HAL
open science

A moving window approach for nonparametric estimation of the conditional tail index

Laurent Gardes, Stéphane Girard

To cite this version:

Laurent Gardes, Stéphane Girard. A moving window approach for nonparametric estimation of the conditional tail index. Journal of Multivariate Analysis, 2008, 99 (10), pp.2368-2388. 10.1016/j.jmva.2008.02.023 . inria-00124637v2

HAL Id: inria-00124637

https://inria.hal.science/inria-00124637v2

Submitted on 13 Jun 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A MOVING WINDOW APPROACH FOR NONPARAMETRIC ESTIMATION OF THE CONDITIONAL TAIL INDEX

Laurent GARDES and Stéphane GIRARD[1]

INRIA Rhône-Alpes, team Mistis
655, avenue de l'Europe, Montbonnot
38334 Saint-Ismier Cedex, France.
{Laurent.Gardes, Stephane.Girard}@inrialpes.fr

Abstract – We present a nonparametric family of estimators for the tail index of a Pareto-type distribution when covariate information is available. Our estimators are based on a weighted sum of the log-spacings between some selected observations. This selection is achieved through a moving window approach on the covariate domain and a random threshold on the variable of interest. Asymptotic normality is proved under mild regularity conditions and illustrated for some weight functions. Finite sample performances are presented on a real data study.

Keywords – Conditional tail index, extreme-values, nonparametric estimation, moving window.

AMS Subject classifications – 62G32, 62G05, 62E20.

1 Introduction

In extreme-value statistics, one of the main problems is the estimation of the tail index associated to a random variable Y. This parameter, denoted by γ, drives the distribution tail heaviness of Y. For instance, when γ is positive, the survival function of Y decreases to zero geometrically, and the larger γ is, the slower is the convergence. Without intending to be exhaustive, we refer to [16, 19, 29] for a presentation of extreme-value methodology in various frameworks and to [11] for an overview of the numerous works dedicated to the estimation of the tail index. Here, we focus on the situation where some covariate information x is recorded simultaneously with the quantity of interest Y. In the general case, the tail heaviness of Y given x depends on x, and thus the tail index is a function $\gamma(x)$ of the covariate. Such situations

[1]Corresponding author

occur for instance in climatology where one may be interested in how climate change over years might affect extreme temperatures. Here, the covariate is univariate (the time). Bivariate examples include the study of extremes rainfall as a function of the geographical location.

Only a few papers address the estimation of conditional tail index. A parametric approach is considered in [31] where a linear trend is fitted to the mean of an extreme-value distribution. We refer to [13] for other examples of parametric models. More recently, Hall and Tajvidi [24] proposed to mix a non-parametric estimation of the trend with a parametric assumption on Y given x. We also refer to [4] where a kind of semi-parametric estimator is introduced for $\gamma(\psi(\beta' x))$ where ψ is a known link function and β is interpreted as a vector of regression coefficients. Fully non-parametric estimators are introduced in [12], where a local polynomial fitting of the extreme-value distribution to the extreme observations is used. In a similar spirit, spline estimators are fitted in [9] through a penalized maximum likelihood method. In both cases, the authors focus on univariate covariates and on the finite sample properties of the estimators. These results are extended in [5] where local polynomials estimators are proposed for multivariate covariates and where their asymptotic properties are established for very regular functions $\gamma(x)$ (at least twice continuously differentiable).

Similarly to these authors, we investigate how to combine nonparametric smoothing techniques with extreme-value methods in order to obtain efficient estimators of $\gamma(x)$. The proposed method is based on a selection, thanks to a moving window approach, of the observations to be used in the estimator of the extreme-value index. This estimator is a weighted sum of the rescaled log-spacings between the selected largest observations. This approach has several advantages. From the theoretical point of view, very few assumptions are made on the regularity of $\gamma(x)$ and on the nature of the covariate. A central limit theorem is established for the proposed estimator, without assuming that x is finite dimensional. As an example, we provide the asymptotic rate of convergence for Lipschitzian functions $\gamma(x)$ and multidimensional covariates x. From the practical point of view, the estimator is easy to compute since it is closed-form and thus does not require optimization procedures.

Our family of nonparametric estimators is defined in Section 2. In Section 3, asymptotic normality properties are established, and links with nonparametric regression are highlighted. The choice of weights is discussed in Section 4. We first present two classical choices of weights extending Hill [26] and Zipf [27, 30] estimators to the conditional case. Next, we address the problem of obtaining minimum variance and/or asymptotically unbiased estimators based on the knowledge of a second order parameter. The practical difficulties arising when this parameter is unknown are also discussed. An illustration on real data is provided in Section 5. Proofs are postponed to Section 6.

2 Estimators of the conditional tail index

Let E be a metric space associated to a metric d. We assume that the conditional distribution function of Y given $x \in E$ is, for all $y > 0$,

$$F(y, x) = 1 - y^{-1/\gamma(x)} L(y, x), \tag{1}$$

where $\gamma(.)$ is an unknown positive function of the covariate x and, for x fixed, $L(., x)$ is a slowly varying function, *i.e.* for $\lambda > 0$,

$$\lim_{y \to \infty} \frac{L(\lambda y, x)}{L(y, x)} = 1.$$

Given a sample $(Y_1, x_1), \ldots, (Y_n, x_n)$ of independent observations from (1), our aim is to build a point-wise estimator of the function γ. More precisely, for a given $t \in E$, we want to estimate $\gamma(t)$, focusing on the case where the design points $x_1, \ldots, x_n$ are non random. To this end, for all $r > 0$, let us denote by $B(t, r)$ the ball centered at point t and with radius r defined by

$$B(t, r) = \{x \in E,\ d(x, t) \leq r\}$$

and let $h_{n,t}$ be a positive sequence tending to zero as n goes to infinity. The proposed estimator uses a moving window approach since it is based on the response variables $Y_i's$ for which the associated covariates $x_i's$ belong to the ball $B(t, h_{n,t})$. The proportion of such design points is thus defined by

$$\varphi(h_{n,t}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{x_i \in B(t, h_{n,t})\}$$

and plays an important role in this study. It describes how the design points concentrate in the neighborhood of t when $h_{n,t}$ goes to zero, similarly to the small ball probability does, see for instance the monograph on functional data analysis [18]. Thus, the nonrandom number of observations in $(0, \infty) \times B(t, h_{n,t})$ is given by $m_{n,t} = n\varphi(h_{n,t})$. Let $\{Z_i(t),\ i = 1, \ldots, m_{n,t}\}$ be the response variables $Y_i's$ for which the associated covariates $x_i's$ belong to the ball $B(t, h_{n,t})$ and let $Z_{1,m_{n,t}}(t) \leq \ldots \leq Z_{m_{n,t},m_{n,t}}(t)$ be the corresponding order statistics. Our family of estimators of $\gamma(t)$ is defined by

$$\hat{\gamma}_n(t, W) = \sum_{i=1}^{k_{n,t}} i \log\left(\frac{Z_{m_{n,t}-i+1,m_{n,t}}(t)}{Z_{m_{n,t}-i,m_{n,t}}(t)}\right) W\left(i/k_{n,t}, t\right) \Bigg/ \sum_{i=1}^{k_{n,t}} W\left(i/k_{n,t}, t\right), \tag{2}$$

where $k_{n,t}$ is a sequence of integers such that $1 \leq k_{n,t} < m_{n,t}$ and $W(., t)$ a function defined on $(0, 1)$ such that $\int_0^1 W(s, t)ds \neq 0$. Thus, without loss of generality, we can assume that $\int_0^1 W(s, t)ds = 1$. Note that this family of estimators is an extension of estimators proposed in [3] in the situation where

there is no covariate information. In this latter case, we also refer to [10] for the definition of kernel estimators based on non-increasing and non-negative functions, and to [21] for a similar work dedicated to Weibull tail-distributions. In [33], Viharos discusses the choice of the weight function to obtain universal asymptotic normality of the corresponding weighted least-squares estimator.

We also introduce the following extended family of estimators:

$$\tilde{\gamma}_n(t,\mu^{\mathrm{W}}) = \sum_{i=1}^{k_{n,t}} i \log\left(\frac{Z_{m_{n,t}-i+1,m_{n,t}}(t)}{Z_{m_{n,t}-i,m_{n,t}}(t)}\right)\mu_{i,n}^{\mathrm{W}}(t) \Bigg/ \sum_{i=1}^{k_{n,t}} \mu_{i,n}^{\mathrm{W}}(t)\ , \qquad (3)$$

where the weights $\mu_{i,n}^{\mathrm{W}}(t)$ are defined by $\mu_{i,n}^{\mathrm{W}}(t) = W(i/k_{n,t},t)(1+o(1))$ uniformly in $i = 1,\ldots,k_{n,t}$.

3 Main results

We first give all the conditions we assume in order to obtain the asymptotic normality of our estimators. In the sequel, we fix $t \in E$ such that $\gamma(t) > 0$.

Assumptions on the conditional distribution. Let $x \in E$ be fixed. Then, model (1) is well known to be equivalent to the so-called first order condition

$$U(y,x) \stackrel{def}{=} \inf\{s; F(s,x) \geq 1 - 1/y\} = y^{\gamma(x)}\ell(y,x), \qquad (4)$$

where, for x fixed, $\ell(.,x)$ is a slowly varying function. The function $U(.,x)$ is said to be regularly varying with index $\gamma(x)$. We refer to [6] for a detailed account on this topic. The conditions are:

(A.1) The conditional cumulative distribution $F(.,t)$ is continuous.

(A.2) There exists positive constants c_U, z_U and $\alpha_U \leq 1$ such that for all $x \in B(t,1)$,

$$\sup_{z \geq z_U}\left|\frac{\log U(z,x)}{\log U(z,t)} - 1\right| \leq c_U d^{\alpha_U}(x,t).$$

(A.3) There exists a negative function $\rho(t)$ and a rate function $b(.,t)$ satisfying $b(y,t) \to 0$ as $y \to \infty$, such that for all $\lambda \geq 1$,

$$\log\left(\frac{\ell(\lambda y,t)}{\ell(y,t)}\right) = b(y,t)\frac{1}{\rho(t)}(\lambda^{\rho(t)} - 1)(1+o(1)),$$

where "o" is uniform in $\lambda \geq 1$ as $y \to \infty$.

Conditions **(A.1)** and **(A.2)** are regularity conditions on the conditional distribution function. The second-order condition **(A.3)** on the slowly varying function is the cornerstone to establish the asymptotic normality of tail index estimators. It is used in [25] to prove the asymptotic normality of the Hill estimator. The second order parameter $\rho(t) < 0$ tunes the rate of convergence of $\ell(\lambda t, x)/\ell(t, x)$ to 1. The closer $\rho(t)$ is to 0, the slower is the convergence. The function $b(., t)$ is usually called the bias function, since it drives the asymptotic behavior of most tail index estimators. It can be shown that necessarily, $b(., t)$ is regularly varying with index $\rho(t)$ (see [22]).

Assumptions on the weights. The next assumption is assumed in Theorem 2.2 of [3], a stronger version of Theorem 2.1, adequate to derive the asymptotic normality of kernel type statistics based on the scaled log-spacings.

(B.1) The function $s \to sW(s, t)$ is absolutely continuous, *i.e.* there exists a function $u(., t)$ defined on $(0, 1)$ such that

$$sW(s,t) = \int_0^s u(\xi, t) d\xi \tag{5}$$

with, for all $j = 1, \ldots, k_{n,t}$,

$$\left| k_{n,t} \int_{(j-1)/k_{n,t}}^{j/k_{n,t}} u(\xi, t) d\xi \right| < g\left(\frac{j}{k_{n,t}+1}, t\right), \tag{6}$$

where $g(., t)$ is a positive continuous function defined on $(0, 1)$ and satisfying

$$\int_0^1 \max(1, \log(1/s)) g(s, t) ds < \infty. \tag{7}$$

(B.2) There exists a constant $\delta > 0$ such that $\int_0^1 |W(s, t)|^{2+\delta} ds < \infty$.

Assumptions on the sequences $k_{n,t}$ and $h_{n,t}$. We assume that $k_{n,t}$ is an intermediate sequence, which is a classical assumption in extreme-value analysis:

(C) $n\varphi(h_{n,t})/k_{n,t} \to \infty$ and $k_{n,t} \to \infty$.

Remark that **(C)** implies $n\varphi(h_{n,t}) \to \infty$ *i.e.* the number of points in $(0, \infty) \times B(t, h_{n,t})$ goes to infinity as the total number of points does.
In order to simplify the notations, let

$$b_{n,t} \stackrel{def}{=} b\left(\frac{n\varphi(h_{n,t})}{k_{n,t}}, t\right)$$

and introduce the rescaled log-spacings

$$C_{i,n}(t) \stackrel{def}{=} i \log \left(\frac{Z_{m_{n,t}-i+1,m_{n,t}}(t)}{Z_{m_{n,t}-i,m_{n,t}}(t)} \right), \; i = 1, \ldots, k_{n,t},$$

such that estimator (2) can be rewritten as

$$\hat{\gamma}_n(t, W) = \sum_{i=1}^{k_{n,t}} C_{i,n}(t) W\left(i/k_{n,t}, t\right) \Bigg/ \sum_{i=1}^{k_{n,t}} W\left(i/k_{n,t}, t\right).$$

Besides, in the following, each vector $\{v_{i,n}, \; i = 1, \ldots, k_{n,t}\}$ is denoted by $\{v_{i,n}\}_i$. Our first main result establishes the exponential regression model for $\{C_{i,n}(t)\}_i$.

Theorem 1 *Suppose* **(A.1)**, **(A.2)**, **(A.3)**, **(B.1)** *and* **(C)** *hold. Then, the random vector* $\{C_{i,n}(t)\}_i$ *has the same distribution as*

$$\left\{ \left[\left(\gamma(t) + b_{n,t} \left(\frac{i}{k_{n,t}+1} \right)^{-\rho(t)} \right) F_i + \beta_{i,n}(t) + o_{\mathrm{P}}\left(b_{n,t}\right) \right] \left(1 + O_{\mathrm{P}}(h_{n,t}^{\alpha_U})\right) \right\}_i,$$

uniformly in $i = 1, \ldots, k_{n,t}$ *with*

$$\frac{1}{k_{n,t}} \sum_{i=1}^{k_{n,t}} W\left(i/k_{n,t}, t\right) \beta_{i,n}(t) = o_{\mathrm{P}}\left(b_{n,t}\right),$$

and where $F_1, \ldots, F_{k_{n,t}}$ *are independent standard exponential variables.*

Similar results can be found in [14] for rescaled log-spacings of Weibull-type random variables, and in [3] in the case of Pareto-type random variables without covariate. See also [15] for approximations of the Hill process by sums of standard exponential random variables. In the conditional case, *i.e.* when covariate information is available, only few results exist. We refer to [17], Theorem 3.5.2, for the approximation of the nearest neighbors distribution using the Hellinger distance and to [20] for the study of their asymptotic distribution. Our second main result establishes the asymptotic normality of our estimators.

Theorem 2 *Suppose* **(A.1)**, **(A.2)**, **(A.3)**, **(B.1)**, **(B.2)** *and* **(C)** *hold. If, moreover,*

$$k_{n,t}^{1/2} b_{n,t} \to \lambda(t) \in \mathbb{R} \text{ and } k_{n,t}^{1/2} h_{n,t}^{\alpha_U} \to 0 \tag{8}$$

then

$$k_{n,t}^{1/2} \left(\hat{\gamma}_n(t, W) - \gamma(t) - b_{n,t} \mathcal{AB}(t, W) \right) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2(t) \mathcal{AV}(t, W)\right), \tag{9}$$

where we have defined

$$\mathcal{AB}(t,W) = \int_0^1 W(s,t)s^{-\rho(t)}ds \text{ and } \mathcal{AV}(t,W) = \int_0^1 W^2(s,t)ds.$$

It appears that the asymptotic bias involves two parts. The first one is given by $b_{n,t}$ and thus depends on the original distribution itself. The second one is given by $\mathcal{AB}(t,W)$. This multiplicative factor can be made small by an appropriate choice of the weighting function W as we may see in the next section. Similarly, the variance term is inversely proportional to $k_{n,t}$, the number of observations used to build the estimator, and the multiplicative coefficient $\gamma^2(t)\mathcal{AV}(t,W)$ can also be adjusted. When $\lambda(t) \neq 0$, the first part of condition (8) forces the bias to be of the same order as the standard-deviation. The second part $k_{n,t}^{1/2}h_{n,t}^{\alpha_U} \to 0$ is due to the functional nature of the tail index to estimate. It imposes the fluctuations of $t \to U(.,t)$ to be negligible compared to the standard deviation of the estimator.
The following result establishes that the estimators in extended family (3) inherits the asymptotic distribution of estimators in family (2).

Corollary 1 *Under the assumptions of Theorem 2,*

$$k_{n,t}^{1/2}\left(\tilde{\gamma}_n(t,\mu^{\mathrm{W}}) - \gamma(t) - b_{n,t}\mathcal{AB}(t,W)\right) \xrightarrow{d} \mathcal{N}\left(0,\gamma^2(t)\mathcal{AV}(t,W)\right). \tag{10}$$

We now propose a precise evaluation of the rate of convergence obtained in Theorem 2 in the particular framework of multidimensional nonparametric regression.

Corollary 2 *Let $E = \mathbb{R}^p$ and suppose* **(B.1)**, **(B.2)** *hold. If, moreover, γ is α-Lipschitzian, the slowly-varying function L in (1) is such that $L(y,x) = 1$ for all $(y,x) \in \mathbb{R}_+ \times \mathbb{R}^p$ and*

$$\liminf_{n\to\infty} \varphi(h_{n,t})/h_{n,t}^p > 0, \tag{11}$$

then the convergences in distribution (9) and (10) hold with rate $n^{\frac{\alpha}{p+2\alpha}}\eta_n$, where $\eta_n \to 0$ arbitrarily slowly.

Condition (11) is an assumption on the multidimensional design and on the distance d. Lemma 3 in Section 6 provides an example of design fulfilling this assumption. Under the condition on the slowly varying function $L(y,x) = 1$ for all $(y,x) \in \mathbb{R}_+ \times \mathbb{R}^p$, estimating $\gamma(x)$ is a nonparametric regression problem since $\gamma(x) = \mathbb{E}(\log Y|X = x)$. Let us highlight that the convergence rate provided by Corollary 2 is, up to the η_n factor, the optimal convergence rate for estimating α-Lipschitzian regression function in $\mathbb{R}^p$, see [32].

4 Discussion on the choice of the weights

In order to illustrate the usefulness of our results, we first provide two examples of weights extending classical extreme index estimators to the presence of covariates. Second, we propose some "optimal" choices of weights in the theoretical situation where the second order parameter $\rho(t)$ is known. Finally, we give some ideas to overcome this restrictive assumption.

4.1 Two classical examples of weights

We first introduce an adaptation of Hill estimator to take into account the covariate information. Considering in (2) the constant weight function $W^{\mathrm{H}}(s,t) = 1$ for all $s \in [0,1]$ yields

$$\hat{\gamma}_n(t, W^{\mathrm{H}}) = \frac{1}{k_{n,t}} \sum_{i=1}^{k_{n,t}} i \log\left(\frac{Z_{m_{n,t}-i+1,m_{n,t}}(t)}{Z_{m_{n,t}-i,m_{n,t}}(t)}\right) \tag{12}$$

which is formally the same expression as in [26]. Clearly, W^{H} satisfies the assumptions **(B.1)** and **(B.2)** and then the asymptotic normality of $\hat{\gamma}_n(t, W^{\mathrm{H}})$ is a direct consequence of Theorem 2.

Corollary 3 *Under* **(A.1)**, **(A.2)**, **(A.3)**, **(C)** *and (8), the convergence in distribution (9) holds for* $\hat{\gamma}_n(t, W^{\mathrm{H}})$ *with* $\mathcal{AB}(t, W^{\mathrm{H}}) = 1/(1-\rho(t))$ *and* $\mathcal{AV}(t, W^{\mathrm{H}}) = 1$.

Similarly, we define a Zipf estimator (proposed simultaneously by Kratz and Resnick [27] and Schultze and Steinebach [30]) adapted to our conditional framework. To this end, let us remark that the pairs

$$\left(\tau_{i,n}(t) \stackrel{def}{=} \sum_{j=i}^{m_{n,t}} \frac{1}{j},\ \ \log(Z_{m_{n,t}-i+1,m_{n,t}}(t))\right),\ i = 1, \ldots, m_{n,t},$$

are approximatively distributed on a line of slope $\gamma(t)$ at least for small values of i and for $h_{n,t}$ close to zero, since $\tau_{i,n}(t)$ is the expectation of $\log(Z_{m_{n,t}-i+1,m_{n,t}}(t))$ when the associated slowly-varying function is constant. Thus, one can propose a least-squares estimator based on the $k_{n,t}$ largest observations :

$$\tilde{\gamma}_n(t, \mu^{\mathrm{Z}}) = \sum_{i=1}^{k_{n,t}} (\tau_{i,n}(t) - \bar{\tau}_n(t)) \log(Z_{m_{n,t}-i+1,m_{n,t}}(t)) \Big/ \sum_{i=1}^{k_{n,t}} (\tau_{i,n}(t) - \bar{\tau}_n(t)) \tau_{i,n}(t), \tag{13}$$

where $\bar{\tau}_n(t) = \frac{1}{k_{n,t}} \sum_{i=1}^{k_{n,t}} \tau_{i,n}(t)$. Since (13) can be rewritten as

$$\tilde{\gamma}_n(t, \mu^{\mathrm{Z}}) = \sum_{i=1}^{k_{n,t}} i \log\left(\frac{Z_{m_{n,t}-i+1,m_{n,t}}(t)}{Z_{m_{n,t}-i,m_{n,t}}(t)}\right) \mu_{i,n}^{\mathrm{Z}}(t) \Big/ \sum_{i=1}^{k_{n,t}} \mu_{i,n}^{\mathrm{Z}}(t),$$

with

$$\mu_{i,n}^{\mathrm{Z}}(t) = \frac{1}{i}\sum_{j=1}^{i}(\tau_{j,n}(t) - \bar{\tau}_n(t)) = -\log\left(i/k_{n,t}\right)(1+o(1)),$$

uniformly in $i = 1, \ldots, k_{n,t}$ (see Section 6 for a proof), it appears that this estimator belongs to the extended family (3) associated to the weight function $W^{\mathrm{Z}}(s,t) = -\log(s)$. Lemma 2 in Section 6 shows that condition **(B.1)** is fulfilled with $g(s,t) = 1 - \log(s)$ and thus Corollary 1 yields

Corollary 4 *Under* **(A.1)**, **(A.2)**, **(A.3)**, **(C)** *and (8), the convergence in distribution (10) holds for* $\tilde{\gamma}_n(t, \mu^{\mathrm{Z}})$ *with* $\mathcal{AB}(t, W^{\mathrm{Z}}) = 1/(1-\rho(t))^2$ *and* $\mathcal{AV}(t, W^{\mathrm{Z}}) = 2$.

4.2 Theoretical choices of weights

In this subsection, three problems are addressed: The construction of asymptotically unbiased estimators, of minimum variance estimators and of minimum variance asymptotically unbiased estimators.

Asymptotically unbiased estimators. We propose to combine two weights functions in order to cancel the asymptotic bias. More precisely, we use the following result, which proof is straightforward.

Proposition 1 *Given two weights functions* $W_1(.,t)$ *and* $W_2(.,t)$ *satisfying* **(B.1)** *and* **(B.2)** *and a function* $\alpha(t)$ *defined on* E*, the weight function* $\alpha(t)W_1(.,t) + (1-\alpha(t))W_2(.,t)$ *also satisfies* **(B.1)** *and* **(B.2)**.

Hence, Theorem 2 entails that the asymptotic bias of the obtained estimator is given by $b_{n,t}\left(\alpha(t)\mathcal{AB}(t,W_1) + (1-\alpha(t))\mathcal{AB}(t,W_2)\right)$. Clearly, if $W_1(.,t) \neq W_2(.,t)$, choosing

$$\alpha(t) = \frac{\mathcal{AB}(t,W_2)}{\mathcal{AB}(t,W_2) - \mathcal{AB}(t,W_1)}, \tag{14}$$

permits to cancel the asymptotic bias. As an example, one can combine the weights of the conditional Hill and Zipf estimators defined respectively by (12) and (13) to obtain an asymptotically unbiased estimator $\hat{\gamma}_n(t, W^{\mathrm{HZ}})$ with

$$W^{\mathrm{HZ}}(s,t) = \frac{1}{\rho(t)} - \left(1 - \frac{1}{\rho(t)}\right)\log(s).$$

The following result is a direct consequence of the above results.

Corollary 5 *Under* **(A.1)**, **(A.2)**, **(A.3)**, **(C)** *and (8), the convergence in distribution (9) holds for* $\hat{\gamma}_n(t, W^{\mathrm{HZ}})$ *with* $\mathcal{AB}(t, W^{\mathrm{HZ}}) = 0$ *and* $\mathcal{AV}(t, W^{\mathrm{HZ}}) = 1 + (1 - 1/\rho(t))^2$.

Minimum variance estimator. It is also of interest to find the weights minimizing the variance. The following result is the key tool to answer this question.

Proposition 2 *Let $t \in E$. The unique continuous function $W(.,t)$ such that $\int_0^1 W(s,t)ds = 1$ and minimizing $\int_0^1 W^2(s,t)ds$ is given by $W(s,t) = 1$ for all $s \in [0,1]$.*

It thus appears that the conditional Hill estimator (12) is the unique minimum variance estimator in (2).

Asymptotically unbiased estimator with minimum variance. Finally, we provide the asymptotically unbiased estimator with minimum variance.

Proposition 3 *Let $t \in E$. The unique continuous function $W(.,t)$ such that $\int_0^1 W(s,t)ds = 1$, $\int_0^1 W(s,t)s^{-\rho(t)}ds = 0$ and minimizing $\int_0^1 W^2(s,t)ds$ is given by*

$$W^{\text{opt}}(s,t) = \frac{\rho(t)-1}{\rho^2(t)}\left(\rho(t) - 1 + (1-2\rho(t))s^{-\rho(t)}\right).$$

Remark that $W^{\text{opt}}(s,t) = \alpha(t)W_1(s,t) + (1-\alpha(t))W_2(s,t)$ with $W_1(s,t) = 1$ for all $s \in (0,1)$, $W_2(s,t) = (1-\rho(t))s^{-\rho(t)}$ and $\alpha(t) = (1-\rho(t))^2/\rho^2(t)$ defined as in (14). From Lemma 2, $W_1(.,t)$ and $W_2(.,t)$ both satisfy **(B.1)**, with $g_1(s,t) = 1$, $g_2(s,t) = (1-\rho(t))^2 s^{-\rho(t)}$ and **(B.2)**. Thus, Proposition 1 and Theorem 2 yield the following corollary:

Corollary 6 *Under* **(A.1)**, **(A.2)**, **(A.3)**, **(C)** *and (8), the convergence in distribution (9) holds for $\hat{\gamma}_n(t, W^{\text{opt}})$ with $\mathcal{AB}(t, W^{\text{opt}}) = 0$ and $\mathcal{AV}(t, W^{\text{opt}}) = (1 - 1/\rho(t))^2$.*

Unsurprisingly, the estimators $\hat{\gamma}_n(t, W^{\text{HZ}})$ and $\hat{\gamma}_n(t, W^{\text{opt}})$ requires the knowledge of the second order parameter $\rho(t)$. The estimation of the function $t \to \rho(t)$ is beyond the scope of this paper, we refer to [1, 2, 23, 7] for estimators of the second order parameter when there is no covariate information. The definition of estimators of the conditional second order parameter is part of our future work as well as the study of the asymptotic properties of the $\gamma(t)$ estimator obtained by plugging the estimation of $\rho(t)$. Here, we limit ourselves to illustrating in the next subsection the effect of using a arbitrary chosen value.

4.3 Practical choice of weights

In this subsection, we study the behavior of the estimators $\hat{\gamma}_n(t, W^{\text{HZ}})$ and $\hat{\gamma}_n(t, W^{\text{opt}})$ in which we replace the second order parameter $\rho(t)$ by a arbi-

trary value $\rho^* < 0$. We then define $\hat{\gamma}_n(t, W^{\mathrm{HZ}}_{\rho^*})$ and $\hat{\gamma}_n(t, W^{\mathrm{opt}}_{\rho^*})$ with respective weights

$$\begin{aligned} W^{\mathrm{HZ}}_{\rho^*}(s,t) &= \frac{1}{\rho^*} - \left(1 - \frac{1}{\rho^*}\right)\log(s), \\ W^{\mathrm{opt}}_{\rho^*}(s,t) &= \frac{\rho^* - 1}{(\rho^*)^2}\left(\rho^* - 1 + (1 - 2\rho^*)s^{-\rho^*}\right). \end{aligned}$$

From the theoretical point of view, the following corollary illustrates the consequence of misspecifying the second order parameter on the asymptotic distribution.

Corollary 7 *Under* **(A.1)**, **(A.2)**, **(A.3)**, **(C)** *and (8), the convergence in distribution (9) holds for* $\hat{\gamma}_n(t, W^{\mathrm{HZ}}_{\rho^*})$ *and* $\hat{\gamma}_n(t, W^{\mathrm{opt}}_{\rho^*})$ *with*

$$\begin{aligned} &\mathcal{AB}(t, W^{\mathrm{HZ}}_{\rho^*}) = \tfrac{\rho^* - \rho(t)}{\rho^*(1-\rho(t))^2}, && \mathcal{AV}(t, W^{\mathrm{HZ}}_{\rho^*}) = 1 + (1 - 1/\rho^*)^2, \\ &\mathcal{AB}(t, W^{\mathrm{opt}}_{\rho^*}) = \tfrac{(1-\rho^*)(\rho^* - \rho(t))}{\rho^*(1-\rho(t))(1-\rho^* - \rho(t))}, && \mathcal{AV}(t, W^{\mathrm{opt}}_{\rho^*}) = (1 - 1/\rho^*)^2. \end{aligned}$$

The proof is a direct consequence of Theorem 2. It appears that a bias is introduced in the asymptotic distribution. Let us also note that the asymptotic bias of the estimators $\hat{\gamma}_n(t, W^{\mathrm{HZ}}_{\rho^*})$ and $\hat{\gamma}_n(t, W^{\mathrm{opt}}_{\rho^*})$ are of same sign. In terms of variance, such a misspecification can allow an improvement since $\rho^* \leq \rho(t)$ yields $\mathcal{AV}(t, W^{\mathrm{opt}}_{\rho^*}) \leq \mathcal{AV}(t, W^{\mathrm{opt}})$ and $\mathcal{AV}(t, W^{\mathrm{HZ}}_{\rho^*}) \leq \mathcal{AV}(t, W^{\mathrm{HZ}})$, see Figure 1. The densities of the asymptotic distributions of $\hat{\gamma}_n(t, W^{\mathrm{HZ}}_{\rho^*})$ are represented for different choices of ρ^* in case of a Burr distribution with extreme-value index $\gamma(t) = 0.3$ and second order parameter $\rho(t) = -1$. Here, $m_{n,t} = 5000$ and $k_{n,t} = 500$ leading to $b_{n,t} \simeq -0.08$. Clearly, choosing a small value of ρ^* is better than choosing a large one. In fact, it is easily seen that $\mathcal{AV}(t, W^{\mathrm{HZ}}_{\rho^*}) \to \mathcal{AV}(t, W^{\mathrm{Z}})$ and $\mathcal{AB}(t, W^{\mathrm{HZ}}_{\rho^*}) \to \mathcal{AB}(t, W^{\mathrm{Z}})$ as $\rho^* \to -\infty$, whereas $\mathcal{AV}(t, W^{\mathrm{HZ}}_{\rho^*}) \to +\infty$ and $\mathcal{AB}(t, W^{\mathrm{HZ}}_{\rho^*}) \to +\infty$ as $\rho^* \to 0$. Similar conclusions hold for $\hat{\gamma}_n(t, W^{\mathrm{opt}})$. The consequences of the misspecification of the second order parameter on the relative efficiency are studied in [8] in the unconditional case.

From the practical point of view, the four estimators $\hat{\gamma}_n(t, W^{\mathrm{H}})$, $\tilde{\gamma}_n(t, \mu^{\mathrm{Z}})$, $\hat{\gamma}_n(t, W^{\mathrm{HZ}})$ and $\hat{\gamma}_n(t, W^{\mathrm{opt}})$ are easily implementable. The remainder of this paragraph is devoted to their comparison. Simple calculations lead to the following partition of the (ρ, ρ^*) plane into 5 areas (see Figure 2) defined as A=$\{\rho(t) < 0,\ \rho(t)/(2 - \rho(t)) \leq \rho^* < 0\}$, where

$$\mathcal{AB}(t, W^{\mathrm{Z}}) \leq \mathcal{AB}(t, W^{\mathrm{H}}) \leq |\mathcal{AB}(t, W^{\mathrm{HZ}}_{\rho^*})| \leq |\mathcal{AB}(t, W^{\mathrm{opt}}_{\rho^*})|,$$

B=$\{\rho(t) < 0,\ (1 - \sqrt{1 - 2\rho(t)})/2 \leq \rho^* \leq \rho(t)/(2 - \rho(t))\}$, where

$$\mathcal{AB}(t, W^{\mathrm{Z}}) \leq |\mathcal{AB}(t, W^{\mathrm{HZ}}_{\rho^*})| \leq \mathcal{AB}(t, W^{\mathrm{H}}) \leq |\mathcal{AB}(t, W^{\mathrm{opt}}_{\rho^*})|,$$

C=$\{\rho(t) < 0,\ \rho(t)/2 \leq \rho^* \leq (1-\sqrt{1-2\rho(t)})/2\}$, where

$$\mathcal{AB}(t, W^{\mathrm{Z}}) \leq |\mathcal{AB}(t, W^{\mathrm{HZ}}_{\rho^*})| \leq |\mathcal{AB}(t, W^{\mathrm{opt}}_{\rho^*})| \leq \mathcal{AB}(t, W^{\mathrm{H}}),$$

D=$\{\rho(t) < 0,\ \rho_1(t) \leq \rho^* \leq \rho(t)/2\} \cup \{\rho(t) < 0,\ \rho^* \leq \rho_2(t)\}$, where

$$|\mathcal{AB}(t, W^{\mathrm{HZ}}_{\rho^*})| \leq \mathcal{AB}(t, W^{\mathrm{Z}}) \leq |\mathcal{AB}(t, W^{\mathrm{opt}}_{\rho^*})| \leq \mathcal{AB}(t, W^{\mathrm{H}}),$$

E=$\{\rho(t) < 0,\ \rho_2(t) \leq \rho^* \leq \rho_1(t)\}$, where

$$|\mathcal{AB}(t, W^{\mathrm{HZ}}_{\rho^*})| \leq |\mathcal{AB}(t, W^{\mathrm{opt}}_{\rho^*})| \leq \mathcal{AB}(t, W^{\mathrm{Z}}) \leq \mathcal{AB}(t, W^{\mathrm{H}}),$$

and with frontier functions

$$\begin{aligned}
\rho_1(t) &= \frac{\rho(t) - 1 - \sqrt{(1-\rho(t))^2 + 4(1-\rho(t))}}{2}, \\
\rho_2(t) &= \frac{(2+\rho(t))(\rho(t)-1) + \sqrt{(2+\rho(t))^2(1-\rho(t))^2 - 4\rho(t)(\rho(t)-1)(\rho(t)-2)}}{2(\rho(t)-2)}.
\end{aligned}$$

Next, concerning the corresponding asymptotic variances, we have:
In the half-plane N ($\rho^* \geq -1-\sqrt{2}$),

$$\mathcal{AV}(t, W^{\mathrm{H}}) \leq \mathcal{AV}(t, W^{\mathrm{Z}}) \leq \mathcal{AV}(t, W^{\mathrm{opt}}_{\rho^*}) \leq \mathcal{AV}(t, W^{\mathrm{HZ}}_{\rho^*}).$$

In the half-plane S ($\rho^* \leq -1-\sqrt{2}$),

$$\mathcal{AV}(t, W^{\mathrm{H}}) \leq \mathcal{AV}(t, W^{\mathrm{opt}}_{\rho^*}) \leq \mathcal{AV}(t, W^{\mathrm{Z}}) \leq \mathcal{AV}(t, W^{\mathrm{HZ}}_{\rho^*}).$$

These inequalities are summarized in Figure 2. For practical reasons, we limit $\rho(t)$ in $[-10, 0]$ and ρ^* in $[-4, 0]$. The dashed line represents the case $\rho^* = \rho(t)$.

5 Illustration on real data

In this section, we propose to illustrate our approach on the daily mean discharges (in cubic meters per second) of the Chelmer river collected by the Springfield gauging station, from 1969 to 2005. These data are provided by the Centre for Ecology and Hydrology (United Kingdom) and are available at `http://www.ceh.ac.uk/data/nrfa`. In this context, the variable of interest Y is the daily flow of the river and the bi-dimensional covariate $x = (x_1, x_2)$ is built as follows: $x_1 \in \{1969, 1970, \ldots, 2005\}$ is the year of measurement and $x_2 \in \{1, 2, \ldots, 365\}$ is the day. The size of the dataset is $n = 13,505$.

The smoothing parameter $h_{n,t}$ as well as the number of upper order statistics $k_{n,t}$ are assumed to be independent of t, they are thus denoted

by h_n and k_n respectively. They are selected by minimizing the following distance between conditional Hill and Zipf estimators:

$$\min_{h_n,k_n} \max_{t\in T} |\hat{\gamma}_n(t, W^{\mathrm{H}}) - \tilde{\gamma}_n(t, \mu^{\mathrm{Z}})|\,,$$

where $T = \{1969, 1970, \ldots, 2005\} \times \{15, 45, \ldots, 345\}$. This heuristics is commonly used in functional estimation and relies on the idea that, for a properly chosen pair (h_n, k_n) both estimates $\hat{\gamma}_n(t, W^{\mathrm{H}})$ and $\tilde{\gamma}_n(t, \mu^{\mathrm{Z}})$ should approximately yield the same value. The selected value of h_n corresponds to a smoothing over 4 years on x_1 and 2 months on x_2. Each ball $B(t, h_n)$, $t \in T$ contains $m_n = n\varphi(h_n) = 1089$ points and $k_n = 54$ rescaled log-spacings are used. This choice of k_n can be validated by computing on each ball $B(t, h_n)$, $t \in T$ the χ^2 distance to the standard exponential distribution. The histogram of these distances is superimposed in Figure 3 to the theoretical density of the corresponding χ^2 distribution. For instance, at level 5%, the χ^2 goodness of fit test rejects the exponential assumption in 5.7% of the balls. Verifying the independence assumption is more difficult. The use of declustering techniques [28] could help to identify independent events in the data record. In this illustration, we limit ourselves to supposing that the Springfield gauging station is close enough from the source of the Chelmer river so that the daily mean discharges are independent from one day to another one.

The resulting conditional Zipf estimator is presented on Figure 4. The obtained values are located in the interval $[0.2, 0.7]$. It appears that the estimated tail index is almost independent of the year but strongly dependent of the day. The heaviest tails are obtained in September, which means that, during this month, extreme flows are more likely than during the rest of year.

6 Proofs

For the sake of simplicity, in the sequel, we note k_t for $k_{n,t}$, b_t for $b_{n,t}$, m_t for $m_{n,t}$ and h_t for $h_{n,t}$.

6.1 Preliminary results

This first lemma provides sufficient conditions on γ and ℓ to obtain **(A.2)**. The following condition (15) controls the variations of $\log \ell$ with respect to its second argument, whereas assumption **(A.3)** controls the variations with respect to the first argument.

Lemma 1 *Assume that the first-order condition (4) holds. If, moreover, there exists positive constants z_ℓ, c_ℓ, c_γ, $\alpha_\gamma \leq 1$ and $\alpha_\ell \leq 1$ such that for all*

$x \in B(t,1)$,

$$\begin{aligned} |\gamma(x)-\gamma(t)| &\leq c_\gamma d^{\alpha_\gamma}(x,t), \\ \sup_{z>z_\ell}\left|\log\left(\frac{\ell(z,x)}{\ell(z,t)}\right)\right| &\leq c_\ell d^{\alpha_\ell}(x,t), \end{aligned} \tag{15}$$

then **(A.2)** *is verified with* $\alpha_U = \min(\alpha_\ell, \alpha_\gamma)$.

Proof – Under (4), we have

$$\frac{\log U(z,x)}{\log U(z,t)} - 1 = \frac{(\gamma(x)-\gamma(t))\log(z) + \log\left(\frac{\ell(z,x)}{\ell(z,t)}\right)}{\log(z)\gamma(t)\left(1+\frac{\log \ell(z,t)}{\gamma(t)\log(z)}\right)}.$$

Using the well-known property of slowly varying functions $\log \ell(z,x)/\log(z) \to 0$ as $z \to \infty$, and taking into account that $\gamma(t) > 0$, it follows that, for z large enough, there exists a constant $c'_\gamma > 0$ such that

$$\begin{aligned} \left|\frac{\log U(z,x)}{\log U(z,t)} - 1\right| &\leq \frac{c'_\gamma}{\gamma(t)} d^{\alpha_\gamma}(x,t) + \left|\log\left(\frac{\ell(z,x)}{\ell(z,t)}\right)\right| \\ &\leq \frac{c'_\gamma}{\gamma(t)} d^{\alpha_\gamma}(x,t) + c_\ell d^{\alpha_\ell}(x,t), \end{aligned}$$

and the conclusion follows. ■

The next lemma provides sufficient conditions on the weights to verify condition **(B.1)**.

Lemma 2 *Let* $W(.,t)$ *be a differentiable function on* $(0,1)$. *If* $sW(s,t) \to 0$ *as* $s \to 0$ *then (5) holds with* $u(s,t) = \partial sW(s,t)/\partial s$. *Furthermore, if there exists a positive and monotone function* $\phi(.,t)$ *defined on* $(0,1)$ *such that* $\max(|u(s,t)|, |W(s,t)|) \leq \phi(s,t)$, $\phi(1,t) < \infty$ *and* $\phi(.,t)$ *is integrable at the origin then (6) and (7) are satisfied.*

Proof – Clearly, since $W(.,t)$ is a differentiable function with $sW(s,t) \to 0$ as $s \to 0$, the function $sW(s,t)$ is absolutely continuous with $u(s,t) = \partial sW(s,t)/\partial s$. Furthermore, for all $j = 2, \ldots, k_t$,

$$\left|k_t \int_{(j-1)/k_t}^{j/k_t} u(\xi,t)d\xi\right| \leq \sup_{s\in[(j-1)/k_t, j/k_t]} \phi(s,t).$$

Since $\phi(.,t)$ is monotone on $(0,1)$, we have:

$$\left|k_t \int_{(j-1)/k_t}^{j/k_t} u(\xi,t)d\xi\right| \leq \begin{cases} \phi\left(\frac{j-1}{k_t},t\right) & \leq \phi\left(\frac{1}{2}\frac{j}{k_t+1},t\right) \quad \text{if } \phi(.,t) \text{ is decreasing,} \\ \phi\left(\frac{j}{k_t},t\right) & \leq \phi\left(2\frac{j}{k_t+1},t\right) \quad \text{if } \phi(.,t) \text{ is increasing.} \end{cases}$$

For $j=1$, we have

$$\left| k_t \int_0^{1/k_t} u(\xi,t)d\xi \right| = \left| W\left(\frac{1}{k_t},t\right)\right| \leq \phi\left(\frac{1}{k_t},t\right) \leq g\left(\frac{1}{k_t+1},t\right),$$

where

$$g(s,t) = \begin{cases} \phi(s/2,t) & \text{if } \phi(.,t) \text{ is decreasing,} \\ \phi(2s,t) & \text{if } \phi(.,t) \text{ is increasing.} \end{cases}$$

As a conclusion, condition (6) is verified. From Cauchy-Schwartz inequality, to prove (7), it only remains to verify that $\int_0^1 g(s,t)ds < +\infty$. This is a consequence of the integrability of $\phi(.,t)$ at the origin. ■

We now provide and example of a multidimensional design points and a distance d satisfying condition (11). In simple words, Lemma 3 states that, if the n covariates are distributed on a "rectangular" grid in $\mathbb{R}^p$, the proportion of points in $B(t,h_{n,t})$ is asymptotically proportional to the volume of this ball. See [18], Lemma 13.13 for a similar result in the random design setting.

Lemma 3 *Let $E=\mathbb{R}^p$, $d(x,t)=\|x-t\|_\infty$ and let G be a p-dimensional cumulative distribution function associated to a density function g such that $g(t)\neq 0$. Assume moreover that G admits independent margins $G_1,\ldots,G_p$ with bounded supports, $n^{1/p}\in\mathbb{N}$, and define the lattice $\mathcal{L}=\{1,2,\ldots,n^{1/p}\}^p$ in $\mathbb{N}^p$. We define the multidimensional design by $\{x_\beta,\ \beta\in\mathcal{L}\}$ where $\beta=(\beta_1,\ldots,\beta_p)\in\mathbb{N}^p$ is a multi-index and such that each coordinate of x_β is given by*

$$(x_\beta)_j \stackrel{def}{=} x_{\beta_j} \stackrel{def}{=} G_j^{-1}\left(\frac{\beta_j-1}{n^{1/p}-1}\right),\ j=1,\ldots,p.$$

Suppose $nh_t^p\to\infty$, then $\varphi(h_t)=(2h_t)^p g(t)(1+o(1))$.

Proof – Using the above definitions, we have

$$\begin{aligned}
\varphi(h_t) &= \frac{1}{n}\sum_{\beta\in\mathcal{L}} \mathbb{I}\{\|x_\beta - t\|_\infty \leq h\} \\
&= \frac{1}{n}\sum_{\beta_1=1}^{n^{1/p}}\cdots\sum_{\beta_p=1}^{n^{1/p}}\prod_{j=1}^{p}\mathbb{I}\{t_j-h_t\leq x_{\beta_j}\leq t_j+h_t\} \\
&= \frac{1}{n}\prod_{j=1}^{p}\sum_{\beta_j=1}^{n^{1/p}}\mathbb{I}\{t_j-h_t\leq x_{\beta_j}\leq t_j+h_t\} \\
&= \frac{1}{n}\prod_{j=1}^{p}\sum_{\beta_j=1}^{n^{1/p}}\mathbb{I}\left\{G_j(t_j-h_t)\leq \frac{\beta_j-1}{n^{1/p}-1}\leq G_j(t_j+h_t)\right\} \\
&= (1-n^{-1/p})^p\prod_{j=1}^{p}\frac{1}{n^{1/p}-1}\sum_{\beta_j=1}^{n^{1/p}}Q_j\left(\frac{\beta_j-1}{n^{1/p}-1}\right), \qquad (16)
\end{aligned}$$

where the following indicator function

$$Q_j(u) = \mathbb{I}\left\{G_j(t_j - h_t) \leq u \leq G_j(t_j + h_t)\right\}$$

is introduced for $u \in [0,1]$. The above Riemann's sums can be approximated as

$$\begin{aligned}
\frac{1}{n^{1/p}-1}\sum_{\beta_j=1}^{n^{1/p}} Q_j\left(\frac{\beta_j - 1}{n^{1/p}-1}\right) &= \int_0^1 Q_j(u)du + O(n^{-1/p}) \\
&= G_j(t_j + h_t) - G_j(t_j - h_t) + O(n^{-1/p}) \\
&= 2h_t g_j(t_j) + o(h_t) + O(n^{-1/p}) \\
&= 2h_t g_j(t_j)(1 + o(1)),
\end{aligned}$$

since we assumed that $g(t) \neq 0$ and $nh_t^p \to \infty$. Replacing in (16), the result follows. ■

6.2 Proofs of main results

Proof of Theorem 1 – Under **(A.1)** we have $\{Z_{m_t-i+1,m_t}(t)\}_i \stackrel{d}{=} \{U(V_{i,m_t}^{-1}, x_i)\}_i$ where $V_{1,m_t} \leq \ldots \leq V_{m_t,m_t}$ are the order statistics associated to the sample $V_1, \ldots, V_{m_t}$ of independent uniform variables. It follows that:

$$\begin{aligned}
\{\log(Z_{m_t-i+1,m_t}(t))\}_i &\stackrel{d}{=} \left\{\log(U(V_{i,m_t}^{-1}, t))\left(1 + \frac{\log(U(V_{i,m_t}^{-1}, x_i))}{\log(U(V_{i,m_t}^{-1}, t))} - 1\right)\right\}_i \\
&\stackrel{def}{=} \left\{\log(U(V_{i,m_t}^{-1}, t))\,(1 + \varepsilon_{n,i})\right\}_i.
\end{aligned}$$

Now, assumption **(C)** entails that for all $i = 1, \ldots, k_t$,

$$V_{i,m_t}^{-1} \geq V_{k_t,m_t}^{-1} = (m_t/k_t)(1 + o_{\mathrm{P}}(1)) \stackrel{\mathrm{P}}{\to} \infty,$$

which implies that, for n large enough, $V_{i,m_t}^{-1} \geq z_U$ for all $i = 1, \ldots, k_t$. Consequently, **(A.2)** implies that

$$\max_{i=1,\ldots,k_t} |\varepsilon_{n,i}| \leq c_U h_t^{\alpha_U},$$

and we thus have $\{\log(Z_{m_t-i+1,m_t}(t))\}_i \stackrel{d}{=} \{\log(U(V_{i,m_t}^{-1}, t))(1 + O_{\mathrm{P}}\left(h_t^{\alpha_U}\right))\}_i$. The end of the proof is then a direct consequence of the following result (see [3], Theorem 2.1 and 2.2 for a proof):

$$\left\{i\log\left(\frac{U(V_{i,m_t}^{-1}, t)}{U(V_{i+1,m_t}^{-1}, t)}\right)\right\}_i = \left\{\left(\gamma(t) + b_t\left(\frac{i}{k_t+1}\right)^{-\rho(t)}\right)F_i + \beta_{i,n}(t) + o_{\mathrm{P}}(b_t)\right\}_i,$$

where $\{F_i\}_i \stackrel{def}{=} \{i\log(V_{i,m_t}^{-1}/V_{i+1,m_t}^{-1})\}_i$ are independent standard exponential variables and with (under **(B.1)**)

$$\sum_{i=1}^{k_t}\left(\frac{1}{i}\int_0^{i/k_t} u(s,t)ds\right)\beta_{i,n}(t) = \frac{1}{k_t}\sum_{i=1}^{k_t} W\left(i/k_t,t\right)\beta_{i,n}(t) = o_{\mathrm{P}}(b_t).$$

■

Proof of Theorem 2 – From Theorem 1, we have

$$\begin{aligned}\left(\sum_{i=1}^{k_t} W(i/k_t,t)\right)\hat{\gamma}_n(t,W) &\stackrel{d}{=} \gamma(t)(1+O_{\mathrm{P}}(h_t^{\alpha_U}))\sum_{i=1}^{k_t} W(i/k_t,t)F_i \\ &+ (1+O_{\mathrm{P}}(h_t^{\alpha_U}))b_t\sum_{i=1}^{k_t} W(i/k_t,t)\left(\frac{i}{k_t+1}\right)^{-\rho(t)} F_i \\ &+ (1+O_{\mathrm{P}}(h_t^{\alpha_U}))\sum_{i=1}^{k_t} W(i/k_t,t)\beta_{i,n}(t) \\ &+ o_{\mathrm{P}}(b_t)\sum_{i=1}^{k_t}|W(i/k_t,t)|.\end{aligned}$$

Introducing

$$T_{1,n} = \sum_{i=1}^{k_t} W(i/k_t,t)(F_i-1),\ T_{2,n} = \sum_{i=1}^{k_t} W(i/k_t,t)\left(\frac{i}{k_t+1}\right)^{-\rho(t)}(F_i-1),$$

$$T_{3,n} = \sum_{i=1}^{k_t} W(i/k_t,t)\beta_{i,n}(t),\ T_{4,n} = b_t\sum_{i=1}^{k_t} W(i/k_t,t)\left(\frac{i}{k_t+1}\right)^{-\rho(t)},$$

$$T_{5,n} = \sum_{i=1}^{k_t} W(i/k_t,t),\ T_{6,n} = \sum_{i=1}^{k_t}|W(i/k_t,t)|,\ T_{7,n} = \left(\sum_{i=1}^{k_t} W^2(i/k_t,t)\right)^{1/2},$$

we obtain the following expansion:

$$\begin{aligned}\frac{T_{5,n}}{T_{7,n}}\left(\hat{\gamma}_n(t,W)-\gamma(t)-\frac{T_{4,n}}{T_{5,n}}\right) &\stackrel{d}{=} \left(\gamma(t)\frac{T_{1,n}}{T_{7,n}} + b_t\frac{T_{2,n}}{T_{7,n}} + \frac{T_{3,n}}{T_{7,n}}\right)(1+O_{\mathrm{P}}(h_t^{\alpha_U})) \\ &+ \left(\frac{T_{4,n}}{T_{7,n}} + \frac{T_{5,n}}{T_{7,n}}\right)O_{\mathrm{P}}(h_t^{\alpha_U}) + \frac{T_{6,n}}{T_{7,n}}o_{\mathrm{P}}(b_t). \quad (17)\end{aligned}$$

Let δ be defined by **(C.2)**. From Lindeberg theorem, a sufficient condition for $T_{1,n}/T_{7,n} \stackrel{d}{\to} \mathcal{N}(0,1)$ is that

$$\sum_{i=1}^{k_t}|W(i/k_t,t)|^{2+\delta}/T_{7,n}^{2+\delta} \to 0. \quad (18)$$

Since, for any integrable function ψ, the following convergence of Riemann sum holds,

$$\frac{1}{k_t}\sum_{i=1}^{k_t}\psi\left(\frac{i}{k_t}\right) \to \int_0^1 \psi(s)ds, \tag{19}$$

it follows that $T_{7,n} = k_t^{1/2}\mathcal{AV}(t,W)^{1/2}(1+o(1))$. Thus, using again (19),

$$\sum_{i=1}^{k_t} |W(i/k_t,t)|^{2+\delta}/T_{7,n}^{2+\delta} = O(k_t^{-\delta/2}),$$

showing that condition (18) is satisfied and

$$T_{1,n}/T_{7,n} \xrightarrow{d} \mathcal{N}(0,1). \tag{20}$$

Next, we focus on the term $T_{2,n}/T_{7,n}$. Remarking that this term is centered, and that its variance is finite, we can conclude that

$$T_{2,n}/T_{7,n} = O_{\mathrm{P}}(1). \tag{21}$$

Theorem 1 shows that

$$T_{3,n}/T_{7,n} = o_{\mathrm{P}}(k_t^{1/2}b_t) = o_{\mathrm{P}}(1). \tag{22}$$

From repeated use of (19), it follows that

$$\begin{aligned}
T_{4,n}/T_{5,n} &= b_t\mathcal{AB}(t,W)(1+o(1)) && (23)\\
T_{4,n}/T_{7,n} &= O(k_t^{1/2}b_t) = O(1) && (24)\\
T_{5,n}/T_{7,n} &= k_t^{1/2}\mathcal{AV}(t,W)^{-1/2}(1+o(1)) && (25)\\
T_{6,n}/T_{7,n} &= O(k_t^{1/2}). && (26)
\end{aligned}$$

Replacing (21)–(26) in (17) yields

$$\begin{aligned}
& k_t^{1/2}\mathcal{AV}(t,W)^{-1/2}\left(\hat{\gamma}_n(t,\mu) - \gamma(t) - b_t\mathcal{AB}(t,W)\right)\\
\stackrel{d}{=}\ & \gamma(t)T_{1,n}/T_{7,n} + O(k_t^{1/2}h_t^{\alpha_U}) + o_P(1),
\end{aligned}$$

and (20) gives the result. ■

Proof of Corollary 1 – The proof consists in remarking that

$$\begin{aligned}
& \tilde{\gamma}_n(t,\mu) - \gamma(t) - b_t\mathcal{AB}(t,W)\\
= \ & \sum_{i=1}^{k_t}\mu_{i,n}(t)\left(C_{i,n}(t) - \gamma(t) - b_t\mathcal{AB}(t,W)\right) \Big/ \sum_{i=1}^{k_t}\mu_{i,n}(t)\\
= \ & (1+o(1))\sum_{i=1}^{k_t} W\left(i/k_t,t\right)\left(C_{i,n}(t) - \gamma(t) - b_t\mathcal{AB}(t,W)\right) \Big/ \sum_{i=1}^{k_t} W\left(i/k_t,t\right)\\
= \ & (1+o(1))\left(\hat{\gamma}_n(t,W) - \gamma(t) - b_t\mathcal{AB}(t,W)\right),
\end{aligned}$$

and the conclusion follows from Theorem 2. ■

Proof of Corollary 2 – Assuming that $L(y,x)=1$ for all $(y,x)\in\mathbb{R}_+\times\mathbb{R}^p$ implies $\ell(y,x)=1$ in (4) and thus **(A.3)** holds with $b(y,t)=0$. Furthermore, **(A.1)** is straightforwardly true and since γ is α-Lipschitzian, Lemma 1 entails that **(A.2)** holds. Choosing $h_{n,t}=n^{-\frac{1}{p+2\alpha}}$ and $k_{n,t}=n^{\frac{2\alpha}{p+2\alpha}}\eta_n^2$, where $\eta_n\to 0$ arbitrarily slowly, condition **(C)** is verified since $nh_{n,t}^p/k_{n,t}\to\infty$ and (11) imply $n\varphi(h_{n,t})/k_{n,t}\to\infty$. As a conclusion, Theorem 2 provides the asymptotic normality of the estimator with convergence rate $n^{\frac{\alpha}{p+2\alpha}}\eta_n$.■

Proof of Corollary 4 – Let us first prove that (13) belongs to the extended family (3). Remarking that

$$\tau_{i,n}(t)=\sum_{j=i}^{k_t}\frac{1}{j}+\sum_{j=k_t+1}^{m_t}\frac{1}{j},$$

estimator (13) can be rewritten as :

$$\sum_{i=1}^{k_t}(\tau_{i,n}(t)-\bar{\tau}_n(t))\log(Z_{m_t-i+1,m_t}(t)/Z_{m_t-k_t,m_t}(t))\Big/\sum_{i=1}^{k_t}(\tau_{i,n}(t)-\bar{\tau}_n(t))\sum_{j=i}^{k_t}\frac{1}{j}. \tag{27}$$

Next, since

$$\log(Z_{m_t-i+1,m_t}(t)/Z_{m_t-k_t,m_t}(t))=\sum_{j=i}^{k_t}\log(Z_{m_t-j+1,m_t}(t)/Z_{m_t-j,m_t}(t)),$$

inverting the sums in (27), it appears that (13) belongs to family (3) with

$$\mu_{i,n}^{\mathrm{Z}}(t)=\frac{1}{i}\sum_{j=1}^{i}(\tau_{j,n}(t)-\bar{\tau}_n(t)).$$

Second, we prove that, uniformly in $i=1,\ldots,k_t$,

$$\mu_{i,n}^{\mathrm{Z}}(t)=-\log\left(i/k_t\right)(1+o(1)). \tag{28}$$

For the sake of simplicity, we introduce the following notation:

$$S_{i,m_t}=\frac{1}{i}\sum_{j=1}^{i}\tau_{j,n}(t)=\frac{1}{i}\sum_{j=1}^{i}\sum_{l=j}^{m_t}\frac{1}{l},\ i=1,\ldots,k_t,$$

so that $\mu_{i,n}^{\mathrm{Z}}(t)=S_{i,m_t}-S_{k_t,m_t}$. Furthermore, for $i=2,\ldots,k_t$,

$$\begin{aligned}S_{i,m_t}&=\frac{1}{i}\sum_{j=1}^{i-1}\sum_{l=j}^{m_t}\frac{1}{l}+\frac{1}{i}\sum_{l=i}^{m_t}\frac{1}{l}\\&=\frac{i-1}{i}S_{i-1,m_t}+\frac{1}{i}\sum_{l=i}^{m_t}\frac{1}{l}=S_{i-1,m_t}-\frac{1}{i}\left(S_{i-1,m_t}-\sum_{l=i}^{m_t}\frac{1}{l}\right),\end{aligned}$$

and remarking that

$$S_{i-1,m_t} - \sum_{l=i}^{m_t} \frac{1}{l} = \frac{1}{i-1} \sum_{j=1}^{i-1} \sum_{l=j}^{i-1} \frac{1}{l} = 1,$$

we obtain the following recursive relation: $S_{i,m_t} = S_{i-1,m_t} - 1/i$ for $i = 2, \ldots, k_t$. We thus have a simplified expression of the weights:

$$\mu_{i,n}^{\mathrm{Z}}(t) = \begin{cases} \sum\limits_{l=i+1}^{k_t} \frac{1}{l} & i = 1, \ldots, k_t - 1, \\ 0 & i = k_t. \end{cases}$$

We are now in position to evaluate the difference between $\mu_{i,n}^{\mathrm{Z}}(t)$ and $-\log(i/k_t)$. For $i = 1, \ldots, k_t - 1$,

$$-\log(i/k_t) = \log\left(\prod_{l=i+1}^{k_t} \frac{l}{l-1}\right) = \sum_{l=i+1}^{k_t} \log\left(1 + \frac{1}{l-1}\right),$$

and consequently,

$$-\log(i/k_t) - \mu_{i,n}^{\mathrm{Z}}(t) = \begin{cases} \sum\limits_{l=i+1}^{k_t} \left(\log\left(1 + \frac{1}{l-1}\right) - \frac{1}{l}\right) & i = 1, \ldots, k_t - 1, \\ 0 & i = k_t. \end{cases} \tag{29}$$

Remarking that for $l \geq 2$ the following inequality holds,

$$0 \leq \log\left(1 + \frac{1}{l-1}\right) - \frac{1}{l} \leq \frac{1}{l^2},$$

we deduce from (29) that for $i = 1, \ldots, k_t - 1$,

$$0 \leq -\log(i/k_t) - \mu_{i,n}^{\mathrm{Z}}(t) \leq \sum_{l=i+1}^{k_t} \frac{1}{l^2}.$$

Furthermore, since

$$\sum_{l=i+1}^{k_t} \frac{1}{l^2} \leq \int_i^{k_t} \frac{1}{x^2} dx = \frac{1}{i} - \frac{1}{k_t},$$

we have for $i = 1, \ldots, k_t - 1$,

$$0 \leq 1 - \mu_{i,n}^{\mathrm{Z}}(t)/\log(k_t/i) \leq -\frac{1}{\log(i/k_t)} \left(\frac{1}{i} - \frac{1}{k_t}\right).$$

Finally, since the sequence

$$h(i) = -\frac{1}{\log(i/k_t)} \left(\frac{1}{i} - \frac{1}{k_t}\right), \ i \in [1, k_t[$$

is decreasing, we have for $i = 1, \ldots, k_t - 1$

$$0 \leq 1 - \mu_{i,n}^{\mathrm{Z}}(t)/\log(k_t/i) \leq \frac{1}{\log(k_t)}\left(1 - \frac{1}{k_t}\right),$$

proving that (28) is true. The end of the proof is a consequence of Corollary 1 and Theorem 2. ■

Proof of Proposition 2 – For all W such that $\int_0^1 W(s,t)ds = 1$, we have

$$\int_0^1 W^2(s,t)ds = 1 + \int_0^1 (W(s,t) - 1)^2 ds,$$

and thus minimizing the left-hand term $\int_0^1 W^2(s,t)dt$ is equivalent to minimizing $\int_0^1 (W(s,t) - 1)^2 ds$. Consequently, the solution of the constrained optimization problem is $W(.,t) = 1$ almost everywhere on $[0,1]$. Since W is assumed to be continuous, the conclusion follows. ■

Proof of Proposition 3 – First, we easily check that the function $W^{\mathrm{opt}}(.,t)$ is continuous, $\int_0^1 W^{\mathrm{opt}}(s,t)ds = 1$ and $\int_0^1 W^{\mathrm{opt}}(s,t)s^{-\rho(t)}ds = 0$. Next, remarking that for all continuous function $W(.,t)$ satisfying $\int_0^1 W(s,t)ds = 1$ and $\int_0^1 W(s,t)s^{-\rho(t)}ds = 0$, we have

$$\int_0^1 W^2(s,t)ds = \left(\frac{\rho(t) - 1}{\rho(t)}\right)^2 + \int_0^1 (W(s,t) - W^{\mathrm{opt}}(s,t))^2 ds,$$

it appears that minimizing the left-hand term $\int_0^1 W^2(s,t)ds$ is equivalent to minimizing $\int_0^1 (W(s,t) - W^{\mathrm{opt}}(s,t))^2 ds$. Since $W(.,t)$ is continuous, the conclusion of the proof is straightforward. ■

Acknowledgment

The authors are indebted to the anonymous referees for their helpful comments and suggestions.

References

[1] Alves, M.I.F., Gomes, M.I. and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter, *Portugaliae Mathematica*, **(60)**, 193–214.

[2] Alves, M.I.F., de Haan, L. and Lin, T. (2003). Estimation of the parameter controlling the speed of convergence in extreme value theory, *Mathematical Methods of Statistics*, **(12)**, 155–176.

[3] Beirlant, J., Dierckx, G., Guillou, A. and Stărică, C. (2002). On exponential representations of log-spacings of extreme order statistics, *Extremes*, **5**, 157–180.

[4] Beirlant, J. and Goegebeur, Y. (2003). Regression with response distributions of Pareto-type, *Computational Statistics and Data Analysis*, **42**, 595–619.

[5] Beirlant, J. and Goegebeur, Y. (2004). Local polynomial maximum likelihood estimation for Pareto-type distributions, *Journal of Multivariate Analysis*, **89**, 97–118.

[6] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular variation*, Encyclopedia of Mathematics and its Applications, **27**, Cambridge University Press.

[7] Caeiro, F. and Gomes, M.I. (2002). Bias-reduction in the estimation of parameters of rare events, *Theory of Stochastic Processes*, **8**, 67–76.

[8] Caeiro, F. and Gomes, M.I. (2002). A class of asymptotically unbiased semi-parametric estimators of the tail index, *Test*, **11**, 345–364.

[9] Chavez-Demoulin, V. and Davison, A.C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society, series C.*, **54**, 207–222.

[10] Csörgö, S., Deheuvels, P. and Mason, D. (1985). Kernel estimates of the tail index of a distribution, *Annals of Statistics*, **13**, 1050–1077.

[11] Csörgö, S. and Viharos, L. (1998). Estimating the tail index, in: B. Szyszkowicz (Ed.), Asymptotic Methods in Probability and Statistics, North-Holland, Amsterdam, 833–881.

[12] Davison, A.C. and Ramesh, N.I. (2000). Local likelihood smoothing of sample extremes, *Journal of the Royal Statistical Society, series B*, **62**, 191–208.

[13] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society, series B*, **52**, 393–442.

[14] Diebolt, J., Gardes, L., Girard, S. and Guillou, A. (2007). Bias-reduced estimators of the Weibull-tail coefficient, *Test*, to appear.

[15] Drees, H., de Haan, L. and Resnick, S. (2000). How to make a Hill plot, *Annals of Statistics*, **28**, 254–274.

[16] Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). *Modelling extremal events*, Springer.

[17] Falk, M., Hüsler, J. and Reiss, R.D. (2004). *Laws of small numbers: Extremes and rare events*, 2nd edition, Birkhäuser.

[18] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, Springer.

[19] Galambos, J. (1987). *The asymptotic theory of extreme order statistics*, R.E. Krieger publishing compagny.

[20] Gangopadhyay, A.K. (1995). A note on the asymptotic behavior of conditional extremes, *Statistics and Probability Letters*, **25**, 163–170.

[21] Gardes, L. and Girard, S. (2007). Estimation of the Weibull tail-coefficient with linear combination of upper order statistics, *Journal of Statistical Planning and Inference*, to appear.

[22] Geluk, J.L. and de Haan, L. (1987). Regular variation, extensions and Tauberian theorems, *Math Centre tracts*, **40**, Centre for Mathematics and Computer Science, Amsterdam.

[23] Gomes, M.I., Martins, M.J. and Neves, M. (2000). Semi-parametric estimation of the second order parameter, asymptotic and finite sample behaviour, *Extremes*, **3**, 207–229.

[24] Hall, P. and Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data, *Statistical Science*, **15**, 153–167.

[25] Häusler, E. and Teugels, J.L. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation, *Annals of Statistics*, **13**, 743–756.

[26] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, **3**, 1163–1174.

[27] Kratz, M. and Resnick, S. (1996). The QQ-estimator and heavy tails, *Stochastic Models*, **12**, 699–724.

[28] Nadarajah, S. (2001). Multivariate declustering techniques, *Environ-Metrics*, **12**, 357–365.

[29] Reiss, R.D. and Thomas, M. (2001). *Statistical analysis of extreme values*, Birkhäuser, Basel.

[30] Schultze, J. and Steinebach, J. (1996). On least squares estimates of an exponential tail coefficient, *Statistics and Decisions*, **14**, 353–372.

[31] Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone (with discussion). *Statistical Science*, **4**, 367–393.

[32] Stone, C. (1982). Optimal global rates of convergence for nonparametric regression, *Annals of Statistics*, **10**, 1040–1053.

[33] Viharos, L. (1999). Weighted least-squares estimators of tail indices, *Probability and Mathematical Statistics*, **19**, 249–265.

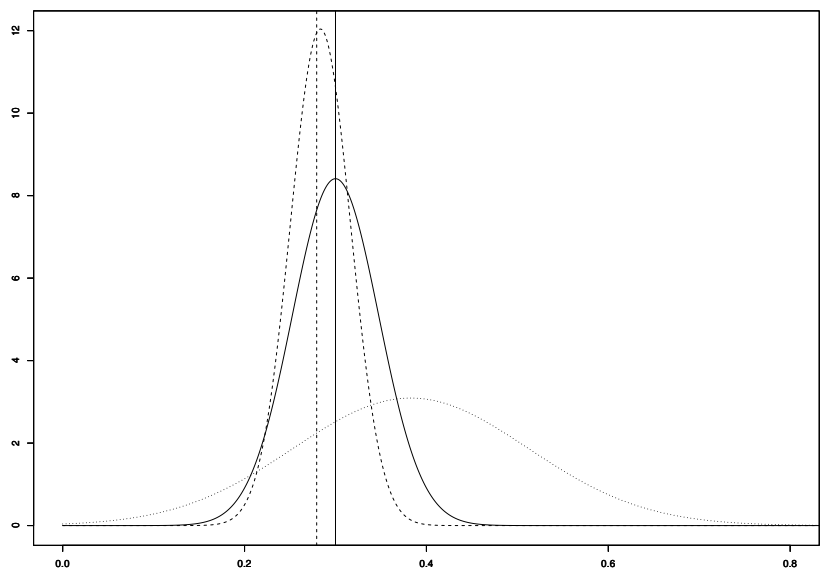

Figure 1: Densities of the asymptotic distributions of $\hat{\gamma}_n(t, W^{\mathrm{HZ}}_{\rho^*})$. Solid curve $\rho^* = -1$, dotted curve $\rho* = -0.2$, dashed curve $\rho^* = -5$, solid vertical line: true value γ, dotted vertical line: $\gamma + b_{n,k}\mathcal{AB}(t, W^{\mathrm{Z}})$, *i.e.*, the mean of the asymptotic distribution when $\rho^* \to -\infty$.

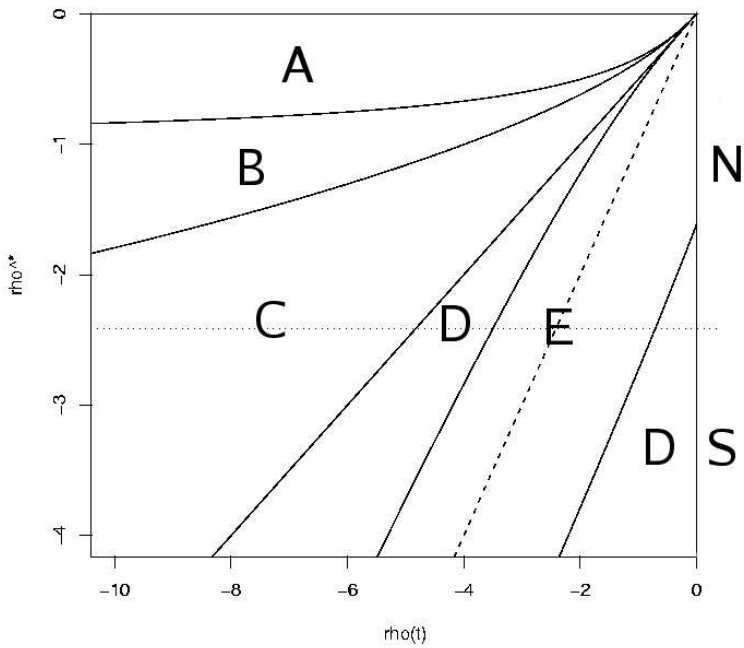

Figure 2: Comparison of the asymptotic bias and variances.

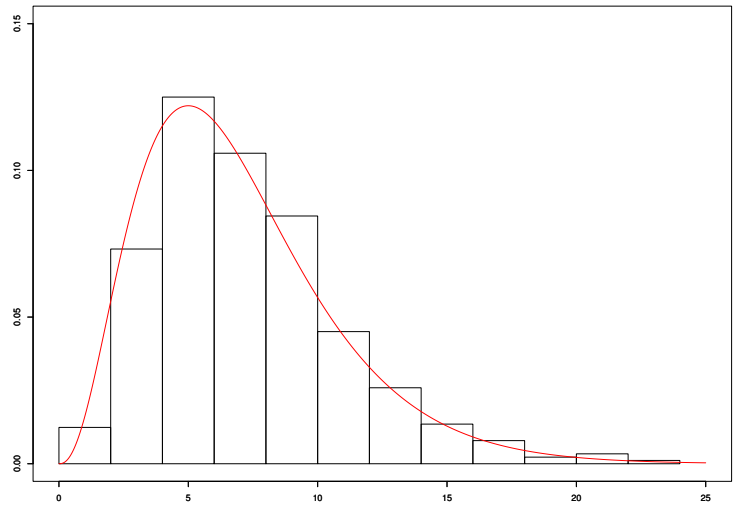

Figure 3: Histogram of the χ^2 distances between the rescaled log-spacings and the standard exponential distribution. The theoretical density of the corresponding χ^2 distribution is superimposed.

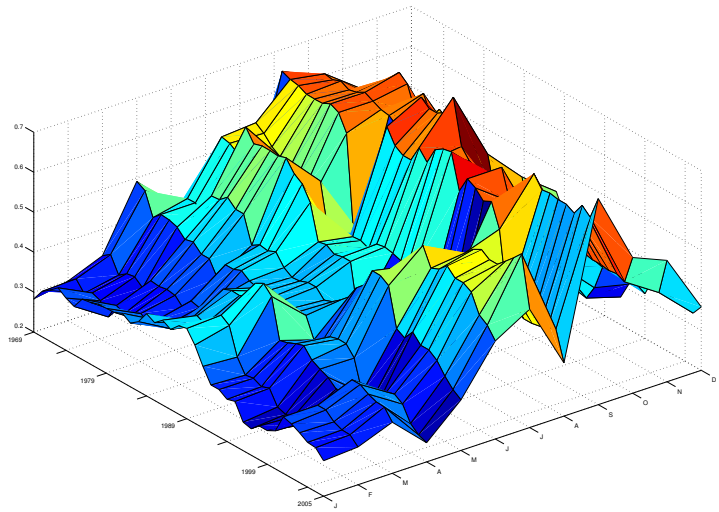

Figure 4: Conditional Zipf estimator $\tilde{\gamma}_n(t, \mu^{\mathrm{Z}})$ of the tail index computed on the real dataset. Two covariates are available: The year ranging from 1969 to 2005 and the day ranging from 1 to 365. For the sake of readability, only the first letter of the corresponding month is represented.