

Nonparametric regression estimation of conditional tails - the random covariate case

Yuri Goegebeur ^{*}
Armelle Guillou [†]
Antoine Schorgen [‡]

September 28, 2012

Abstract

We present families of nonparametric estimators for the conditional tail index of a Pareto-type distribution in presence of random covariates. These families are constructed from locally weighted sums of power transformations of excesses over a high threshold. The asymptotic properties of the proposed estimators are derived under some assumptions on the conditional response distribution, the weight function and the density function of the covariates. We also introduce bias-corrected versions of the estimators for the conditional tail index, and propose in this context a consistent estimator for the second order tail parameter. The finite sample performance of some specific examples from our classes of estimators is illustrated with a small simulation experiment.

Keywords: tail index, Pareto-type distribution, regression, kernel statistic, bias-correction.

1 Introduction

In the area of extreme value statistics, the tail index of a distribution function assumes a central position, and therefore a vast literature has been dedicated to the estimation of this parameter. We refer to Beirlant *et al.* (2004), and de Haan and Ferreira (2006) for recent accounts on univariate tail index estimation. Obtaining an estimate for the tail index is in general a first step in an extreme value analysis. For instance, when interest is in the estimation of extreme quantiles, i.e. quantiles of order α_n , with $\alpha_n > 1 - 1/n$, where n denotes the sample size, one has to extrapolate beyond the available data by means of an extreme value model depending on the tail index. In this paper we will consider tail index estimation for heavy tailed models when a random covariate X is recorded simultaneously with the variable of interest Y .

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

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

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

Let $F(y; x)$ denote the conditional distribution function of the response variable Y given $X = x$, and let g denote the density function of X . We assume that the conditional response distribution is of Pareto-type, i.e. there exists a positive function $\gamma(x)$ such that $\bar{F}(y; x) := 1 - F(y; x)$ is of the form

$$\bar{F}(y; x) = y^{-1/\gamma(x)}\ell(y; x), \quad y > 0, \quad (1)$$

where ℓ is a slowly varying function at infinity, so

$$\lim_{y \rightarrow \infty} \frac{\ell(\lambda y; x)}{\ell(y; x)} = 1, \quad \text{for all } \lambda > 0. \quad (2)$$

The tail function $\gamma(x)$ describes the tail heaviness of the conditional response distribution, and is a function that needs to be adequately estimated from the available data.

The estimation of the tail index in a regression context with fixed, nonrandom covariates, has been addressed in the recent extreme value literature, though not to the same extent as univariate tail index estimation **with independent and identically distributed random variables**. Parametric models based on the generalized extreme value distribution or the generalized Pareto distribution, extended to regression models by assuming specific functional forms for each of their parameters, were proposed in Smith (1989), and Davison and Smith (1990). Beirlant and Goegebeur (2003) introduced a semi-parametric model by extending the representation for scaled-log spacings of successive order statistics of Beirlant *et al.* (1999) to the regression context. Fully nonparametric methods like local estimation and penalized likelihood estimation can be found in e.g. Davison and Ramesh (2000), Hall and Tajvidi (2000), Pauli and Coles (2001), Beirlant and Goegebeur (2004), Chavez-Demoulin and Davison (2005), Gardes and Girard (2008, 2010) and Goegebeur and de Wet (2012). Despite these contributions for the fixed covariate case, conditional tail index estimation with random covariates received only little attention. Wang and Tsai (2009) considered maximum likelihood estimation within the Hall subclass of the Pareto-type models (Hall, 1982) where the tail index is related to the covariates through a log link function. Daouia *et al.* (2011) studied the estimation of extreme quantiles under a conditional Pareto-type model with random covariates, and used a fixed number of such quantile estimates in classical estimators for γ , like the Pickands (Pickands, 1975) and the Hill (Hill, 1975) estimators. The aim of this paper is to introduce flexible classes of estimators for $\gamma(x)$ in the random covariate case, by considering weighted sums of power transformations of excesses over a high threshold.

The remainder of this paper is organized as follows. In Section 2 we introduce our main kernel statistic and derive its asymptotic properties under some conditions. Based on these results we introduce two classes of estimators for the conditional tail index $\gamma(x)$ and establish their limiting distributions. Bias-correction of the estimators for $\gamma(x)$ is studied in Section 3, where we also introduce a consistent estimator for the second order rate parameter. The finite sample behavior of our estimators is illustrated in Section 4 by means of a small simulation experiment. The proofs of all results are deferred to the appendix.

2 Construction of the estimator and asymptotic properties

Let (X_i, Y_i) , $i = 1, \dots, n$, be independent **realizations** of the random vector $(X, Y) \in \mathbb{R}^p \times \mathbb{R}_0^+$, where X has a distribution with joint density function g , and the conditional survival function of Y given $X = x$ is of the form (1).

The idea is to construct estimators for the conditional tail index $\gamma(x)$ on the basis of statistics of the form

$$T_n^{(s,t)}(x, K) := \frac{1}{n} \sum_{i=1}^n K_h^s(x - X_i) (\ln Y_i - \ln \omega_n)_+^t \mathbf{1}\{Y_i > \omega_n\}, \quad (3)$$

where $s \geq 1$, $t \geq 0$, $K_h(x) := K(x/h)/h^p$, K is a joint density function on \mathbb{R}^p , $h = h_n$ is a non-random sequence with $h_n \rightarrow 0$ for $n \rightarrow \infty$, $\mathbf{1}\{A\}$ is the indicator function on the event A , and ω_n is a local non-random threshold sequence for estimation with $\omega_n \rightarrow \infty$ for $n \rightarrow \infty$. **Here, a local threshold means a threshold depending on the point x in the covariate space where the estimation takes place, though the threshold is constant in a neighborhood of x .** Clearly, the growth of the threshold should be related to n and as such also to h , in order to ensure that within the window observations are available for the estimation. The required growth conditions will become clear when studying the asymptotic behavior of the statistic. The bandwidth parameter h is taken to be equal in each dimension of the covariate for mathematical convenience, though such an assumption may not always be in line with the spread of the observations in the covariate space. The proofs of the results can be adjusted to the use of e.g. product kernels, involving a separate bandwidth for each component of X , but given that the selection of bandwidth parameters is challenging in the extreme value context, we propose to stick to a single bandwidth parameter, and to accommodate possible different diffusions of the covariates by a rescaling to an identity covariance matrix. We refer to Givens and Hoeting (2005) for more details on this. The bandwidth can be selected locally or globally. The asymptotic results that follow are for a specific location x in the covariate space, and as such the bandwidth is local. One could for instance minimize the asymptotic mean squared error of the estimator for $\gamma(x)$ with respect to the bandwidth, which would give a bandwidth that takes the density value $g(x)$ into account, so as such the method is locally adaptive. Note that for $t = 0$ the statistic simplifies to

$$T_n^{(s,0)}(x, K) := \frac{1}{n} \sum_{i=1}^n K_h^s(x - X_i) \mathbf{1}\{Y_i > \omega_n\},$$

which is proportional to the kernel estimator for $\bar{F}(y; x)$, proposed in Daouia *et al.* (2011).

In order to obtain the limiting distribution of $T_n^{(s,t)}(x, K)$ we have to impose a further condition on the behavior of the tail of $\bar{F}(y; x)$, the second order condition.

Second order condition (\mathcal{R}_2): *There exists a positive function $\gamma(x)$, a function $\rho(x) < 0$ and a rate function $b(\cdot; x)$ with $b(y; x) \rightarrow 0$ for $y \rightarrow \infty$, of constant sign for large values of y , such that for all $v > 0$*

$$\lim_{y \rightarrow \infty} \frac{\frac{\bar{F}(vy; x) - v^{-1/\gamma(x)} \bar{F}(y; x)}{\bar{F}(y; x)}}{b(y; x)} = v^{-1/\gamma(x)} \frac{v^{\rho(x)/\gamma(x)} - 1}{\gamma(x)\rho(x)}. \quad (4)$$

This condition is not too restrictive and is commonly accepted within extreme value statistics. In fact, after some straightforward simplification of (4), one easily sees that condition (\mathcal{R}_2) makes the convergence in (2) explicit. We refer to Bingham *et al.* (1987) and de Haan and Ferreira (2006) for further details.

We also need some Lipschitz conditions. In these, for all $x, z \in \mathbb{R}^p$, the Euclidean distance between x and z is denoted by $d(x, z)$.

(\mathcal{G}) *There exists $c_g > 0$ such that $|g(x) - g(z)| \leq c_g d(x, z)$.*

(\mathcal{F}) *There exists $c_{\bar{F}} > 0$ and $y_0 > 1$ such that*

$$\sup_{y \geq y_0} \left| \frac{\ln \bar{F}(y; x)}{\ln \bar{F}(y; z)} - 1 \right| \leq c_{\bar{F}} d(x, z).$$

Note that (i) (\mathcal{F}) controls the oscillation of $\ln \bar{F}(y; x)$ with respect to its second argument x , and (ii) the Lipschitz norms in conditions (\mathcal{G}) and (\mathcal{F}) could be taken differently, without complicating the proofs of the results considerably. Finally, we impose a condition on the kernel function K .

(\mathcal{K}) *K is a bounded density function on \mathbb{R}^p , with support Ω included in the unit hypersphere of \mathbb{R}^p .*

We start by introducing two lemmas that will enable us to obtain the asymptotic expansion of $\mathbb{E}[T_n^{(s,t)}(x, K)]$ under the above conditions. Let

$$m_n(x; t) := \mathbb{E}[(\ln Y - \ln \omega_n)_+^t \mathbf{1}\{Y > \omega_n\} | X = x].$$

Lemma 1 *Assume (\mathcal{R}_2), then for $\omega_n \rightarrow \infty$ we have that*

$$m_n(x; t) = \begin{cases} \gamma^t(x) \Gamma(t+1) \bar{F}(\omega_n; x) \left\{ 1 + \frac{b(\omega_n; x)}{\gamma(x)\rho(x)} \left[\frac{1}{(1-\rho(x))^t} - 1 \right] (1 + o(1)) \right\}, & t > 0 \\ \bar{F}(\omega_n; x), & t = 0. \end{cases}$$

Now consider

$$\tilde{m}_n(x; t, s) := \mathbb{E}[K_h^s(x - X) (\ln Y - \ln \omega_n)_+^t \mathbf{1}\{Y > \omega_n\}].$$

Lemma 2 *Assume (\mathcal{R}_2), (\mathcal{G}), (\mathcal{F}) and (\mathcal{K}). Then, for all $x \in \mathbb{R}^p$ such that $g(x) > 0$ we have for $\omega_n \rightarrow \infty$ and $h \rightarrow 0$ with $h \ln \omega_n \rightarrow 0$ that*

$$\tilde{m}_n(x; t, s) = \frac{m_n(x; t)}{h^{(s-1)p}} g(x) \|K\|_s^s (1 + O(h \ln \omega_n)).$$

Note that

$$\begin{aligned}\mathbb{E}[T_n^{(s,t)}(x, K)] &= \tilde{m}_n(x; t, s) \\ &= \mathbb{E}[K_h^s(x - X)m_n(X; t)].\end{aligned}$$

From Lemma 1 and 2 we have that for $t > 0$,

$$\begin{aligned}\mathbb{E}(T_n^{(1,t)}(x, K)) &= \gamma^t(x)g(x)\Gamma(t+1)\bar{F}(\omega_n; x) \\ &\times \left\{ 1 + \frac{b(\omega_n; x)}{\gamma(x)\rho(x)} \left[\frac{1}{(1-\rho(x))^t} - 1 \right] (1 + o(1)) + O(h \ln \omega_n) \right\}\end{aligned}$$

and

$$\mathbb{E}(T_n^{(1,0)}(x, L)) = g(x)\bar{F}(\omega_n; x)(1 + O(h \ln \omega_n)),$$

where L is assumed to satisfy (\mathcal{K}) , which motivates the following estimators for $\gamma(x)$

$$\begin{aligned}\hat{\gamma}_n^{(1)}(x, t, K, L) &:= \left(\frac{T_n^{(1,t)}(x, K)}{\Gamma(t+1)T_n^{(1,0)}(x, L)} \right)^{1/t}, \text{ for } t > 0, \\ \hat{\gamma}_n^{(2)}(x, t, K, L) &:= \frac{T_n^{(1,t+1)}(x, K)}{(t+1)T_n^{(1,t)}(x, L)}, \text{ for } t \geq 0.\end{aligned}$$

Note that $\hat{\gamma}_n^{(1)}(x, 1, K, L) = \hat{\gamma}_n^{(2)}(x, 0, K, L)$.

Now we study the asymptotic behavior of a vector of statistics of the form (3), when appropriately normalized. Denote

$$\mathbb{T}'_n := \frac{1}{\bar{F}(\omega_n; x)g(x)} [T_n^{(1,t_1)}(x, K_1), \dots, T_n^{(1,t_J)}(x, K_J)],$$

and let Σ be a $(J \times J)$ **covariance matrix** with elements

$$\sigma_{j,k} := \gamma^{t_j+t_k}(x) \|K_j K_k\|_1 \Gamma(t_j + t_k + 1).$$

Theorem 1 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample of independent and identically distributed (i.i.d.) random vectors, and assume (\mathcal{R}_2) , (\mathcal{G}) , (\mathcal{F}) and that K_1, \dots, K_J satisfy (\mathcal{K}) . For all $x \in \mathbb{R}^p$ where $g(x) > 0$, we have that if $h \rightarrow 0$, $\omega_n \rightarrow \infty$ with $h \ln \omega_n \rightarrow 0$, $nh^p \bar{F}(\omega_n; x) \rightarrow \infty$, then*

$$\sqrt{nh^p \bar{F}(\omega_n; x)g(x)} [\mathbb{T}_n - \mathbb{E}(\mathbb{T}_n)] \xrightarrow{\mathcal{D}} N_J(0, \Sigma) \quad \text{as } n \rightarrow \infty.$$

We now derive the limiting distribution of the estimator $\hat{\gamma}_n^{(1)}(x, t, K, L)$ and $\hat{\gamma}_n^{(2)}(x, t, K, L)$.

Theorem 2 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample of i.i.d. random vectors, and assume (\mathcal{R}_2) , (\mathcal{G}) , (\mathcal{F}) and kernel functions K and L satisfying (\mathcal{K}) . For all $x \in \mathbb{R}^p$ where $g(x) > 0$,*

we have that if $h \rightarrow 0$, $\omega_n \rightarrow \infty$ with $nh^p \bar{F}(\omega_n; x) \rightarrow \infty$, $\sqrt{nh^p \bar{F}(\omega_n; x)} b(\omega_n; x) \rightarrow \lambda$ and $\sqrt{nh^p \bar{F}(\omega_n; x)} h \ln \omega_n \rightarrow 0$, **then for** $t > 0$

$$\sqrt{nh^p \bar{F}(\omega_n; x)} g(x) [\hat{\gamma}_n^{(1)}(x, t, K, L) - \gamma(x)] \xrightarrow{\mathcal{D}} N \left(\lambda \frac{\sqrt{g(x)}}{t\rho(x)} \left(\frac{1}{(1-\rho(x))^t} - 1 \right), \frac{\gamma^2(x)}{t^2 \Gamma^2(t+1)} [\Gamma(2t+1) \|K\|_2^2 + \Gamma^2(t+1) \|L\|_2^2 - 2\Gamma^2(t+1) \|KL\|_1] \right) \text{ as } n \rightarrow \infty.$$

Theorem 3 Under the same assumptions as in Theorem 2, we have **for** $t \geq 0$

$$\sqrt{nh^p \bar{F}(\omega_n; x)} g(x) [\hat{\gamma}_n^{(2)}(x, t, K, L) - \gamma(x)] \xrightarrow{\mathcal{D}} N \left(\lambda \frac{\sqrt{g(x)}}{(1-\rho(x))^{t+1}}, \frac{\gamma^2(x) \Gamma(2t+1)}{(t+1) \Gamma^2(t+1)} [2(2t+1) \|K\|_2^2 + (t+1) \|L\|_2^2 - 2(2t+1) \|KL\|_1] \right) \text{ as } n \rightarrow \infty.$$

For $K = L$ the asymptotic variance expressions simplify and are given by

$$\mathbb{A} \text{Var}(\hat{\gamma}_n^{(1)}(x, t, K, K)) = \frac{\gamma^2(x) \|K\|_2^2}{t^2} \left(\frac{\Gamma(2t+1)}{\Gamma^2(t+1)} - 1 \right) =: \gamma^2(x) \|K\|_2^2 v_1^2, \quad (5)$$

and

$$\mathbb{A} \text{Var}(\hat{\gamma}_n^{(2)}(x, t, K, K)) = \gamma^2(x) \|K\|_2^2 \frac{\Gamma(2t+1)}{\Gamma^2(t+1)} =: \gamma^2(x) \|K\|_2^2 v_2^2. \quad (6)$$

In Figure 1 (a) we show the components of the asymptotic standard deviation v_1 **in (5)** (solid line) and v_2 **in (6)** (dashed line) as a function of t . Clearly, no estimator performs uniformly best in t , though the estimator $\hat{\gamma}_n^{(1)}(x, t, K, K)$ shows the better performance over a wide range of values for t . Note that the asymptotic variance of the estimator $\hat{\gamma}_n^{(1)}(x, t, K, K)$ is minimized for $t = 1$ whereas for $\hat{\gamma}_n^{(2)}(x, t, K, K)$ the minimum variance is reached for $t = 0$. **Concerning** the asymptotic bias, it can be proven that the estimator $\hat{\gamma}_n^{(2)}(x, t, K, K)$ performs uniformly better than $\hat{\gamma}_n^{(1)}(x, t, K, K)$. This is illustrated in Figure 1 (b) where we show the **multiplicative** components of the asymptotic bias $((1-\rho(x))^{-t} - 1)/(t\rho(x))$ (solid line) and $(1-\rho(x))^{-(t+1)}$ (dashed line) for the case where $\rho(x) = -1$, as a function of t .

As mentioned before, $\hat{\gamma}_n^{(1)}(x, 1, K, L) = \hat{\gamma}_n^{(2)}(x, 0, K, L)$. In this case, setting $K = L = \mathbf{1}\{S\}/|S|$, where S is the p dimensional unit hypersphere with volume $|S| = 2\pi^{p/2}/(p\Gamma(p/2))$, one obtains a local Hill-type estimator, denoted $\hat{\gamma}_{H,n}(x)$, **whose asymptotic distribution** is given in the following corollary.

Corollary 1 Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample of *i.i.d.* random vectors, and assume (\mathcal{R}_2) , (\mathcal{G}) and (\mathcal{F}) . For all $x \in \mathbb{R}^p$ where $g(x) > 0$, we have that if $h \rightarrow 0$, $\omega_n \rightarrow \infty$ with $nh^p \bar{F}(\omega_n; x) \rightarrow \infty$, $\sqrt{nh^p \bar{F}(\omega_n; x)} b(\omega_n; x) \rightarrow \lambda$ and $\sqrt{nh^p \bar{F}(\omega_n; x)} h \ln \omega_n \rightarrow 0$, then

$$\sqrt{nh^p \bar{F}(\omega_n; x)} g(x) [\hat{\gamma}_{H,n}(x) - \gamma(x)] \xrightarrow{\mathcal{D}} N \left(\lambda \frac{\sqrt{g(x)}}{1-\rho(x)}, \frac{\gamma^2(x)}{|S|} \right) \text{ as } n \rightarrow \infty.$$

3 Bias-corrected estimation of $\gamma(x)$

In this section we will introduce bias-corrected estimators for $\gamma(x)$ based on $\hat{\gamma}_n^{(1)}(x, t, K, L)$ and $\hat{\gamma}_n^{(2)}(x, t, K, L)$, as well as a consistent estimator for the second order parameter $\rho(x)$.

A first possibility to obtain a bias-corrected estimator for γ consists in constructing an appropriately chosen weighted sum of two estimators $\hat{\gamma}_n^{(1)}(x, t_1, K_1, L_1)$ and $\hat{\gamma}_n^{(1)}(x, t_2, K_2, L_2)$ **with** $t_1, t_2 > 0$, and therefore we start by establishing the joint asymptotic behavior **of these estimators**.

Let $\boldsymbol{\mu}^{(1)}$ be a (2×1) vector with elements

$$\mu_i^{(1)} := \frac{\sqrt{g(x)}}{t_i \rho(x)} \left(\frac{1}{(1 - \rho(x))^{t_i}} - 1 \right), \quad i = 1, 2,$$

and $\tilde{\boldsymbol{\Sigma}}^{(1)}$ a (2×2) matrix with elements

$$\begin{aligned} \tilde{\sigma}_{i,i}^{(1)} &:= \frac{\gamma^2(x)}{[t_i \Gamma(t_i + 1)]^2} [\Gamma(2t_i + 1) \|K_i\|_2^2 + \Gamma^2(t_i + 1) \|L_i\|_2^2 - 2\Gamma^2(t_i + 1) \|K_i L_i\|_1], \quad i = 1, 2, \\ \tilde{\sigma}_{1,2}^{(1)} &:= \frac{\gamma^2(x)}{t_1 t_2 \Gamma(t_1 + 1) \Gamma(t_2 + 1)} [\Gamma(t_1 + t_2 + 1) \|K_1 K_2\|_1 + \Gamma(t_1 + 1) \Gamma(t_2 + 1) \|L_1 L_2\|_1 \\ &\quad - \Gamma(t_1 + 1) \Gamma(t_2 + 1) \|K_1 L_2\|_1 - \Gamma(t_1 + 1) \Gamma(t_2 + 1) \|K_2 L_1\|_1]. \end{aligned}$$

Theorem 4 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample of i.i.d. random vectors, and assume (\mathcal{R}_2) , (\mathcal{G}) , (\mathcal{F}) and kernel functions K_1, K_2, L_1 and L_2 satisfying (\mathcal{K}) . For all $x \in \mathbb{R}^p$ where $g(x) > 0$, if $h \rightarrow 0$, $\omega_n \rightarrow \infty$ with $nh^p \bar{F}(\omega_n; x) \rightarrow \infty$, $\sqrt{nh^p \bar{F}(\omega_n; x)} b(\omega_n; x) \rightarrow \lambda$ and $\sqrt{nh^p \bar{F}(\omega_n; x)} h \ln \omega_n \rightarrow 0$, **we have for** $t_1, t_2 > 0$*

$$\sqrt{nh^p \bar{F}(\omega_n; x) g(x)} \begin{bmatrix} \hat{\gamma}_n^{(1)}(x, t_1, K_1, L_1) - \gamma(x) \\ \hat{\gamma}_n^{(1)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} \xrightarrow{\mathcal{D}} N_2(\lambda \boldsymbol{\mu}^{(1)}, \tilde{\boldsymbol{\Sigma}}^{(1)}) \quad \text{as } n \rightarrow \infty.$$

Consider now a weighted sum of two estimators with $t_1 \neq t_2$ **and** $\alpha \in \mathbb{R}$:

$$\begin{aligned} \hat{\gamma}_n^{(1)}(x, \alpha) &:= \hat{\gamma}_n^{(1)}(x, \alpha, t_1, t_2, K_1, K_2, L_1, L_2) \\ &:= \alpha \hat{\gamma}_n^{(1)}(x, t_1, K_1, L_1) + (1 - \alpha) \hat{\gamma}_n^{(1)}(x, t_2, K_2, L_2). \end{aligned}$$

From Theorem 4 we obtain the asymptotic bias of such weighted estimators to be

$$\alpha \lambda \frac{\sqrt{g(x)}}{t_1 \rho(x)} \left(\frac{1}{(1 - \rho(x))^{t_1}} - 1 \right) + (1 - \alpha) \lambda \frac{\sqrt{g(x)}}{t_2 \rho(x)} \left(\frac{1}{(1 - \rho(x))^{t_2}} - 1 \right),$$

which can be eliminated by taking the weight α as

$$\alpha_{BC}^{(1)}(\rho(x)) := \frac{1}{1 - \beta [1 - \rho(x)]^{t_2 - t_1}} \quad \text{with} \quad \beta := \frac{t_2}{t_1} \frac{1 - [1 - \rho(x)]^{t_1}}{1 - [1 - \rho(x)]^{t_2}}.$$

The result is formalized in the following corollary. Let $\boldsymbol{\alpha}'_1 = [\alpha_{BC}^{(1)}(\rho(x)), 1 - \alpha_{BC}^{(1)}(\rho(x))]$.

Corollary 2 *Under the conditions of Theorem 4 we have that*

$$\sqrt{nh^p \bar{F}(\omega_n; x) g(x)} [\hat{\gamma}_n^{(1)}(x, \alpha_{BC}^{(1)}(\rho(x))) - \gamma(x)] \xrightarrow{\mathcal{D}} N(0, \alpha_1' \tilde{\Sigma}^{(1)} \alpha_1) \quad \text{as } n \rightarrow \infty.$$

The elimination of bias by the construction of an appropriately weighted sum of two estimators is also referred to as the generalized Jackknife methodology; see Gray and Schucany (1972) for further details.

We now turn to the construction of a bias-corrected estimator based on a weighted sum of two estimators of the type $\hat{\gamma}^{(2)}(x, t, K, L)$.

Let $\boldsymbol{\mu}^{(2)}$ be a (2×1) vector with elements

$$\mu_i^{(2)} := \frac{\sqrt{g(x)}}{(1 - \rho(x))^{t_i+1}}, \quad i = 1, 2,$$

and $\tilde{\Sigma}^{(2)}$ a (2×2) matrix with elements

$$\begin{aligned} \tilde{\sigma}_{i,i}^{(2)} &:= \frac{\gamma^2(x) \Gamma(2t_i + 1)}{(t_i + 1) \Gamma^2(t_i + 1)} [2(2t_i + 1) \|K_i\|_2^2 + (t_i + 1) \|L_i\|_2^2 - 2(2t_i + 1) \|K_i L_i\|_1], \quad i = 1, 2, \\ \tilde{\sigma}_{1,2}^{(2)} &:= \frac{\gamma^2(x) \Gamma(t_1 + t_2 + 1)}{\Gamma(t_1 + 2) \Gamma(t_2 + 2)} [(t_1 + t_2 + 1)(t_1 + t_2 + 2) \|K_1 K_2\|_1 - (t_2 + 1)(t_1 + t_2 + 1) \|K_1 L_2\|_1 \\ &\quad - (t_1 + 1)(t_1 + t_2 + 1) \|K_2 L_1\|_1 + (t_1 + 1)(t_2 + 1) \|L_1 L_2\|_1]. \end{aligned}$$

Theorem 5 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample of i.i.d. random vectors, and assume (\mathcal{R}_2) , (\mathcal{G}) , (\mathcal{F}) and kernel functions K_1, K_2, L_1 and L_2 satisfying (\mathcal{K}) . For all $x \in \mathbb{R}^p$ where $g(x) > 0$, if $h \rightarrow 0$, $\omega_n \rightarrow \infty$ with $nh^p \bar{F}(\omega_n; x) \rightarrow \infty$, $\sqrt{nh^p \bar{F}(\omega_n; x) b(\omega_n; x)} \rightarrow \lambda$ and $\sqrt{nh^p \bar{F}(\omega_n; x) h \ln \omega_n} \rightarrow 0$, we have for $t_1, t_2 \geq 0$*

$$\sqrt{nh^p \bar{F}(\omega_n; x) g(x)} \left[\begin{array}{c} \hat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) - \gamma(x) \\ \hat{\gamma}_n^{(2)}(x, t_2, K_2, L_2) - \gamma(x) \end{array} \right] \xrightarrow{\mathcal{D}} N_2(\lambda \boldsymbol{\mu}^{(2)}, \tilde{\Sigma}^{(2)}) \quad \text{as } n \rightarrow \infty.$$

For a weighted estimator

$$\begin{aligned} \hat{\gamma}_n^{(2)}(x, \alpha) &:= \hat{\gamma}_n^{(2)}(x, \alpha, t_1, t_2, K_1, K_2, L_1, L_2) \\ &:= \alpha \hat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) + (1 - \alpha) \hat{\gamma}_n^{(2)}(x, t_2, K_2, L_2), \end{aligned}$$

with $t_1 \neq t_2$ and $\alpha \in \mathbb{R}$, we obtain from Theorem 5 that its asymptotic bias is given by

$$\alpha \lambda \frac{\sqrt{g(x)}}{(1 - \rho(x))^{t_1+1}} + (1 - \alpha) \lambda \frac{\sqrt{g(x)}}{(1 - \rho(x))^{t_2+1}},$$

which vanishes if α is taken as

$$\alpha_{BC}^{(2)}(\rho(x)) := \frac{1}{1 - (1 - \rho(x))^{t_2 - t_1}}.$$

Let $\alpha_2' = [\alpha_{BC}^{(2)}(\rho(x)), 1 - \alpha_{BC}^{(2)}(\rho(x))]$.

Corollary 3 *Under the conditions of Theorem 5 we have that*

$$\sqrt{nh^p \bar{F}(\omega_n; x) g(x)} [\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\rho(x))) - \gamma(x)] \xrightarrow{\mathcal{D}} N(0, \alpha_2' \tilde{\Sigma}^{(2)} \alpha_2) \quad \text{as } n \rightarrow \infty.$$

From the above discussion it is clear that the bias-correcting weights $\alpha_{BC}^{(1)}(\rho(x))$ and $\alpha_{BC}^{(2)}(\rho(x))$ depend on the second order parameter $\rho(x)$, which is typically unknown, and therefore we introduce next a consistent estimator for this parameter. In the univariate context, several estimators have been proposed that show good practical performance, for instance, the estimators by Gomes *et al.* (2002), Fraga Alves *et al.* (2003), Ciuperca and Mercadier (2010) and Goegebeur *et al.* (2010).

Let $\mathbb{K} := (K_1, K_2, K_3)$ and $\mathbb{L} := (L_1, L_2, L_3)$ be kernel functions satisfying (\mathcal{K}) . Inspired by Fraga Alves *et al.* (2003) and **Caeiro and Gomes (2006)**, we introduce the following class of estimators for $\rho(x)$, parameterized in terms of a tuning parameter $\tau \in \mathbb{R}$:

$$\hat{\rho}_n(x; \mathbb{K}, \mathbb{L}, \tau) := \frac{3(R_n(x; \mathbb{K}, \mathbb{L}, \tau) - 1)}{R_n(x; \mathbb{K}, \mathbb{L}, \tau) - 3},$$

provided $1 \leq R_n(x; \mathbb{K}, \mathbb{L}, \tau) < 3$, where

$$R_n(x; \mathbb{K}, \mathbb{L}, \tau) := \frac{\left(\frac{T_n^{(1,1)}(x, K_1)}{T_n^{(1,0)}(x, L_1)} \right)^\tau - \left(\frac{T_n^{(1,2)}(x, K_2)}{2T_n^{(1,0)}(x, L_2)} \right)^{\tau/2}}{\left(\frac{T_n^{(1,2)}(x, K_2)}{2T_n^{(1,0)}(x, L_2)} \right)^{\tau/2} - \left(\frac{T_n^{(1,3)}(x, K_3)}{6T_n^{(1,0)}(x, L_3)} \right)^{\tau/3}}$$

and with the notation $a^{b\tau} := b \ln a$ whenever $\tau = 0$. The consistency of $\hat{\rho}_n(x; \mathbb{K}, \mathbb{L}, \tau)$ for $\rho(x)$ is established in the following theorem.

Theorem 6 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sample of i.i.d. random vectors, and assume (\mathcal{R}_2) , (\mathcal{G}) , (\mathcal{F}) and kernel functions K_1, K_2, K_3, L_1, L_2 and L_3 satisfying (\mathcal{K}) . For all $x \in \mathbb{R}^p$ where $g(x) > 0$, if $h \rightarrow 0$, $\omega_n \rightarrow \infty$ with $\sqrt{nh^p \bar{F}(\omega_n; x) b(\omega_n; x)} \rightarrow \infty$ and $h \ln \omega_n = o(b(\omega_n; x))$, we have that $\hat{\rho}_n(x; \mathbb{K}, \mathbb{L}, \tau) \xrightarrow{\mathbb{P}} \rho(x)$ as $n \rightarrow \infty$.*

The condition $\sqrt{nh^p \bar{F}(\omega_n; x) b(\omega_n; x)} \rightarrow \infty$ in Theorem 6 is quite natural: for instance in the classical framework of extreme value index estimation, where the rate of convergence in the random threshold case is \sqrt{k} , we have to impose $\sqrt{kb}(n/k) \rightarrow \infty$ to obtain consistency of estimators for the second order parameter ρ . Here we have to replace \sqrt{k} by the rate of convergence of our estimator. The condition $h \ln \omega_n = o(b(\omega_n; x))$ is new and specific for the local estimation of $\rho(x)$, and is needed to make the term involving $b(\omega_n; x)$ the dominant one in the asymptotic expansions. The condition $h \ln \omega_n = o(b(\omega_n; x))$ for the estimation of $\rho(x)$ is generally not more restrictive compared to that used in the estimation of $\gamma(x)$: for instance in Theorems 2 and 3 one requires $\sqrt{nh^p \bar{F}(\omega_n; x) b(\omega_n; x)} \rightarrow \lambda$ and $\sqrt{nh^p \bar{F}(\omega_n; x) h \ln \omega_n} \rightarrow 0$, so in case $\lambda \neq 0$ one also has that $h \ln \omega_n = o(b(\omega_n; x))$.

The next two lemma's state that replacing $\rho(x)$ in $\hat{\gamma}_n^{(1)}(x, \alpha_{BC}^{(1)}(\rho(x)))$ and $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\rho(x)))$ by a consistent estimator $\hat{\rho}_n(x)$ does not change the limiting distribution of the bias-corrected estimators.

Corollary 4 *Under the conditions of Theorem 4, and with $\hat{\rho}_n(x)$ being a consistent estimator for $\rho(x)$, we have that*

$$\sqrt{nh^p \bar{F}(\omega_n; x) g(x)} [\hat{\gamma}_n^{(1)}(x, \alpha_{BC}^{(1)}(\hat{\rho}_n(x))) - \gamma(x)] \xrightarrow{\mathcal{D}} N(0, \boldsymbol{\alpha}'_1 \tilde{\boldsymbol{\Sigma}}^{(1)} \boldsymbol{\alpha}_1) \quad \text{as } n \rightarrow \infty.$$

Corollary 5 *Under the conditions of Theorem 5, and with $\hat{\rho}_n(x)$ being a consistent estimator for $\rho(x)$, we have that*

$$\sqrt{nh^p \bar{F}(\omega_n; x) g(x)} [\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\hat{\rho}_n(x))) - \gamma(x)] \xrightarrow{\mathcal{D}} N(0, \boldsymbol{\alpha}'_2 \tilde{\boldsymbol{\Sigma}}^{(2)} \boldsymbol{\alpha}_2) \quad \text{as } n \rightarrow \infty.$$

The corollaries follow from a straightforward application of Taylor's theorem to $\hat{\gamma}_n^{(1)}(x, \alpha_{BC}^{(1)}(\hat{\rho}_n(x)))$ and $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\hat{\rho}_n(x)))$, respectively, and for brevity we omit their proofs.

4 Simulation results

In this section we illustrate our results by a small simulation study. For the practical implementation of our estimators we have to determine the bandwidth parameter h and the threshold ω_n , where we take, as usual in extreme value statistics, the latter as the $(k+1)$ -th largest response observation in the ball $B(x, h)$. We propose two algorithms for the purpose of selecting (h, k) : (i) an oracle strategy and (ii) a completely data driven method.

As proposed in Daouia *et al.* (2011), the oracle strategy consists in selecting (h, k) as follows

$$(h_o, k_o) := \arg \min_{h \in \mathcal{H}_o, k \in \mathcal{K}_o} \Delta(\hat{\gamma}(\cdot), \gamma(\cdot)), \quad (7)$$

where \mathcal{H}_o and \mathcal{K}_o are grids of values for h and k , respectively, and

$$\Delta^2(\hat{\gamma}(\cdot), \gamma(\cdot)) := \frac{1}{M} \sum_{m=1}^M (\hat{\gamma}(z_m) - \gamma(z_m))^2,$$

where z_1, \dots, z_M are regularly spaced in the covariate space. Note that this method requires knowledge of the function $\gamma(x)$, which is unknown in practical situations.

To solve this issue we introduce our second approach which is completely data driven, and based on a two steps procedure. First we select the bandwidth parameter h using the following cross-validation criterion

$$h_{cv} := \arg \min_{h \in \mathcal{H}} \sum_{i=1}^n \sum_{j=1}^n \left(\mathbf{1}\{Y_i \leq Y_j\} - \hat{F}_{n,-i}(Y_j | X_i) \right)^2,$$

where \mathcal{H} is a grid of values for h and

$$\hat{F}_{n,-i}(y|x) := \frac{\sum_{j=1, j \neq i}^n K_h(x - X_j) \mathbf{1}\{Y_j \leq y\}}{\sum_{j=1, j \neq i}^n K_h(x - X_j)}.$$

This criterion was introduced in Yao (1999), implemented by Gannoun *et al.* (2002), and considered in an extreme value context by Daouia *et al.* (2011, 2012). Using this bandwidth we determine in the second step the number of extremes k with a stability criterion of the estimates for $\gamma(x)$. The algorithm that is used is as follows:

For all x under consideration

- compute the estimates for $\gamma(x)$ with $k = 5, \dots, m_x - 1$ (m_x is the number of observations in $B(x, h)$),
- split the range of k into several blocks of same size,
- calculate the standard deviation of the estimates for $\gamma(x)$ in each block,
- the block with minimal standard deviation determines the k to be used (in particular we take the median of the k values in that block).

Note that in the data driven method h is selected globally while k is selected locally.

In the simulations we consider the estimator $\hat{\gamma}_n^{(2)}(x, 0, K, L)$ and its bias-corrected version $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\rho(x)))$, with $t_1 = 0$ and $t_2 = 1$. All kernel functions are taken as the bi-quadratic kernel function

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

We assume that the conditional distribution of Y given $X = x$ is the following Burr distribution:

$$1 - F(y; x) = \left(1 + y^{-\rho(x)/\gamma(x)}\right)^{1/\rho(x)},$$

where $\rho(x) < 0$ is the second order parameter, and X is uniformly distributed on $(0, 1)$. We fix $\rho(x) = -1$ and the function $\gamma(\cdot)$ is, as in Daouia *et al.* (2011), taken as

$$\gamma(x) = \frac{1}{2} \left(\frac{1}{10} + \sin(\pi x) \right) \left(\frac{11}{10} - \frac{1}{2} \exp \left(-64 \left(x - \frac{1}{2} \right)^2 \right) \right).$$

We simulate $N = 100$ samples of size $n = 1000$.

In Figure 2 we show the boxplots of the 100 realizations of **the estimates of $\gamma(x)$ and $\rho(x)$** for different values of x , as obtained with the oracle method. The minimization (7) is performed on a grid of $h \in [0.05, 0.5]$ and of $k \in \{2, \dots, m_x - 1\}$, with $M = 35$. For the bias-corrected estimator for $\gamma(x)$, different values for $\rho(x)$ have been used namely, the true value $\rho(x) = -1$, a misspecification at $\rho(x) = -2$ and -5 , and an estimation of $\rho(x)$ using $\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, \tau)$ with $\tau = 0.5$.

The value 0.5 for the tuning parameter τ can be motivated by the simulation results reported in Fraga Alves *et al.* (2003) for the univariate i.i.d. context, where it gave a reasonable performance for a wide range of models. In case $\rho(x)$ is estimated, we use an oracle strategy, similar to the one discussed above, with a grid of $h \in [0.05, 0.5]$, but this time $k \in \{m_x/4, \dots, m_x - 1\}$. As expected, the bias-corrected estimator $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\rho(x)))$ outperforms the uncorrected version $\hat{\gamma}_n^{(2)}(x, 0, K, K)$ in terms of bias, even with a mis-specification of $\rho(x)$. The estimator for $\rho(x)$ has a reasonable performance, despite the small local sample sizes. In Table 1 we compare the errors $\Delta^2(\hat{\gamma}(\cdot), \gamma(\cdot))$, averaged over the 100 simulated data sets. Also in terms of these, $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\rho(x)))$ typically works better than $\hat{\gamma}_n^{(2)}(x, 0, K, K)$, **except when $\rho(x)$ is estimated. In this case, the variability of the estimator for $\rho(x)$ may compensate for the bias-correction by $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)))$, leading to a slightly larger error.**

In Figure 3 we show the corresponding results obtained with the data driven method. **Concerning** the estimation of $\gamma(x)$, we select the optimal value of the bandwidth h in a set $\mathcal{H} = \{h_1 \leq \dots \leq h_{10}\}$ where the points h_1, \dots, h_{10} are regularly spaced in $[0.05, 0.5]$, and the selection of k is along the above described algorithm with a block size of 40. For the estimation of $\rho(x)$, we select h **as above with leave-one-out** cross-validation, but we set $k = m_x^{0.975}$, see e.g. Caeiro *et al.* (2009). Clearly, all implementations of the bias-corrected estimator outperform the uncorrected estimator in terms of bias, even with a severe mis-specification at $\rho(x) = -5$. The superior performance is also reflected in the errors reported in Table 1. Note in particular that the error of $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\rho(x)))$ with an estimation of $\rho(x)$ is close to the one with the true value $\rho(x) = -1$. Also, they are only slightly larger in the data driven setting than in the oracle framework.

With the data driven method, the variance of the bias-corrected estimates seems to be smaller under a mis-specification of $\rho(x)$ compared to when the true value is used. A possible explanation for this observation is as follows. If one uses the true value of $\rho(x)$, then the bias-corrected estimators for $\gamma(x)$ are asymptotically unbiased, according to Corollaries 2 and 3. If one mis-specifies $\rho(x)$ then one typically loses the bias correction, though the asymptotic variance of the estimator is still given by the expressions in Corollaries 2 and 3, though with $\rho(x)$ replaced by its mis-specified value, and this variance is smaller when $|\rho(x)|$ is large, see Figure 4 where we show the asymptotic standard deviation of the implemented bias-corrected estimator for $\gamma(x)$ as a function of $\rho(x)$. A similar phenomenon is observed in the univariate i.i.d. case: the asymptotic variance of an asymptotically unbiased estimator is typically given by $\gamma^2(1 - \rho)^2/\rho^2$, which is a decreasing function of $|\rho|$. This variance expression remains valid if one mis-specifies ρ at ρ^* , giving $\gamma^2(1 - \rho^*)^2/\rho^{*2}$ as asymptotic variance. This could also explain the results obtained when $\rho(x)$ is replaced by its estimator since the estimator for $\rho(x)$ suffers somehow from a negative bias over the complete range of x . The oracle results do not show the dependency of the variance of the bias-corrected estimator on the second order parameter as clearly, but this could be due to a different implementation: in the oracle method both h and k are

chosen globally, whereas in the data-driven method h is chosen globally and k locally.

From the simulation results, the variance of the tail index estimates tends to decrease near the boundaries. Note that at the theoretical level, the asymptotic variance of the proposed tail index estimators is increasing in $\gamma(x)$. We refer to Theorems 2 and 3 for general local kernel estimators, and to Corollary 1 for the case of the local Hill estimator. These findings also apply for the bias-corrected estimators, as is clear from Theorems 4 and 5, and associated corollaries. In finite samples, this can also be seen in Figures 2 and 3, where the whiskers are generally shorter for the smaller values of $\gamma(x)$. The smallest values of $\gamma(x)$ occur in this simulation setting near the boundaries, and it seems that the smaller asymptotic variance is thus more important than possible boundary effects, that could e.g. lead to smaller local values of k near the boundaries.

Finally, we comment on the behavior of the local estimator for $\rho(x)$. From panel (f) in Figures 2 and 3 it is clear that the proposed estimator suffers from some negative bias. Concerning this issue, it should be noted that the original estimator for ρ proposed by Fraga Alves *et al.* (2003) in the i.i.d. univariate case shows, when implemented with $\tau = 0.5$, often a negative bias in cases where $-2 < \rho < 0$. We refer to the simulation results in the Fraga Alves *et al.* (2003) paper for further details. In Goegebeur *et al.* (2010), several estimators for the second order parameter ρ were compared in a simulation experiment, and from the results reported there it seems that they all suffer from a negative bias when $-2 < \rho < 0$. These findings are consistent with the results in the present paper where the estimator for $\rho(x)$ also shows a negative bias over the whole range of x -values. That the bias seems to be more severe at the boundaries is harder to explain. On the one hand, the leading terms of the asymptotic bias of the estimator for $\rho(x)$ will probably show a dependence on the tail index $\gamma(x)$, as was also observed in the univariate i.i.d. case (we refer again to the above mentioned papers). In fact, from the Figures 2 and 3 one can see already that the bias of $\rho(x)$ indeed follows the behavior of $\gamma(x)$. On the other hand, the typical boundary bias that appears in local estimation can be playing here. We would though want to mention that the smallest and largest values in our grid for x were 0.1 and 0.9, respectively, whereas the distribution of the covariates was $U(0,1)$, meaning that we stayed somehow away from the boundary. Whatever the reason, we see that in the data-driven method, which is the practically relevant method, despite the negative bias in the estimation of $\rho(x)$, the boundary bias of the bias-corrected estimator for $\gamma(x)$ is not worse compared to the case where the true value $\rho(x) = -1$ was used.

Appendix: Proofs

Proof of Lemma 1

First consider the case $t > 0$. From the definition of conditional expectation one easily obtains

$$\begin{aligned} m_n(x; t) &= t \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{1}{y} \bar{F}(y; x) dy \\ &= t \bar{F}(\omega_n; x) \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y \bar{F}(\omega_n; x)} dy \\ &= t \bar{F}(\omega_n; x) \int_1^{\infty} (\ln z)^{t-1} \frac{\bar{F}(\omega_n z; x)}{z \bar{F}(\omega_n; x)} dz. \end{aligned}$$

Concerning the ratio $\bar{F}(\omega_n z; x)/\bar{F}(\omega_n; x)$ we invoke condition (\mathcal{R}_2) . From Theorem B.2.18 in de Haan and Ferreira (2006), see also Drees (1998), we have that for a function b_0 , possibly different from the function b , though with $b_0(\omega_n; x) \sim b(\omega_n; x)$, $\omega_n \rightarrow \infty$, and for each $\varepsilon, \delta > 0$ there exists a $\omega_{n,0}$ such that for $\omega_n \geq \omega_{n,0}$, $z > 1$,

$$\left| \gamma^2(x) \frac{z^{1/\gamma(x)} \frac{\bar{F}(\omega_n z; x)}{\bar{F}(\omega_n; x)} - 1}{b_0(\omega_n; x)} - \frac{z^{\rho(x)/\gamma(x)} - 1}{\rho(x)/\gamma(x)} \right| \leq \varepsilon z^{\rho(x)/\gamma(x) + \delta}. \quad (8)$$

Motivated by (8), we write

$$\begin{aligned} m_n(x; t) &= t \bar{F}(\omega_n; x) \left\{ \int_1^{\infty} (\ln z)^{t-1} z^{-1/\gamma(x)-1} dz \right. \\ &\quad + \frac{b_0(\omega_n; x)}{\gamma^2(x)} \int_1^{\infty} (\ln z)^{t-1} z^{-1/\gamma(x)-1} \frac{z^{\rho(x)/\gamma(x)} - 1}{\rho(x)/\gamma(x)} dz \\ &\quad \left. + \frac{b_0(\omega_n; x)}{\gamma^2(x)} \int_1^{\infty} (\ln z)^{t-1} z^{-1/\gamma(x)-1} \left(\gamma^2(x) \frac{z^{1/\gamma(x)} \frac{\bar{F}(\omega_n z; x)}{\bar{F}(\omega_n; x)} - 1}{b_0(\omega_n; x)} - \frac{z^{\rho(x)/\gamma(x)} - 1}{\rho(x)/\gamma(x)} \right) dz \right\} \\ &=: t \bar{F}(\omega_n; x) \left\{ I_1 + \frac{b_0(\omega_n; x)}{\gamma^2(x)} (I_2 + I_3) \right\}. \end{aligned}$$

For I_1 and I_2 one obtains easily

$$\begin{aligned} I_1 &= \gamma^t(x) \Gamma(t), \\ I_2 &= \frac{\gamma^{t+1}(x) \Gamma(t)}{\rho(x)} \left[\frac{1}{(1 - \rho(x))^t} - 1 \right]. \end{aligned}$$

Concerning I_3 we use (8). Hence, for all $\varepsilon > 0$ we have that there exists an $\omega_{n,0}$ such that for $\omega_n > \omega_{n,0}$

$$\begin{aligned} |I_3| &\leq \varepsilon \int_1^\infty (\ln z)^{t-1} z^{-(1-\rho(x))/\gamma(x)+\delta-1} dz \\ &= \varepsilon \int_0^\infty \exp(-((1-\rho(x))/\gamma(x) - \delta)u) u^{t-1} du \\ &= \varepsilon \frac{\Gamma(t)}{\left(\frac{1-\rho(x)}{\gamma(x)} - \delta\right)^t}, \end{aligned}$$

provided $0 < \delta < (1-\rho(x))/\gamma(x)$. We have thus that $I_3 = o(1)$ for $\omega_n \rightarrow \infty$. The result for $t > 0$ is then obtained by collecting the terms.

The case $t = 0$ follows trivially from the definition of conditional expectation.

Proof of Lemma 2

From the rule of repeated expectations we obtain

$$\begin{aligned} \tilde{m}_n(x; t, s) &= \mathbb{E}[K_h^s(x - X)m_n(X; t)] \\ &= \int_{\mathbb{R}^p} K_h^s(x - u)m_n(u; t)g(u)du \\ &= \frac{\|K\|_s^s}{h^{(s-1)p}} \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} m_n(x - hz; t)g(x - hz)dz. \end{aligned}$$

In case $t > 0$ we have

$$\tilde{m}_n(x; t, s) = \frac{t\|K\|_s^s}{h^{(s-1)p}} \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \int_{\omega_n}^\infty (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x - hz)}{y} dy g(x - hz)dz,$$

and, consequently

$$\begin{aligned} &\left| \tilde{m}_n(x; t, s) - \frac{m_n(x; t)}{h^{(s-1)p}} g(x)\|K\|_s^s \right| \\ &\leq \frac{t\|K\|_s^s}{h^{(s-1)p}} \left\{ \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \int_{\omega_n}^\infty (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} \left| \frac{\bar{F}(y; x - hz)}{\bar{F}(y; x)} - 1 \right| dy g(x)dz \right. \\ &\quad + \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \int_{\omega_n}^\infty (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} dy |g(x - hz) - g(x)|dz \\ &\quad \left. + \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \int_{\omega_n}^\infty (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} \left| \frac{\bar{F}(y; x - hz)}{\bar{F}(y; x)} - 1 \right| dy |g(x - hz) - g(x)|dz \right\} \\ &=: \frac{t\|K\|_s^s}{h^{(s-1)p}} \{I_4 + I_5 + I_6\}. \end{aligned}$$

Concerning I_4 , let

$$\tilde{I}_4 := \int_{\omega_n}^\infty (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} \left| \frac{\bar{F}(y; x - hz)}{\bar{F}(y; x)} - 1 \right| dy,$$

note that

$$\frac{\bar{F}(y; x - hz)}{\bar{F}(y; x)} = \exp \left[\ln \bar{F}(y; x) \left(\frac{\ln \bar{F}(y; x - hz)}{\ln \bar{F}(y; x)} - 1 \right) \right],$$

and, by using (\mathcal{F}) and the property that $\ln \ell(y; x) / \ln y \rightarrow 0$ as $y \rightarrow \infty$, we have that if ω_n is sufficiently large then

$$|\ln \bar{F}(y; x)| \left| \frac{\ln \bar{F}(y; x - hz)}{\ln \bar{F}(y; x)} - 1 \right| \leq Ch \ln y,$$

for some constant $C > 0$. Now use the well known inequality $|\exp(w) - 1| \leq |w| + \frac{1}{2}(1 + \exp(|w|))w^2$ to obtain for ω_n sufficiently large

$$\begin{aligned} \tilde{I}_4 &\leq Ch \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} \ln y \, dy \\ &\quad + \frac{C^2 h^2}{2} \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} \ln^2 y \, dy \\ &\quad + \frac{C^2 h^2}{2} \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} \exp(Ch \ln y) \ln^2 y \, dy \\ &=: \tilde{I}_{4,1} + \tilde{I}_{4,2} + \tilde{I}_{4,3}. \end{aligned}$$

From Lemma 1 we have that

$$\begin{aligned} \tilde{I}_{4,1} &= \frac{m_n(x; t)}{t} O(h \ln \omega_n), \\ \tilde{I}_{4,2} &= \frac{m_n(x; t)}{t} O(h^2 \ln^2 \omega_n). \end{aligned}$$

Concerning $\tilde{I}_{4,3}$, by the substitution $v = y/\omega_n$ and a straightforward expansion

$$\begin{aligned} \tilde{I}_{4,3} &= \frac{C^2 h^2 \exp(Ch \ln \omega_n)}{2} \left\{ \ln^2 \omega_n \int_1^{\infty} (\ln v)^{t-1} \frac{\bar{F}(\omega_n v; x)}{v} \exp(Ch \ln v) \, dv \right. \\ &\quad + \int_1^{\infty} (\ln v)^{t+1} \frac{\bar{F}(\omega_n v; x)}{v} \exp(Ch \ln v) \, dv \\ &\quad \left. + 2 \ln \omega_n \int_1^{\infty} (\ln v)^t \frac{\bar{F}(\omega_n v; x)}{v} \exp(Ch \ln v) \, dv \right\} \\ &=: \frac{C^2 h^2 \exp(Ch \ln \omega_n)}{2} \left\{ \ln^2 \omega_n \tilde{I}_{4,3,1} + \tilde{I}_{4,3,2} + 2 \ln \omega_n \tilde{I}_{4,3,3} \right\}. \end{aligned}$$

Since all the integrals in the right hand side of the expression are of the same form, we will only consider the first of them. Again, inspired by (8), we write

$$\begin{aligned}
\tilde{I}_{4,3,1} &= \bar{F}(\omega_n; x) \left\{ \int_1^\infty (\ln v)^{t-1} v^{Ch-1-1/\gamma(x)} dv \right. \\
&\quad + \frac{b_0(\omega_n; x)}{\gamma^2(x)} \int_1^\infty (\ln v)^{t-1} v^{Ch-1-1/\gamma(x)} \frac{v^{\rho(x)/\gamma(x)} - 1}{\rho(x)/\gamma(x)} dv \\
&\quad \left. + \frac{b_0(\omega_n; x)}{\gamma^2(x)} \int_1^\infty (\ln v)^{t-1} v^{Ch-1-1/\gamma(x)} \left(\gamma^2(x) \frac{v^{1/\gamma(x)} \frac{\bar{F}(\omega_n v; x)}{\bar{F}(\omega_n; x)} - 1}{b_0(\omega_n; x)} - \frac{v^{\rho(x)/\gamma(x)} - 1}{\rho(x)/\gamma(x)} \right) dv \right\} \\
&=: \bar{F}(\omega_n; x) \left\{ \tilde{I}_{4,3,1,1} + \frac{b_0(\omega_n; x)}{\gamma^2(x)} (\tilde{I}_{4,3,1,2} + \tilde{I}_{4,3,1,3}) \right\}.
\end{aligned}$$

Taking into account that Ch can be made arbitrarily small we have that

$$\begin{aligned}
\tilde{I}_{4,3,1,1} &= \gamma^t(x) \Gamma(t) + o(1), \\
\tilde{I}_{4,3,1,2} &= \frac{\gamma^{t+1}(x) \Gamma(t)}{\rho(x)} \left[\frac{1}{(1 - \rho(x))^t} - 1 \right] + o(1).
\end{aligned}$$

Concerning $\tilde{I}_{4,3,1,3}$ we assume that $Ch < \delta'$. Then, again, by (8) we have that for all $\varepsilon > 0$ there exists an $\omega_{n,0}$ such that for $\omega_n > \omega_{n,0}$

$$\begin{aligned}
|\tilde{I}_{4,3,1,3}| &\leq \varepsilon \int_1^\infty (\ln v)^{t-1} v^{-(1-\rho(x))/\gamma(x) + \delta' + \delta - 1} dv \\
&= \varepsilon \frac{\Gamma(t)}{\left(\frac{1-\rho(x)}{\gamma(x)} - \delta' - \delta \right)^t},
\end{aligned}$$

provided $0 < \delta + \delta' < (1 - \rho(x))/\gamma(x)$. Hence $\tilde{I}_{4,3,1,3} = o(1)$ under the conditions of the lemma. Thus

$$\begin{aligned}
\tilde{I}_{4,3,1} &= \bar{F}(\omega_n; x) \gamma^t(x) \Gamma(t) \left\{ 1 + o(1) + \frac{b(\omega_n; x)}{\gamma(x) \rho(x)} \left[\frac{1}{(1 - \rho(x))^t} - 1 \right] (1 + o(1)) \right\} \\
&= \frac{m_n(x; t)}{t} (1 + o(1)).
\end{aligned}$$

Collecting the above results gives that

$$\tilde{I}_{4,3} = \frac{m_n(x; t)}{t} O(h^2 \ln^2 \omega_n).$$

Combined,

$$I_4 = \frac{1}{t} g(x) m_n(x; t) O(h \ln \omega_n),$$

using the fact that $K^s(\cdot)/\|K\|_s^s$ satisfies (\mathcal{K}) .

Concerning I_5 , we use (\mathcal{G}) and (\mathcal{K}) to obtain

$$\begin{aligned} I_5 &\leq c_g \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} d(0, hz) dz \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\bar{F}(y; x)}{y} dy \\ &= \frac{1}{t} g(x) m_n(x; t) O(h). \end{aligned}$$

Similarly, by using (\mathcal{F}) , (\mathcal{G}) and (\mathcal{K}) we have that, for ω_n sufficiently large,

$$I_6 = \frac{1}{t} g(x) m_n(x; t) O(h^2 \ln \omega_n).$$

Combining the above establishes the result for the case $t > 0$.

We now turn to the case $t = 0$. We have

$$\tilde{m}_n(x; 0, s) = \frac{\|K\|_s^s}{h^{(s-1)p}} \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \bar{F}(\omega_n; x - hz) g(x - hz) dz.$$

Further,

$$\begin{aligned} &\left| \tilde{m}_n(x; 0, s) - \frac{m_n(x; 0)}{h^{(s-1)p}} g(x) \|K\|_s^s \right| \\ &\leq \frac{\|K\|_s^s}{h^{(s-1)p}} \left\{ \bar{F}(\omega_n; x) g(x) \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \left| \frac{\bar{F}(\omega_n; x - hz)}{\bar{F}(\omega_n; x)} - 1 \right| dz \right. \\ &\quad + \bar{F}(\omega_n; x) \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} |g(x - hz) - g(x)| dz \\ &\quad \left. + \bar{F}(\omega_n; x) \int_{\Omega} \frac{K^s(z)}{\|K\|_s^s} \left| \frac{\bar{F}(\omega_n; x - hz)}{\bar{F}(\omega_n; x)} - 1 \right| |g(x - hz) - g(x)| dz \right\}. \end{aligned}$$

The result for the case $t = 0$ can now be established along a **similar line of arguments** as used for the case $t > 0$.

Proof of Theorem 1

To prove Theorem 1 we make use of the Cramér-Wold device (see e.g. van der Vaart, 1998, p 16), **according to which it is sufficient to prove that for all $\xi \in \mathbb{R}^J$, $\|\xi\| \neq 0$, we have that**

$$\Lambda_n := \sqrt{nh^p \bar{F}(\omega_n; x) g(x)} \xi' [\mathbb{T}_n - \mathbb{E}(\mathbb{T}_n)] \xrightarrow{\mathcal{D}} N(0, \xi' \Sigma \xi).$$

From straightforward calculations we obtain

$$\begin{aligned}
\Lambda_n &= \sum_{i=1}^n \sqrt{\frac{h^p}{n\bar{F}(\omega_n; x)g(x)}} \left[\sum_{j=1}^J \xi_j K_{j,h}(x - X_i) (\ln Y_i - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_i > \omega_n\} \right. \\
&\quad \left. - \mathbb{E} \left(\sum_{j=1}^J \xi_j K_{j,h}(x - X_i) (\ln Y_i - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_i > \omega_n\} \right) \right] \\
&=: \sum_{i=1}^n V_i.
\end{aligned}$$

Note that V_1, \dots, V_n are i.i.d. random variables, and hence $\mathbb{V}\text{ar}(\Lambda_n) = n\mathbb{V}\text{ar}(V_1)$. Now

$$\begin{aligned}
\mathbb{V}\text{ar}(V_1) &= \frac{h^p}{n\bar{F}(\omega_n; x)g(x)} \sum_{j=1}^J \sum_{k=1}^J \xi_j \xi_k \text{Cov} \left(K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\}, \right. \\
&\quad \left. K_{k,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_k} \mathbf{1}\{Y_1 > \omega_n\} \right) \\
&=: \frac{h^p}{n\bar{F}(\omega_n; x)g(x)} \sum_{j=1}^J \sum_{k=1}^J \xi_j \xi_k \mathbb{C}_{j,k},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{C}_{j,k} &= \mathbb{E} \left[K_{j,h}(x - X_1) K_{k,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j+t_k} \mathbf{1}\{Y_1 > \omega_n\} \right] \\
&\quad - \mathbb{E} \left[K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\} \right] \mathbb{E} \left[K_{k,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_k} \mathbf{1}\{Y_1 > \omega_n\} \right] \\
&= \frac{\|K_j K_k\|_1}{h^p} \mathbb{E} \left[\frac{1}{h^p \|K_j K_k\|_1} K_j \left(\frac{x - X_1}{h} \right) K_k \left(\frac{x - X_1}{h} \right) (\ln Y_1 - \ln \omega_n)_+^{t_j+t_k} \mathbf{1}\{Y_1 > \omega_n\} \right] \\
&\quad - \mathbb{E} \left[K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\} \right] \mathbb{E} \left[K_{k,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_k} \mathbf{1}\{Y_1 > \omega_n\} \right].
\end{aligned}$$

Using Lemma 1 and 2 we obtain that

$$\mathbb{C}_{j,k} = \frac{\|K_j K_k\|_1}{h^p} \gamma^{t_j+t_k}(x) \bar{F}(\omega_n; x) g(x) \Gamma(t_j + t_k + 1) (1 + o(1)),$$

and consequently $\mathbb{V}\text{ar}(\Lambda_n) = \xi' \Sigma \xi (1 + o(1))$.

To establish the asymptotic normality of Λ_n we verify Lyapounov's criterion for triangular arrays of random variables, see e.g. Billingsley (1995). In the present context this simplifies to proving

that $\sum_{i=1}^n \mathbb{E}|V_i|^3 = n\mathbb{E}|V_1|^3 \rightarrow 0$. We have

$$\begin{aligned} \mathbb{E}|V_1|^3 &\leq \left(\frac{h^p}{n\bar{F}(\omega_n; x)g(x)} \right)^{3/2} \left\{ \mathbb{E} \left[\left(\sum_{j=1}^J |\xi_j| K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\} \right)^3 \right] \right. \\ &\quad + 3\mathbb{E} \left[\left(\sum_{j=1}^J |\xi_j| K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\} \right)^2 \right] \\ &\quad \times \mathbb{E} \left[\sum_{j=1}^J |\xi_j| K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\} \right] \\ &\quad \left. + 4 \left[\mathbb{E} \left(\sum_{j=1}^J |\xi_j| K_{j,h}(x - X_1) (\ln Y_1 - \ln \omega_n)_+^{t_j} \mathbf{1}\{Y_1 > \omega_n\} \right) \right]^3 \right\}. \end{aligned}$$

Again by using Lemma 1 and 2 we obtain that

$$\mathbb{E}|V_1|^3 = O \left(\left(n \sqrt{nh^p \bar{F}(\omega_n; x)} \right)^{-1} \right), \quad (9)$$

and hence, under the conditions of the theorem $n\mathbb{E}|V_1|^3 \rightarrow 0$.

Proof of Theorem 2

Let $r_n := \sqrt{nh^p \bar{F}(\omega_n; x)g(x)}$. Write

$$\begin{aligned} r_n &\left(\frac{T_n^{(1,t)}(x, K)}{\Gamma(t+1)T_n^{(1,0)}(x, L)} - \gamma^t(x) \right) \\ &= \frac{\bar{F}(\omega_n; x)g(x)}{T_n^{(1,0)}(x, L)} \left\{ r_n \left[\frac{T_n^{(1,t)}(x, K)}{\Gamma(t+1)\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(1,t)}(x, K)}{\Gamma(t+1)\bar{F}(\omega_n; x)g(x)} \right) \right] \right. \\ &\quad - \gamma^t(x)r_n \left[\frac{T_n^{(1,0)}(x, L)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(1,0)}(x, L)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ &\quad \left. + r_n \left[\frac{\mathbb{E}(T_n^{(1,t)}(x, K)) - \gamma^t(x)\Gamma(t+1)\mathbb{E}(T_n^{(1,0)}(x, L))}{\Gamma(t+1)\bar{F}(\omega_n; x)g(x)} \right] \right\} \\ &=: \frac{\bar{F}(\omega_n; x)g(x)}{T_n^{(1,0)}(x, L)} \{T_1 + T_2 + T_3\}. \end{aligned}$$

From Theorem 1, we have

$$T_1 + T_2 \xrightarrow{D} N \left(0, \frac{\gamma^{2t}(x)}{\Gamma^2(t+1)} [\Gamma(2t+1)\|K\|_2^2 + \Gamma^2(t+1)\|L\|_2^2 - 2\Gamma^2(t+1)\|KL\|_1] \right).$$

Concerning T_3 , by using Lemma 1 and 2

$$T_3 = r_n \gamma^t(x) \left\{ \frac{b(\omega_n; x)}{\gamma(x)\rho(x)} \left(\frac{1}{(1-\rho(x))^t} - 1 \right) (1 + o(1)) + O(h \ln \omega_n) \right\},$$

and hence under our assumptions, $T_3 \rightarrow \lambda \sqrt{g(x)} \frac{\gamma^{t-1}(x)}{\rho(x)} \left(\frac{1}{(1-\rho(x))^t} - 1 \right)$.

Further, from Theorem 1 we have that $T_n^{(1,0)}(x, L) / (\bar{F}(\omega_n; x)g(x)) = 1 + o_{\mathbb{P}}(1)$.

Combined, the above gives that

$$r_n \left(\frac{T_n^{(1,t)}(x, K)}{\Gamma(t+1)T_n^{(1,0)}(x, L)} - \gamma^t(x) \right) \xrightarrow{\mathcal{D}} N \left(\lambda \sqrt{g(x)} \frac{\gamma^{t-1}(x)}{\rho(x)} \left(\frac{1}{(1-\rho(x))^t} - 1 \right), \frac{\gamma^{2t}(x)}{\Gamma^2(t+1)} \left[\Gamma(2t+1)\|K\|_2^2 + \Gamma^2(t+1)\|L\|_2^2 - 2\Gamma^2(t+1)\|KL\|_1 \right] \right).$$

Finally, after a straightforward application of the δ -method, Theorem 2 follows.

Proof of Theorem 3

First write

$$\begin{aligned} & r_n \left(\frac{T_n^{(1,t+1)}(x, K)}{(t+1)T_n^{(1,t)}(x, L)} - \gamma(x) \right) \\ &= \frac{\bar{F}(\omega_n; x)g(x)}{T_n^{(1,t)}(x, L)} \left\{ r_n \left[\frac{T_n^{(1,t+1)}(x, K)}{(t+1)\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(1,t+1)}(x, K)}{(t+1)\bar{F}(\omega_n; x)g(x)} \right) \right] \right. \\ & \quad - \gamma(x)r_n \left[\frac{T_n^{(1,t)}(x, L)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(1,t)}(x, L)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ & \quad \left. + r_n \left[\frac{\mathbb{E}(T_n^{(1,t+1)}(x, K)) - \gamma(x)(t+1)\mathbb{E}(T_n^{(1,t)}(x, L))}{(t+1)\bar{F}(\omega_n; x)g(x)} \right] \right\} \\ &=: \frac{\bar{F}(\omega_n; x)g(x)}{T_n^{(1,t)}(x, L)} \{T_4 + T_5 + T_6\}. \end{aligned}$$

From Theorem 1, we have

$$T_4 + T_5 \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{\gamma^{2t+2}(x)\Gamma(2t+1)}{t+1} \left[2(2t+1)\|K\|_2^2 + (t+1)\|L\|_2^2 - 2(2t+1)\|KL\|_1 \right] \right).$$

Concerning T_6 , using Lemma 1 and 2, we have

$$T_6 = r_n \left\{ \frac{\gamma^t(x)\Gamma(t+1)b(\omega_n; x)}{(1-\rho(x))^{t+1}} (1 + o(1)) + O(h \ln \omega_n) \right\},$$

and hence under our assumptions, we have $T_6 \rightarrow \lambda \sqrt{g(x)} \frac{\gamma^t(x)\Gamma(t+1)}{(1-\rho(x))^{t+1}}$.

Now, according to Theorem 1, $T_n^{(1,t)}(x, L) / (\bar{F}(\omega_n; x)g(x)) = \gamma^t(x)\Gamma(t+1) + o_{\mathbb{P}}(1)$. This achieves the proof of Theorem 3.

Proof of Theorem 4

First we will show that under the conditions of the theorem

$$r_n \begin{bmatrix} \frac{T_n^{(1,t_1)}(x,K_1)}{\Gamma(t_1+1)T_n^{(1,0)}(x,L_1)} - \gamma^{t_1}(x) \\ \frac{T_n^{(1,t_2)}(x,K_2)}{\Gamma(t_2+1)T_n^{(1,0)}(x,L_2)} - \gamma^{t_2}(x) \end{bmatrix} \xrightarrow{\mathcal{D}} N_2(\boldsymbol{\delta}, \boldsymbol{\Omega}) \quad (10)$$

whereafter the result will follow by a straightforward application of the δ -method. Using the Cramér-Wold device, to establish (10) **it is sufficient to show that for all** $\xi \in \mathbb{R}^2$, $\|\xi\| \neq 0$, **we** have that

$$\xi' r_n \begin{bmatrix} \frac{T_n^{(1,t_1)}(x,K_1)}{\Gamma(t_1+1)T_n^{(1,0)}(x,L_1)} - \gamma^{t_1}(x) \\ \frac{T_n^{(1,t_2)}(x,K_2)}{\Gamma(t_2+1)T_n^{(1,0)}(x,L_2)} - \gamma^{t_2}(x) \end{bmatrix} \xrightarrow{\mathcal{D}} N(\xi' \boldsymbol{\delta}, \xi' \boldsymbol{\Omega} \xi). \quad (11)$$

From Lemma 1 and 2, and Theorem 1 we have that $T_n^{(1,0)}(x, L_i)/(\bar{F}(\omega_n; x)g(x)) = 1 + o_{\mathbb{P}}(1)$, $i = 1, 2$, and consequently, **by the line of arguments** of Theorem 2 we obtain

$$\begin{aligned} \xi' r_n \begin{bmatrix} \frac{T_n^{(1,t_1)}(x,K_1)}{\Gamma(t_1+1)T_n^{(1,0)}(x,L_1)} - \gamma^{t_1}(x) \\ \frac{T_n^{(1,t_2)}(x,K_2)}{\Gamma(t_2+1)T_n^{(1,0)}(x,L_2)} - \gamma^{t_2}(x) \end{bmatrix} &= \xi_1 r_n \frac{T_n^{(1,t_1)}(x, K_1) - \gamma^{t_1}(x)\Gamma(t_1+1)T_n^{(1,0)}(x, L_1)}{\Gamma(t_1+1)\bar{F}(\omega_n; x)g(x)} \\ &\quad + \xi_2 r_n \frac{T_n^{(1,t_2)}(x, K_2) - \gamma^{t_2}(x)\Gamma(t_2+1)T_n^{(1,0)}(x, L_2)}{\Gamma(t_2+1)\bar{F}(\omega_n; x)g(x)} \\ &\quad + o_{\mathbb{P}}(1), \end{aligned}$$

which can be further decomposed as

$$\begin{aligned} \xi' r_n \begin{bmatrix} \frac{T_n^{(1,t_1)}(x,K_1)}{\Gamma(t_1+1)T_n^{(1,0)}(x,L_1)} - \gamma^{t_1}(x) \\ \frac{T_n^{(1,t_2)}(x,K_2)}{\Gamma(t_2+1)T_n^{(1,0)}(x,L_2)} - \gamma^{t_2}(x) \end{bmatrix} &= \frac{\xi_1}{\Gamma(t_1+1)} r_n \left[\frac{T_n^{(1,t_1)}(x, K_1)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(1,t_1)}(x, K_1)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ &\quad - \xi_1 \gamma^{t_1}(x) r_n \left[\frac{T_n^{(1,0)}(x, L_1)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(1,0)}(x, L_1)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ &\quad + \frac{\xi_2}{\Gamma(t_2+1)} r_n \left[\frac{T_n^{(1,t_2)}(x, K_2)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(1,t_2)}(x, K_2)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ &\quad - \xi_2 \gamma^{t_2}(x) r_n \left[\frac{T_n^{(1,0)}(x, L_2)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(1,0)}(x, L_2)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ &\quad + \xi_1 r_n \frac{\mathbb{E}(T_n^{(1,t_1)}(x, K_1)) - \gamma^{t_1}(x)\Gamma(t_1+1)\mathbb{E}(T_n^{(1,0)}(x, L_1))}{\Gamma(t_1+1)\bar{F}(\omega_n; x)g(x)} \\ &\quad + \xi_2 r_n \frac{\mathbb{E}(T_n^{(1,t_2)}(x, K_2)) - \gamma^{t_2}(x)\Gamma(t_2+1)\mathbb{E}(T_n^{(1,0)}(x, L_2))}{\Gamma(t_2+1)\bar{F}(\omega_n; x)g(x)} \\ &\quad + o_{\mathbb{P}}(1) \\ &=: T_7 + T_8 + T_9 + T_{10} + T_{11} + T_{12} + o_{\mathbb{P}}(1). \end{aligned}$$

From Theorem 1 we have that

$$T_7 + T_8 + T_9 + T_{10} \xrightarrow{\mathcal{D}} N(0, \xi' \mathbf{\Omega} \xi),$$

where $\mathbf{\Omega}$ is a (2×2) matrix with elements

$$\begin{aligned} \omega_{i,i} &:= \frac{\gamma^{2t_i}(x)}{\Gamma^2(t_i + 1)} [\Gamma(2t_i + 1) \|K_i\|_2^2 + \Gamma^2(t_i + 1) \|L_i\|_2^2 - 2\Gamma^2(t_i + 1) \|K_i L_i\|_1], \quad i = 1, 2, \\ \omega_{1,2} &:= \frac{\gamma^{t_1+t_2}(x)}{\Gamma(t_1 + 1)\Gamma(t_2 + 1)} [\Gamma(t_1 + t_2 + 1) \|K_1 K_2\|_1 + \Gamma(t_1 + 1)\Gamma(t_2 + 1) \|L_1 L_2\|_1 \\ &\quad - \Gamma(t_1 + 1)\Gamma(t_2 + 1) \|K_1 L_2\|_1 - \Gamma(t_1 + 1)\Gamma(t_2 + 1) \|K_2 L_1\|_1], \end{aligned}$$

and by Lemma 1 and 2

$$\begin{aligned} T_{11} &\rightarrow \xi_1 \delta_1 := \xi_1 \lambda \frac{\sqrt{g(x)} \gamma^{t_1-1}(x)}{\rho(x)} \left(\frac{1}{(1 - \rho(x))^{t_1}} - 1 \right) \\ T_{12} &\rightarrow \xi_2 \delta_2 := \xi_2 \lambda \frac{\sqrt{g(x)} \gamma^{t_2-1}(x)}{\rho(x)} \left(\frac{1}{(1 - \rho(x))^{t_2}} - 1 \right). \end{aligned}$$

Thus, we have established (11), with $\boldsymbol{\delta}' = (\delta_1, \delta_2)$, and, by the Cramér-Wold device, also (10). Now, denoting

$$\mathbf{D} := \begin{bmatrix} \gamma^{1-t_1}(x)/t_1 & 0 \\ 0 & \gamma^{1-t_2}(x)/t_2 \end{bmatrix},$$

a straightforward application of the δ -method gives

$$r_n \begin{bmatrix} \hat{\gamma}_n^{(1)}(x, t_1, K_1, L_1) - \gamma(x) \\ \hat{\gamma}_n^{(1)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} \xrightarrow{\mathcal{D}} N_2(\mathbf{D}\boldsymbol{\delta}, \mathbf{D}\mathbf{\Omega}\mathbf{D}),$$

which is the desired result.

Proof of Theorem 5

To establish the bivariate asymptotic normality we invoke the Cramér-Wold device, **according to which it is sufficient that for all** $\xi \in \mathbb{R}^2$, $\|\xi\| \neq 0$, we have that

$$\xi' r_n \begin{bmatrix} \hat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) - \gamma(x) \\ \hat{\gamma}_n^{(2)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} \xrightarrow{\mathcal{D}} N_2(\lambda \xi' \boldsymbol{\mu}^{(2)}, \xi' \tilde{\boldsymbol{\Sigma}}^{(2)} \xi).$$

First write

$$\begin{aligned} &\xi' r_n \begin{bmatrix} \hat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) - \gamma(x) \\ \hat{\gamma}_n^{(2)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} \\ &= \frac{\bar{F}(\omega_n; x)g(x)}{T_n^{(1,t_1)}(x, L_1)} \xi_1 r_n \frac{T_n^{(1,t_1+1)}(x, K_1) - \gamma(x)(t_1 + 1)T_n^{(1,t_1)}(x, L_1)}{(t_1 + 1)\bar{F}(\omega_n; x)g(x)} \\ &\quad + \frac{\bar{F}(\omega_n; x)g(x)}{T_n^{(1,t_2)}(x, L_2)} \xi_2 r_n \frac{T_n^{(1,t_2+1)}(x, K_2) - \gamma(x)(t_2 + 1)T_n^{(1,t_2)}(x, L_2)}{(t_2 + 1)\bar{F}(\omega_n; x)g(x)}. \end{aligned}$$

From Lemma 1 and 2, and Theorem 1 we have that

$$\frac{T_n^{(1,t_i)}(x, L_i)}{\bar{F}(\omega_n; x)g(x)} = \gamma^{t_i}(x)\Gamma(t_i + 1)(1 + o_{\mathbb{P}}(1)), \quad i = 1, 2,$$

and hence, by using **the line of arguments** of Theorem 3, we obtain that

$$\begin{aligned} \xi' r_n \begin{bmatrix} \hat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) - \gamma(x) \\ \hat{\gamma}_n^{(2)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} &= \xi_1 r_n \frac{T_n^{(1,t_1+1)}(x, K_1) - \gamma(x)(t_1 + 1)T_n^{(1,t_1)}(x, L_1)}{\gamma^{t_1}(x)\Gamma(t_1 + 2)\bar{F}(\omega_n; x)g(x)} \\ &\quad + \xi_2 r_n \frac{T_n^{(1,t_2+1)}(x, K_2) - \gamma(x)(t_2 + 1)T_n^{(1,t_2)}(x, L_2)}{\gamma^{t_2}(x)\Gamma(t_2 + 2)\bar{F}(\omega_n; x)g(x)} \\ &\quad + o_{\mathbb{P}}(1), \end{aligned}$$

which can be rewritten as

$$\begin{aligned} &\xi' r_n \begin{bmatrix} \hat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) - \gamma(x) \\ \hat{\gamma}_n^{(2)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} \\ &= \frac{\xi_1}{\gamma^{t_1}(x)\Gamma(t_1 + 2)} r_n \left[\frac{T_n^{(1,t_1+1)}(x, K_1)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(1,t_1+1)}(x, K_1)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ &\quad - \frac{\xi_1 \gamma(x)(t_1 + 1)}{\gamma^{t_1}(x)\Gamma(t_1 + 2)} r_n \left[\frac{T_n^{(1,t_1)}(x, L_1)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(1,t_1)}(x, L_1)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ &\quad + \frac{\xi_2}{\gamma^{t_2}(x)\Gamma(t_2 + 2)} r_n \left[\frac{T_n^{(1,t_2+1)}(x, K_2)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(1,t_2+1)}(x, K_2)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ &\quad - \frac{\xi_2 \gamma(x)(t_2 + 1)}{\gamma^{t_2}(x)\Gamma(t_2 + 2)} r_n \left[\frac{T_n^{(1,t_2)}(x, L_2)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(1,t_2)}(x, L_2)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\ &\quad + \xi_1 r_n \frac{\mathbb{E}(T_n^{(1,t_1+1)}(x, K_1)) - \gamma(x)(t_1 + 1)\mathbb{E}(T_n^{(1,t_1)}(x, L_1))}{\gamma^{t_1}(x)\Gamma(t_1 + 2)\bar{F}(\omega_n; x)g(x)} \\ &\quad + \xi_2 r_n \frac{\mathbb{E}(T_n^{(1,t_2+1)}(x, K_2)) - \gamma(x)(t_2 + 1)\mathbb{E}(T_n^{(1,t_2)}(x, L_2))}{\gamma^{t_2}(x)\Gamma(t_2 + 2)\bar{F}(\omega_n; x)g(x)} + o_{\mathbb{P}}(1) \\ &=: T_{13} + T_{14} + T_{15} + T_{16} + T_{17} + T_{18} + o_{\mathbb{P}}(1). \end{aligned}$$

From Theorem 1 we have that

$$T_{13} + T_{14} + T_{15} + T_{16} \xrightarrow{\mathcal{D}} N(0, \xi' \tilde{\Sigma}^{(2)} \xi),$$

and, by Lemma 1 and 2,

$$T_{17} \rightarrow \xi_1 \lambda \frac{\sqrt{g(x)}}{(1 - \rho(x))^{t_1+1}}, \quad \text{and} \quad T_{18} \rightarrow \xi_2 \lambda \frac{\sqrt{g(x)}}{(1 - \rho(x))^{t_2+1}}.$$

By combining the above results we have that

$$\xi' r_n \begin{bmatrix} \hat{\gamma}_n^{(2)}(x, t_1, K_1, L_1) - \gamma(x) \\ \hat{\gamma}_n^{(2)}(x, t_2, K_2, L_2) - \gamma(x) \end{bmatrix} \xrightarrow{\mathcal{D}} N_2(\lambda \xi' \boldsymbol{\mu}^{(2)}, \xi' \tilde{\Sigma}^{(2)} \xi),$$

from which the theorem follows.

Proof of Theorem 6

We only give the proof of the result for the case where $\tau \neq 0$. The case $\tau = 0$ can be obtained in an analogous way and is for brevity omitted from the paper.

Using Lemma 1 and 2, and Theorem 1, we obtain

$$\frac{T_n^{(1,1)}(x, K_1)}{T_n^{(1,0)}(x, L_1)} = \gamma(x) \left\{ 1 + \frac{b(\omega_n; x)}{\gamma(x)\rho(x)} \left[\frac{1}{1 - \rho(x)} - 1 \right] (1 + o(1)) + O(h \ln \omega_n) + O_{\mathbb{P}} \left(\frac{1}{r_n} \right) \right\},$$

and, by Taylor's theorem,

$$\left(\frac{T_n^{(1,1)}(x, K_1)}{T_n^{(1,0)}(x, L_1)} \right)^{\tau} = \gamma^{\tau}(x) \left\{ 1 + \frac{\tau b(\omega_n; x)}{\gamma(x)\rho(x)} \left[\frac{1}{1 - \rho(x)} - 1 \right] (1 + o(1)) + O(h \ln \omega_n) + O_{\mathbb{P}} \left(\frac{1}{r_n} \right) \right\}.$$

Similarly

$$\begin{aligned} \left(\frac{T_n^{(1,2)}(x, K_2)}{2T_n^{(1,0)}(x, L_2)} \right)^{\tau/2} &= \gamma^{\tau}(x) \left\{ 1 + \frac{\tau b(\omega_n; x)}{2\gamma(x)\rho(x)} \left[\frac{1}{(1 - \rho(x))^2} - 1 \right] (1 + o(1)) \right. \\ &\quad \left. + O(h \ln \omega_n) + O_{\mathbb{P}} \left(\frac{1}{r_n} \right) \right\}, \\ \left(\frac{T_n^{(1,3)}(x, K_3)}{6T_n^{(1,0)}(x, L_3)} \right)^{\tau/3} &= \gamma^{\tau}(x) \left\{ 1 + \frac{\tau b(\omega_n; x)}{3\gamma(x)\rho(x)} \left[\frac{1}{(1 - \rho(x))^3} - 1 \right] (1 + o(1)) \right. \\ &\quad \left. + O(h \ln \omega_n) + O_{\mathbb{P}} \left(\frac{1}{r_n} \right) \right\}. \end{aligned}$$

We now easily obtain that under the conditions of the theorem

$$\begin{aligned} \frac{\left(\frac{T_n^{(1,1)}(x, K_1)}{T_n^{(1,0)}(x, L_1)} \right)^{\tau} - \left(\frac{T_n^{(1,2)}(x, K_2)}{2T_n^{(1,0)}(x, L_2)} \right)^{\tau/2}}{b(\omega_n; x)} &\xrightarrow{\mathbb{P}} -\tau \gamma^{\tau-1}(x) \frac{\rho(x)}{2(1 - \rho(x))^2}, \\ \frac{\left(\frac{T_n^{(1,2)}(x, K_2)}{2T_n^{(1,0)}(x, L_2)} \right)^{\tau/2} - \left(\frac{T_n^{(1,3)}(x, K_3)}{6T_n^{(1,0)}(x, L_3)} \right)^{\tau/3}}{b(\omega_n; x)} &\xrightarrow{\mathbb{P}} -\tau \gamma^{\tau-1}(x) \frac{\rho(x)(3 - \rho(x))}{6(1 - \rho(x))^3}, \end{aligned}$$

and hence

$$R_n(x; \mathbb{K}, \mathbb{L}, \tau) \xrightarrow{\mathbb{P}} \frac{3(1 - \rho(x))}{3 - \rho(x)}.$$

Finally, the consistency of $\hat{\rho}_n(x; \mathbb{K}, \mathbb{L}, \tau)$ follows from a straightforward application of the continuous mapping theorem.

Acknowledgement

The authors are very grateful to the two anonymous referees and the associate editor for their very constructive comments on the paper. Their suggestions have definitely improved the exposition of the material.

References

- [1] Beirlant, J., Dierckx, G., Goegebeur, Y., Matthys, G., 1999. Tail index estimation and an exponential regression model. *Extremes* 2, 177–200.
- [2] Beirlant, J., Goegebeur, Y., 2003. Regression with response distributions of Pareto-type. *Computational Statistics and Data Analysis* 42, 595–619.
- [3] Beirlant, J., Goegebeur, Y., 2004. Local polynomial maximum likelihood estimation for Pareto-type distributions. *Journal of Multivariate Analysis* 89, 97–118.
- [4] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. *Statistics of Extremes - Theory and Applications*. Wiley Series in Probability and Statistics.
- [5] Billingsley, P., 1995. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics.
- [6] Bingham, N.H., Goldie, C.M., Teugels, J.L., 1987. *Regular Variation*. Cambridge University Press, Cambridge.
- [7] **Caeiro, F., Gomes, M.I., 2006. A new class of estimators of a "scale" second order parameter. *Extremes* 9, 193–211.**
- [8] Caeiro, F., Gomes, M.I., Rodrigues, L.H., 2009. Reduced-bias tail index estimators under a third order framework. *Communications in Statistics – Theory and Methods* 38, 1019–1040.
- [9] Chavez-Demoulin, V., Davison, A.C., 2005. Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society Series C* 54, 207–222.
- [10] Ciuperca, G., Mercadier, C., 2010. Semi-parametric estimation for heavy tailed distributions. *Extremes* 13, 55–87.
- [11] Daouia, A., Gardes, L., Girard, S., 2012. On kernel smoothing for extremal quantile regression. **Bernoulli, to appear.**
- [12] Daouia, A., Gardes, L., Girard, S., Lekina, A., 2011. Kernel estimators of extreme level curves. *Test* 20, 311–333.
- [13] Davison, A.C., Ramesh, N.I., 2000. Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society Series B* 62, 191–208.
- [14] Davison, A.C., Smith, R.L., 1990. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society Series B* 52, 393–442.

- [15] de Haan, L., Ferreira, A., 2006. *Extreme Value Theory: An Introduction*. Springer.
- [16] **Drees, H., 1998.** On smooth statistical tail functionals. *Scandinavian Journal of Statistics* 25, 187–210.
- [17] Fraga Alves, M.I., Gomes, M.I., de Haan, L., 2003. A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* 60, 193–213.
- [18] Gannoun, A., Girard, S., Guinot, C., Saracco, J., 2002. Reference **curves** based on non-parametric quantile regression. *Statistics in Medicine* 21, 3119–3135.
- [19] Gardes, L., Girard, S., 2008. A moving window approach for nonparametric estimation of the conditional tail index. *Journal of Multivariate Analysis* 99, 2368–2388.
- [20] Gardes, L., Girard, S., 2010. Conditional extremes from heavy-tailed distributions: an application to the estimation of extreme rainfall return levels. *Extremes* 13, 177–204.
- [21] **Givens, G.H., Hoeting, J.A., 2005. *Computational Statistics*. Wiley.**
- [22] Goegebeur, Y., Beirlant, J., de Wet, T., 2010. Kernel estimators for the second order parameter in extreme value statistics. *Journal of Statistical Planning and Inference* 140, 2632–2652.
- [23] Goegebeur, Y., de Wet, T., 2012. Local estimation of the second order parameter in extreme value statistics and local unbiased estimation of the tail index. *Communications in Statistics - Theory and Methods* **21, 3575–3607.**
- [24] Gomes, M.I., de Haan, L., Peng, L., 2002. Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes* 5, 387–414.
- [25] Gray, H.L., Schucany, W.R., 1972. *The Generalized Jackknife Statistic*. Marcel Dekker.
- [26] Hall, P., 1982. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society Series B* 44, 37–42.
- [27] Hall, P., Tajvidi, N., 2000. Nonparametric analysis of temporal trend when fitting parametric models to extreme value data. *Statistical Science* 15, 153–167.
- [28] Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 3, 1163–1174.
- [29] Pauli, F., Coles, S., 2001. Penalized likelihood inference in extreme value analysis. *Journal of Applied Statistics* 28, 547–560.
- [30] Pickands, J., 1975. Statistical inference using extreme order statistics. *Annals of Statistics* 3, 119–131.
- [31] Smith, R.L., 1989. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science* 4, 367–393.

- [32] van der Vaart, A.W., 1998. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics.
- [33] Wang, H., Tsai, C.L., 2009. Tail index regression. *Journal of the American Statistical Association* 104, 1233–1240.
- [34] Yao, Q., 1999. Conditional predictive regions for stochastic processes. Technical report, **Institute of Mathematics and Statistics**, University of Kent at Canterbury.

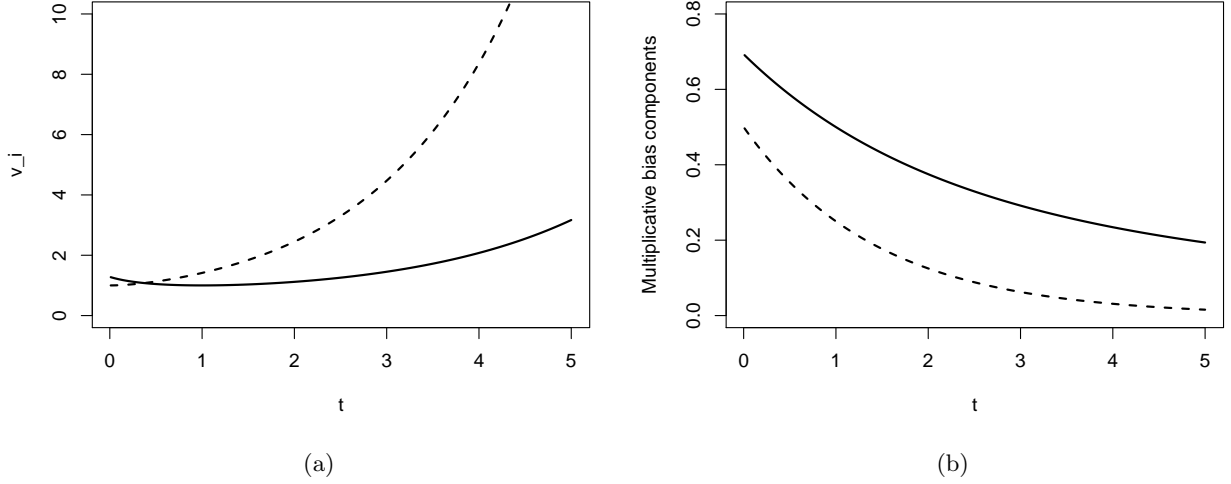
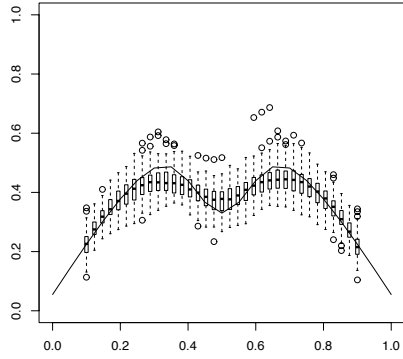


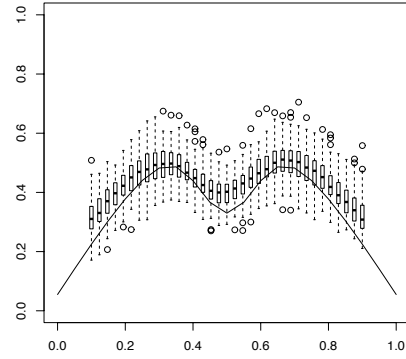
Figure 1: (a) Components of the asymptotic standard deviation v_1 (solid line) and v_2 (dashed line) as a function of t , (b) **multiplicative** components of the asymptotic bias $((1 - \rho(x))^{-t} - 1)/(t\rho(x))$ (solid line) and $(1 - \rho(x))^{-(t+1)}$ (dashed line), for $\rho(x) = -1$, as a function of t .

Estimator	Oracle	Data-driven
$\hat{\gamma}_n^{(2)}(x, 0, K, K)$	0.00555	0.01930
$\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-1))$	0.00266	0.00608
$\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-2))$	0.00468	0.00648
$\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-5))$	0.00532	0.00729
$\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)))$	0.00691	0.00621

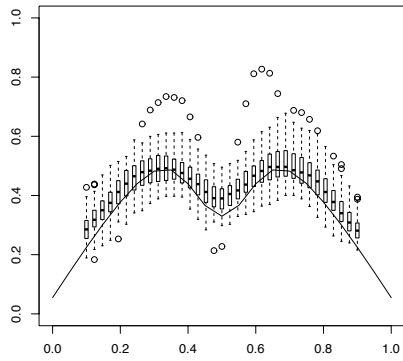
Table 1: Average of $\Delta^2(\hat{\gamma}(\cdot), \gamma(\cdot))$ for $\hat{\gamma}_n^{(2)}(x, 0, K, K)$ and $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}(\rho(x)))$ with oracle and data driven methods.



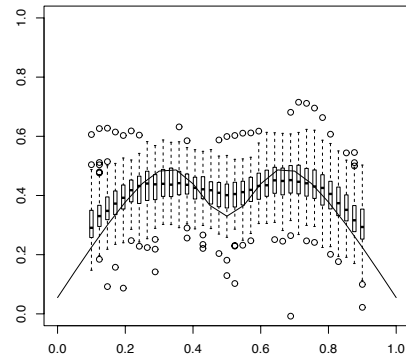
(a) Benchmark: $\rho(x) = -1$ (True Value)



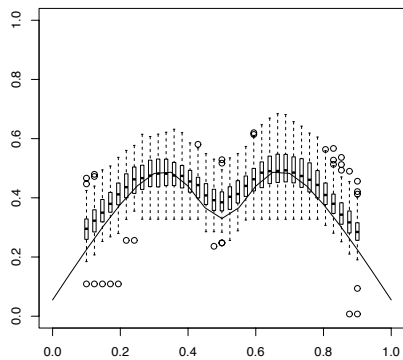
(b) No bias correction



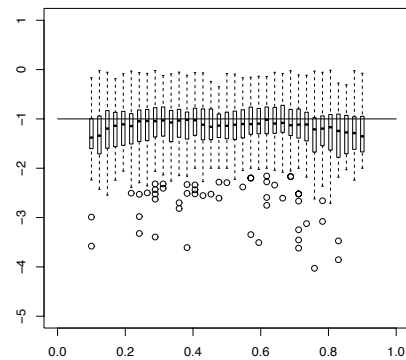
(c) Mis-specified: $\rho(x) = -2$



(d) Using Estimated Second Order Parameter

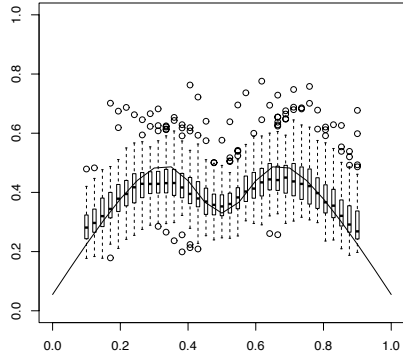


(e) Mis-specified: $\rho(x) = -5$

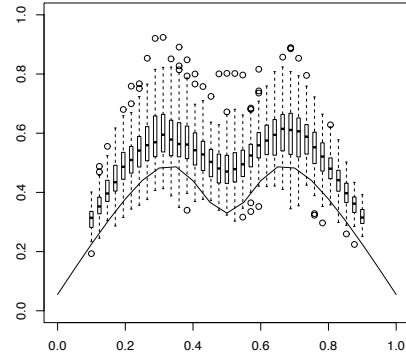


(f) Estimated Second Order Parameter

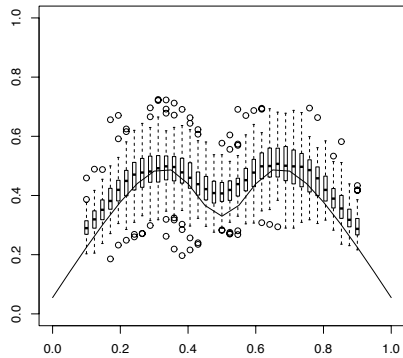
Figure 2: Burr simulation, oracle criterion: boxplots of (a) $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-1))$, (b) $\hat{\gamma}_n^{(2)}(x, 0, K, K)$, (c) $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-2))$, (d) $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)))$, (e) $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-5))$ and (f) $\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)$.



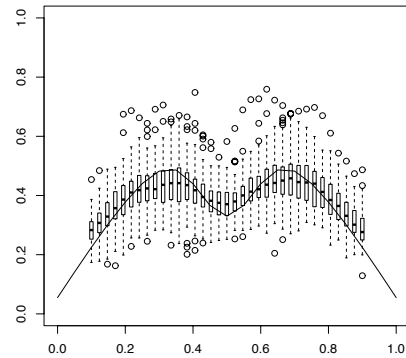
(a) Benchmark: $\rho(x) = -1$ (True Value)



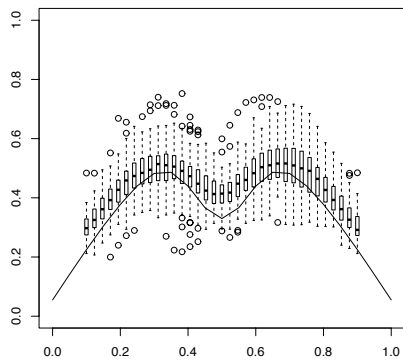
(b) No bias correction



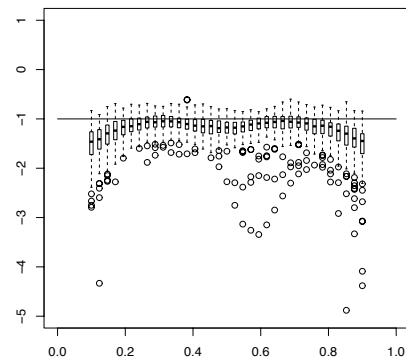
(c) Mis-specified: $\rho(x) = -2$



(d) Using Estimated Second Order Parameter



(e) Mis-specified: $\rho(x) = -5$



(f) Estimated Second Order Parameter

Figure 3: Burr simulation, data driven criterion: boxplots of (a) $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-1))$, (b) $\hat{\gamma}_n^{(2)}(x, 0, K, K)$, (c) $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-2))$, (d) $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)))$, (e) $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(-5))$ and (f) $\hat{\rho}_n(x; \mathbb{K}, \mathbb{K}, 0.5)$.

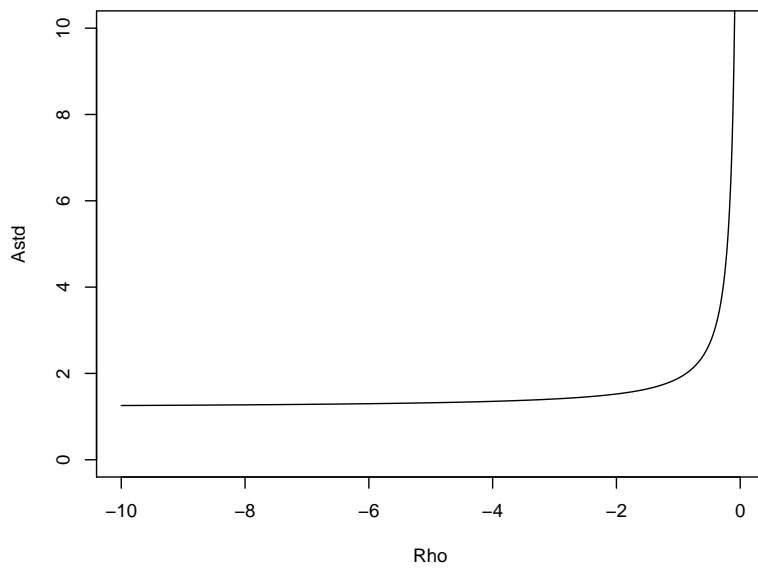


Figure 4: Asymptotic standard deviation of $\hat{\gamma}_n^{(2)}(x, \alpha_{BC}^{(2)}(\rho))$ as a function of ρ .