{aligned}
\{A_{n/k,y} \cup A_{n/k,y'}\} \setminus \{A_{n/k,y} \cap A_{n/k,y'}\} &\subset \left\{ \bigcup_{j=1}^d A_{n/k,y_j \vee y'_j,j} \right\} \cap \left\{ \bigcap_{j=1}^d A_{n/k,y_j \wedge y'_j,j}^c \right\} \\
&\subset \bigcup_{j=1}^d \left\{ A_{n/k,y_j \vee y'_j,j} \cap A_{n/k,y_j \wedge y'_j,j}^c \right\} \\
&= \bigcup_{j=1}^d \left\{ \frac{k}{n}(y_j \wedge y'_j) \leq 1 - F_j(Y^{(j)}|X) \leq \frac{k}{n}(y_j \vee y'_j) \right\}.
\end{aligned}$$

Returning now to (A.7), we have

$$\begin{aligned}
&\mathbb{P}(\{A_{n/k,y} \cup A_{n/k,y'}\} \setminus \{A_{n/k,y} \cap A_{n/k,y'}\} | X = x') \\
&\leq \sum_{j=1}^d \mathbb{P}\left(\frac{k}{n}(y_j \wedge y'_j) \leq 1 - F_j(Y^{(j)}|X) \leq \frac{k}{n}(y_j \vee y'_j) | X = x'\right) \leq \frac{k}{n} \sum_{j=1}^d |y_j - y'_j|.
\end{aligned}$$

Thus, defining

$$\rho_{x_0}(y, y') = \sum_{j=1}^d |y_j - y'_j|,$$

also called the Manhattan distance on \mathbb{R}^d , which is clearly a semimetric making $[0, T]^d$ totally bounded, we have proven (A.3).

We define now the envelope functions

$$F_n(u, z) := \sqrt{\frac{n}{k}} h^p K_h(x_0 - z) \mathbb{1}_{\{1 - F_1(u_1|z) \leq k/nT \text{ or } \dots \text{ or } 1 - F_d(u_d|z) \leq k/nT\}}.$$

With $y_T := \underbrace{(T, \dots, T)}_{d \text{ times}}$, assertion (A.4) results from Lemma 2.1 since

$$\begin{aligned} PF_n^2 &= \int_{S_K} K(u)^2 \frac{n}{k} \mathbb{P}(A_{n/k, y_T} | X = x_0 - hu) f(x_0 - hu) du \\ &= L(y_T | x_0) f(x_0) \|K\|_2^2 + o(1). \end{aligned}$$

For (A.5), note that we have $\{F_n > \varepsilon\sqrt{n}\} = \{(F_n/\varepsilon\sqrt{n})^\xi > 1\}$ for any $\xi > 0$, thus

$$\begin{aligned} PF_n^2 \mathbb{1}_{\{F_n > \varepsilon\sqrt{n}\}} &\leq \frac{1}{\varepsilon^\xi n^{\xi/2}} PF_n^{2+\xi} \\ &= \frac{1}{\varepsilon^\xi n^{\xi/2}} \left(\frac{n}{kh^p}\right)^{\xi/2} \int_{S_K} K(u)^{2+\xi} \frac{n}{k} \mathbb{P}(A_{n/k, y_T} | X = x_0 - hu) f(x_0 - hu) du \\ &= \left(\varepsilon\sqrt{kh^p}\right)^{-\xi} \left(\|K\|_{2+\xi}^{2+\xi} L(y_T | x_0) f(x_0) + o(1)\right), \end{aligned} \tag{A.8}$$

where the right-hand side converges towards 0 since $kh^p \rightarrow \infty$ and K satisfies Assumption (\mathcal{K}_1) .

Finally, it remains to prove (A.6). Define the following class of functions on $\mathbb{R}^d \times \mathbb{R}^p$

$$\begin{aligned} \tilde{\mathcal{F}} &:= \left\{ (u, z) \rightarrow \mathbb{1}_{\{1-F_1(u_1|z) \leq y_1 \text{ or } \dots \text{ or } 1-F_d(u_d|z) \leq y_d\}}, y \in \mathbb{R}_+^d \right\} \\ &= \left\{ (u, z) \rightarrow \mathbb{1}_{\{1-\cdot \leq y_1 \text{ or } \dots \text{ or } 1-\cdot \leq y_d\}} \circ (F_1, \dots, F_d)(u, z), y \in \mathbb{R}_+^d \right\}. \end{aligned}$$

Let's focus for a moment on the class of functions on $[0, 1]^d$

$$\left\{ u \rightarrow \mathbb{1}_{\{1-u_1 \leq y_1 \text{ or } \dots \text{ or } 1-u_d \leq y_d\}}, y \in \mathbb{R}_+^d \right\}.$$

Since this is a family of indicator functions, it is a VC -class if and only if the family of sets associated to the indicator functions is a VC -class of sets. The latter sets can be easily represented as the union of d VC -classes of sets and thus it is also a VC -class of sets (see Lemma 2.6.17 (iii) in van der Vaart & Wellner, 1996). Next, according to Lemma 2.6.18 (vii) in van der Vaart & Wellner (1996), it follows that $\tilde{\mathcal{F}}$ is a VC -class with VC -index V fixed. Define now

$$\tilde{\mathcal{F}}_n := \left\{ (u, z) \rightarrow \mathbb{1}_{\{1-F_1(u_1|z) \leq k/n y_1 \text{ or } \dots \text{ or } 1-F_d(u_d|z) \leq k/n y_d\}}, y \in [0, T]^d \right\},$$

and the envelope function $\tilde{F}_n(u, z) := \mathbb{1}_{\{1-F_1(u_1|z) \leq k/n T \text{ or } \dots \text{ or } 1-F_d(u_d|z) \leq k/n T\}}$. The previous arguments for $\tilde{\mathcal{F}}$ remain, thus we have that $\tilde{\mathcal{F}}_n$ is also a VC -class with VC -index V . According

to Theorem 2.6.7 in van der Vaart & Wellner (1996), there exists a universal constant C such that for \mathcal{Q} the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^p$ and any $0 < \tau < 1$

$$\sup_{Q \in \mathcal{Q}} N(\tilde{\mathcal{F}}_n, L_2(Q), \tau \|\tilde{F}_n\|_{Q,2}) \leq CV(16e)^V \left(\frac{1}{\tau}\right)^{2(V-1)}.$$

Next, we retrieve \mathcal{F}_n by multiplying the previous family with one single function, i.e.

$$\mathcal{F}_n = \{z \rightarrow \sqrt{nh^p/k}K_h(x_0 - z)\} \times \tilde{\mathcal{F}}_n.$$

Since only one ball is needed to cover the class $\{z \rightarrow \sqrt{nh^p/k}K_h(x_0 - z)\}$ whatever the measure $Q \in \mathcal{Q}$, according to the last inequality in the proof of Theorem 2.10.20 in van der Vaart & Wellner (1996)

$$\sup_{Q \in \mathcal{Q}} N(\mathcal{F}_n, L_2(Q), \tau \|F_n\|_{Q,2}) \leq CV(16e)^V \left(\frac{1}{\tau}\right)^{2(V-1)} := L \left(\frac{1}{\tau}\right)^V.$$

Thus, (A.6) is established since for any sequence $\delta_n \searrow 0$ and n large enough, we have

$$J(\delta_n, \mathcal{F}_n, L_2) \leq \int_0^{\delta_n} \sqrt{\log(L) - V \log(\tau)} d\tau = o(1).$$

Finally, we consider the process (3). Straightforward calculations give the following decomposition

$$\begin{aligned} & \sqrt{kh^p} \left(\frac{\hat{T}_k(y|x_0)}{\hat{f}_n(x_0)} - L(y|x_0) - \alpha_{x_0} \left(\frac{n}{k}\right) M_{x_0}(y) \right) \\ &= \frac{\sqrt{kh^p}}{f(x_0)} \left(\hat{T}_k(y|x_0) - f(x_0) \left[L(y|x_0) + \alpha_{x_0} \left(\frac{n}{k}\right) M_{x_0}(y) \right] \right) - \sqrt{kh^p} \frac{\hat{T}_k(y|x_0)}{\hat{f}_n(x_0)f(x_0)} (\hat{f}_n(x_0) - f(x_0)). \end{aligned}$$

Note that for the first term in the right-hand side of the above display we have just established the weak convergence as a stochastic process (apart from the factor $1/f(x_0)$), whereas for the second term we have essentially to study $\sqrt{kh^p}(\hat{f}_n(x_0) - f(x_0))$. The latter can be rewritten as

$$\sqrt{kh^p}(\hat{f}_n(x_0) - f(x_0)) = \sqrt{\frac{k}{n}} \sqrt{nh^p}(\hat{f}_n(x_0) - f(x_0)).$$

Under our assumptions on K and f one can easily verify that $\sqrt{nh^p}(\hat{f}_n(x_0) - f(x_0)) = O_{\mathbb{P}}(1)$ (see e.g. Parzen, 1962) and hence the theorem follows.

A.4 Proof of Theorem 2.2

Let

$$\mathcal{I}_n := \{g_{\theta, \delta, n} : \theta \in \Theta, \delta \in H\},$$

where for $\theta \in \Theta := [0, T]^d$, and $\delta \in H := \{\delta = (\delta_1, \dots, \delta_d); \delta_j : \mathbb{R} \times S_X \rightarrow \mathbb{R}\}$ with

$$\begin{aligned} g_{\theta, \delta, n}(u, z) &:= \sqrt{\frac{n}{k}} h^p K_h(x_0 - z) q_{\theta, \delta, n}(u, z) \\ &:= \sqrt{\frac{n}{k}} h^p K_h(x_0 - z) \mathbb{1}_{\{1 - \delta_1(u_1, z) \leq k/n \theta_1 \text{ or } \dots \text{ or } 1 - \delta_d(u_d, z) \leq k/n \theta_d\}}. \end{aligned}$$

For convenience, denote $\delta_n := (\hat{F}_{n,1}, \dots, \hat{F}_{n,d})$ and $\delta_0 := (F_1, \dots, F_d)$. According to Lemma 2.3, $r_n^{-1} |\delta_n - \delta_0|$ converges in probability towards the null function $H_0 := \{0\}$ in H endowed with the norm $\|\delta\|_H := \sum_{i=1}^d \|\delta_i\|_\infty$. In order to apply Theorem 2.3 in van der Vaart & Wellner (2007), we have now to show

Assertion 1: $\sup_{\theta \in \Theta} \sqrt{n} P G_n(\theta, a_n) \rightarrow 0$ for every $a_n \rightarrow 0$,

Assertion 2: $\sup_{\theta \in \Theta} |\mathbb{G}_n G_n(\theta, a)| \xrightarrow{\mathbb{P}} 0$, for every $a > 0$,

where $G_n(\theta, a)$ is an envelope function for the class

$$\mathcal{G}_n(\theta, a) := \{g_{\theta, \delta_0 + r_n \delta, n} - g_{\theta, \delta_0, n} : \delta \in H, \|\delta\|_H \leq a\}.$$

Proof of Assertion 1. Using the ideas of the proof of Theorem 2.1, for any $\delta \in H$ such that $\|\delta\|_H \leq a$

$$\begin{aligned} |q_{\theta, \delta_0 + r_n \delta, n} - q_{\theta, \delta_0, n}|(u, z) &\leq \mathbb{1}_{\{k/n \theta_1 - r_n a \leq 1 - F_1(u_1 | z) \leq k/n \theta_1 + r_n a \text{ or } \dots \text{ or } k/n \theta_d - r_n a \leq 1 - F_d(u_d | z) \leq k/n \theta_d + r_n a\}} \\ &=: \mathbb{1}_{B_{n, \theta, a}}(u, z). \end{aligned}$$

Thus, we set $G_n(\theta, a)(u, z) := \sqrt{(nh^p/k)} K_h(x_0 - z) \mathbb{1}_{B_{n, \theta, a}}(u, z)$ and we have

$$\sqrt{n} P G_n(\theta, a_n) = n \sqrt{\frac{h^p}{k}} \int_{S_K} K(u) \mathbb{P}(B_{n, \theta, a_n} | X = x_0 - hu) f(x_0 - hu) du,$$

with

$$\begin{aligned} \mathbb{P}(B_{n, \theta, a_n} | X = x_0 - hu) &\leq \sum_{j=1}^d \mathbb{P}\left(k/n \theta_j - r_n a_n \leq 1 - F_j(Y^{(j)} | X) \leq k/n \theta_j + r_n a_n | X = x_0 - hu\right) \\ &\leq 2d r_n a_n. \end{aligned} \tag{A.9}$$

Hence,

$$\sup_{\theta \in \Theta} \sqrt{n} PG_n(\theta, a_n) \leq 2dn \sqrt{\frac{h^p}{k}} r_n a_n (f(x_0) + o(1)) \rightarrow 0, \quad (\text{A.10})$$

and the assertion follows.

Proof of Assertion 2. The idea is to apply Lemma 2.2 in van der Vaart & Wellner (2007). Now we work with the class of functions $\{G_n(\theta, a), \theta \in [0, T]^d\}$, for any $a > 0$ with the envelope function

$$E_n(u, z) := \sqrt{\frac{n}{k}} h^p K_h(x_0 - z) \mathbb{1}_{\{1 - F_1(u_1|z) \leq k/nT + r_na \text{ OR } \dots \text{ OR } 1 - F_d(u_d|z) \leq k/nT + r_na\}}.$$

Consequently, we have first to prove that

$$\sup_{\theta \in \Theta} PG_n(\theta, a)^2 \rightarrow 0, \quad (\text{A.11})$$

$$PE_n^2 = O(1), \quad (\text{A.12})$$

$$PE_n^2 \mathbb{1}_{\{E_n \geq \epsilon \sqrt{n}\}} \rightarrow 0 \text{ for every } \epsilon > 0. \quad (\text{A.13})$$

For what concerns condition (A.11), we have according to (A.9)

$$\begin{aligned} PG_n(\theta, a)^2 &= \int_{S_K} K(u)^2 \frac{n}{k} \mathbb{P}(B_{n,\theta,a} | X = x_0 - hu) f(x_0 - hu) du \\ &\leq 2d \frac{n}{k} r_n a \int_{S_K} K(u)^2 f(x_0 - hu) du \\ &= 2d \frac{n}{k} r_n a (f(x_0) \|K\|_2^2 + o(1)). \end{aligned}$$

We have that $r_n n/k \rightarrow 0$ since $n\sqrt{h^p/k} r_n$ converges and $kh^p \rightarrow \infty$, and as such (A.11) is established.

By the first order condition, we have

$$\begin{aligned} PE_n^2 &= \int_{S_K} K(u)^2 \frac{n}{k} \mathbb{P}(A_{n/k, y_T + (n/k)r_n y_a} | X = x_0 - hu) f(x_0 - hu) du \\ &= L(y_T | x_0) f(x_0) \|K\|_2^2 + o(1), \end{aligned}$$

where $y_a := (a, \dots, a) \in \mathbb{R}_+^d$ and (A.12) follows.

Now we verify condition (A.13). For any $\xi > 0$ we obtain the following inequality

$$\begin{aligned} PE_n^2 \mathbb{1}_{\{E_n \geq \epsilon \sqrt{n}\}} &\leq \frac{1}{\epsilon^\xi n^{\xi/2}} PE_n^{2+\xi} \\ &\leq \frac{1}{\epsilon^\xi} \frac{d}{(kh^p)^{\xi/2}} \left(T + r_n \frac{n}{k} a\right) (\|K\|_{2+\xi}^{2+\xi} f(x_0) + o(1)), \end{aligned}$$

which tends to zero under the assumptions of the theorem.

It remains to show that

$$J(d_n, \{G_n(\theta, a) : \theta \in \Theta\}, L_2) \longrightarrow 0 \text{ for all } d_n \searrow 0.$$

To deal with the uniform entropy integral, we can reuse the lines of proof of Theorem 2.1 by considering the following class of functions on $[0, 1]^d$

$$\left\{ u \rightarrow \mathbb{1}_{\{y_1 \leq 1-u_1 \leq y_2 \text{ or } \dots \text{ or } y_{2d-1} \leq 1-u_d \leq y_{2d}\}}, y_1 < y_2, \dots, y_{2d-1} < y_{2d} \right\},$$

which is a VC -class since the class of sets associated to the indicator functions is a VC -class of sets as the union of d VC -classes of sets. This allows us to prove that there exist positive constants C and V such that

$$\sup_{Q \in \mathcal{Q}} N(\{G_n(\theta, a) : \theta \in \Theta\}, L_2(Q), \tau \|E_n\|_{Q,2}) \leq C \left(\frac{1}{\tau}\right)^V,$$

from which the last assertion follows. This achieves the proof of Theorem 2.2.

A.5 Proof of Theorem 2.3

Due to the decomposition (4), we have to prove that

$$\sup_{y \in [0, T]^d} \sqrt{kh^p} \left| \mathbb{E} \left[\check{T}_k(y|x_0) - \hat{T}_k(y|x_0) \right] \right| = o(1).$$

According to the notation in the proof of Theorem 2.2, note that $\sqrt{kh^p} \mathbb{E} \left[\left| \check{T}_k - \hat{T}_k \right| \right] (y|x_0)$ equals

$$\begin{aligned} & \sqrt{n} \mathbb{E} \left[\left| \frac{1}{n} \sum_{i=1}^n \left[\sqrt{\frac{n}{k}} h^p K_h(x_0 - X_i) \mathbb{1}_{\{1-\hat{F}_{n,1}(Y_i^{(1)}|X_i) \leq k/n y_1 \text{ or } \dots \text{ or } 1-\hat{F}_{n,d}(Y_i^{(d)}|X_i) \leq k/n y_d\}} \right. \right. \right. \\ & - \left. \left. \sqrt{\frac{n}{k}} h^p K_h(x_0 - X_i) \mathbb{1}_{\{1-F_1(Y_i^{(1)}|X_i) \leq k/n y_1 \text{ or } \dots \text{ or } 1-F_d(Y_i^{(d)}|X_i) \leq k/n y_d\}} \right] \right| \right] \\ & \leq \sqrt{n} \mathbb{E} [|g_{y, \delta_n, n}(Y_1, X_1) - g_{y, \delta_0, n}(Y_1, X_1)|] \\ & \leq \sqrt{n} PG_n(y, a), \end{aligned}$$

for n large enough, since with probability tending to 1, $\delta_n \in \delta_0 + r_n \mathcal{B}(0, a)$ where $\mathcal{B}(0, a) := \{\delta : \|\delta\|_H \leq a\}$, and by using the Skorohod representation. This implies that

$$\sup_{y \in [0, T]^d} \sqrt{kh^p} \mathbb{E} \left[\left| \check{T}_k - \hat{T}_k \right| \right] (y|x_0) \leq \sup_{y \in [0, T]^d} \sqrt{n} PG_n(y, a) \rightarrow 0,$$

by Assertion 1 since it is clear that $a_n \rightarrow 0$ can be replaced by any fixed value a in (A.10) and conclude with the fact that $nr_n\sqrt{h^p/k} \rightarrow 0$.

Finally,

$$\left\{ \sqrt{kh^p} \left(\frac{\check{T}_k(y|x_0)}{\hat{f}_n(x_0)} - L(y|x_0) - \alpha_{x_0} \left(\frac{n}{k} \right) M_{x_0}(y) \right), y \in [0, T]^d \right\},$$

can be handled using the same arguments as those at the end of the proof of Theorem 2.1.

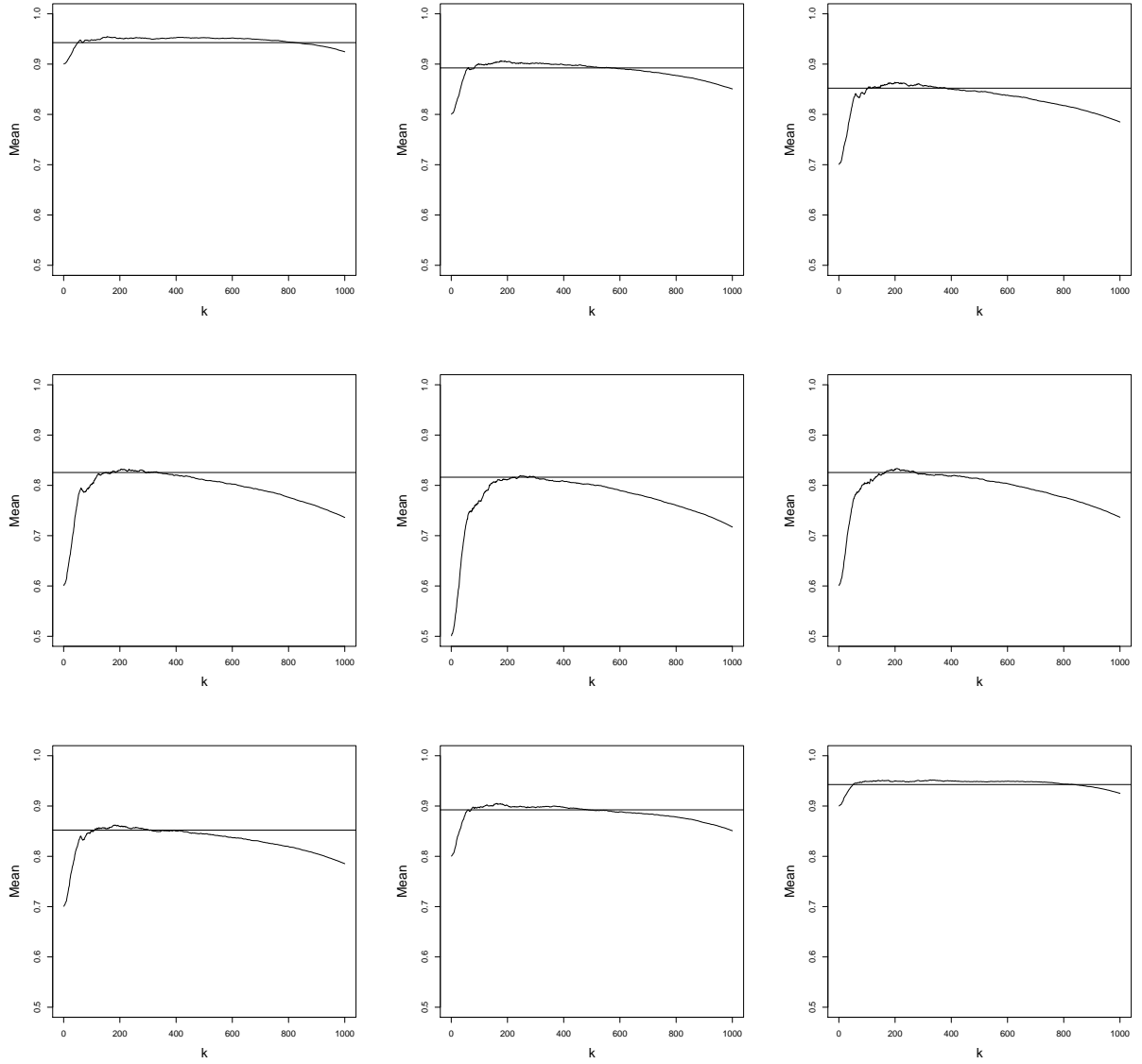


Figure 1: Model 1: Sample mean of $\bar{L}_k(y|0.2)$ as a function of k with, from left to right and up to down, $y = y_t$, $t = 1, \dots, 9$, respectively. The true value of the parameter is represented by a full horizontal black line.

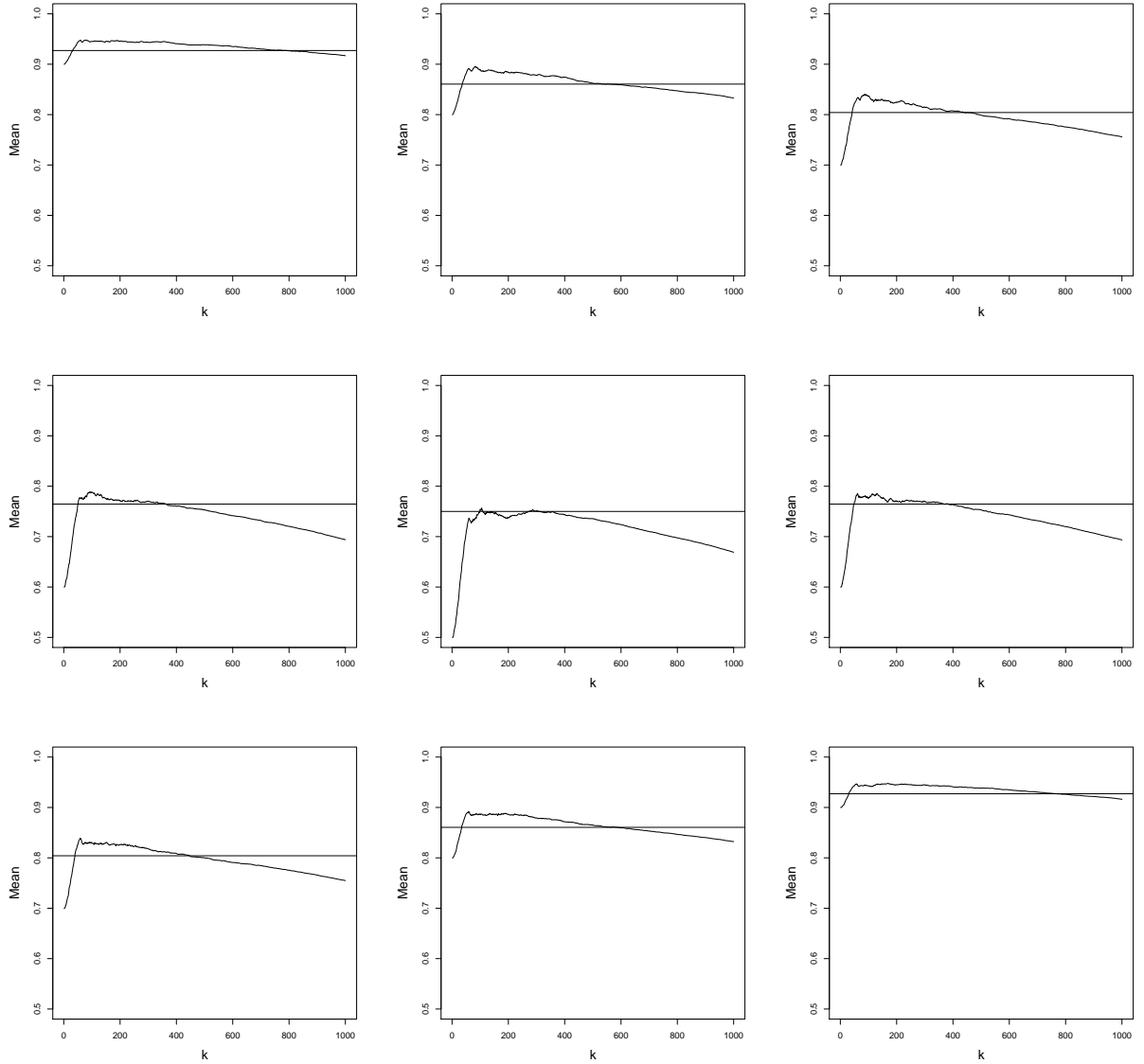


Figure 2: Model 1: Sample mean of $\bar{L}_k(y|0.5)$ as a function of k with, from left to right and up to down, $y = y_t$, $t = 1, \dots, 9$, respectively. The true value of the parameter is represented by a full horizontal black line.

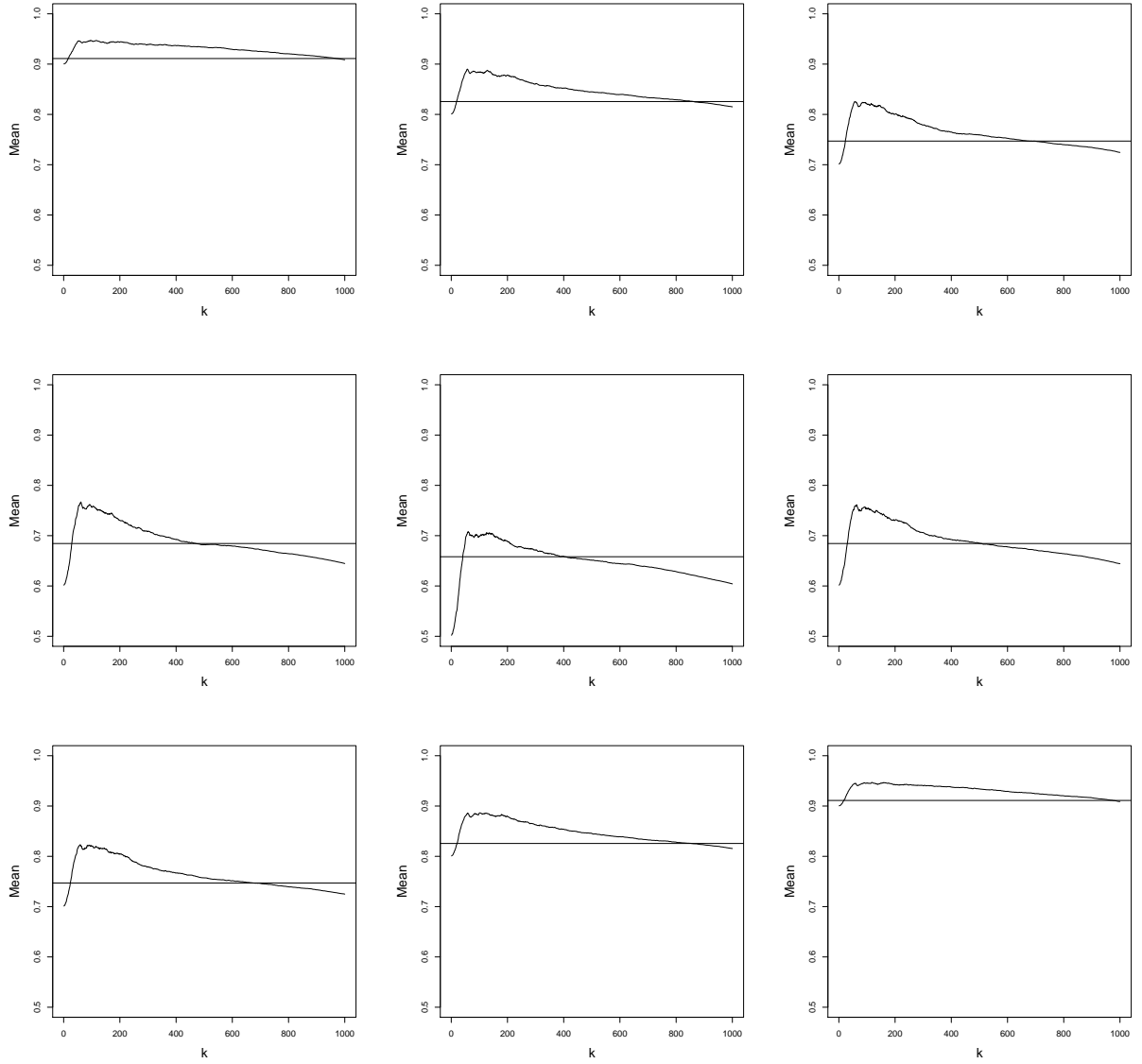


Figure 3: Model 1: Sample mean of $\bar{L}_k(y|0.8)$ as a function of k with, from left to right and up to down, $y = y_t$, $t = 1, \dots, 9$, respectively. The true value of the parameter is represented by a full horizontal black line.

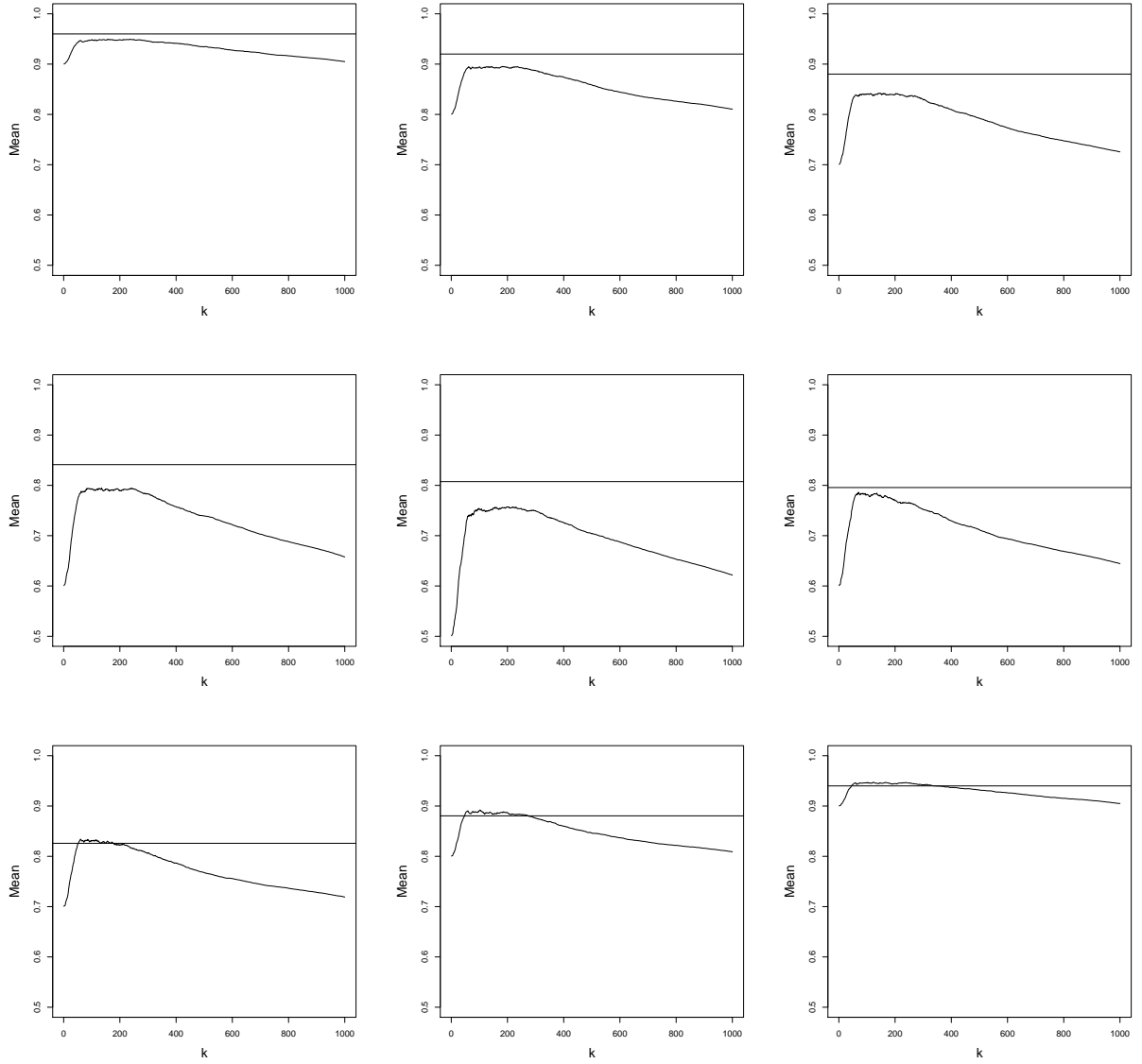


Figure 4: Model 2: Sample mean of $\bar{L}_k(y|0.2)$ as a function of k with, from left to right and up to down, $y = y_t$, $t = 1, \dots, 9$, respectively. The true value of the parameter is represented by a full horizontal black line.

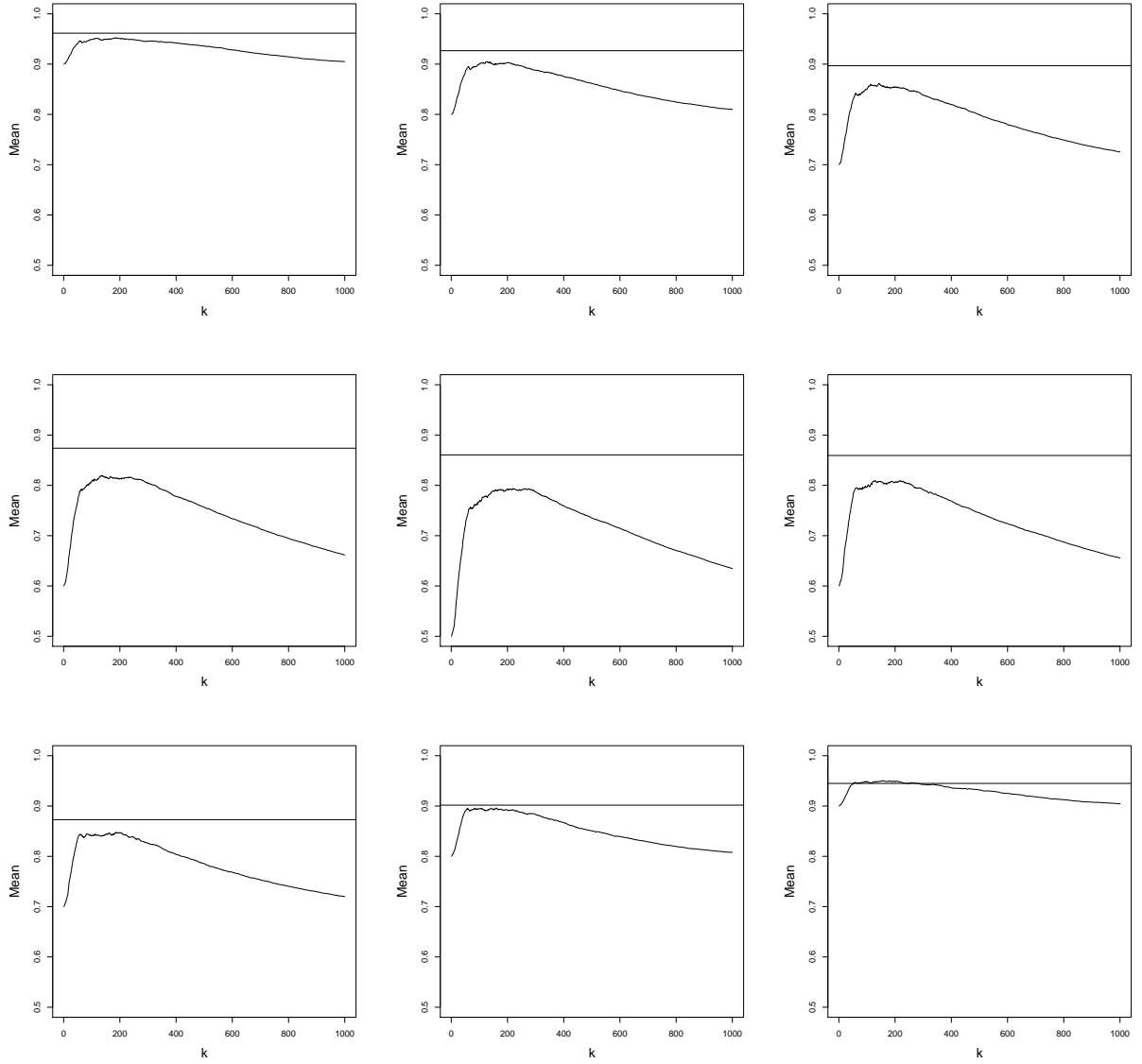


Figure 5: Model 2: Sample mean of $\bar{L}_k(y|0.5)$ as a function of k with, from left to right and up to down, $y = y_t$, $t = 1, \dots, 9$, respectively. The true value of the parameter is represented by a full horizontal black line.

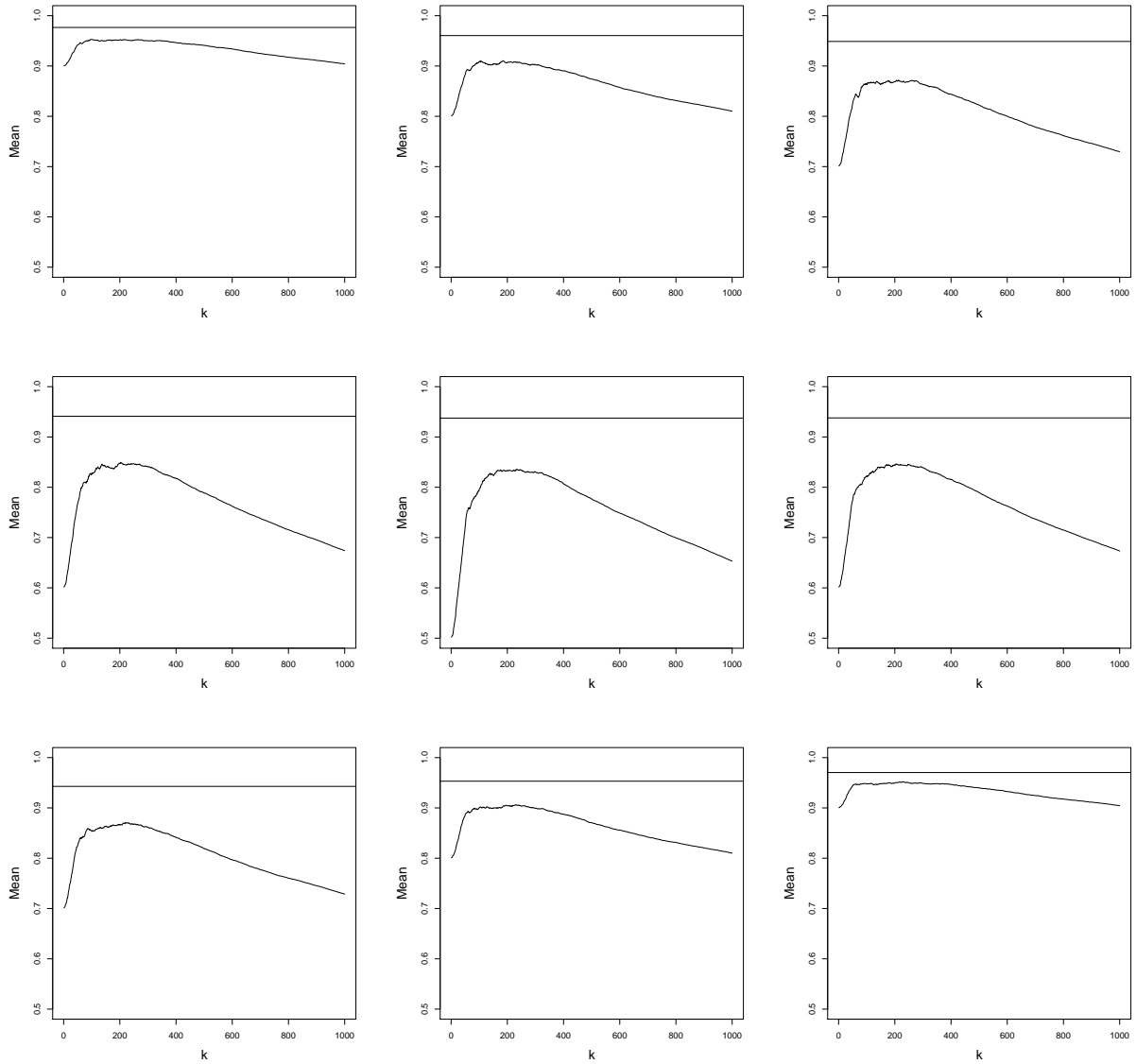


Figure 6: Model 2: Sample mean of $\bar{L}_k(y|0.8)$ as a function of k with, from left to right and up to down, $y = y_t$, $t = 1, \dots, 9$, respectively. The true value of the parameter is represented by a full horizontal black line.

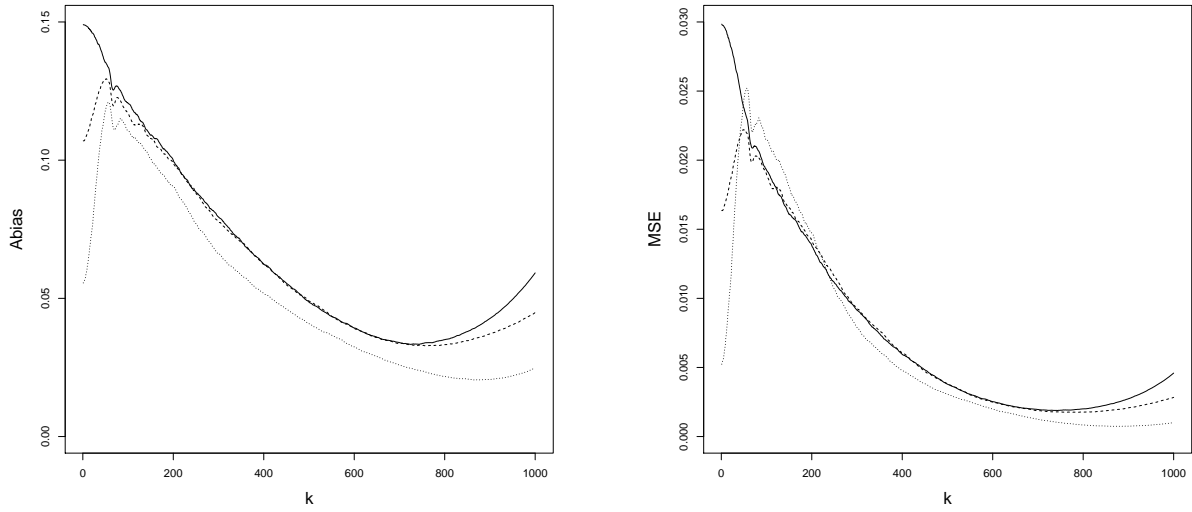


Figure 7: Model 1: Absolute bias and MSE as a function of k for different covariate positions $x = 0.2$ (full line), 0.5 (dashed line), 0.8 (dotted line).

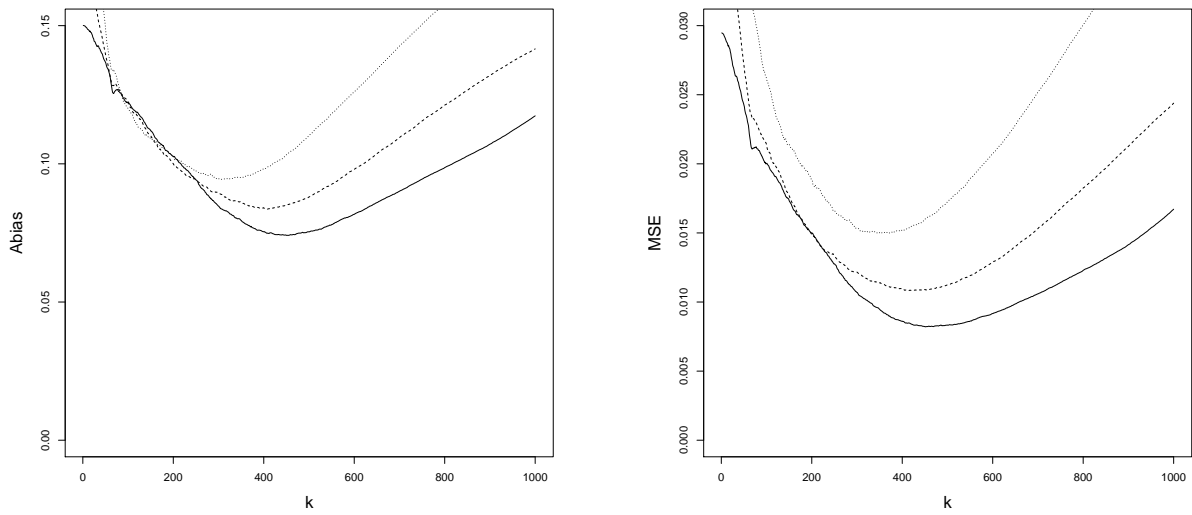


Figure 8: Model 2: Absolute bias and MSE as a function of k for different covariate positions $x = 0.2$ (full line), 0.5 (dashed line), 0.8 (dotted line).

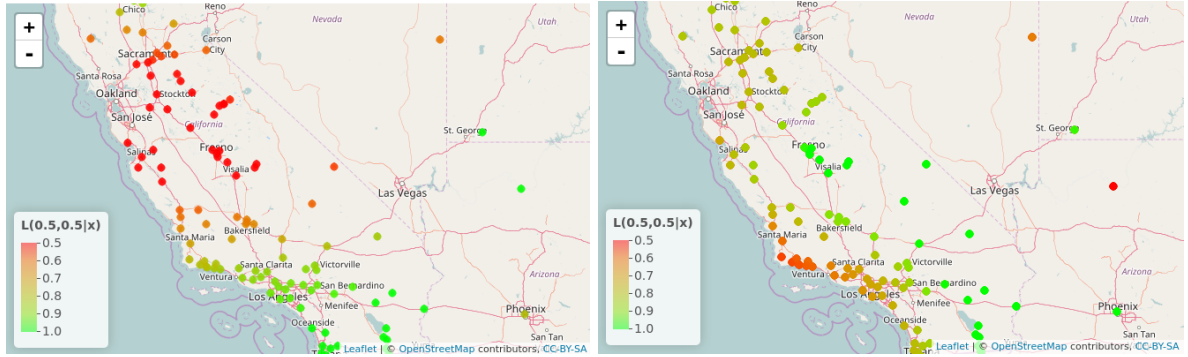


Figure 9: Air pollution data: Estimates median $\{\bar{L}_k(0.5, 0.5|x), k = n/4, \dots, n/2\}$ of $L(0.5, 0.5|x)$ for stations in California on January 15, 2007 (left) and June 15, 2007 (right).

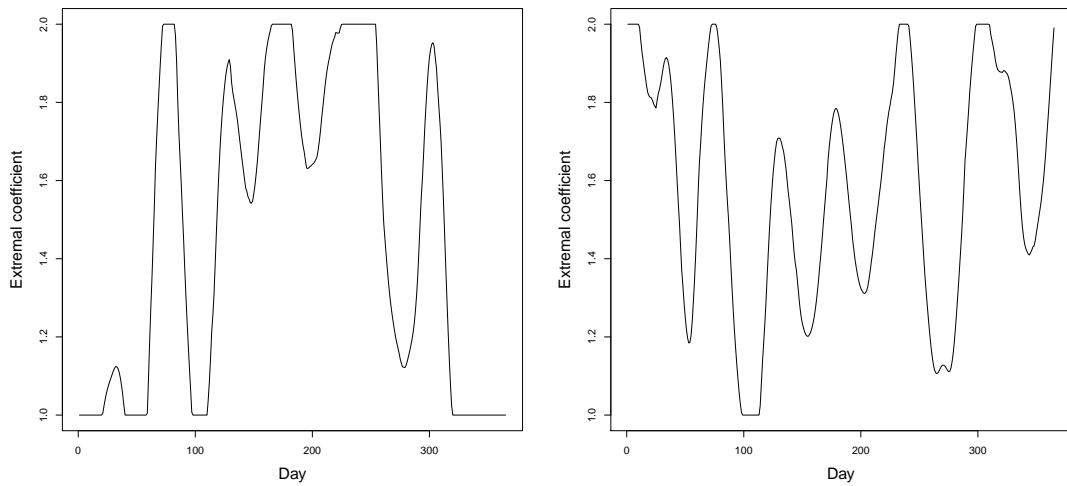


Figure 10: Air pollution data: Time plot of the estimate $2 \times \text{median}\{\bar{L}_k(0.5, 0.5|x), k = n/4, \dots, n/2\}$ of the conditional extremal coefficient at Fresno (left) and Los Angeles (right) over the year 2007.

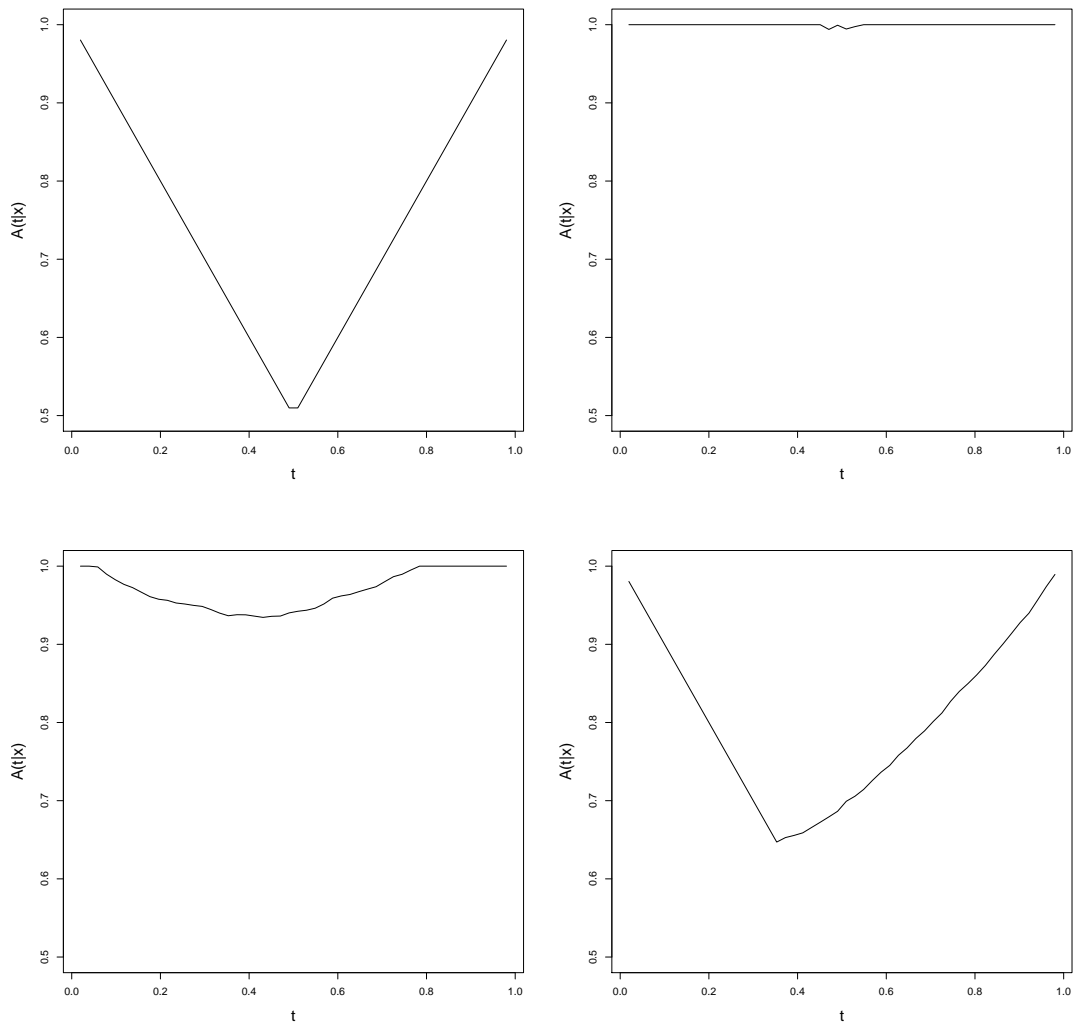


Figure 11: Air pollution data: Estimate of the Pickands dependence function median $\{\bar{L}_k(t, 1 - t|x), k = n/4, \dots, n/2\}$ for Fresno (top) and Los Angeles (bottom) on January 15, 2007 (first column) and June 15, 2007 (second column).