A local moment type estimator for the extreme value index in regression with random covariates

Yuri Goegebeur[1], Armelle Guillou[2]* and Michael Osmann[3]

[1]*Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark*
[2]*Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France*
[3]*Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark*

Abstract: This paper deals with the nonparametric estimation of the conditional extreme value index of a response in presence of random covariates. In particular, it is assumed that the conditional response distribution belongs to the max-domain of attraction of the extreme value distribution, and its index is estimated locally within a narrow neighbourhood of the point of interest in the covariate space. The moment estimator, originally introduced in Dekkers, Einmahl, & de Haan (1989), is adjusted to the local estimation context, and its asymptotic properties are investigated under some mild conditions on the response distribution, the density function of the covariates, the kernel function, and for appropriately chosen sequences of bandwidth and threshold parameters. The finite sample performance of the proposed estimator is evaluated by means of an extensive simulation study where a comparison with alternatives from the recent literature is provided. We also illustrate the practical applicability of the estimator on the world catalogue of earthquake magnitudes. *The Canadian Journal of Statistics* xx: 1–25; 20?? © 20?? Statistical Society of Canada

Résumé: Nous considérons dans cet article l'estimation non paramétrique de l'indice de queue conditionnel en présence de covariables aléatoires. Sous l'hypothèse que la loi conditionnelle des réponses appartient au domaine d'attraction d'une loi des extrêmes, nous estimons son indice de queue localement dans le voisinage d'un point de l'espace des covariables. L'estimateur des moments introduit par Dekkers, Einmahl, & de Haan (1989) est adapté au contexte de l'estimation locale. Ses propriétés asymptotiques sont établies sous des hypothèses convenables sur la loi conditionnelle, la densité des covariables, le noyau et pour un choix approprié de la fenêtre et du seuil. Nous illustrons le comportement à distance finie de notre estimateur sur simulations et le comparons à plusieurs alternatives introduites récemment dans la littérature. L'utilisation pratique sur des données de magnitudes de séismes est également proposée. *La revue canadienne de statistique* xx: 1–25; 20?? © 20?? Société statistique du Canada

1. INTRODUCTION

In extreme value theory the estimation of the extreme value index assumes a central position. Estimation of the extreme value index γ in the univariate framework has been considered extensively, and a vast literature has been dedicated to it. See for instance Beirlant et al. (2004), and de Haan & Ferreira (2006), for recent accounts on this topic. We consider here a regression

---

* *Author to whom correspondence may be addressed.*
 E-mail: armelle.guillou@math.unistra.fr

setting where the response variable Y is recorded along with a *random* covariate X. Let the conditional distribution function of Y given $X = x$ be $F(y; x)$ and set $\overline{F}(y; x) := 1 - F(y; x)$. We assume that $F(y; x)$ belongs to the max-domain of attraction of the generalized extreme value distribution, i.e. there exists a constant $\gamma(x)$ and a positive rate function $a(.; x)$, such that

$$\lim_{t \to \infty} \frac{U(ty; x) - U(t; x)}{a(t; x)} = D_{\gamma(x)}(y), \tag{1}$$

with

$$D_{\gamma(x)}(y) := \int_1^y u^{\gamma(x)-1} du$$

for every $y > 0$, where $U(t; x)$ is the tail quantile function defined as

$$U(t; x) := F^{\leftarrow}\left(1 - \frac{1}{t}; x\right) := \inf\left\{y : F(y; x) \geq 1 - \frac{1}{t}\right\}, \quad t > 1.$$

The parameter $\gamma(x)$, called the conditional extreme value index, gives important information about the tail heaviness of $\overline{F}(y; x)$, distinguishing between heavy tails ($\gamma(x) > 0$), moderate tails ($\gamma(x) = 0$) and light tails ($\gamma(x) < 0$). In the present paper we introduce a nonparametric estimator for $\gamma(x)$ based on local estimation within a narrow neighbourhood of the point of interest in the covariate space. In particular we will adjust the moment estimator, originally proposed by Dekkers, Einmahl, & de Haan (1989) as estimator for the extreme value index in the univariate context, to the setting of local estimation. Unlike classical regression analysis with a focus on the estimation of the mean of the conditional response distribution, we consider here a regression problem where interest is in estimating the conditional extreme value index. There are many practical instances where it is more relevant to study the tail of the conditional response distribution rather than the mean. For instance, in an excess-of-loss reinsurance contract the reinsurer intervenes if the claims faced by the ceding companies exceed some contractually specified threshold by paying the claim amount in excess of the threshold. This threshold is typically large and therefore it is for reinsurers of crucial importance to have an accurate estimate of the tail of the claim size distribution. Linking the tail of this claim size distribution to independent variables allows them to obtain a more adequate assessment of the risks involved, and consequently a better determination of the premium levels. In fact, for heavy tailed response distributions the conditional expectation might not exist. Also, estimation of the conditional extreme value index is generally a first step in an extreme value analysis, since e.g. estimation of extreme quantiles or small tail probabilities will require an estimate for $\gamma(x)$.

To illustrate the practical applicability of our method we will consider the world catalogue of earthquakes which contains information about earthquakes that have happened between 1976 and the present. Accurate modeling of the tail of the earthquake energy distribution has clearly a considerable practical relevance since severe earthquakes cause much damage and losses. With our method we can link the tail of this earthquake energy distribution to local factors, which allows us to differentiate the risks geographically. This information is useful for e.g. engineers in order to determine the strength of structures like buildings, bridges and nuclear reactors. Other applications concern the study of claim sizes in insurance as a function of risk factors, estimation of the tail of the diamond value distribution conditional on the variables size and colour, the analysis of survival at extreme durations, to name but a few.

The estimation of the extreme value index with fixed, i.e. nonrandom, covariates has been investigated rather extensively in the recent extreme value literature, and we refer to Chapter 6 in Coles (2001) and Chapter 7 in Beirlant et al. (2004), and the references therein, for an overview of the available methodology. Less attention though has been paid to the random covariate case, despite its practical interest, and most of the available methods are situated in the class of the Pareto-type distributions, corresponding to $\gamma(x) > 0$. Wang & Tsai (2009) and Wang & Li (2013) developed parametric estimation methods based on a setting where $\gamma(x) = \exp(-x^T\theta)$, with θ a $\mathbb{R}^p-$ vector of regression coefficients, or based on the assumption that the Box-Cox transformation of the conditional quantile function of the response variable is linear in the covariate x. Nonparametric kernel methods were introduced by Daouia et al. (2011), who used a fixed number of extreme conditional quantile estimators to estimate the conditional extreme value index, for instance using the Hill (Hill, 1975) and Pickands (Pickands, 1975) estimators, whereas Goegebeur, Guillou, & Schorgen (2013) developed a nonparametric and asymptotically unbiased estimator based on weighted exceedances over a high threshold. The estimation of the extreme value index for the complete max-domain of attraction in presence of random covariates has hardly been considered. We are only aware of two papers. The first is by Daouia, Gardes, & Girard (2013), generalizing the methodology of Daouia et al. (2011), where a fixed number of kernel estimators for extreme conditional quantiles was plugged in a refined Pickands estimator (Drees, 1995) for the extreme value index, and its asymptotic properties were established, though assuming that the distribution function is twice differentiable. The second paper by Stupfler (2013) proposes a simple estimator, based on a direct adaptation of the moment estimator, but using a data- driven local threshold. This more complex setting makes the technical issues more difficult, but does not necessarily improve the performance of the estimator in practice as illustrated in our Section 3. Also, the Stupfler (2013) estimator does not weight the excesses over a threshold by a kernel function, giving less weight to observations more distant from the point of interest.

The remainder of this paper is organized as follows. In the next section we introduce the estimator and study its asymptotic properties under some mild regularity conditions. The finite sample performance of the proposed estimator is evaluated in an extensive simulation study in Section 3 where a comparison with alternatives proposed in the literature is provided. In Section 4 we illustrate the applicability of the method on the world catalogue of earthquake magnitudes. Section 5 concludes the paper. The proofs of all our results are detailed in Supplementary Material that is available online.

2. THE LOCAL MOMENT ESTIMATOR AND ASYMPTOTIC PROPERTIES

Let (X_i, Y_i), $i = 1, \ldots, n$, be independent copies of the random vector $(X, Y) \in \mathbb{R}^p \times \mathbb{R}_+$, where the conditional distribution of Y given $X = x$ satisfies (Eq. 1). The basic building block for our estimator is the statistic

$$T_n^{(t)}(x, K) := \frac{1}{n} \sum_{i=1}^n K_{h_n}(x - X_i)(\ln Y_i - \ln \omega_n)_+^t \mathbf{1}\{Y_i > \omega_n\}, t = 0, 1, \ldots, 6, \quad (2)$$

where $K_{h_n}(x) := K(x/h_n)/h_n^p$, K is a joint density on $\mathbb{R}^p$, h_n is a positive non-random sequence satisfying $h_n \to 0$ as $n \to \infty$, $(x)_+ := \max(0, x)$, $\mathbf{1}\{A\}$ denotes the indicator function of the event A and ω_n is a local non-random threshold sequence satisfying $\omega_n \to y^*(x)$ for $n \to \infty$, where $y^*(x)$ indicates the right endpoint of $F(y; x)$ defined as $y^*(x) := \sup\{y : F(y; x) < 1\}$. This statistic was introduced and studied by Goegebeur,

Guillou, & Schorgen (2013) in the framework of conditional Pareto-type tails, and will also serve as our basic building block in the creation of the local moment estimator.

In order to obtain the limiting distribution of the statistic (Eq. 2) we need to impose a second order condition on the tail behaviour of $F(y; x)$, specifying the rate of convergence in (Eq. 1). For more details about second order conditions we refer to e.g. Bingham, Goldie, & Teugels (1987) and de Haan & Stadtmüller (1996).

Assumption $(\mathcal{R})$ *There exists constants* $\gamma(x) \in \mathbb{R}$ *and* $\rho(x) \leq 0$, *a positive rate function* $a(.;x)$ *and a function* $A(.;x)$ *not changing sign ultimately, with* $A(t;x) \to 0$ *for* $t \to \infty$, *such that*

$$\lim_{t \to \infty} \frac{\frac{U(ty;x) - U(t;x)}{a(t;x)} - D_{\gamma(x)}(y)}{A(t;x)} = D_{\gamma(x),\rho(x)}(y), \qquad (3)$$

with

$$D_{\gamma(x),\rho(x)}(y) := \int_1^y s^{\gamma(x)-1} \int_1^s u^{\rho(x)-1} du ds$$

for all $y > 0$.

This second order condition is widely accepted in the extreme value literature, and is not very restrictive. Note that $(\mathcal{R})$ implies that $|A|$ is regularly varying with index $\rho(x)$, i.e. $\lim_{t \to \infty} \frac{A(ty;x)}{A(t;x)} = y^{\rho(x)}, \forall y > 0$, (see e.g. de Haan & Stadtmüller, 1996).

Since our main statistic (Eq. 2) is expressed in terms of log-excesses, we need a reformulation of (Eq. 3) in terms of $\ln U(y; x)$. To this aim we introduce some notation. We refer to Fraga Alves et al. (2007) for more details. Let

$$\overline{A}(t;x) := \frac{a(t;x)}{U(t;x)} - \gamma_+(x), \quad \gamma_+(x) := \max(0, \gamma(x))$$

and

$$\ell(x) := \lim_{t \to \infty} \left(U(t;x) - \frac{a(t;x)}{\gamma(x)} \right) \in \mathbb{R} \quad \text{for } 0 < \gamma(x) < -\rho(x).$$

Then, according to Theorem 2.1 in Fraga Alves et al. (2007), we have that

$$\overline{A}(t;x) \to 0 \text{ and } \frac{\overline{A}(t;x)}{A(t;x)} \to c, \quad \text{for } t \to \infty,$$

if $\rho(x) \neq \gamma(x)$, where

$$c = \begin{cases} 0, & \gamma(x) < \rho(x) \leq 0 \\ \frac{\gamma(x)}{\gamma(x)+\rho(x)}, & 0 \leq -\rho(x) < \gamma(x) \text{ or } (0 < \gamma(x) < -\rho(x) \text{ and } \ell(x) = 0) \\ \pm\infty, & \gamma(x) + \rho(x) = 0 \text{ or } (0 < \gamma(x) < -\rho(x) \text{ and } \ell(x) \neq 0) \text{ or } \rho(x) < \gamma(x) \leq 0 \end{cases} .$$

With this notation we set

$$B(t;x) := \begin{cases} \overline{A}(t;x), & c = \pm\infty \\ \frac{\gamma(x)}{\gamma(x)+\rho(x)} A(t;x), & c = \frac{\gamma(x)}{\gamma(x)+\rho(x)} \\ A(t;x), & \text{otherwise} \end{cases} .$$

As a first step in the development of our local moment estimator we study the asymptotic behaviour of (Eq. 2) under Assumption $(\mathcal{R})$. To this aim, in Lemma 1, we establish the asymptotic expansion of the following conditional expectation

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

Lemma 1. *The conditional expectation $m(\omega_n, 0; x)$ can be explicitly computed as*

$$m(\omega_n, 0; x) = \overline{F}(\omega_n; x).$$

Assume now $(\mathcal{R})$ with $\rho(x) \neq \gamma(x)$. Then for $\omega_n \to y^*(x)$ we can expand $m(\omega_n, 1; x)$ as follows:

$$m(\omega_n, 1; x) = \begin{cases} \overline{F}(\omega_n; x)\left[\gamma(x) + \frac{1}{1-\widetilde{\rho}(x)} B\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)(1+o(1))\right], & \gamma(x) > 0 \\ \\ \overline{F}(\omega_n;x) \frac{a^*\left(\frac{1}{\overline{F}(\omega_n;x)};x\right)}{U\left(\frac{1}{\overline{F}(\omega_n;x)};x\right)}\left[\frac{1}{1-\gamma(x)} + b^{(1)}_{\gamma(x),\rho(x)} B\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)(1+o(1))\right], & \gamma(x) \leq 0 \end{cases},$$

where

$$\widetilde{\rho}(x) := \begin{cases} \rho(x), & c = \frac{\gamma(x)}{\gamma(x)+\rho(x)} \\ -\gamma(x), & c = \pm\infty \end{cases},$$

$$a^*(t;x) := \begin{cases} a(t;x)\left(1 - \frac{B(t;x)}{\gamma(x)}\right), & \gamma(x) < \rho(x) = 0 \\ a(t;x)\left(1 - \frac{B(t;x)}{\rho(x)}\right), & \gamma(x) < \rho(x) < 0 \\ a(t;x)\left(1 + \frac{2B(t;x)}{\gamma(x)}\right), & \rho(x) < \gamma(x) < 0 \\ a(t;x), & \rho(x) < \gamma(x) = 0 \end{cases},$$

and

$$b^{(1)}_{\gamma(x),\rho(x)} := \begin{cases} \frac{1}{\gamma(x)(1-\gamma(x))^2}, & \gamma(x) < \rho(x) = 0 \\ \frac{1}{\rho(x)(1-\gamma(x)-\rho(x))}, & \gamma(x) < \rho(x) < 0 \\ -\frac{2-3\gamma(x)}{\gamma(x)(1-\gamma(x))(1-2\gamma(x))}, & \rho(x) < \gamma(x) < 0 \\ -1, & \rho(x) < \gamma(x) = 0 \end{cases}.$$

Under the same assumptions, $m(\omega_n, 2; x)$ can also be expanded as:

$$m(\omega_n, 2; x) = \begin{cases} \overline{F}(\omega_n; x)\left[2\gamma^2(x) + \frac{2\gamma(x)(2-\widetilde{\rho}(x))}{(1-\widetilde{\rho}(x))^2}B\left(\frac{1}{\overline{F}(\omega_n; x)}; x\right)(1+o(1))\right], & \gamma(x) > 0 \\[3em] \overline{F}(\omega_n; x)\left(\frac{a^*\left(\frac{1}{\overline{F}(\omega_n; x)}; x\right)}{U\left(\frac{1}{\overline{F}(\omega_n; x)}; x\right)}\right)^2\left[\frac{2}{(1-\gamma(x))(1-2\gamma(x))} + b^{(2)}_{\gamma(x),\rho(x)}B\left(\frac{1}{\overline{F}(\omega_n; x)}; x\right)(1+o(1))\right], & \gamma(x) \le 0 \end{cases},$$

where

$$b^{(2)}_{\gamma(x),\rho(x)} := \begin{cases} \frac{2(2-3\gamma(x))}{\gamma(x)(1-\gamma(x))^2(1-2\gamma(x))^2}, & \gamma(x) < \rho(x) = 0 \\ \frac{2(2-2\gamma(x)-\rho(x))}{\rho(x)(1-\gamma(x))(1-\gamma(x)-\rho(x))(1-2\gamma(x)-\rho(x))}, & \gamma(x) < \rho(x) < 0 \\ -\frac{8-18\gamma(x)}{\gamma(x)(1-\gamma(x))(1-2\gamma(x))(1-3\gamma(x))}, & \rho(x) < \gamma(x) < 0 \\ -6, & \rho(x) < \gamma(x) = 0 \end{cases}.$$

We now assume that the random vector X has density $g(x)$ for $x \in \mathbb{R}^p$. This density function is assumed to follow a Hölder condition. Let $\|\cdot\|$ be the Euclidean norm on $\mathbb{R}^p$.

Assumption $(\mathcal{G})$ *There exists $c_g > 0$ and $\eta_g > 0$ such that $|g(x) - g(z)| \le c_g\|x - z\|^{\eta_g}$ for all $x, z \in \mathbb{R}^p$.*

Concerning the kernel function K we introduce the following condition which is standard in local estimation. It is also used in Daouia, Gardes, & Girard (2013) and Goegebeur, Guillou, & Schorgen (2013).

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

Besides Assumption $(\mathcal{R})$, which describes the tail behaviour of $U(.;x)$, we also need a condition to control the oscillation of $U(y;x)$ when considered as a function of the covariate x. This condition is formulated in terms of the conditional expectation $m(\omega_n, t; x)$ and implies in particular the continuity of $\gamma(.)$.

Assumption $(\mathcal{F})$ *The conditional expectation $m(\omega_n, t; x)$ satisfies that, for $\omega_n \to y^*(x)$, $h_n \to 0$,*

$$\Phi(\omega_n, h_n; x) := \sup_{t\in\{0,1,\ldots,6\}}\sup_{z\in\Omega}\left|\frac{m(\omega_n, t; x - zh_n)}{m(\omega_n, t; x)} - 1\right| \to 0 \text{ if } n \to \infty.$$

To deal with the randomness in X we consider now the unconditional expectation

$$\widetilde{m}_n(K, t; x) := \mathbb{E}\left[K_{h_n}(x - X)(\ln Y - \ln \omega_n)_+^t \mathbf{1}\{Y > \omega_n\}\right],$$

which corresponds in fact with the expectation of $T_n^{(t)}(x, K)$, since the summands in (Eq. 2) are independent and identically distributed random variables. Lemma 2 states the asymptotic expansion of $\widetilde{m}_n(K, t; x)$.

Lemma 2. *Assume $(\mathcal{R})$ with $\rho(x) \neq \gamma(x)$, $(\mathcal{G})$, $(\mathcal{K})$ and $(\mathcal{F})$. For all $x \in \mathbb{R}^p$ where $g(x) > 0$ we have for $n \to \infty$ with $h_n \to 0$ and $\omega_n \to y^*(x)$,*

$$\widetilde{m}_n(K, t; x) = m(\omega_n, t; x) g(x) \left(1 + O(h_n^{\eta_g}) + O(\Phi(\omega_n, h_n; x))\right).$$

Note that in the case where $t = 0$, the result of Lemma 2 can in fact be obtained without assuming $(\mathcal{R})$.

As our next step in the construction of an estimator for the extreme value index in a regression context, we need to establish the asymptotic normality of a vector of appropriately normalized statistics of the form (Eq. 2). This is done in Theorem 1. Inspired by Lemmas 1 and 2, we define

$$\widetilde{T}_n^{(t)}(x, K) := \begin{cases} \frac{1}{\overline{F}(\omega_n; x)g(x)} T_n^{(t)}(x, K), & \gamma(x) > 0 \\[2em] \left(\frac{U\left(\frac{1}{\overline{F}(\omega_n; x)}; x\right)}{a^*\left(\frac{1}{\overline{F}(\omega_n; x)}; x\right)}\right)^t \frac{1}{\overline{F}(\omega_n; x)g(x)} T_n^{(t)}(x, K), & \gamma(x) \leq 0 \end{cases}$$

and

$$\mathbb{T}_n' := \left[\widetilde{T}_n^{(0)}(x, K_0), \widetilde{T}_n^{(1)}(x, K_1), \widetilde{T}_n^{(2)}(x, K_2)\right].$$

Theorem 1. *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent copies of the random vector (X, Y), and let g denote the density function of X. Assume $(\mathcal{R})$ with $\rho(x) \neq \gamma(x)$, $(\mathcal{G})$ and $(\mathcal{F})$ are satisfied and that the kernel functions K_0, K_1 and K_2 satisfy $(\mathcal{K})$. For all $x \in \mathbb{R}^p$ where $g(x) > 0$ we have that if $h_n \to 0$, $\omega_n \to y^*(x)$ and $n h_n^p \overline{F}(\omega_n; x) \to \infty$ for $n \to \infty$, then*

$$\sqrt{n h_n^p \overline{F}(\omega_n; x)}\left[\mathbb{T}_n - \mathbb{E}(\mathbb{T}_n)\right] \xrightarrow{D} N_3(0, \Sigma),$$

where the elements of Σ are given by

$$\Sigma_{j,k} := \begin{cases} \frac{(j+k)!\|K_j K_k\|_1 \gamma^{j+k}(x)}{g(x)}, & \gamma(x) > 0 \\[1.5em] \frac{(j+k)!\|K_j K_k\|_1}{g(x)\Pi_{i=0}^{j+k}(1 - i\gamma(x))}, & \gamma(x) \leq 0 \end{cases}, \quad j, k = 0, 1, 2.$$

Looking at the moment estimator from the univariate framework (Dekkers, Einmahl, & de Haan, 1989), and the results we have obtained so far, we introduce our local moment estimator

$$\widehat{\gamma}_n(x) := \frac{T_n^{(1)}(x, K_1)}{T_n^{(0)}(x, K_0)} + 1 - \frac{1}{2}\left(1 - \frac{\left(\frac{T_n^{(1)}(x, K_1)}{T_n^{(0)}(x, K_0)}\right)^2}{\frac{T_n^{(2)}(x, K_2)}{T_n^{(0)}(x, K_0)}}\right)^{-1}, \tag{4}$$

for kernel functions K_0, K_1 and K_2 satisfying our Assumption $(\mathcal{K})$. Indeed, using the leading terms of the asymptotic expansions for $\mathbb{E}(T_n^{(j)}(x, K_j))$, $j = 0, 1, 2$, as given by Lemmas 1 and 2, one easily motivates intuitively that $\widehat{\gamma}_n(x)$ is an estimator for $\gamma(x)$. We focus here on an adjustment of the moment estimator, but a similar idea could as well have been applied to other estimators for $\gamma(x) \in \mathbb{R}$, like e.g. the probability weighted moment estimator (Hosking & Wallis, 1987) and the mixed moment estimator (Fraga Alves et al., 2009). Using the result of Theorem 1 we can now obtain the limiting distribution of (Eq. 4), when properly normalized.

Theorem 2. *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent copies of the random vector (X, Y), and let g denote the density function of X. Assume $(\mathcal{R})$ with $\rho(x) \neq \gamma(x)$, $(\mathcal{G})$ and $(\mathcal{F})$ are satisfied and that the kernel functions K_0, K_1 and K_2 satisfy $(\mathcal{K})$. For all $x \in \mathbb{R}^p$ where $g(x) > 0$ we have that if $h_n \to 0$, $\omega_n \to y^*(x)$ and $nh_n^p\overline{F}(\omega_n; x) \to \infty$ for $n \to \infty$ with $\sqrt{nh_n^p\overline{F}(\omega_n; x)}B\left(\frac{1}{\overline{F}(\omega_n; x)}; x\right) \to \lambda(x)$ for some constant $\lambda(x) \in \mathbb{R}$, $nh_n^{p+2\eta_g}\overline{F}(\omega_n; x) \to 0$ and $nh_n^p\overline{F}(\omega_n; x)\Phi^2(\omega_n, h_n; x) \to 0$, then*

$$\sqrt{nh_n^p\overline{F}(\omega_n; x)}\left(\widehat{\gamma}_n(x) - \gamma(x)\right) \xrightarrow{D} N\left(\lambda(x)\mu, V'\Sigma V\right),$$

where

$$\mu = \begin{cases} \frac{\gamma(x)(1-\widetilde{\rho}(x))+\widetilde{\rho}(x)}{\gamma(x)(1-\widetilde{\rho}(x))^2}, & \gamma(x) > 0 \\ \\ (1 - 2\gamma(x))(1 - \gamma(x))^2\left[-2b^{(1)}_{\gamma(x),\rho(x)} + \frac{1}{2}(1 - 2\gamma(x))b^{(2)}_{\gamma(x),\rho(x)}\right] + \frac{\mathbf{1}\{\rho(x)<\gamma(x)\leq 0\}}{1-\gamma(x)}, & \gamma(x) \leq 0 \end{cases}$$

and

$$V := \begin{bmatrix} 1 - \gamma(x) \\ 1 - \frac{2}{\gamma(x)} \\ \frac{1}{2\gamma^2(x)} \end{bmatrix}$$

when $\gamma(x) > 0$, while

$$V := \begin{bmatrix} (1 - \gamma(x))(1 - 2\gamma(x)) \\ -2(1 - \gamma(x))^2(1 - 2\gamma(x)) \\ \frac{1}{2}(1 - \gamma(x))^2(1 - 2\gamma(x))^2 \end{bmatrix}$$

when $\gamma(x) \leq 0$.

The bias component μ in Theorem 2 can be calculated as

$$\mu = \begin{cases} \frac{(1-\gamma(x))(1-2\gamma(x))}{(1-\gamma(x)-\rho(x))(1-2\gamma(x)-\rho(x))} & \gamma(x) < \rho(x) \leq 0 \\ -\frac{\gamma(x)(1+\gamma(x))}{(1-\gamma(x))(1-3\gamma(x))} & \rho(x) < \gamma(x) \leq 0 \\ \frac{\gamma(x)(1-\rho(x))+\rho(x)}{\gamma(x)(1-\rho(x))^2} & 0 \leq -\rho(x) < \gamma(x) \text{ or } (0 < \gamma(x) < -\rho(x) \text{ and } \ell(x) = 0) \\ \frac{\gamma(x)}{(1+\gamma(x))^2} & \gamma(x) + \rho(x) = 0 \text{ or } (0 < \gamma(x) < -\rho(x) \text{ and } \ell(x) \neq 0) \end{cases},$$

and thus it corresponds with the bias of the moment estimator in the univariate context, as given in Theorem 3.5.4 in de Haan & Ferreira (2006), if one takes the slightly different definition of the function Q there into account. For the special case where the three kernel functions are equal, the asymptotic variance expression simplifies and is given by

$$V'\Sigma V = \begin{cases} \frac{\|K\|_2^2(1+\gamma^2(x))}{g(x)} & \gamma(x) > 0 \\ \frac{\|K\|_2^2(1-\gamma(x))^2(1-2\gamma(x))(1-\gamma(x)+6\gamma^2(x))}{g(x)(1-3\gamma(x))(1-4\gamma(x))} & \gamma(x) \leq 0 \end{cases},$$

where K denotes the common kernel function. Note that apart from the scaling factor $\|K\|_2^2/g(x)$ this variance expression coincides with the asymptotic variance of the moment estimator. As

expected, the asymptotic variance of our estimator is inversely proportional to $g(x)$.

3. SIMULATION EXPERIMENT

In the simulation experiment we compare our local moment estimator with several others estimators proposed in the literature: Daouia, Gardes, & Girard (2013) and Stupfler (2013) which both are valid in the complete max-domain of attraction, Goegebeur, Guillou, & Schorgen (2013) which is only valid in the Fréchet case ($\gamma(x) > 0$) and two methods proposed in the parametric framework, Wang & Tsai (2009) and Wang & Li (2013), also only valid in the Fréchet case. We briefly describe all these methods below.

3.1. Benchmark estimators proposed in the literature
- Daouia, Gardes, & Girard's (2013) estimator is valid in the complete max-domain of attraction. This class of estimators for $\gamma(x)$ is defined by

$$\widetilde{\gamma}_J^{RP}(x) := \frac{1}{\ln r} \sum_{j=1}^{J-2} \pi_j \ln \left(\frac{\widehat{q}_n\left(\tau_j \alpha_n; x\right) - \widehat{q}_n\left(\tau_{j+1} \alpha_n; x\right)}{\widehat{q}_n\left(\tau_{j+1} \alpha_n; x\right) - \widehat{q}_n\left(\tau_{j+2} \alpha_n; x\right)} \right),$$

with $J \geq 3$, $r, \alpha_n \in (0,1)$, and for $j = 1, \ldots, J$, $\tau_j = r^{j-1}$, π_j a sequence of weights summing to one, and

$$\widehat{q}_n(\alpha; x) := \inf\left\{ y : \widehat{\overline{F}}_n(y; x) \leq \alpha \right\},$$

with

$$\widehat{\overline{F}}_n(y; x) := \frac{\sum_{i=1}^n K_{h_n}(x - X_i) \mathbf{1}\{Y_i > y\}}{\sum_{i=1}^n K_{h_n}(x - X_i)}.$$

In particular they propose using $J = 3$ or $J = 4$ with $r = \frac{1}{J}$, while the weights π_j are chosen as constant weights $\pi_1 = \ldots = \pi_{J-2} = \frac{1}{J-2}$ or linear weights $\pi_j = \frac{2j}{(J-1)(J-2)}$ for $j = 1, \ldots, J-2$. With these settings we get the two estimators (see Daouia, Gardes, & Girard, 2013, p. 2563)

$$\widetilde{\gamma}_J^{RP,1}(x) := \frac{1}{(J-2)\ln r} \ln \left(\frac{\widehat{q}_n\left(\tau_1 \alpha_n; x\right) - \widehat{q}_n\left(\tau_2 \alpha_n; x\right)}{\widehat{q}_n\left(\tau_{J-1} \alpha_n; x\right) - \widehat{q}_n\left(\tau_J \alpha_n; x\right)} \right)$$

and

$$\widetilde{\gamma}_J^{RP,2}(x) := \frac{2}{(J-1)(J-2)\ln r} \sum_{j=1}^{J-2} \ln \left(\frac{\widehat{q}_n\left(\tau_j \alpha_n; x\right) - \widehat{q}_n\left(\tau_{j+1} \alpha_n; x\right)}{\widehat{q}_n\left(\tau_{J-1} \alpha_n; x\right) - \widehat{q}_n\left(\tau_J \alpha_n; x\right)} \right).$$

Note that for $J = 3$, the two estimators are identical.
- Stupfler's (2013) estimator, also valid in the complete max-domain of attraction, is given by

$$\widehat{\gamma}_S(x) := M_n^{(1)}(x, k_x, h_n) + 1 - \frac{1}{2}\left(1 - \frac{\left[M_n^{(1)}(x, k_x, h_n)\right]^2}{M_n^{(2)}(x, k_x, h_n)} \right)^{-1},$$

with

$$M_n^{(j)}(x, k_x, h_n) := \frac{1}{k_x} \sum_{i=1}^{k_x} \left[\log \left(Z_{n^*-i+1,n^*} \right) - \log \left(Z_{n^*-k_x,n^*} \right) \right]^j, \quad j = 1, 2,$$

where n^* is the number of observations in the ball $B(x, h_n)$, $k_x \in \{1, \ldots, n^* - 1\}$ and $Z_{1,n^*} \leq \ldots \leq Z_{n^*,n^*}$ are the order statistics associated to the response variables for which the corresponding covariates belong to this ball.

- Goegebeur, Guillou, & Schorgen's (2013) estimator is valid only in the Fréchet domain of attraction. It is defined as

$$\widehat{\gamma}_{GGS}(x) := \frac{T_n^{(1)}(x, K)}{T_n^{(0)}(x, K)}.$$

Also the following bias-corrected version is introduced

$$\widehat{\gamma}_{GGS}^{BC}(x) := \frac{1}{\widehat{\rho}(x)} \frac{T_n^{(1)}(x, K)}{T_n^{(0)}(x, K)} + \left(1 - \frac{1}{\widehat{\rho}(x)} \right) \frac{T_n^{(2)}(x, K)}{2 T_n^{(1)}(x, K)},$$

where $\widehat{\rho}(x)$ is defined as follows:

$$\widehat{\rho}(x) := 3 \frac{R_n(x, K) - 1}{R_n(x, K) - 3} \quad \text{if } 1 \leq R_n(x, K) < 3$$

where

$$R_n(x, K) = \frac{\left(\frac{T_n^{(1)}(x,K)}{T_n^{(0)}(x,K)} \right)^{\frac{1}{2}} - \left(\frac{T_n^{(2)}(x,K)}{2 T_n^{(0)}(x,K)} \right)^{\frac{1}{4}}}{\left(\frac{T_n^{(2)}(x,K)}{2 T_n^{(0)}(x,K)} \right)^{\frac{1}{4}} - \left(\frac{T_n^{(3)}(x,K)}{6 T_n^{(0)}(x,K)} \right)^{\frac{1}{6}}}.$$

This estimator is consistent for $\rho(x)$ according to Theorem 6 in Goegebeur, Guillou, & Schorgen (2013). In case $R_n(x, K) \notin [1, 3)$, $\widehat{\rho}(x)$ is set to the value -1.

- Wang & Li's (2013) estimator, $\widehat{\gamma}_{WL}(x)$, valid in the Fréchet domain of attraction, and defined in a parametric setting where the conditional quantile function of the response variable is assumed to be of the form

$$F^{\leftarrow}(\tau; x) = \Lambda_\lambda^{\leftarrow} \left(\alpha(\tau) + x^T \theta(\tau) \right), \quad \tau \in [1 - \varepsilon, 1],$$

with ε some small constant, $\alpha(\tau) \in \mathbb{R}$, $\theta(\tau) \in \mathbb{R}^p$, and where $\Lambda_\lambda(\cdot)$ denotes the family of power transformations

$$\Lambda_\lambda(y) = \begin{cases} \ln(y) & \lambda = 0 \\ \frac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \end{cases}.$$

- Wang & Tsai's (2009) estimator, $\widehat{\gamma}_{WT}(x)$, defined in the parametric setting where $\gamma(x) = \exp \left(-x^T \theta \right)$ with θ a $\mathbb{R}^p-$ vector of regression coefficients.

3.2. Practical implementation of the estimators

For our local moment estimator, we have to choose the bandwidth h_n and the threshold ω_n, where we for the latter use the $(k + 1)$-th largest response in the ball $B(x, h_n)$. There is, in fact,

some discrepancy between the theory, established in case of a fixed, i.e. non-random, threshold and practical use with a data-driven one. But this is usual in extreme value statistics, see for instance, Smith (1987) and Davison & Smith (1990), in the framework of GPD modelling of excesses, or Wang & Tsai (2009) and Goegebeur, Guillou, & Schorgen (2013), for Pareto-type models. A similar issue is in fact present when selecting the bandwidth parameter in classical nonparametric density estimation and regression where a non-random bandwidth is used in theoretical deviations, though in practice a data-based one is considered (at least one looks at the range of the data in order to get an idea about plausible bandwidth values). In all cases the kernel functions are chosen as the biquadratic kernel function

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

except one time where it will be the triquadratic function $\widetilde{K}(x) = \frac{35}{32} \left(1 - x^2\right)^3 \mathbf{1}\{x \in [-1, 1]\}$. Selection of (h_n, k) is done using a completely data-driven method, which selects h_n and k in two steps. First, the bandwidth parameter h_n is selected using a cross validation criterion introduced by Yao (1999), implemented by Gannoun et al. (2002) and considered in an extreme value context by Daouia et al. (2011) or Daouia, Gardes, & Girard (2013). It proceeds by selecting h_n by

$$h_c := \operatorname*{argmin}_{h_n \in \mathcal{H}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\mathbf{1} \left\{ Y_i \leq Y_j \right\} - \widehat{F}_{n,-i} \left(Y_j; X_i \right) \right)^2, \tag{5}$$

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

$$\widehat{F}_{n,-i} \left(y; x \right) := \frac{\sum_{k=1, k \neq i}^{n} K_{h_n} \left(x - X_k \right) \mathbf{1} \left\{ Y_k \leq y \right\}}{\sum_{k=1, k \neq i}^{n} K_{h_n} \left(x - X_k \right)}.$$

Next for each z_ℓ, where $z_1, \ldots, z_L$ are points regularly spaced in the covariate space, we do the following

- Compute $\widehat{\gamma}_n(z_\ell)$ for $k = 5, 6, \ldots, k_{\max}$, where $k_{\max}$ is chosen appropriately.
- Split the estimates $\widehat{\gamma}_n(z_\ell)$ into blocks of size $\lfloor \sqrt{k_{\max}} \rfloor$.
- For each block we compute the standard deviation of the estimates for $\gamma(z_\ell)$. For the block with the smallest standard deviation, we select the median of the estimates.

This methodology has been used with $\mathcal{H} = \{0.05, 0.075, \ldots, 0.3\}$ for our estimator, Daouia, Gardes, & Girard (2013) and Goegebeur, Guillou, & Schorgen (2013) ones. For our local moment estimator $k_{\max} = \lfloor n^*/2 \rfloor$, while $k_{\max} = n^* - 1$ for the Goegebeur, Guillou, & Schorgen (2013) estimators and $\widetilde{\gamma}_J^{RP}(x)$ is computed for $\alpha_n = k/n^*$ where $k = 5, \ldots, n^* - 1$ for the Daouia, Gardes, & Girard (2013) estimators. The remaining benchmarks estimators are all implemented with the approach described in the papers where they were published.

We compare the estimators on five conditional distributions of Y given $X = x$, at least when those can be applied. The conditional distributions we consider are

- The $reversed\, Burr(\eta(x), \tau(x), \lambda(x))$ distribution, left-truncated at 0 and with right endpoint $y^*(x)$,

$$F(y; x) = 1 - \left(\frac{\eta(x) + y^*(x)^{-\tau}}{\eta(x) + (y^*(x) - y)^{-\tau(x)}} \right)^{\lambda(x)}, \quad 0 < y < y^*(x); \lambda(x), \eta(x), \tau(x) > 0,$$

for which $\gamma(x) = -1/(\lambda(x)\tau(x))$ and $\rho(x) = -1/\lambda(x)$. Here we always use $\eta(x) = 3$ and $y^*(x) = 5$, while we consider the cases

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

with $\lambda(x)$ fixed at the values $\lambda(x) = 0.5, 1, 1.5$, or $\tau(x)$ fixed at the values $\tau(x) = 0.5, 1.5, 2, 2.5$, and the case

$$\gamma(x) = -\frac{1}{4}\left(\frac{1}{10} + \sin(\pi x)\right)\left(\frac{11}{10} - \frac{1}{2}\exp\left(-64\left(x - \frac{1}{2}\right)^2\right)\right),$$

with $\eta(x) = 3$ and $\lambda(x) = 2$. Note that the function $\gamma(x)$ is chosen differently in the last case, in order to satisfy the requirement $\gamma(x) \neq \rho(x)$.

- The $strict\,Weibull(\lambda(x), \tau(x))$ distribution,

$$1 - F(y; x) = e^{-\lambda(x)y^{\tau(x)}}, \quad y > 0; \lambda(x), \tau(x) > 0,$$

for which $\gamma(x) = 0$ and $\rho(x) = 0$. Note that this distribution does not fit in to our framework since $\gamma(x) = \rho(x)$, but we include it to see how our estimator performs when the assumptions are violated. We consider the cases

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

with $\tau(x)$ fixed at some constant, and

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

with $\lambda(x)$ fixed at some constant.

- The $Burr(\eta(x), \tau(x), \lambda(x))$ distribution,

$$1 - F(y; x) = \left(\frac{\eta(x)}{\eta(x) + y^{\tau(x)}}\right)^{\lambda(x)}, \quad y > 0; \eta(x), \tau(x), \lambda(x) > 0,$$

for which $\gamma(x) = 1/(\lambda(x)\tau(x))$ and $\rho(x) = -1/\lambda(x)$. We consider the case

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

with $\eta(x) = 1$ and different but fixed values of $\lambda(x)$.

The distribution of X in all these three first cases is chosen as the $Unif(0, 1)$ distribution. We consider also two cases where the dimension of the covariate space is 2. In these settings, X is uniformly distributed on $[-1, 1] \times [-1, 1]$ and we assume that the framework of Wang & Tsai (2009) or Wang & Li (2013) can be applied. More specifically, we consider the two additional cases:

- A distribution with conditional quantile function given by

$$F^{\leftarrow}(\tau; x) = \exp\left(\alpha(\tau) + x^T \theta(\tau)\right),$$

  where

$$\alpha(\tau) = 2 + \frac{1}{2}\left(\tau - 1 - \ln(1 - \tau)\right)$$

  and

$$\theta(\tau) = \begin{pmatrix} 1 + \frac{1}{4}\left(\tau - 1 - \ln(1 - \tau)\right) \\ 1 \end{pmatrix}.$$

  In that case, $\gamma(x) = \frac{1}{2} + x^T\sigma$ where $\sigma = (0.25, 0)^T$ and $\rho(x) = -1$.

- The $Burr\left(1, \frac{2}{\gamma(x)}, \frac{1}{2}\right)$ distribution with

$$\gamma(x) = \exp(-x^T \theta),$$

  where $\theta = (0.1, 0.2)^T$.

For all distributions we simulate $N = 500$ samples of size $n = 1000$ and as measures of efficiency, we compute the absolute bias

$$\text{Bias}\left(\widehat{\gamma}(\cdot)\right) = \frac{1}{L}\sum_{\ell=1}^{L}\left|\frac{1}{N}\sum_{j=1}^{N}\left(\widehat{\gamma}_j\left(z_\ell\right) - \gamma\left(z_\ell\right)\right)\right|$$

and the mean squared error

$$\text{MSE}\left(\widehat{\gamma}(\cdot)\right) = \frac{1}{LN}\sum_{\ell=1}^{L}\sum_{j=1}^{N}\left(\widehat{\gamma}_j\left(z_\ell\right) - \gamma\left(z_\ell\right)\right)^2$$

with the z_ℓ's being $L = 41$ points equidistantly spaced in $[0.1, 0.9]$ and $\widehat{\gamma}_j\left(z_\ell\right)$ being the estimate from simulation run j at covariate value z_ℓ.

3.3. The results

First, we compare in Figure 1 the asymptotic variance of the local moment estimator with the asymptotic variance of $\widetilde{\gamma}_J^{RP,1}(x)$ and $\widetilde{\gamma}_J^{RP,2}(x)$ for $J = 3, 4$, in the case where all kernel functions are chosen to be equal. In the plot we have fixed $\|K\|_2^2/g(x) = 1$, since this term appears in the expression of the asymptotic variance for all the estimators. It appears that for $\gamma(x) > -0.6$ the local moment estimator has the smallest asymptotic variance, while $\widetilde{\gamma}_4^{RP,1}(x)$ is best for $-2 < \gamma(x) < -0.6$ and this estimator also seems to have the smallest asymptotic variance of the benchmark estimators over a wide range of values for $\gamma(x)$.

Then, we compare our local moment estimator with all the benchmark methods when those can be applied. These results are summarized in Tables 1-7, from which we can draw the following conclusions:

- Among the estimators proposed by Daouia, Gardes, & Girard (2013) no estimator performs uniformly best in terms of bias and MSE, though $\widetilde{\gamma}_4^{RP,1}(x)$ seems to have an overall good performance;
- In case of the Reversed Burr distribution, our estimator is almost all the time the best one compared to the one in Stupfler (2013), and Daouia, Gardes, & Girard (2013) (see Tables 1 and 2);
- For the Strict Weibull distribution, which does not satisfy the assumptions of our theorems, our estimator is still competitive with the Daouia, Gardes, & Girard (2013) or Stupfler (2013) ones. It thus seems to be robust with respect to a violation of the assumptions (see Tables 3 and 4);
- In case of the Burr distribution, our estimator outperforms the one of Stupfler (2013) in terms of bias, whereas a reverse conclusion applies in terms of MSE. Compared to Daouia, Gardes, & Girard (2013), our estimator is far superior, but this can be expected from the asymptotic variance. For such a distribution, we can also apply the estimator in Goegebeur, Guillou, & Schorgen (2013) which is the best one if we use the bias-corrected version (see Table 5) but this can be easily explained by the fact that this estimator can only be applied in this context of $\gamma(x) > 0$. Also as can be seen from this table, the choice of the kernel has little effect on the estimation of $\gamma(x)$ for our local moment estimator;
- In the specific case where Wang & Li (2013) can be applied, it is indeed the best estimator in terms of bias and MSE, but we can see in Table 6 that our estimator is also competitive, in particular in terms of MSE and it outperforms Stupfler's estimator in terms of bias and MSE. Remark also the bad behaviour of the bias-corrected estimator proposed by Goegebeur, Guillou, & Schorgen (2013) which can be explained by the fact that the estimate of the second order parameter $\rho(x)$ is close to zero, whereas a null true value of the parameter is excluded in their approach;
- Regarding the specific case where Wang & Tsai (2009) can be applied (see Table 7), their method is superior to the rest of the estimators, but again this can be expected since the parametric setting for which it has been developed is the suited one. Among the remaining estimators, our local moment estimator performs the best in terms of bias and MSE.

To conclude we can observe that our estimator is always very competitive. It can be outperformed by other estimators in particular settings where those are constructed to work well (optimal): for instance in some specific parametric frameworks (Wang & Tsai, 2009; Wang & Li, 2013), in the Fréchet domain of attraction (Goegebeur, Guillou, & Schorgen, 2013), but this is because these estimators are only valid in these contexts. However our estimator is very general, not restrictive to a specific domain of attraction nor a specific parametric setting and it often outperforms the Stupfler (2013) estimator and the Daouia, Gardes, & Girard (2013) one, which are also very general.

4. DATA ANALYSIS: LOCAL ESTIMATION OF THE SEISMIC MOMENT DISTRIBUTION

In this section we illustrate the practical applicability of the method for estimating the tail of the seismic energy distribution. Accurate modeling of the tail of the seismic moment distribution is clearly of crucial importance, since large earthquakes cause heavy losses. We use the Global Centroid Moment Tensor database, formerly known as the Harvard CMT catalog, that is accessible at `http://www.globalcmt.org/CMTsearch.html` (Dziewonski, Chou, & Woodhouse, 1981; Ekström, Nettles, & Dziewonski, 2012). This database contains information about, among others, longitude, latitude and seismic moment of earthquakes that have occurred

between 1976 and the present. The variable of main interest is the earthquake's seismic moment (measured in dyne-centimeters) and as covariate we use the location of the earthquake (given in latitude and longitude). We want to study the tail behaviour at a specific, fixed, location, but for the estimation of the conditional extreme value index we have to take into account that earthquakes happen at a random location. Hence, this dataset is very well suited for our local moment estimator. As the points in the covariate space where we want to estimate the extreme value index, we use locations where an earthquake has already happened. In order to determine the neighbourhood of these locations, we compute the distance in kilometres to every other earthquake position using the formula

$$d = R \cos^{-1} \left(\cos(\psi_1) \cos(\psi_2) \cos(\phi_1 - \phi_2) + \sin(\psi_1) \sin(\psi_2) \right),$$

which gives the spherical distance between two points with longitude and latitude (ϕ_1, ψ_1) and (ϕ_2, ψ_2), respectively (see e.g. Weisstein, 2003). Here it is assumed that the earth is a perfect sphere, with radius $R = 6371 km$ (mean radius of the earth). In the analysis we ignored isolated earthquakes, i.e. earthquakes for which there is no neighbouring earthquake within a radius of $200 km$. The bandwidth h_n is chosen by applying the cross validation criterion (Eq. 5) on a grid of $\mathcal{H} = \{200, 300, \ldots, 2000\}$ (measured in km). Here we use in all cases the biquadratic kernel function

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

This leads us to use a bandwidth of $h_n = 400 km$. Next, the threshold is selected locally in the same fashion as for the simulations. A plot of the local estimates of the extreme value index of the seismic energy distribution can be seen in Figure 2. For the mid-ocean ridges, the tail of the seismic moment distribution tends to be lighter than for the other areas. Here, we typically observe $\gamma(x) \le 1$, while $\gamma(x) > 1$ seems to be more common at other places. This was also observed by Pisarenko & Sornette (2003), Okal & Romanowich (1994) and Kagan (1997, 1999). From the analysis, the tail heaviness of the seismic moment distribution seems to be largest at the northern part of Japan, Indonesia, southern part of Mexico, and various places along the western coast of South America.

Previously, Pisarenko & Sornette (2003) have estimated the extreme value index of the seismic moment distribution with the generalized Pareto distribution, fitted to exceedances over a high threshold by means of the maximum likelihood method. To take the spatial differences of the seismic parameters into account they used the Flinn-Engdahl regionalization, which allowed them to identify 14 zones (out of 50 in the original Flinn-Engdahl regionalization), and all data within a zone were considered to be coming from the same distribution. Our approach is clearly more flexible since it is based on local estimation: for a given location we consider the data of all earthquakes that have occurred in a given radius to estimate the extreme value index, and these areas are much smaller than the zones considered in Pisarenko & Sornette (2003). Also, we do not need predefined zones, and the size of the neighbourhood is selected in an automatic data-driven way. Overall we could though say that the estimates are in line with those reported in Pisarenko & Sornette (2003).

5. CONCLUSION

We considered the estimation of the conditional extreme value index in the presence of random covariates. An adaptation of the moment estimator proposed by Dekkers, Einmahl, & de Haan (1989) was proposed. Its asymptotic properties were established and its finite sample perfor-
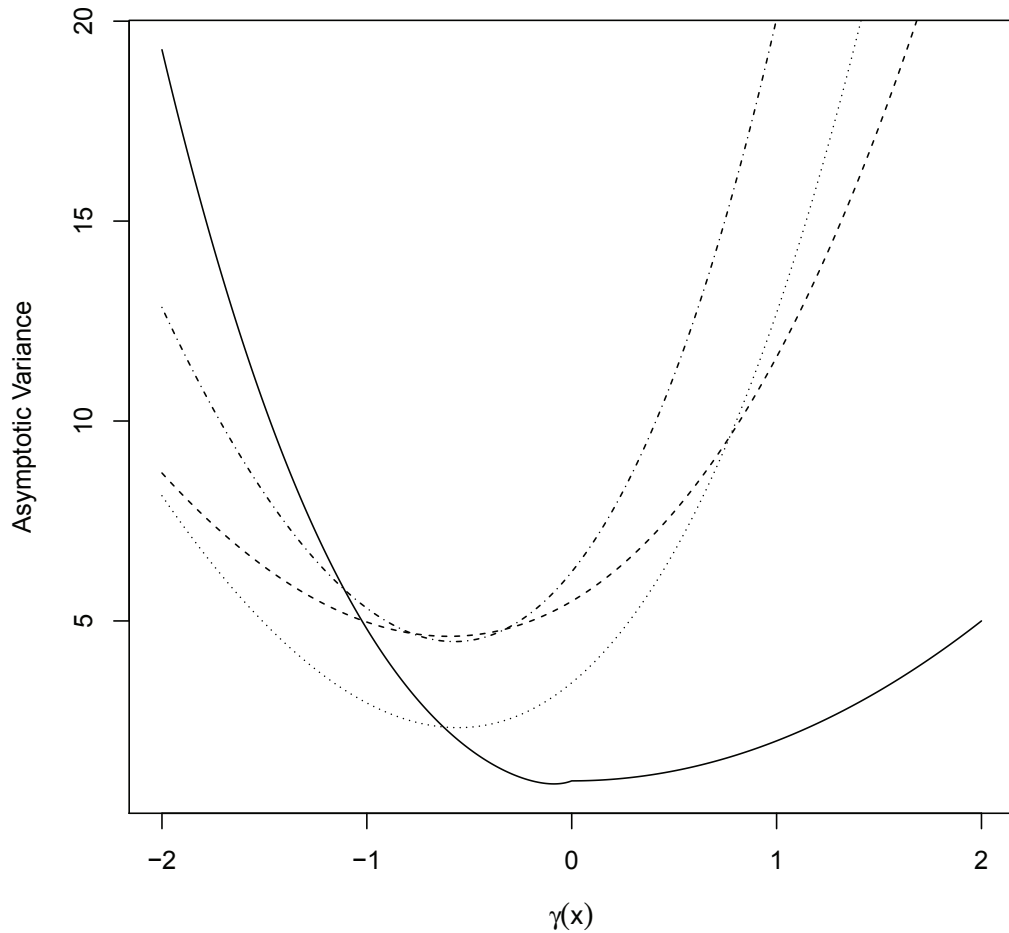
FIGURE 1: Comparison of four estimators from Section 3: Asymptotic variance of $\widehat{\gamma}_n(x)$ (solid), $\widetilde{\gamma}_3^{RP,1}(x)$ (dashed), $\widetilde{\gamma}_4^{RP,1}(x)$ (dotted) and $\widetilde{\gamma}_4^{RP,2}(x)$ (dashed-dotted).

mance illustrated on a simulation study and a real data analysis. Also a comparison with several other methods proposed in the literature has been provided. In future research we will focus on the development of estimators for high quantiles or the right endpoint (if finite).
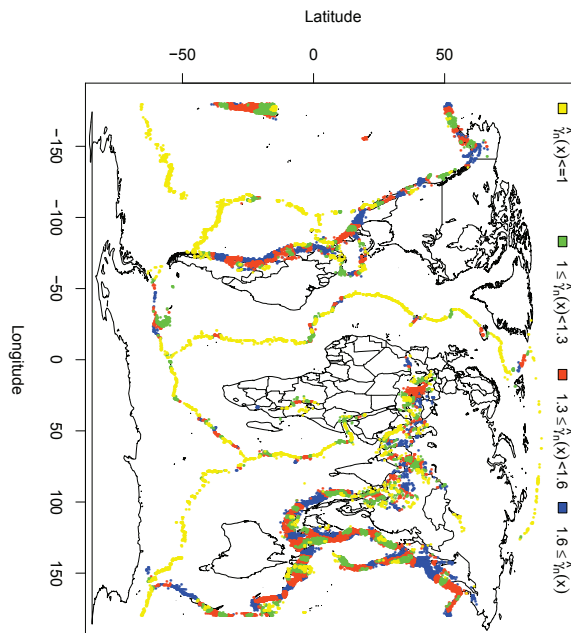
FIGURE 2: Estimation of tail index for earthquake data: Local estimates of $\gamma(x)$ at locations where earthquakes have been observed.

TABLE 1: The $reversed\ Burr(\eta(x), \tau(x), \lambda(x))$ distribution. Performance of $\widehat{\gamma}_n(x), \widehat{\gamma}_S(x), \widetilde{\gamma}_3^{RP,1}(x)$, $\widetilde{\gamma}_4^{RP,1}(x)$ and $\widetilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$\lambda(x)$	$\widehat{\gamma}_n(x)$		$\widehat{\gamma}_S(x)$		$\widetilde{\gamma}_3^{RP,1}(x)$		$\widetilde{\gamma}_4^{RP,1}(x)$		$\widetilde{\gamma}_4^{RP,2}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	0.0409	0.0399	0.1192	0.0357	0.2602	0.1463	0.1300	0.0619	0.1369	0.0959
1.0	0.1962	0.0936	0.2955	0.1238	0.5238	0.3757	0.2567	0.1108	0.2520	0.1319
1.5	0.3310	0.1829	0.4289	0.2348	0.7217	0.6467	0.3724	0.1855	0.3656	0.1971
2.0	0.4442	0.2849	0.5399	0.3578	0.8834	0.9290	0.4749	0.2743	0.4647	0.2751

TABLE 2: The $reversed\ Burr(\eta(x), \tau(x), \lambda(x))$ distribution. Performance of $\widehat{\gamma}_n(x), \widehat{\gamma}_S(x), \widetilde{\gamma}_3^{RP,1}(x)$, $\widetilde{\gamma}_4^{RP,1}(x)$ and $\widetilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$\tau(x)$	$\widehat{\gamma}_n(x)$		$\widehat{\gamma}_S(x)$		$\widetilde{\gamma}_3^{RP,1}(x)$		$\widetilde{\gamma}_4^{RP,1}(x)$		$\widetilde{\gamma}_4^{RP,2}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	0.7594	0.7451	1.0801	1.2953	1.3488	1.9372	0.8906	0.8467	0.8362	0.7661
1.5	0.3714	0.2352	0.5297	0.3684	0.8551	0.8844	0.4427	0.2592	0.4332	0.2718
2.0	0.2696	0.1500	0.3941	0.2207	0.6859	0.6011	0.3408	0.1750	0.3333	0.1945
2.5	0.1984	0.1033	0.3035	0.1441	0.5704	0.4417	0.2750	0.1310	0.2701	0.1549

TABLE 3: The $strict\,Weibull(\lambda(x), \tau(x))$ distribution. Performance of $\widehat{\gamma}_n(x)$, $\widehat{\gamma}_S(x)$, $\widetilde{\gamma}_3^{RP,1}(x)$, $\widetilde{\gamma}_4^{RP,1}(x)$ and $\widetilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$\lambda(x)$	$\widehat{\gamma}_n(x)$		$\widehat{\gamma}_S(x)$		$\widetilde{\gamma}_3^{RP,1}(x)$		$\widetilde{\gamma}_4^{RP,1}(x)$		$\widetilde{\gamma}_4^{RP,2}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	0.1593	0.0535	0.1185	0.0330	0.4521	0.2708	0.2628	0.1064	0.2558	0.1320
1.0	0.1771	0.0598	0.1446	0.0474	0.4949	0.3095	0.2796	0.1144	0.2709	0.1365
1.5	0.2017	0.0754	0.1646	0.0627	0.5373	0.3579	0.3174	0.1430	0.3083	0.1702
2.0	0.2173	0.0865	0.2177	0.0903	0.5468	0.3749	0.3411	0.1651	0.3364	0.2024

TABLE 4: The $strict\,Weibull(\lambda(x), \tau(x))$ distribution. Performance of $\widehat{\gamma}_n(x)$, $\widehat{\gamma}_S(x)$, $\widetilde{\gamma}_3^{RP,1}(x)$, $\widetilde{\gamma}_4^{RP,1}(x)$ and $\widetilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$\tau(x)$	$\widehat{\gamma}_n(x)$		$\widehat{\gamma}_S(x)$		$\widetilde{\gamma}_3^{RP,1}(x)$		$\widetilde{\gamma}_4^{RP,1}(x)$		$\widetilde{\gamma}_4^{RP,2}(x)$	
0.5	0.6728	0.4931	0.6818	0.5093	0.7640	0.6967	0.4213	0.2592	0.3635	0.2577
1.0	0.1168	0.0423	0.2112	0.0582	0.0178	0.0564	0.0363	0.0446	0.0597	0.0817
1.5	0.0731	0.0371	0.0730	0.0156	0.2933	0.1416	0.2131	0.0875	0.2204	0.1284
2.0	0.1680	0.0629	0.2057	0.0567	0.4363	0.2517	0.2953	0.1308	0.2958	0.1688

TABLE 5: The $Burr(\eta(x), \tau(x), \lambda(x))$ distribution. Performance of $\widehat{\gamma}_n(x)$, $\widehat{\gamma}_n^{\widetilde{K}}(x)$ where the triquadratic kernel is used, $\widehat{\gamma}_S(x)$, $\widetilde{\gamma}_3^{RP,1}(x)$, $\widetilde{\gamma}_4^{RP,1}(x)$, $\widetilde{\gamma}_4^{RP,2}(x)$, $\widehat{\gamma}_{GGS}(x)$ and $\widehat{\gamma}_{GGS}^{BC}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$\lambda(x)$	$\widehat{\gamma}_n(x)$		$\widehat{\gamma}_n^{\widetilde{K}}(x)$		$\widehat{\gamma}_S(x)$		$\widetilde{\gamma}_3^{RP,1}(x)$		$\widetilde{\gamma}_4^{RP,1}(x)$		$\widetilde{\gamma}_4^{RP,2}(x)$		$\widehat{\gamma}_{GGS}(x)$		$\widehat{\gamma}_{GGS}^{BC}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	0.0344	0.0194	0.0341	0.0189	0.0738	0.0135	0.1550	0.0978	0.0703	0.0705	0.0702	0.1286	0.0541	0.0065	0.0287	0.0084
1.0	0.0323	0.0201	0.0326	0.0195	0.0808	0.0166	0.2382	0.1243	0.1013	0.0665	0.0921	0.1138	0.1411	0.0278	0.0283	0.0072
1.5	0.0119	0.0221	0.0121	0.0217	0.0883	0.0199	0.2322	0.1185	0.1095	0.0676	0.1009	0.1190	0.2365	0.0731	0.0394	0.0093
2.0	0.0426	0.0283	0.0436	0.0278	0.1070	0.0273	0.1614	0.1013	0.0892	0.0710	0.0917	0.1237	0.3403	0.1486	0.0809	0.0163

TABLE 6: The parametric setting where the conditional quantile function of the response variable is of the form $F^{\leftarrow}(\tau; x) = \exp\left(\alpha(\tau) + x^T\theta(\tau)\right)$, for the specific values of $\alpha(\tau)$ and $\theta(\tau)$ given in Section 3.2. Comparison between $\widehat{\gamma}_n(x)$, $\widehat{\gamma}_S(x)$, $\widetilde{\gamma}_3^{RP,1}(x)$, $\widetilde{\gamma}_4^{RP,1}(x)$, $\widetilde{\gamma}_4^{RP,2}(x)$, $\widehat{\gamma}_{GGS}(x)$, $\widehat{\gamma}_{GGS}^{BC}(x)$ and $\widehat{\gamma}_{WL}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

	Bias	MSE
$\widehat{\gamma}_n(x)$	0.1468	0.0636
$\widehat{\gamma}_S(x)$	0.2371	0.0970
$\widetilde{\gamma}_3^{RP,1}(x)$	0.1020	0.1179
$\widetilde{\gamma}_4^{RP,1}(x)$	0.1033	0.1169
$\widetilde{\gamma}_4^{RP,2}(x)$	0.1033	0.1785
$\widehat{\gamma}_{GGS}(x)$	0.638	0.7049
$\widehat{\gamma}_{GGS}^{BC}(x)$	1.0037	24954.332
$\widehat{\gamma}_{WL}(x)$	0.066	0.0175

TABLE 7: The parametric setting of the $Burr\left(1, \frac{2}{\gamma(x)}, \frac{1}{2}\right)$ distribution with $\gamma(x) = \exp(-x^T\theta)$, where $\theta = (0.1, 0.2)^T$. Comparison between $\widehat{\gamma}_n(x)$, $\widehat{\gamma}_S(x)$, $\widetilde{\gamma}_3^{RP,1}(x)$, $\widetilde{\gamma}_4^{RP,1}(x)$, $\widetilde{\gamma}_4^{RP,2}(x)$, $\widehat{\gamma}_{GGS}(x)$, $\widehat{\gamma}_{GGS}^{BC}(x)$ and $\widehat{\gamma}_{WT}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

	Bias	MSE
$\widehat{\gamma}_n(x)$	0.0318	0.0244
$\widehat{\gamma}_S(x)$	0.0933	0.0320
$\widetilde{\gamma}_3^{RP,1}(x)$	0.1127	0.1011
$\widetilde{\gamma}_4^{RP,1}(x)$	0.0902	0.1105
$\widetilde{\gamma}_4^{RP,2}(x)$	0.0905	0.1694
$\widehat{\gamma}_{GGS}(x)$	0.1503	0.0419
$\widehat{\gamma}_{GGS}^{BC}(x)$	0.0477	0.0381
$\widehat{\gamma}_{WT}(x)$	0.0076	0.0063

BIBLIOGRAPHY

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

Bingham, N. H., Goldie, C. M., & Teugels, J. L. (1987). *Regular Variation*, Cambridge University Press, Cambridge.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer.

Daouia, A., Gardes, L., & Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19, 2557–2589.

Daouia, A., Gardes, L., Girard, S., & Lekina, A. (2011). Kernel estimators of extreme level curves. *Test*, 20, 311–333.

Davison, A. C. & Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 52, 3, 393-442.

de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer.

de Haan, L. & Stadtmüller, U. (1996). Generalized regular variation of second order. *Journal of the Australian Mathematical Society (Series A)*, 61, 381–395.

Dekkers, A. L. M., Einmahl, J. H. J., & de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 17, 1833–1855.

Drees, H. (1995). Refined Pickands estimators of the extreme value index. *Annals of Statistics*, 23, 2059–2080.

Dziewonski, A. M., Chou, T. A., & Woodhouse, J. H. (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research*, 86, 2825–2852.

Ekström, G., Nettles, M., & Dziewonski, A. M. (2012). The global CMT project 2004-2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, 200–201, 1–9.

Fraga Alves, M. I., Gomes, M. I., de Haan, L., & Neves, C. (2007). A note on second order conditions in extreme value theory: linking general and heavy tail conditions. *REVSTAT – Statistical Journal*, 5, 285–304.

Fraga Alves, M. I., Gomes, M. I., de Haan, L., & Neves, C. (2009). Mixed moment estimator and location invariant alternatives. *Extremes*, 12, 149–185.

Gannoun, A., Girard, S., Guinot, C., & Saracco, J. (2002). Reference ranges based on nonparametric quantile regression. *Statistics in Medicine*, 21, 3119–3135.

Goegebeur, Y., Guillou, A., & Schorgen, A. (2013). Nonparametric regression estimation of conditional tails - the random covariate case. *Statistics*, DOI: 10.1080/02331888.2013.800064.

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

Hosking, J. R. M. & Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29, 339–349.

Kagan, Y. Y. (1997). Seismic moment-frequency relation for shallow earthquakes: regional comparison. *Journal of Geophysical Research*, 102, 2835–2852.

Kagan, Y. Y. (1999). Universality of seismic moment-frequency relation. *Pure and Applied Geophysics*, 155, 537–573.

Okal, E. A. & Romanowich, B. A. (1994). On the variation of $b-$values with earthquake size. *Physics of the Earth and Planetary Interiors*, 87, 55–76.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3, 119–131.

Pisarenko, V. F. & Sornette, D. (2003). Characterization of the frequency of extreme earthquake events by the generalized Pareto distribution. *Pure and Applied Geophysics*, 160, 2343–2364.

Smith, R. L. (1987). Estimating tails of probability distributions. *Annals of Statistics*, 15, 3, 1174-1207.

Stupfler, G. (2013). A moment estimator for the conditional extreme-value index. *Electronic Journal of Statistics*, 7, 2298–2353.

Wang, H. & Li, D. (2013). Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association*, 108, 1062–1074.

Wang, H. & Tsai, C. L. (2009). Tail index regression. *Journal of the American Statistical Association*, 104, 1233–1240.

Weisstein, E. W. (2003). *CRC Concise Encyclopedia of Mathematics (Second edition)*, Chapman and Hall.

Yao, Q. (1999). Conditional predictive regions for stochastic processes. *Technical report*, University of Kent at Canterbury.