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

Yuri Goegebeur, Armelle Guillou, Michael Osmann

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

Yuri Goegebeur *
Armelle Guillou †
Michael Osmann ‡

Key words and phrases: Local estimation; max-domain of attraction; second order condition; tail index.

MSC 2010: Primary 62G32; secondary 62G05, 62G30.

Abstract. This paper deals with the nonparametric estimation of the conditional tail index 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 tail index is estimated locally within a narrow neighborhood 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 and an alternative from the recent extreme value literature are evaluated with a small simulation study. We also illustrate the practical applicability of the estimator on the world catalogue of earthquake magnitudes.

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 une alternative introduite récemment dans la littérature. L'utilisation pratique sur des données de magnitudes de séismes est également proposée.

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

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

‡Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: mosma@imada.sdu.dk).

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 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

$$D_{\gamma(x)}(y) := \int_1^y u^{\gamma(x)-1} du \tag{2}$$

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

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

The parameter $\gamma(x)$, called the conditional extreme value index or also conditional tail index, gives important information about the tail heaviness of $\overline{F}(y; x)$. It allows to distinguish 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 neighborhood 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 tail index. There are many practical instances where it is more relevant to study the tail of the conditional response distribution rather than the mean. In fact, for heavy-tailed response distributions the conditional expectation might not exist. Also, estimation of the conditional tail 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 present. Accurate modeling of the tail of the earthquake energy distribution has clearly a large practical relevance since severe earthquakes cause a lot of damage and losses. With our method we can link the tail of this earthquake energy distribution to local factors, which allows 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 color, the analysis of survival at extreme durations, to name but a few.

The estimation of the tail 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) developed a parametric estimation method based on maximum likelihood in the Hall subclass of Pareto-type models (Hall, 1982). Nonparametric kernel methods were introduced in Daouia et al. (2011), who used a fixed number of extreme conditional quantile estimators to estimate the conditional tail 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 tail index for the complete max-domain of attraction in presence of random covariates has hardly been considered. We are only aware of the attempt made by Daouia, Gardes, & Girard (2013), generalizing the methodology of Daouia et al. (2011). In particular, 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 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 estimator and an alternative procedure from the recent extreme value literature are evaluated with a small simulation experiment in Section 3. In Section 4 we illustrate the applicability of the method on the world catalogue of earthquake magnitudes. The proofs of all results are postponed to Section 5.

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 (1). The basic building block for our estimator is the statistic

$$T_n^{(t)}(x, K) := \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(x - X_i) \left(\ln Y_i - \ln \omega_n\right)_+^t \mathbf{1}\left\{Y_i > \omega_n\right\}, t = 0, 1, \ldots, 6, \tag{3}$$

where $K_{h_n}(x) := \frac{K\left(\frac{x}{h_n}\right)}{h_n^p}$, 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 (3) we need to impose a second order condition on the tail behavior of $F(y;x)$, specifying the rate of convergence in (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*

$$\lim_{t\to\infty} \frac{\frac{U(ty;x)-U(t;x)}{a(t;x)} - D_{\gamma(x)}(y)}{A(t;x)} = D_{\gamma(x),\rho(x)}(y), \tag{4}$$

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 (3) is expressed in terms of log-excesses, we need a reformulation of (4) 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 behavior of (3) 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 *Case (i), $t = 0$:*

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

Assume $(\mathcal{R})$ with $\rho(x) \neq \gamma(x)$, then for $\omega_n \to y^*(x)$ we have that:
Case (ii), $t = 1$:

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

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

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

Case (iii), $t = 2$:

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

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 some norm on $\mathbb{R}^p$.

Assumption $(\mathcal{G})$ *There exists $c_g > 0$ and $\eta_g > 0$ such that $|g(x) - g(z)| \leq c_g \|x - z\|^{\eta_g}$ 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 $(\mathcal{K})$ *K is a bounded density 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 behavior 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 $m(\omega_n, t; x)$.

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 (3) 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 $\omega_n \to y^*(x)$,*

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

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 (3). This is done in Theorem 1. Inspired by Lemmas 1 and 2, we define

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

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 $nh_n^p \overline{F}(\omega_n;x) \to \infty$ for $n \to \infty$, then*

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

where the elements of Σ are given by

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

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

$$\widehat{\gamma}_n(x) := \frac{T_n^{(1)}(x,K_1)}{T_n^{(0)}(x,K_0)} + 1 - \frac{1}{2} \left(1 - \frac{\left(\frac{T_n^{(1)}(x,K_1)}{T_n^{(0)}(x,K_0)} \right)^2}{\frac{T_n^{(2)}(x,K_2)}{T_n^{(0)}(x,K_0)}} \right)^{-1}, \tag{5}$$

for kernel functions K_0, K_1 and K_2. 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 (5), 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 the estimators proposed in Daouia, Gardes, & Girard (2013). They consider the class of estimators for $\gamma(x)$, defined by

$$\tilde{\gamma}_J^{RP}(x) := \frac{1}{\ln r} \sum_{j=1}^{J-2} \pi_j \ln \left(\frac{\hat{q}_n\left(\tau_j \alpha_n; x\right) - \hat{q}_n\left(\tau_{j+1} \alpha_n; x\right)}{\hat{q}_n\left(\tau_{j+1} \alpha_n; x\right) - \hat{q}_n\left(\tau_{j+2} \alpha_n; x\right)} \right), \tag{6}$$

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

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

with

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

In particular they propose to use $J = 3$ 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

$$\tilde{\gamma}_J^{RP,1} := \frac{1}{(J-2)\ln r} \ln \left(\frac{\hat{q}_n\left(\tau_1 \alpha_n; x\right) - \hat{q}_n\left(\tau_2 \alpha_n; x\right)}{\hat{q}_n\left(\tau_{J-1} \alpha_n; x\right) - \hat{q}_n\left(\tau_J \alpha_n; x\right)} \right) \tag{7}$$

and

$$\tilde{\gamma}_J^{RP,2}(x) := \frac{2}{(J-1)(J-2)\ln r} \sum_{j=1}^{J-2} \ln \left(\frac{\hat{q}_n\left(\tau_j \alpha_n; x\right) - \hat{q}_n\left(\tau_{j+1} \alpha_n; x\right)}{\hat{q}_n\left(\tau_{J-1} \alpha_n; x\right) - \hat{q}_n\left(\tau_J \alpha_n; x\right)} \right). \tag{8}$$

Note that for $J = 3$, the two estimators are identical. In Figure 1 we compare the asymptotic variance of the local moment estimator, with the asymptotic variance of $\tilde{\gamma}_J^{RP,1}(x)$ and $\tilde{\gamma}_J^{RP,2}(x)$ for $J = 3, 4$, in the case where all kernel functions are chosen to be equal. In the plot we have fixed $\frac{\|K\|_2^2}{g(x)} = 1$, 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 $\tilde{\gamma}_4^{RP,1}(x)$ 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 $\gamma(x)$.

For the practical implementation of the local moment estimator, we have to choose the bandwidth h_n and the threshold ω_n, where we use for the latter the $(k+1)$-th largest response in the ball $B(x, h_n)$. In all cases the kernel functions are chosen as the biquadratic kernel function

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

Selection of (h_n, k) is done using (i) an oracle strategy and (ii) a completely data driven method.

The oracle strategy was proposed in Gardes, Girard, & Lekina (2010) and consists in selecting (h_n, k) as follows:

$$(h_0, k_0) := \underset{h_n,k}{\operatorname{argmin}} \Delta\left(\hat{\gamma}_n(\cdot), \gamma(\cdot)\right), \tag{9}$$

with

$$\Delta^2\left(\hat{\gamma}_n(\cdot), \gamma(\cdot)\right) := \frac{1}{L}\sum_{\ell=1}^{L}\left(\hat{\gamma}_n\left(z_\ell\right) - \gamma\left(z_\ell\right)\right)^2, \tag{10}$$

where $z_1, \ldots, z_L$ are points regularly spaced in the covariate space. This method requires the knowledge of the function $\gamma(x)$, which is unknown in practical situations. To deal with this we also use a two step method, which is completely data driven. 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 consists in selecting h_n by

$$h_c := \underset{h_n \in \mathcal{H}}{\operatorname{argmin}} \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\mathbf{1}\left\{Y_i \leq Y_j\right\} - \hat{F}_{n,-i}\left(Y_j; X_i\right)\right)^2, \tag{11}$$

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

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

Next for each z_ℓ we do the following

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

Note that in the oracle approach both h_n and k are selected globally, while in the data driven approach h_n is selected globally and k is selected locally.

We compare the estimators on three conditional distributions of Y given $X = x$. The conditional distributions we consider are

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

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

for which $\gamma(x) = -\frac{1}{\lambda(x)\tau(x)}$ and $\rho(x) = -\frac{1}{\lambda(x)}$. 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 $\lambda(x) = 2$. Note that the function $\gamma(x)$ is chosen differently in the last case, in order to satisfy the requirement $\gamma(x) \neq \rho(x)$. Such type of functions $\gamma(x)$ was also considered in Daouia et al. (2011), though in the framework of Pareto-type distributions. They are differentiable with several stationary points.

- 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 $\gamma(x) = \frac{1}{\lambda(x)\tau(x)}$ and $\rho(x) = -\frac{1}{\lambda(x)}$. 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 is in all cases chosen as the $Unif(0,1)$ distribution. For all distributions we simulate $N = 500$ samples of size $n = 1000$ and as measures of efficiency, we compute the absolute bias for a given estimator $\hat{\gamma}(\cdot)$

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

and the mean squared error

$$\text{MSE}\left(\hat{\gamma}(\cdot)\right) = \frac{1}{L}\sum_{\ell=1}^{L}\left(\hat{\gamma}\left(z_\ell\right) - \gamma\left(z_\ell\right)\right)^2$$

in each simulation, with the z_ℓ's being $L = 41$ points equidistantly spaced in $[0.1, 0.9]$. The values for bias and MSE we report are averaged over the $N = 500$ simulations.

The benchmark estimators of Daouia, Gardes, & Girard (2013) are also examined using the oracle approach as well as the data driven approach. For the oracle approach, we use α_n instead of k in (9), while in the data driven approach, $\tilde{\gamma}_J^{RP}(x)$ is computed for $\alpha_n = \frac{k}{n^*}$ where $k = 2, \ldots, n^* - 1$ and n^* is the number of observations in the ball $B\left(x, h_n\right)$. In the oracle method we search for h_n in the grid $\mathcal{H} = \{0.05, 0.075, \ldots, 0.3\}$ for all estimators, while k is chosen in $k = 2, \ldots, 100$ for the local moment estimator, and α_n is chosen in the grid $\{0.05, \ldots, 0.95\}$ for the benchmark estimators. For the data driven approach we also use $\mathcal{H} = \{0.05, 0.075, \ldots, 0.3\}$ for the cross validation for all estimators, and we choose $k_{\max} = \left\lfloor \frac{n^*}{2} \right\rfloor$ for the local moment estimator.

The results of the simulations can be seen in Table 1 to Table 10. In the tables the figures in bold indicate the estimator with the smallest bias and MSE, and the numbers in parenthesis are the standard errors of the estimates. Regarding the tables, we have the following general remarks:

- As expected, the results for the data driven method are worse than those for the oracle method.
- Among the estimators proposed by Daouia, Gardes, & Girard (2013) no estimator performs uniformly best in terms of bias and MSE, though $\tilde{\gamma}_4^{RP,1}(x)$ seems to have an overall good performance.
- In the cases where the conditions of our theorems are satisfied (reversed Burr and Burr), our local moment estimator outperforms the benchmark estimators in terms of bias and MSE, both with the oracle method and the data driven method.
- For the Burr distribution, where $\gamma(x) > 0$, the local moment estimator is far superior to the benchmark estimators, but this was also to be expected since the asymptotic variance of the local moment estimator is much smaller than that of the benchmark estimators in this case.

- 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) estimators. Our estimator thus seems to be robust with respect to a violation of the assumptions.

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 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 tail 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 tail index, we use locations where an earthquake has already happened. In order to determine the neighborhood of these locations, we compute the distance in km 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 $R = 6371km$ (mean radius of the earth). In the analysis we ignored isolated earthquakes, i.e. earthquakes for which there is no neighboring earthquake within a radius of $200km$. The bandwidth h_n is chosen by applying the cross validation criterion (11) 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 $h_n = 400km$. Next, the threshold is selected locally in the same fashion as for the simulations. A plot of the local estimates of the tail 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 $\gamma(x) \leq 1$, 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 tail 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 tail 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 neighborhood 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 PROOFS

Proof of Lemma 1

Case (i) is trivial, so we only have to consider cases (ii) and (iii). Throughout the proof we make use of the notation $p_n := F(\omega_n; x)$ and we assume that V is uniformly distributed on $(0, 1)$.

With this notation, we can write

$$\begin{aligned} m(\omega_n, t; x) &= \mathbb{E}\left[\left(\ln U\left(\frac{1}{1-V}; x\right) - \ln U\left(\frac{1}{1-p_n}; x\right)\right)_+^t\right] \\ &= \int_{p_n}^1 \left(\ln U\left(\frac{1}{1-v}; x\right) - \ln U\left(\frac{1}{1-p_n}; x\right)\right)^t dv \\ &= (1-p_n)\int_1^\infty \left(\ln U\left(\frac{1}{1-p_n} z; x\right) - \ln U\left(\frac{1}{1-p_n}; x\right)\right)^t \frac{1}{z^2} dz. \end{aligned}$$

From here we have to consider all the cases separately. We start with $\gamma(x) > 0$, where we make use of the notation

$$R_n(z; x) := \frac{\ln U\left(\frac{1}{1-p_n} z; x\right) - \ln U\left(\frac{1}{1-p_n}; x\right) - \gamma(x) \ln z}{B\left(\frac{1}{1-p_n}; x\right)} - D_{\widetilde{\rho}(x)}(z),$$

where $D_{\widetilde{\rho}(x)}(z)$ was defined in (2). Inspired by Lemma 4.4 from Fraga Alves et al. (2009) we rewrite $m(\omega_n, t; x)$ for $t = 1$ as

$$\begin{aligned} m(\omega_n, 1; x) &= (1-p_n)\left[\gamma(x)\int_1^\infty \ln z \frac{1}{z^2} dz + B\left(\frac{1}{1-p_n}; x\right)\int_1^\infty D_{\widetilde{\rho}(x)}(z)\frac{1}{z^2} dz \right. \\ &\quad \left. + B\left(\frac{1}{1-p_n}; x\right)\int_1^\infty R_n(z; x)\frac{1}{z^2} dz\right] \\ &=: (1-p_n)\left[T_1 + T_2 + T_3\right]. \end{aligned}$$

We easily find that $T_1 = \gamma(x)$, while $T_2 = B\left(\frac{1}{1-p_n}; x\right)\frac{1}{1-\widetilde{\rho}(x)}$. Regarding T_3 we use Lemma 4.4 from Fraga Alves et al. (2009), according to which, for all $0 < \varepsilon < 1$ there exists $\omega_{n,0}$ such that for all $\omega_n \geq \omega_{n,0}$

$$\left|\int_1^\infty R_n\left(z;x\right)\frac{1}{z^2}dz\right| \leq \varepsilon \int_1^\infty z^{\widetilde{\rho}(x)+\varepsilon-2}dz.$$

Hence, we conclude that $T_3 = o\left(B\left(\frac{1}{1-p_n}; x\right)\right)$. Combining the terms concludes the proof for $\gamma(x) > 0$ with $t = 1$.

Continuing with the case $\gamma(x) > 0$ for $t = 2$, it follows that

$$\begin{aligned}
m\left(\omega_n, 2; x\right) =& \left(1-p_n\right)\left[\gamma^2(x)\int_1^\infty \ln^2 z\frac{1}{z^2}dz + 2\gamma(x)B\left(\frac{1}{1-p_n}; x\right)\int_1^\infty D_{\widetilde{\rho}(x)}(z)\ln z\frac{1}{z^2}dz\right.\\
&+ B^2\left(\frac{1}{1-p_n}; x\right)\int_1^\infty D^2_{\widetilde{\rho}(x)}(z)\frac{1}{z^2}dz + 2B^2\left(\frac{1}{1-p_n}; x\right)\int_1^\infty D_{\widetilde{\rho}(x)}(z)R_n(z;x)\frac{1}{z^2}dz\\
&\left.+ B^2\left(\frac{1}{1-p_n}; x\right)\int_1^\infty R^2_n(z;x)\frac{1}{z^2}dz + 2\gamma(x)B\left(\frac{1}{1-p_n}; x\right)\int_1^\infty R_n(z;x)\ln z\frac{1}{z^2}dz\right]\\
=:& \left(1-p_n\right)\left[\widetilde{T}_1 + \widetilde{T}_2 + \widetilde{T}_3 + \widetilde{T}_4 + \widetilde{T}_5 + \widetilde{T}_6\right].
\end{aligned}$$

Here, we find $\widetilde{T}_1 = 2\gamma^2(x)$, while $\widetilde{T}_2 = 2\gamma(x)\frac{2-\widetilde{\rho}(x)}{\left(1-\widetilde{\rho}(x)\right)^2}B\left(\frac{1}{1-p_n}; x\right)$. Concerning the term $\widetilde{T}_3$ it follows that

$$\widetilde{T}_3 = o\left(B\left(\frac{1}{1-p_n}; x\right)\right)$$

since $\int_1^\infty D^2_{\widetilde{\rho}(x)}(z)\frac{1}{z^2}dz < \infty$. All the terms $\widetilde{T}_4$, $\widetilde{T}_5$ and $\widetilde{T}_6$ are also found to be $o\left(B\left(\frac{1}{1-p_n}; x\right)\right)$ using arguments similar to the one used for the term T_3. Combining all the terms establishes the result for $\gamma(x) > 0$ with $t = 2$.

For the cases where $\gamma(x) \leq 0$ we let

$$D^*_{\gamma(x),\rho(x)}(z) := \begin{cases} z^{\gamma(x)}\ln(z), & \gamma(x) < \rho(x) = 0\\ \frac{z^{\gamma(x)+\rho(x)}-1}{\gamma(x)+\rho(x)}, & \gamma(x) < \rho(x) < 0\\ \frac{z^{2\gamma(x)}-1}{2\gamma(x)}, & \rho(x) < \gamma(x) < 0\\ \ln^2 z, & \rho(x) < \gamma(x) = 0 \end{cases},$$

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

and

$$\rho^*(x) := \begin{cases} \rho(x), & \gamma(x) < \rho(x) \leq 0\\ \gamma(x), & \rho(x) < \gamma(x) \leq 0 \end{cases}.$$

From the way the function $B(.;x)$ is defined, we note that $A^*(.;x)$ is regularly varying with index $\rho^*(x)$ (see Fraga Alves et al., 2007).

Furthermore, we define the three error terms

$$R_{0,n}(z;x) := -\ln\left[1-\left(1-\frac{U\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n}z;x\right)}\right)\right]-\left(1-\frac{U\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n}z;x\right)}\right)-\frac{1}{2}\left(1-\frac{U\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n}z;x\right)}\right)^2,$$

$$R_{1,n}(z;x) := \frac{\frac{U\left(\frac{1}{1-p_n};x\right)}{a^*\left(\frac{1}{1-p_n};x\right)}\left(1-\frac{U\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n}z;x\right)}\right)-D_{\gamma(x)}(z)}{A^*\left(\frac{1}{1-p_n};x\right)}-D^*_{\gamma(x),\rho(x)}(z)$$

and

$$R_{2,n}(z;x) := \frac{\left(\frac{U\left(\frac{1}{1-p_n};x\right)}{a^*\left(\frac{1}{1-p_n};x\right)}\left(1-\frac{U\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n}z;x\right)}\right)\right)^2-D^2_{\gamma(x)}(z)}{A^*\left(\frac{1}{1-p_n};x\right)}-2D_{\gamma(x)}(z)D^*_{\gamma(x),\rho(x)}(z).$$

For all the cases where $\gamma(x)\le 0$, we note that we can write

$$\begin{aligned} m(\omega_n,1;x) &= (1-p_n)\left[\int_1^\infty\left(1-\frac{U\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n}z;x\right)}\right)\frac{1}{z^2}dz\right.\\ &\qquad\left.+\frac{1}{2}\int_1^\infty\left(1-\frac{U\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n}z;x\right)}\right)^2\frac{1}{z^2}dz+\int_1^\infty R_{0,n}(z;x)\frac{1}{z^2}dz\right]\\ &=: (1-p_n)\left[I_1+I_2+I_3\right] \end{aligned}$$

and

$$\begin{aligned} m(\omega_n,2;x) &= (1-p_n)\left[\int_1^\infty\left(1-\frac{U\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n}z;x\right)}\right)^2\frac{1}{z^2}dz+\int_1^\infty\left(1-\frac{U\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n}z;x\right)}\right)^3\frac{1}{z^2}dz\right.\\ &\quad+\frac{1}{4}\int_1^\infty\left(1-\frac{U\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n}z;x\right)}\right)^4\frac{1}{z^2}dz+2\int_1^\infty\left(1-\frac{U\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n}z;x\right)}\right)R_{0,n}(z;x)\frac{1}{z^2}dz\\ &\quad\left.+\int_1^\infty\left(1-\frac{U\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n}z;x\right)}\right)^2R_{0,n}(z;x)\frac{1}{z^2}dz+\int_1^\infty R^2_{0,n}(z;x)\frac{1}{z^2}dz\right]\\ &=: (1-p_n)\left[\widetilde{I}_1+\widetilde{I}_2+\widetilde{I}_3+\widetilde{I}_4+\widetilde{I}_5+\widetilde{I}_6\right]. \end{aligned}$$

If we consider the term I_1 first, we rewrite this as

$$\begin{aligned} I_1 =& \frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\left[\int_1^\infty D_{\gamma(x)}(z)\frac{1}{z^2}dz + A^*\left(\frac{1}{1-p_n};x\right)\int_1^\infty D^*_{\gamma(x),\rho(x)}(z)\frac{1}{z^2}dz\right. \\ &\left. + A^*\left(\frac{1}{1-p_n};x\right)\int_1^\infty R_{1,n}(z;x)\frac{1}{z^2}dz\right] \\ =:& \frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\left[I_{11}+I_{12}+I_{13}\right]. \end{aligned}$$

It is easily seen that $I_{11} = \frac{1}{1-\gamma(x)}$, while

$$I_{12} = \begin{cases} \frac{1}{(1-\gamma(x))^2}A^*\left(\frac{1}{1-p_n};x\right), & \gamma(x) < \rho(x) = 0 \\ \frac{1}{1-\gamma(x)-\rho(x)}A^*\left(\frac{1}{1-p_n};x\right), & \gamma(x) < \rho(x) < 0 \\ \frac{1}{1-2\gamma(x)}A^*\left(\frac{1}{1-p_n};x\right), & \rho(x) < \gamma(x) < 0 \\ 2A^*\left(\frac{1}{1-p_n};x\right), & \rho(x) < \gamma(x) = 0 \end{cases}.$$

Concerning the remainder term I_{13}, we use Lemma 4.4 from Fraga Alves et al. (2009), which states that for all $\varepsilon > 0$ there exists $\omega_{n,0}$ such that for all $\omega_n \geq \omega_{n,0}$,

$$|R_{1,n}(z;x)| \leq \varepsilon z^{\gamma(x)+\rho^*(x)+\varepsilon}.$$

By choosing ε such that $\varepsilon < 1-\gamma(x)-\rho^*(x)$, we find that

$$I_{13} = o\left(A^*\left(\frac{1}{1-p_n};x\right)\right).$$

So, by changing from the function A^* to the function B, it follows that

$$I_1 = \frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\left[\frac{1}{1-\gamma(x)} + c^{(1)}_{\gamma(x),\rho(x)}B\left(\frac{1}{1-p_n};x\right)(1+o(1))\right],$$

where

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

The term I_2 can also be written as

$$\begin{aligned} I_2 =& \frac{1}{2}\left(\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\right)^2 \left[\int_1^\infty D^2_{\gamma(x)}(z)\frac{1}{z^2}dz \right. \\ & \left. + 2A^*\left(\frac{1}{1-p_n};x\right)\int_1^\infty D_{\gamma(x)}(z)D^*_{\gamma(x),\rho(x)}(z)\frac{1}{z^2}dz + A^*\left(\frac{1}{1-p_n};x\right)\int_1^\infty R_{2,n}(z;x)\frac{1}{z^2}dz\right] \\ =:& \frac{1}{2}\left(\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\right)^2 \left[I_{21}+I_{22}+I_{23}\right]. \end{aligned}$$

Here, we have that $I_{21} = \frac{2}{(1-\gamma(x))(1-2\gamma(x))}$, while

$$I_{22} = \begin{cases} \frac{2(2-3\gamma(x))}{(1-\gamma(x))^2(1-2\gamma(x))^2}A^*\left(\frac{1}{1-p_n};x\right), & \gamma(x) < \rho(x) = 0 \\ \frac{2(2-2\gamma(x)-\rho(x))}{(1-\gamma(x))(1-\gamma(x)-\rho(x))(1-2\gamma(x)-\rho(x))}A^*\left(\frac{1}{1-p_n};x\right), & \gamma(x) < \rho(x) < 0 \\ \frac{2(2-3\gamma(x))}{(1-\gamma(x))(1-2\gamma(x))(1-3\gamma(x))}A^*\left(\frac{1}{1-p_n};x\right), & \rho(x) < \gamma(x) < 0 \\ 12A^*\left(\frac{1}{1-p_n};x\right), & \rho(x) < \gamma(x) = 0 \end{cases}.$$

The remainder term I_{23} is dealt with using Lemma 4.4 from Fraga Alves et al. (2009), which ensures that for all $\varepsilon > 0$ there exists $\omega_{n,0}$ such that for all $\omega_n \geq \omega_{n,0}$,

$$|R_{2,n}(z;x)| \leq \varepsilon z^{2\gamma(x)+\rho^*(x)+\varepsilon}.$$

From this it follows by choosing $\varepsilon < 1 - 2\gamma(x) - \rho^*(x)$, that

$$I_{23} = o\left(A^*\left(\frac{1}{1-p_n};x\right)\right).$$

If we combine the terms and replace the function A^* by the function B, then we get

$$I_2 = \left(\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\right)^2 \left[\frac{1}{(1-\gamma(x))(1-2\gamma(x))} + c^{(2)}_{\gamma(x),\rho(x)}B\left(\frac{1}{1-p_n};x\right)(1+o(1))\right]$$

where

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

For the case $t = 1$ we can simplify the expression for I_2 as

$$I_2 = \begin{cases} \frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}o\left(B\left(\frac{1}{1-p_n};x\right)\right), & \gamma(x) < \rho(x) \leq 0 \\ \frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\frac{1}{(1-\gamma(x))(1-2\gamma(x))}B\left(\frac{1}{1-p_n};x\right)(1+o(1)), & \rho(x) < \gamma(x) \leq 0 \end{cases}.$$

The remainder term I_3 can be handled with the inequality

$$0 < -\ln(1-y) - y - \frac{y^2}{2} < \frac{y^3}{3(1-y)}, \quad 0 < y < 1,$$

which allows us to write

$$\begin{aligned}
|I_3| &\leq \frac{1}{3}\left(\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\right)^3 \int_1^\infty \left(\frac{U\left(\frac{1}{1-p_n};x\right)}{U\left(z\frac{1}{1-p_n};x\right)}\right)^2 \left(\frac{U\left(z\frac{1}{1-p_n};x\right) - U\left(\frac{1}{1-p_n};x\right)}{a^*\left(\frac{1}{1-p_n};x\right)}\right)^3 \frac{1}{z^2}dz \\
&= O\left(\left(\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\right)^3\right),
\end{aligned}$$

where the last step is obtained by an application of Drees' inequality (Drees, 1998); see also Theorem B.2.18 in de Haan & Ferreira (2006).

Combining the results for I_1, I_2 and I_3 establishes the result for $m\left(\omega_n, 1; x\right)$.

Concerning $m\left(\omega_n, 2; x\right)$, we notice that $\widetilde{I}_1 = 2I_2$, so

$$\widetilde{I}_1 = \left(\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\right)^2 \left[\frac{2}{(1-\gamma(x))(1-2\gamma(x))} + 2c^{(2)}_{\gamma(x),\rho(x)} B\left(\frac{1}{1-p_n};x\right)(1+o(1))\right],$$

while

$$\begin{aligned}
\widetilde{I}_2 &= \left(\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\right)^3 \int_1^\infty \left(D_{\gamma(x)}(z) + A^*\left(\frac{1}{1-p_n};x\right) D^*_{\gamma(x),\rho(x)}(z)\right. \\
&\quad \left. + A^*\left(\frac{1}{1-p_n};x\right) R_{1,n}(z;x)\right)^3 \frac{1}{z^2}dz \\
&= \left(\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\right)^3 \int_1^\infty D^3_{\gamma(x)}(z)\frac{1}{z^2}dz + o\left(\left(\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\right)^3\right) \\
&= \frac{6}{(1-3\gamma(x))(1-2\gamma(x))(1-\gamma(x))}\left(\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\right)^3 (1+o(1)).
\end{aligned}$$

Using the relationship between $\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}$ and $B\left(\frac{1}{1-p_n};x\right)$, we obtain

$$\widetilde{I}_2 = \begin{cases} \left(\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\right)^2 o\left(B\left(\frac{1}{1-p_n};x\right)\right), & \gamma(x) < \rho(x) \leq 0 \\ \left(\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\right)^2 \frac{6}{(1-3\gamma(x))(1-2\gamma(x))(1-\gamma(x))} B\left(\frac{1}{1-p_n};x\right)(1+o(1)), & \rho(x) < \gamma(x) \leq 0 \end{cases}.$$

All the terms $\widetilde{I}_3$, $\widetilde{I}_4$, $\widetilde{I}_5$ and $\widetilde{I}_6$ are easily seen to be $o\left(\left(\frac{a^*\left(\frac{1}{1-p_n};x\right)}{U\left(\frac{1}{1-p_n};x\right)}\right)^2 B\left(\frac{1}{1-p_n};x\right)\right)$, which when combined with $\widetilde{I}_1$ and $\widetilde{I}_2$ establishes the result.

5.1 Proof of Lemma 2

Use of the rule of repeated expectation allows us to write

$$\begin{aligned} \widetilde{m}_n(K,t;x) &= \mathbb{E}\left[K_{h_n}(x-X)m\left(\omega_n,t;X\right)\right] \\ &= \int_{\mathbb{R}^p} K_{h_n}(x-v)m\left(\omega_n,t;v\right)g(v)dv \\ &= \int_{\Omega} K(z)m\left(\omega_n,t;x-h_nz\right)g(x-h_nz)dz. \end{aligned}$$

Now, we easily obtain

$$\begin{aligned} |\widetilde{m}_n(K,t;x) - m\left(\omega_n,t;x\right)g(x)| \leq & m\left(\omega_n,t;x\right)\int_{\Omega} K(z)\left|g\left(x-h_nz\right)-g(x)\right|dz \\ & + g(x)\int_{\Omega} K(z)\left|m\left(\omega_n,t;x-h_nz\right)-m\left(\omega_n,t;x\right)\right|dz \\ & + \int_{\Omega} K(z)\left|g\left(x-h_nz\right)-g(x)\right|\left|m\left(\omega_n,t;x-h_nz\right)-m\left(\omega_n,t;x\right)\right|dz \\ =: & \bar{T}_1+\bar{T}_2+\bar{T}_3. \end{aligned}$$

From Assumptions $(\mathcal{G})$ and $(\mathcal{K})$ we obtain that

$$\begin{aligned} \bar{T}_1 &\leq m\left(\omega_n,t;x\right)c_g\int_{\Omega} K(z)\|h_nz\|^{\eta_g}dz \\ &= O\left(m\left(\omega_n,t;x\right)h_n^{\eta_g}\right), \end{aligned}$$

while assumptions $(\mathcal{K})$ and $(\mathcal{F})$ give us that

$$\begin{aligned} \bar{T}_2 &\leq g(x)m\left(\omega_n,t;x\right)\Phi\left(\omega_n,h_n;x\right)\int_{\Omega} K(z)dz \\ &= O\left(m\left(\omega_n,t;x\right)\Phi\left(\omega_n,h_n;x\right)\right). \end{aligned}$$

Similar arguments lead to

$$\bar{T}_3 = O\left(m\left(\omega_n,t;x\right)\Phi\left(\omega_n,h_n;x\right)h_n^{\eta_g}\right),$$

so collecting all the terms establishes Lemma 2.

5.2 Proof of Theorem 1

Let $r_n := \sqrt{nh_n^p \overline{F}(\omega_n; x)}$ and

$$\xi_n := r_n \delta' \left(\mathbb{T}_n - \mathbb{E}\left[\mathbb{T}_n\right] \right),$$

where $\delta \in \mathbb{R}^3$. In order to prove the theorem we use the Cramér-Wold device (see e.g. van der Vaart, 1998, p. 16), according to which it is sufficient to verify that

$$\xi_n \xrightarrow{D} N_3(0, \delta' \Sigma \delta)$$

for all $\delta \in \mathbb{R}^3$. We let the components of δ be denoted by δ_0, δ_1 and δ_2.

In the case $\gamma(x) > 0$ we write

$$\begin{aligned}
\xi_n =& \sum_{i=1}^{n} \sqrt{\frac{h_n^p}{n\overline{F}(\omega_n; x)}} \frac{1}{g(x)} \Bigg(\sum_{j=0}^{2} \delta_j K_{j,h_n}(x - X_i) \left(\ln Y_i - \ln \omega_n\right)_+^j \mathbf{1}\left\{Y_i > \omega_n\right\} \\
& - \mathbb{E}\left[\sum_{j=0}^{2} \delta_j K_{j,h_n}(x - X_i) \left(\ln Y_i - \ln \omega_n\right)_+^j \mathbf{1}\left\{Y_i > \omega_n\right\} \right] \Bigg) \\
=: & \sum_{i=1}^{n} W_i.
\end{aligned}$$

Observe that the W_i, $i = 1, \ldots, n$, are independent and identically distributed random variables. We can hence find the variance of ξ_n as

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\xi_n\right) &= n \mathbb{V}\mathrm{ar}\left(W_1\right) \\
&= \frac{h_n^p}{\overline{F}(\omega_n; x)\, g^2(x)} \sum_{j=0}^{2} \sum_{k=0}^{2} \delta_j \delta_k C_{j,k},
\end{aligned}$$

where

$$\begin{aligned}
C_{j,k} :=& \mathbb{E}\left[K_{j,h_n}(x - X_1) K_{k,h_n}(x - X_1) \left(\ln Y_1 - \ln \omega_n\right)_+^{j+k} \mathbf{1}\left\{Y_1 > \omega_n\right\} \right] \\
& - \mathbb{E}\left[K_{j,h_n}(x - X_1) \left(\ln Y_1 - \ln \omega_n\right)_+^{j} \mathbf{1}\left\{Y_1 > \omega_n\right\} \right] \\
& \times \mathbb{E}\left[K_{k,h_n}(x - X_1) \left(\ln Y_1 - \ln \omega_n\right)_+^{k} \mathbf{1}\left\{Y_1 > \omega_n\right\} \right].
\end{aligned}$$

Using an approach similar to the one used in the proof of Lemmas 1 and 2, we find

$$\mathbb{E}\left[T_n^{(t)}(x, K) \right] = t! \overline{F}(\omega_n; x)\, g(x) \gamma^t(x) (1 + o(1)), \quad t = 3, 4, 5, 6,$$

and hence

$$\begin{aligned}
C_{j,k} =& \frac{\|K_j K_k\|_1}{h_n^p} \mathbb{E}\left[\frac{1}{h_n^p \|K_j K_k\|_1} K_j\left(\frac{x-X_1}{h_n}\right) K_k\left(\frac{x-X_1}{h_n}\right)(\ln Y_1 - \ln\omega_n)_+^{j+k} \mathbf{1}\{Y_1 > \omega_n\}\right] \\
& - \mathbb{E}\left[K_{j,h_n}(x-X_1)(\ln Y_1 - \ln\omega_n)_+^{j} \mathbf{1}\{Y_1 > \omega_n\}\right] \\
& \times \mathbb{E}\left[K_{k,h_n}(x-X_1)(\ln Y_1 - \ln\omega_n)_+^{k} \mathbf{1}\{Y_1 > \omega_n\}\right] \\
=& (j+k)! \frac{\|K_j K_k\|_1}{h_n^p} \gamma^{j+k}(x) \overline{F}(\omega_n; x)\, g(x)(1+o(1)).
\end{aligned}$$

It then follows that the variance of ξ_n is

$$\mathbb{V}\mathrm{ar}(\xi_n) = \delta' \Sigma \delta (1+o(1)).$$

The asymptotic normality is established by verifying Lyapounov's criterion for triangular arrays, see e.g. Billingsley (1995, p. 362), which in our case corresponds to verifying

$$\sum_{i=1}^{n} \mathbb{E}\left[|W_i|^3\right] = n\mathbb{E}\left[|W_1|^3\right] \to 0$$

as $n \to \infty$. We have

$$\begin{aligned}
\mathbb{E}\left[|W_1|^3\right] \leq & \left(\frac{h_n^p}{n\overline{F}(\omega_n; x)}\right)^{\frac{3}{2}} \frac{1}{g^3(x)} \Bigg\{ \mathbb{E}\left[\left(\sum_{j=0}^{2} |\delta_j|\, K_{j,h_n}(x-X_1)(\ln Y_1 - \ln\omega_n)_+^{j} \mathbf{1}\{Y_1 > \omega_n\}\right)^3\right] \\
& + 3\mathbb{E}\left[\left(\sum_{j=0}^{2} |\delta_j|\, K_{j,h_n}(x-X_1)(\ln Y_1 - \ln\omega_n)_+^{j} \mathbf{1}\{Y_1 > \omega_n\}\right)^2\right] \\
& \times \mathbb{E}\left[\sum_{j=0}^{2} |\delta_j|\, K_{j,h_n}(x-X_1)(\ln Y_1 - \ln\omega_n)_+^{j} \mathbf{1}\{Y_1 > \omega_n\}\right] \\
& + 4\left(\mathbb{E}\left[\sum_{j=0}^{2} |\delta_j|\, K_{j,h_n}(x-X_1)(\ln Y_1 - \ln\omega_n)_+^{j} \mathbf{1}\{Y_1 > \omega_n\}\right]\right)^3 \Bigg\} \\
=& O\left(\frac{1}{n\sqrt{nh_n^p \overline{F}(\omega_n; x)}}\right).
\end{aligned}$$

Under the conditions of the theorem it follows that $nE\left[|W_1|^3\right] \to 0$ as $n \to \infty$. This concludes the proof for $\gamma(x) > 0$.

Concerning $\gamma(x) \leq 0$, we write

$$\begin{aligned}\xi_n =& \sum_{i=1}^{n} \sqrt{\frac{h_n^p}{n\overline{F}\left(\omega_n; x\right)}} \frac{1}{g(x)} \left(\sum_{j=0}^{2} \delta_j K_{j,h_n}\left(x - X_i\right) \left(\ln Y_i - \ln \omega_n\right)_+^j \mathbf{1}\left\{Y_i > \omega_n\right\} \left(\frac{U\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)}{a^*\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)} \right)^j \right. \\ & \left. - \mathbb{E}\left[\sum_{j=0}^{2} \delta_j K_{j,h_n}\left(x - X_i\right) \left(\ln Y_i - \ln \omega_n\right)_+^j \mathbf{1}\left\{Y_i > \omega_n\right\} \left(\frac{U\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)}{a^*\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)} \right)^j \right] \right) \\ =:& \sum_{i=1}^{n} W_i.\end{aligned}$$

Again the W_i, $i = 1, \ldots, n$, are independent and identically distributed random variables, so the variance of ξ_n can be obtained as

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(\xi_n\right) &= n\mathbb{V}\mathrm{ar}\left(W_1\right) \\ &= \frac{h_n^p}{\overline{F}\left(\omega_n; x\right) g^2(x)} \sum_{j=0}^{2} \sum_{k=0}^{2} \delta_j \delta_k C_{j,k},\end{aligned}$$

where

$$\begin{aligned}C_{j,k} :=& \mathbb{E}\left[K_{j,h_n}\left(x - X_1\right) K_{k,h_n}\left(x - X_1\right) \left(\ln Y_1 - \ln \omega_n\right)_+^{j+k} \mathbf{1}\left\{Y_1 > \omega_n\right\} \left(\frac{U\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)}{a^*\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)} \right)^{j+k} \right] \\ & - \mathbb{E}\left[K_{j,h_n}\left(x - X_1\right) \left(\ln Y_1 - \ln \omega_n\right)_+^{j} \mathbf{1}\left\{Y_1 > \omega_n\right\} \left(\frac{U\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)}{a^*\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)} \right)^{j} \right] \\ & \times \mathbb{E}\left[K_{k,h_n}\left(x - X_1\right) \left(\ln Y_1 - \ln \omega_n\right)_+^{k} \mathbf{1}\left\{Y_1 > \omega_n\right\} \left(\frac{U\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)}{a^*\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)} \right)^{k} \right].\end{aligned}$$

We find that

$$\mathbb{E}\left[T_n^{(t)}(x, K)\right] = \overline{F}\left(\omega_n; x\right) \left(\frac{a^*\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)}{U\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)} \right)^t g(x) \frac{t!}{\Pi_{i=0}^{t}(1 - i\gamma(x))}(1 + o(1)), \quad t = 3, 4, 5, 6,$$

using the same approach as was used in the proof of Lemmas 1 and 2. Hence, the term $C_{j,k}$ can be written as

$$C_{j,k} = \frac{(j+k)!}{\Pi_{i=0}^{j+k}(1 - i\gamma(x))} \frac{\left\|K_j K_k\right\|_1}{h_n^p} \overline{F}\left(\omega_n; x\right) g(x)(1 + o(1))$$

and

$$\mathbb{V}\mathrm{ar}\left(\xi_n\right) = \delta' \Sigma \delta (1 + o(1)).$$

The asymptotic normality is established analogously to the case $\gamma(x) > 0$.

Proof of Theorem 2

Concerning the case $\gamma(x) > 0$, let

$$\psi := \begin{bmatrix} 1 \\ \gamma(x) \\ 2\gamma^2(x) \end{bmatrix}.$$

We have

$$\begin{aligned} r_n\left(\mathbb{T}_n - \psi\right) &= r_n\left(\mathbb{T}_n - \mathbb{E}\left[\mathbb{T}_n\right]\right) + r_n\left(\mathbb{E}\left[\mathbb{T}_n\right] - \psi\right) \\ &=: L_1 + L_2. \end{aligned}$$

From Theorem 1 we know that $L_1 \xrightarrow{D} N_3(0, \Sigma)$, while

$$L_2 \to \zeta,$$

where

$$\zeta := \frac{\lambda(x)}{(1-\widetilde{\rho}(x))^2} \begin{bmatrix} 0 \\ 1-\widetilde{\rho}(x) \\ 2\gamma(x)\left(2-\widetilde{\rho}(x)\right) \end{bmatrix}$$

according to the assumptions of the theorem. Hence

$$r_n\left(\mathbb{T}_n - \psi\right) \xrightarrow{D} N_3(\zeta, \Sigma).$$

The result then follows by applying the delta method since we have

$$r_n\left(\widehat{\gamma}_n(x) - \gamma(x)\right) = r_n\left(f_1\left(\widetilde{T}_n^{(0)}\left(x, K_0\right), \widetilde{T}_n^{(1)}\left(x, K_1\right), \widetilde{T}_n^{(2)}\left(x, K_2\right)\right) - f_1\left(1, \gamma(x), 2\gamma^2(x)\right)\right),$$

where $f_1(u, v, w) := \frac{v}{u} + 1 - \frac{1}{2}\left(1 - \frac{v^2}{uw}\right)^{-1}$.

In case $\gamma(x) \leq 0$, let

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

Then, for

$$\begin{aligned} r_n\left(\mathbb{T}_n - \psi\right) &= r_n\left(\mathbb{T}_n - \mathbb{E}\left[\mathbb{T}_n\right]\right) + r_n\left(\mathbb{E}\left[\mathbb{T}_n\right] - \psi\right) \\ &=: M_1 + M_2, \end{aligned}$$

we know that $M_1 \xrightarrow{D} N_3(0, \Sigma)$ according to Theorem 1, while

$$M_2 \to \zeta,$$

where

$$\zeta := \lambda(x) \begin{bmatrix} 0 \\ b^{(1)}_{\gamma(x),\rho(x)} \\ b^{(2)}_{\gamma(x),\rho(x)} \end{bmatrix}$$

according to the assumptions of the theorem. Using (5) we have

$$\widehat{\gamma}_n(x) = \frac{a^*\left(\frac{1}{\overline{F}(\omega_n;x)};x\right)}{U\left(\frac{1}{\overline{F}(\omega_n;x)};x\right)}\frac{\widetilde{T}_n^{(1)}(x,K_1)}{\widetilde{T}_n^{(0)}(x,K_0)} + 1 - \frac{1}{2}\left(1 - \frac{\left[\widetilde{T}_n^{(1)}(x,K_1)\right]^2}{\widetilde{T}_n^{(0)}(x,K_0)\,\widetilde{T}_n^{(2)}(x,K_2)}\right)^{-1},$$

so by Theorem 1 it follows that

$$\begin{aligned} r_n\left(\widehat{\gamma}_n(x) - \gamma(x)\right) =& r_n \frac{a^*\left(\frac{1}{\overline{F}(\omega_n;x)};x\right)}{U\left(\frac{1}{\overline{F}(\omega_n;x)};x\right)}\frac{\mathbb{E}\left[\widetilde{T}_n^{(1)}(x,K_1)\right] + O_{\mathbb{P}}\left(\frac{1}{r_n}\right)}{\mathbb{E}\left[\widetilde{T}_n^{(0)}(x,K_0)\right] + O_{\mathbb{P}}\left(\frac{1}{r_n}\right)} \\ &+ r_n\Bigg[f_2\left(\widetilde{T}_n^{(0)}(x,K_0),\widetilde{T}_n^{(1)}(x,K_1),\widetilde{T}_n^{(2)}(x,K_2)\right) \\ &- f_2\left(1,\frac{1}{1-\gamma(x)},\frac{2}{(1-\gamma(x))(1-2\gamma(x))}\right)\Bigg], \end{aligned} \tag{12}$$

where $f_2(u,v,w) := -\frac{1}{2}\left(1 - \frac{v^2}{uw}\right)^{-1}$.

Then, combining Lemmas 1 and 2 with Theorem 1 and the delta method, we deduce that

$$r_n\left(\widehat{\gamma}_n(x) - \gamma(x)\right) \xrightarrow{D} N\left(\lambda(x)\widetilde{\mu}, V'\Sigma V\right),$$

for

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

where the term $\frac{\mathbf{1}\{\rho(x)<\gamma(x)\leq 0\}}{1-\gamma(x)}$ comes from the first term in the right-hand side of (12).

References

[1] Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J., 2004. *Statistics of Extremes – Theory and Applications*. Wiley.

[2] Billingsley, P., 1995. *Probability and Measure (Third edition)*. John Wiley & Sons.

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

[4] Coles, S., 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer.

[5] Daouia, A., Gardes, L., & Girard, S., 2013. On kernel smoothing for extremal quantile regression. *Bernoulli*, to appear.

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

[7] de Haan, L. & Ferreira, A., 2006. *Extreme Value Theory: An Introduction*. Springer.

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

[9] 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.

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

[11] Drees, H., 1998. On smooth statistical tail functionals. *Scandinavian Journal of Statistics*, 25, 187–210.

[12] 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.

[13] 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.

[14] 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.

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

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

[17] Gardes, L., Girard, S., & Lekina, A., 2010. Functional nonparametric estimation of conditional extreme quantiles. *Journal of Multivariate Analysis*, 101, 419–433.

[18] Goegebeur, Y., Guillou, A., & Schorgen, A., 2013. Nonparametric regression estimation of conditional tails - the random covariate case. *Statistics*, to appear.

[19] Hall, P., 1982. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society Series B*, 44, 37–42.

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

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

[22] Kagan, Y.Y., 1997. Seismic Moment-frequency relation for shallow earthquakes: regional comparison. *Journal of Geophysical Research*, 102, 2835–2852.

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

[24] 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.

[25] Pickands, J., 1975. Statistical inference using extreme order statistics. *Annals of Statistics*, 3, 119–131.

[26] 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.

[27] van der Vaart, A.W., 1998. *Asymptotic Statistics.* Cambridge University Press, Cambridge.

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

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

[30] Yao, Q., 1999. Conditional predictive regions for stochastic processes. *Technical report*, University of Kent at Canterbury.

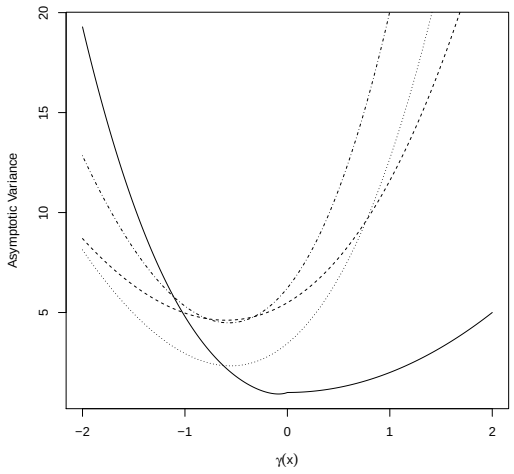

Figure 1: Asymptotic variance of $\hat{\gamma}_n(x)$ (solid), $\tilde{\gamma}_3^{RP,1}(x)$ (dashed), $\tilde{\gamma}_4^{RP,1}(x)$ (dotted) and $\tilde{\gamma}_4^{RP,2}(x)$ (dashed-dotted).

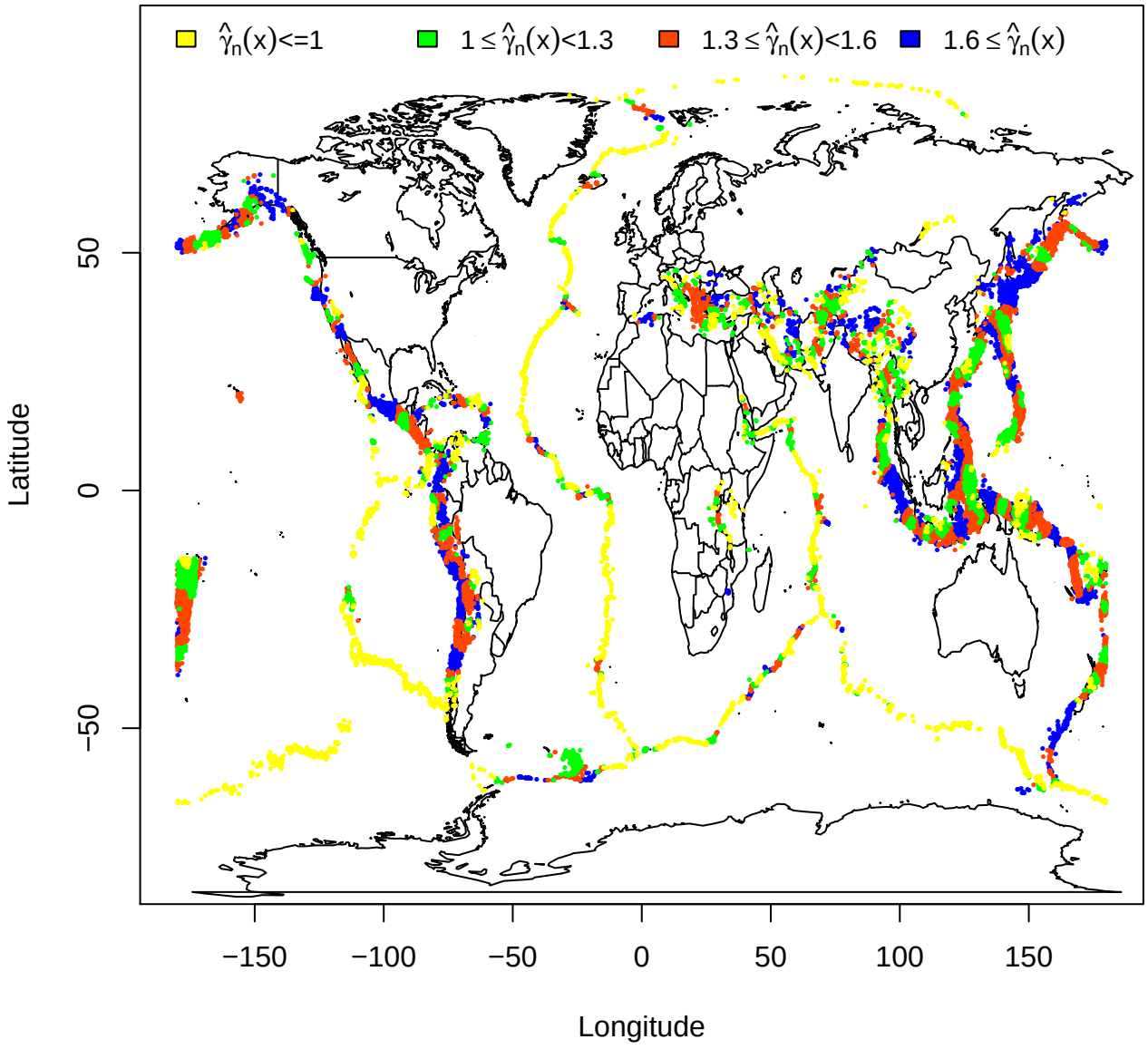

Figure 2: Local estimates of $\gamma(x)$ at locations where earthquakes have been observed.

$Reversed\,Burr(\eta(x), \tau(x), \lambda(x))$

$\lambda(x)$	$\hat{\gamma}_n(x)$		$\tilde{\gamma}_3^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,2}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	**0.0900** (0.0011)	**0.0133** (0.0003)	0.1189 (0.0017)	0.0239 (0.0007)	0.1027 (0.0017)	0.0171 (0.0005)	0.1128 (0.0018)	0.0211 (0.0006)
1	**0.1242** (0.0015)	**0.0253** (0.0006)	0.1781 (0.0023)	0.0510 (0.0012)	0.1668 (0.0025)	0.0415 (0.0010)	0.1675 (0.0025)	0.0443 (0.0011)
1.5	**0.1654** (0.0022)	**0.0449** (0.0011)	0.2357 (0.0032)	0.0866 (0.0020)	0.2278 (0.0032)	0.0722 (0.0017)	0.2201 (0.0032)	0.0725 (0.0018)
2	**0.1472** (0.0023)	**0.0353** (0.0010)	0.2250 (0.0028)	0.0765 (0.0017)	0.2362 (0.0034)	0.0746 (0.0018)	0.2242 (0.0034)	0.0721 (0.0018)

Table 1: Oracle method: performance of $\hat{\gamma}_n(x)$, $\tilde{\gamma}_3^{RP,1}(x)$, $\tilde{\gamma}_4^{RP,1}(x)$ and $\tilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$Reversed\,Burr(\eta(x), \tau(x), \lambda(x))$

$\lambda(x)$	$\hat{\gamma}_n(x)$		$\tilde{\gamma}_3^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,2}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	**0.1531** (0.0017)	**0.0398** (0.0009)	0.3220 (0.0028)	0.1513 (0.0024)	0.2168 (0.0026)	0.0831 (0.0022)	0.2574 (0.0031)	0.1196 (0.0032)
1	**0.2416** (0.0024)	**0.0916** (0.0017)	0.5458 (0.0038)	0.3872 (0.0048)	0.3105 (0.0026)	0.1388 (0.0025)	0.3248 (0.0033)	0.1666 (0.0041)
1.5	**0.3582** (0.0029)	**0.1843** (0.0027)	0.7362 (0.0042)	0.6586 (0.0069)	0.4130 (0.0029)	0.2154 (0.0031)	0.4116 (0.0033)	0.2274 (0.0038)
2	**0.4039** (0.0032)	**0.2168** (0.0031)	0.8213 (0.0045)	0.7942 (0.0080)	0.4843 (0.0033)	0.2840 (0.0042)	0.4775 (0.0037)	0.2901 (0.0048)

Table 2: Data driven method: performance of $\hat{\gamma}_n(x)$, $\tilde{\gamma}_3^{RP,1}(x)$, $\tilde{\gamma}_4^{RP,1}(x)$ and $\tilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$Reversed\,Burr(\eta(x), \tau(x), \lambda(x))$

$\tau(x)$	$\hat{\gamma}_n(x)$		$\tilde{\gamma}_3^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,2}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	**0.3725** (0.0055)	**0.2156** (0.0051)	0.4702 (0.0053)	0.3086 (0.0054)	0.5146 (0.0050)	0.3156 (0.0051)	0.4744 (0.0055)	0.2925 (0.0054)
1.5	**0.1746** (0.0023)	**0.0500** (0.0013)	0.2519 (0.0034)	0.0978 (0.0022)	0.2410 (0.0036)	0.0838 (0.0021)	0.2359 (0.0035)	0.0845 (0.0022)
2	**0.1375** (0.0018)	**0.0314** (0.0008)	0.2035 (0.0026)	0.0661 (0.0015)	0.1921 (0.0030)	0.0552 (0.0015)	0.1905 (0.0030)	0.0568 (0.0015)
2.5	**0.1220** (0.0016)	**0.0251** (0.0006)	0.1800 (0.0023)	0.0523 (0.0012)	0.1612 (0.0026)	0.0404 (0.0011)	0.1638 (0.0026)	0.0434 (0.0012)

Table 3: Oracle method: performance of $\hat{\gamma}_n(x)$, $\tilde{\gamma}_3^{RP,1}(x)$, $\tilde{\gamma}_4^{RP,1}(x)$ and $\tilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$Reversed\,Burr(\eta(x), \tau(x), \lambda(x))$

$\tau(x)$	$\hat{\gamma}_n(x)$		$\tilde{\gamma}_3^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,2}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	**0.7655** (0.0047)	**0.7454** (0.0083)	1.3225 (0.0041)	1.8623 (0.0110)	0.8997 (0.0028)	0.8653 (0.0054)	0.8457 (0.0034)	0.7880 (0.0064)
1.5	**0.4097** (0.0032)	**0.2442** (0.0033)	0.8691 (0.0043)	0.8955 (0.0078)	0.5017 (0.0031)	0.3148 (0.0040)	0.4925 (0.0035)	0.3207 (0.0047)
2	**0.3145** (0.0028)	**0.1514** (0.0024)	0.7148 (0.0041)	0.6265 (0.0062)	0.3972 (0.0030)	0.2133 (0.0034)	0.4003 (0.0035)	0.2343 (0.0045)
2.5	**0.2562** (0.0024)	**0.1047** (0.0018)	0.5910 (0.0038)	0.4506 (0.0050)	0.3393 (0.0027)	0.1655 (0.0027)	0.3523 (0.0033)	0.1911 (0.0036)

Table 4: Data driven method: performance of $\hat{\gamma}_n(x)$, $\tilde{\gamma}_3^{RP,1}(x)$, $\tilde{\gamma}_4^{RP,1}(x)$ and $\tilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$Strict\,Weibull(\lambda(x), \tau(x))$

$\tau(x)$	$\hat{\gamma}_n(x)$		$\tilde{\gamma}_3^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,2}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	0.2044 (0.0030)	0.0653 (0.0018)	0.2727 (0.0032)	0.1177 (0.0025)	**0.1743** (0.0024)	**0.0492** (0.0012)	0.1934 (0.0026)	0.0598 (0.0015)
1	0.0825 (0.0010)	0.0110 (0.0003)	0.0901 (0.0012)	0.0135 (0.0003)	**0.0688** (0.0010)	**0.0078** (0.0002)	0.0889 (0.0014)	0.0132 (0.0004)
1.5	**0.0830** (0.0014)	**0.0118** (0.0004)	0.1319 (0.0020)	0.0277 (0.0007)	0.1475 (0.0024)	0.0313 (0.0009)	0.1468 (0.0026)	0.0334 (0.0011)
2	**0.1106** (0.0021)	**0.0205** (0.0007)	0.1762 (0.0025)	0.0477 (0.0012)	0.2048 (0.0030)	0.0557 (0.0014)	0.1948 (0.0031)	0.0549 (0.0012)

Table 5: Oracle method: performance of $\hat{\gamma}_n(x)$, $\tilde{\gamma}_3^{RP,1}(x)$, $\tilde{\gamma}_4^{RP,1}(x)$ and $\tilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$Strict\,Weibull(\lambda(x), \tau(x))$

$\tau(x)$	$\hat{\gamma}_n(x)$		$\tilde{\gamma}_3^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,2}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	0.6783 (0.0032)	0.5012 (0.0042)	0.7332 (0.0043)	0.6409 (0.0064)	**0.3681** (0.0027)	**0.1819** (0.0024)	0.3535 (0.0028)	0.1827 (0.0029)
1	**0.1731** (0.0016)	**0.0433** (0.0007)	0.1909 (0.0021)	0.0594 (0.0013)	0.1771 (0.0024)	0.0582 (0.0019)	0.2362 (0.0030)	0.1016 (0.0031)
1.5	**0.1480** (0.0017)	**0.0376** (0.0009)	0.3228 (0.0028)	0.1456 (0.0024)	0.2893 (0.0030)	0.1231 (0.0028)	0.3273 (0.0038)	0.1746 (0.0044)
2	**0.1994** (0.0021)	**0.0638** (0.0013)	0.4481 (0.0031)	0.2533 (0.0032)	0.3667 (0.0031)	0.1777 (0.0033)	0.3896 (0.0039)	0.2264 (0.0050)

Table 6: Data driven method: performance of $\hat{\gamma}_n(x)$, $\tilde{\gamma}_3^{RP,1}(x)$, $\tilde{\gamma}_4^{RP,1}(x)$ and $\tilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$Strict\,Weibull(\lambda(x), \tau(x))$

$\lambda(x)$	$\hat{\gamma}_n(x)$		$\tilde{\gamma}_3^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,2}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	**0.0897** (0.0017)	**0.0140** (0.0005)	0.2727 (0.0032)	0.1177 (0.0025)	0.1743 (0.0024)	0.0492 (0.0012)	0.1934 (0.0026)	0.0598 (0.0015)
1	0.0919 (0.0018)	0.0144 (0.0005)	0.0901 (0.0012)	0.0135 (0.0003)	**0.0688** (0.0010)	**0.0078** (0.0002)	0.0889 (0.0014)	0.0132 (0.0004)
1.5	**0.0922** (0.0018)	**0.0144** (0.0006)	0.1319 (0.0020)	0.0277 (0.0007)	0.1475 (0.0024)	0.0313 (0.0009)	0.1468 (0.0026)	0.0334 (0.0011)
2	**0.0922** (0.0019)	**0.0147** (0.0006)	0.1762 (0.0025)	0.0477 (0.0012)	0.2048 (0.0030)	0.0557 (0.0014)	0.1948 (0.0031)	0.0549 (0.0015)

Table 7: Oracle method: performance of $\hat{\gamma}_n(x)$, $\tilde{\gamma}_3^{RP,1}(x)$, $\tilde{\gamma}_4^{RP,1}(x)$ and $\tilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$Strict\,Weibull(\lambda(x), \tau(x))$

$\lambda(x)$	$\hat{\gamma}_n(x)$		$\tilde{\gamma}_3^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,2}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	**0.1851** (0.0022)	**0.0542** (0.0013)	0.4650 (0.0034)	0.2705 (0.0036)	0.3264 (0.0031)	0.1441 (0.0030)	0.3371 (0.0036)	0.1699 (0.0041)
1	**0.2095** (0.0022)	**0.0677** (0.0013)	0.5220 (0.0033)	0.3313 (0.0036)	0.3684 (0.0031)	0.1756 (0.0031)	0.3775 (0.0038)	0.2071 (0.0046)
1.5	**0.2225** (0.0023)	**0.0749** (0.0015)	0.5474 (0.0035)	0.3607 (0.0041)	0.3875 (0.0032)	0.1902 (0.0034)	0.3931 (0.0038)	0.2189 (0.0046)
2	**0.2372** (0.0024)	**0.0858** (0.0016)	0.5586 (0.0034)	0.3774 (0.0040)	0.4172 (0.0032)	0.2197 (0.0035)	0.4258 (0.0039)	0.2553 (0.0049)

Table 8: Data driven method: performance of $\hat{\gamma}_n(x)$, $\tilde{\gamma}_3^{RP,1}(x)$, $\tilde{\gamma}_4^{RP,1}(x)$ and $\tilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$Burr(\eta(x), \tau(x), \lambda(x))$

$\lambda(x)$	$\hat{\gamma}_n(x)$		$\tilde{\gamma}_3^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,2}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	**0.0853** (0.0013)	**0.0119** (0.0003)	0.1166 (0.0015)	0.0221 (0.0005)	0.1177 (0.0018)	0.0223 (0.0007)	0.1331 (0.0020)	0.0289 (0.0009)
1	**0.0884** (0.0014)	**0.0128** (0.0004)	0.1282 (0.0017)	0.0264 (0.0006)	0.1305 (0.0021)	0.0272 (0.0008)	0.1420 (0.0023)	0.0327 (0.0010)
1.5	**0.0884** (0.0014)	**0.0129** (0.0004)	0.1245 (0.0016)	0.0245 (0.0006)	0.1273 (0.0021)	0.0257 (0.0008)	0.1393 (0.0022)	0.0312 (0.0010)
2	**0.0923** (0.0015)	**0.0144** (0.0005)	0.1186 (0.0017)	0.0224 (0.0006)	0.1145 (0.0018)	0.0209 (0.0007)	0.1314 (0.0020)	0.0277 (0.0008)

Table 9: Oracle method: performance of $\hat{\gamma}_n(x)$, $\tilde{\gamma}_3^{RP,1}(x)$, $\tilde{\gamma}_4^{RP,1}(x)$ and $\tilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

$Burr(\eta(x), \tau(x), \lambda(x))$

$\lambda(x)$	$\hat{\gamma}_n(x)$		$\tilde{\gamma}_3^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,1}(x)$		$\tilde{\gamma}_4^{RP,2}(x)$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.5	**0.1106** (0.0015)	**0.0200** (0.0006)	0.2551 (0.0027)	0.0999 (0.0021)	0.2262 (0.0031)	0.0849 (0.0026)	0.2931 (0.0038)	0.1460 (0.0042)
1	**0.1128** (0.0016)	**0.0208** (0.0006)	0.2948 (0.0031)	0.1251 (0.0024)	0.2310 (0.0032)	0.0856 (0.0027)	0.2830 (0.0039)	0.1344 (0.0041)
1.5	**0.1206** (0.0016)	**0.0234** (0.0006)	0.2913 (0.0030)	0.1234 (0.0023)	0.2397 (0.0033)	0.0905 (0.0025)	0.2950 (0.0040)	0.1440 (0.0041)
2	**0.1413** (0.0017)	**0.0315** (0.0007)	0.2641 (0.0027)	0.1098 (0.0021)	0.2513 (0.0032)	0.1026 (0.0027)	0.3110 (0.0040)	0.1637 (0.0045)

Table 10: Data driven method: performance of $\hat{\gamma}_n(x)$, $\tilde{\gamma}_3^{RP,1}(x)$, $\tilde{\gamma}_4^{RP,1}(x)$ and $\tilde{\gamma}_4^{RP,2}(x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.