A local moment type estimator for an extreme quantile in regression with random covariates

Yuri Goegebeur *
Armelle Guillou †
Michael Osmann ‡

Abstract. A conditional extreme quantile estimator is proposed in the presence of random covariates. It is based on an adaptation of the moment estimator introduced by Dekkers *et al.* (1989) in the classical univariate setting and thus it is valid in the domain of attraction of the extreme value distribution, i.e. whatever the sign of the extreme value index is. Asymptotic normality of the estimator is established under suitable assumptions, and its finite sample behaviour is evaluated with a small simulation study, where a comparison with an alternative estimator already proposed in the literature is provided. An illustration to a real dataset concerning the world catalogue of earthquake magnitudes is also proposed.

Key words and phrases: Extreme quantile; local estimation; max-domain of attraction.

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

1 Introduction

In many statistical applications it is necessary to make inferences about the tail of a distribution, where little data is available. For instance, one is interested in the probability that the maximum of n random variables exceeds a given threshold or, vice versa, one wants to determine a level such that the exceedance probability is below a given small value. Estimation of such a high quantile is not an easy task but the applications are numerous. For instance, in hydrology, engineers are interested in estimating the height of a dike such that the probability of a flooding in a given year is less than a very small probability p (cf. Dekkers and de Haan, 1989).
In this paper, we consider this problem of estimating a high quantile when some random covariate X is recorded simultaneously with the variable of interest Y. We study thus a regression problem, though here the primary interest is not in the conditional expectation, but rather in extreme conditional quantiles. More specifically, we assume that the conditional response distribution belongs to the max-domain of attraction of the extreme value distribution. In this case, the tail

*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).

index depends on the covariate x, say $\gamma(x)$, and is referred to in the following as the conditional tail index. This index is estimated locally within a narrow neighborhood of the point of interest in the covariate space. The estimation of this index $\gamma(x)$ has been recently studied in the literature: we can mention for instance Wang and Tsai (2009) who used a maximum likelihood approach, Daouia *et al.* (2011) who considered a fixed number of nonparametric conditional quantile estimators to estimate $\gamma(x)$ or Gardes and Stupfler (2013) who suggested a smoothed local Hill estimator. All these papers focus on the case where $\gamma(x) > 0$. The extension to the general case, $\gamma(x) \in \mathbb{R}$, has been considered in Daouia *et al.* (2013), who generalized Daouia *et al.* (2011), and Stupfler (2013) and Goegebeur *et al.* (2014) where an adjustment of the moment estimator, originally proposed by Dekkers *et al.* (1989), to this setting of local estimation has been proposed. Using the estimator introduced by Goegebeur *et al.* (2014), thus based on a non random threshold but by weighting the exceedances over the threshold by a kernel function, we propose an extreme conditional quantile estimator. We compare it with an alternative approach introduced by Daouia *et al.* (2013), based on Pickands-type estimators.
To illustrate our methodology, we will consider the world catalogue of earthquakes which contains information about earthquakes that have happened between 1976 and present. In this context, accurate estimation of extreme quantiles of the earthquake energy distribution is clearly of practical relevance since severe earthquakes may cause damages to certain structures and as such entail serious losses. With our method we can link the tail of this energy distribution to local factors, which allows us to differentiate the risks geographically. Such 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 in order to obtain a better determination of the premium levels or the estimation of the tail of the diamond value distribution conditional on the variables size and colour, to name but a few.
Our paper is organized as follows. In Section 2, we introduce our estimator and the required assumptions in order to establish its main asymptotic properties, which are stated in Section 3. A small simulation study together with a comparison with Daouia *et al.*'s (2013) estimator is provided in Section 4. Section 5 illustrates the methodology on the world catalogue of earthquake magnitudes. All the proofs are postponed to the appendix.

2 Extreme quantile estimator and assumptions

Let (X_i, Y_i), $i = 1, \ldots, n$, denote independent copies of the random vector $(X, Y) \in \mathbb{R}^p \times \mathbb{R}_+$, where the conditional tail quantile function of Y given $X = x$, denoted

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

satisfies

$$\lim_{t \to \infty} \frac{U(ty; x) - U(t; x)}{a(t; x)} = D_{\gamma(x)}(y) := \int_1^y u^{\gamma(x)-1} du, \quad y > 0, \tag{1}$$

where $a(.; x)$ is a positive auxiliary function. Consider a level $\beta_n \to 0$ such that $\frac{\overline{F}(\omega_n; x)}{\beta_n} \to \infty$ as $n \to \infty$ where $\overline{F}(y; x) := 1 - F(y; x)$ and ω_n is a local non-random threshold sequence satisfying

$\omega_n \to y^*(x)$ for $n \to \infty$, where $y^*(x) := \sup\{y : F(y;x) < 1\}$ is the right endpoint of $F(y;x)$. Motivated by convergence (1), our extreme quantile estimator has the classical following form (see e.g. de Haan and Ferreira, 2006; Dekkers *et al.*, 1989)

$$\widehat{U}\left(\frac{1}{\beta_n};x\right) := \omega_n + \widehat{a}_n(x)\frac{\left(\frac{\widehat{\overline{F}}(\omega_n;x)}{\beta_n}\right)^{\widehat{\gamma}_n(x)} - 1}{\widehat{\gamma}_n(x)} \tag{2}$$

i.e. it requires an estimation of the conditional tail index $\gamma(x)$, of the scale parameter $a\left(\frac{1}{\overline{F}(\omega_n;x)};x\right)$ and of the tail $\overline{F}(\omega_n;x)$. All these estimators will be based on the following basic building block

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

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 ln denotes the natural logarithm. Recall that $\omega_n \to y^*(x)$ for $n \to \infty$. Although this condition reduces to $\omega_n \to \infty$ when $\gamma(x) > 0$, it is more problematic in the case $\gamma(x) < 0$ since it requires that the nonrandom threshold sequence (ω_n) converges to the finite conditional right endpoint $y^*(x)$, which is unknown. To circumvent this issue in practice, we choose for ω_n, as usual in the extreme value literature, the $(k+1)$'th largest response for which the covariate is contained in the ball $B(x;h_n)$, i.e. a random threshold.

Assuming that X has a density function $g(x)$ for $x \in \mathbb{R}^p$, we define an estimator for the conditional survival function $\overline{F}(\omega_n;x)$ by using the classical kernel density estimator proposed by Parzen (1962) in the univariate framework:

$$\widehat{\overline{F}}(\omega_n;x) := \frac{T_n^{(0)}(x,K)}{\widehat{g}_n(x)} := \frac{T_n^{(0)}(x,K)}{\frac{1}{n}\sum_{i=1}^n K_{h_n}(x-X_i)}. \tag{4}$$

Next, inspired by the scale estimator in the univariate framework (see de Haan and Ferreira, 2006; Dekkers *et al.*, 1989), we introduce the local scale estimator

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

with kernel functions K_0, K_1 and K_2. Finally, concerning the estimation of the conditional extreme value index, we propose to use the estimator recently introduced in Goegebeur *et al.* (2014) in the same setting, that is

$$\widehat{\gamma}_n(x) = \frac{T_n^{(1)}(x,K_1)}{T_n^{(0)}(x,K_0)} + 1 - \frac{1}{2}\left(1 - \frac{\left(\frac{T_n^{(1)}(x,K_1)}{T_n^{(0)}(x,K_0)}\right)^2}{\frac{T_n^{(2)}(x,K_2)}{T_n^{(0)}(x,K_0)}}\right)^{-1}. \tag{6}$$

Our aim in the next section is to establish the asymptotic normality of the extreme quantile estimator (2) using the estimators for the three parameters defined in (4), (5) and (6). As usual in extreme value theory, such a result requires some conditions, in particular the following second-order condition, which specifies the rate of convergence of the left-hand side in (1) to its limit $D_{\gamma(x)}(y)$.

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*

$$\lim_{t\to\infty} \frac{\frac{U(ty;x)-U(t;x)}{a(t;x)} - D_{\gamma(x)}(y)}{A(t;x)} = \int_1^y s^{\gamma(x)-1} \int_1^s u^{\rho(x)-1} du ds, \tag{7}$$

for all $y > 0$.

Since our basic statistic (3) is expressed in terms of log-excesses, we need a reformulation of (7) in terms of $\ln U(y;x)$. To this aim we introduce some notation (see 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

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

while for $\gamma(x) > 0$ we let

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

and for $\gamma(x) \leq 0$ we define an asymptotically equivalent version of $a(.;x)$

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

Now we need some additional conditions due to the regression framework. First, the density function $g(x)$ is assumed to satisfy a Hölder-type condition.

Assumption $(\mathcal{G})$ *There exist $c_g > 0$ and $\eta_g > 0$ such that $|g(x) - g(z)| \leq c_g \|x - z\|^{\eta_g}$ for all $x, z \in \mathbb{R}^p$, where $\|\cdot\|$ is some norm on $\mathbb{R}^p$.*

Then, the kernel functions have to satisfy the following assumption which is now standard in local estimation, see e.g. Daouia *et al.* (2013) or Goegebeur *et al.* (2014).

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

Finally, we need to control the oscillation of $U(t;x)$ when considered as a function of the covariate x. The formulation of this condition requires the introduction of the conditional expectation

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

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

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

Note that this assumption implies in particular the continuity of $\gamma(.)$. It is written in terms of conditional expectations because in this elegant way it covers the three subclasses of the max-domain of attraction, corresponding to $\gamma(x) < 0$, $\gamma(x) = 0$ and $\gamma(x) > 0$. However, as illustrated in Appendix B for $\gamma(x) > 0$, this condition can be satisfied by imposing further conditions on the functions involved as well as on the sequences h_n and ω_n.

We have now all the ingredients to prove the asymptotic normality of our extreme quantile estimator (2).

3 Asymptotic properties

Instead of working with the main statistic (3), we work with a standardized version

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

which constitutes the key element of the vector

$$\mathbb{T}_n^\top := \left[\widetilde{T}_n^{(0)}(x,K_0), \widetilde{T}_n^{(1)}(x,K_1), \widetilde{T}_n^{(2)}(x,K_2)\right].$$

As a preliminary result, we establish the joint asymptotic normality of $\mathbb{T}_n$ under suitable conditions.

Theorem 1 *Assume $(\mathcal{R})$ with $\rho(x)\neq\gamma(x)$, $(\mathcal{G})$ and $(\mathcal{F})$ and that the kernel functions K_0, K_1 and K_2 satisfy $(\mathcal{K})$. Under the additional conditions, as $n\to\infty$,*

- $h_n \to 0$,
- $\omega_n \to y^*(x)$,
- $nh_n^p\overline{F}(\omega_n;x) \to \infty$,
- $\sqrt{nh_n^p\overline{F}(\omega_n;x)}B\left(\frac{1}{\overline{F}(\omega_n;x)};x\right) \to \lambda(x)$ *for some constant* $\lambda(x)\in\mathbb{R}$,
- $nh_n^p\overline{F}(\omega_n;x)\max\left\{h_n^{2\eta_g}, \Phi^2(\omega_n,h_n;x)\right\} \to 0$,

for all $x\in\mathbb{R}^p$ where $g(x)>0$ we have that

$$\sqrt{nh_n^p\overline{F}(\omega_n;x)}\left[\mathbb{T}_n-\psi\right] \xrightarrow{D} (P,Q,R)^\top \sim N_3(\zeta,\Sigma),$$

where

$$\psi^\top := \begin{cases} \left[1,\gamma(x),2\gamma^2(x)\right], & \gamma(x)>0 \\ \left[1,\frac{1}{1-\gamma(x)},\frac{2}{(1-\gamma(x))(1-2\gamma(x))}\right], & \gamma(x)\le 0 \end{cases},$$

the elements of Σ are given by

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

and

$$\zeta^\top := \begin{cases} \frac{\lambda(x)}{(1-\widetilde{\rho}(x))^2}\left[0, 1-\widetilde{\rho}(x), 2\gamma(x)\left(2-\widetilde{\rho}(x)\right)\right], & \gamma(x) > 0 \\ \\ \lambda(x)\left[0, b^{(1)}_{\gamma(x),\rho(x)}, b^{(2)}_{\gamma(x),\rho(x)}\right], & \gamma(x) \leq 0 \end{cases},$$

with

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

and

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

The proof of Theorem 1 is omitted since it can be easily deduced from the one of Theorem 2 in Goegebeur *et al.* (2014). Theorem 1 can be used in order to prove the convergence in distribution of our estimator of the conditional survival function $\overline{F}(\omega_n; x)$.

Lemma 1 *Under the conditions of Theorem 1, we have that*

$$\sqrt{nh_n^p\overline{F}(\omega_n;x)}\left[\frac{\widehat{\overline{F}}(\omega_n;x)}{\overline{F}(\omega_n;x)} - 1\right] \stackrel{D}{\to} P.$$

Note that this result is in fact valid without assumption $(\mathcal{R})$.

Since the local scale estimator $\widehat{a}_n(x)$ and the extreme quantile estimator $\widehat{U}\left(\frac{1}{\beta_n}; x\right)$ depend on the local non-random threshold ω_n, we need the following auxiliary lemma, which under $(\mathcal{R})$ ensures that in some sense a second-order condition is also valid for $U\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right) - \omega_n$.

Lemma 2 *Let $F(y;x)$ be a distribution function with right endpoint $y^*(x)$, where the tail quantile function $U(t;x)$ satisfies $(\mathcal{R})$. Then*

$$\lim_{t\to y^*(x)} \frac{U\left(\frac{1}{\overline{F}(t;x)}; x\right) - t}{a\left(\frac{1}{\overline{F}(t;x)}; x\right) A\left(\frac{1}{\overline{F}(t;x)}; x\right)} = 0.$$

Note that this lemma immediately implies the slightly weaker result

$$\lim_{t\to y^*(x)} \frac{U\left(\frac{1}{\overline{F}(t;x)}; x\right) - t}{a\left(\frac{1}{\overline{F}(t;x)}; x\right) B\left(\frac{1}{\overline{F}(t;x)}; x\right)} = 0,$$

which is in fact the one that we need. The asymptotic normality of our local scale estimator $\widehat{a}_n(x)$ can now be established using Theorem 1 and Lemma 2.

Theorem 2 *Under the conditions of Theorem 1, we have that*

$$\sqrt{nh_n^p \overline{F}(\omega_n; x)} \left[\frac{\widehat{a}_n(x)}{a\left(\frac{1}{\overline{F}(\omega_n;x)}; x\right)} - 1 \right] \xrightarrow{D} \Lambda$$

where

$$\Lambda := \begin{cases} -2P + \frac{3}{\gamma(x)}Q - \frac{1}{2\gamma^2(x)}R - \frac{\lambda(x)}{\gamma(x)}, & \gamma(x) > 0 \\ -2(1-\gamma(x))P + (1-\gamma(x))(3-4\gamma(x))Q - \frac{1}{2}(1-\gamma(x))(1-2\gamma(x))^2 R + \lambda(x) b^{(3)}_{\gamma(x),\rho(x)}, & \gamma(x) \le 0 \end{cases}$$

with

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

The joint asymptotic normality of $\widehat{\overline{F}}(\omega_n; x)$, $\widehat{a}_n(x)$ and $\widehat{\gamma}_n(x)$, when properly normalized, is easily established in Theorem 3 by noting that all the three estimators converge in distribution to linear combinations of the three jointly normally distributed random variables P, Q and R from Theorem 1.

Theorem 3 *Under the assumptions of Theorem 1, we have that*

$$\sqrt{nh_n^p \overline{F}(\omega_n; x)} \left[\begin{array}{c} \frac{\widehat{\overline{F}}(\omega_n;x)}{\overline{F}(\omega_n;x)} - 1 \\ \frac{\widehat{a}_n(x)}{a\left(\frac{1}{\overline{F}(\omega_n;x)};x\right)} - 1 \\ \widehat{\gamma}_n(x) - \gamma(x) \end{array} \right] \xrightarrow{D} \begin{pmatrix} P \\ \Lambda \\ \Gamma \end{pmatrix}$$

with

$$\begin{pmatrix} P \\ \Lambda \\ \Gamma \end{pmatrix} \sim N_3\left(\lambda(x)\mu, C\Sigma C^\top\right)$$

where

$$\mu = \frac{1}{\gamma(x)\left(1-\widetilde{\rho}(x)\right)^2} \left[\begin{array}{c} 0 \\ -\widetilde{\rho}^2(x) \\ \gamma(x) + \widetilde{\rho}(x) - \gamma(x)\widetilde{\rho}(x) \end{array} \right]$$

and

$$C = \begin{pmatrix} 1 & 0 & 0 \\ -2 & \frac{3}{\gamma(x)} & -\frac{1}{2\gamma^2(x)} \\ 1-\gamma(x) & 1 - \frac{2}{\gamma(x)} & \frac{1}{2\gamma^2(x)} \end{pmatrix}$$

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

$$\mu = \left[\begin{array}{c} 0 \\ (1-\gamma(x))(3-4\gamma(x)) b^{(1)}_{\gamma(x),\rho(x)} - \frac{1}{2}(1-\gamma(x))(1-2\gamma(x))^2 b^{(2)}_{\gamma(x),\rho(x)} + b^{(3)}_{\gamma(x),\rho(x)} \\ -2(1-\gamma(x))^2(1-2\gamma(x)) b^{(1)}_{\gamma(x),\rho(x)} + \frac{1}{2}(1-\gamma(x))^2(1-2\gamma(x))^2 b^{(2)}_{\gamma(x),\rho(x)} + \frac{\mathbf{1}\{\rho(x)<\gamma(x)\le 0\}}{1-\gamma(x)} \end{array} \right]$$

and

$$C=\begin{pmatrix} 1 & 0 & 0 \\ -2(1-\gamma(x)) & (1-\gamma(x))(3-4\gamma(x)) & -\frac{1}{2}(1-\gamma(x))(1-2\gamma(x))^2 \\ (1-\gamma(x))(1-2\gamma(x)) & -2(1-\gamma(x))^2(1-2\gamma(x)) & \frac{1}{2}(1-\gamma(x))^2(1-2\gamma(x))^2 \end{pmatrix}$$

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

For details about establishing the asymptotic normality of $\widehat{\gamma}_n(x)$ we refer to Goegebeur *et al.* (2014). Now, we are ready to state our main result, that is the asymptotic normality of our extreme quantile estimator (2). To this aim, we first introduce the notations

$$q_\gamma(t) := \int_1^t s^{\gamma-1}\ln s ds$$

for $t>1$ and $\gamma_-(x) := \min\{0,\gamma(x)\}$.

Theorem 4 *Under the assumptions of Theorem 1, and further assuming that* $\rho(x)<0$ *or* ($\rho(x)=0$ *and* $\gamma(x)<0$) *and*

- $\frac{\overline{F}(\omega_n;x)}{\beta_n}\to\infty$,
- $\ln\left(\frac{\overline{F}(\omega_n;x)}{\beta_n}\right)=o\left(\sqrt{nh_n^p\overline{F}\left(\omega_n;x\right)}\right)$,

we have that

$$\sqrt{nh_n^p\overline{F}\left(\omega_n;x\right)}\,\frac{\widehat{U}\left(\frac{1}{\beta_n};x\right)-U\left(\frac{1}{\beta_n};x\right)}{q_{\gamma(x)}\left(\frac{\overline{F}(\omega_n;x)}{\beta_n}\right)a\left(\frac{1}{\overline{F}(\omega_n;x)};x\right)}\xrightarrow{D}\Gamma-\gamma_-(x)\Lambda-\lambda(x)\frac{\gamma_-(x)}{\gamma_-(x)+\rho(x)}\mathbf{1}\{\gamma(x)<\rho(x)\leq 0\},$$

i.e.

$$\sqrt{nh_n^p\overline{F}\left(\omega_n;x\right)}\,\frac{\widehat{U}\left(\frac{1}{\beta_n};x\right)-U\left(\frac{1}{\beta_n};x\right)}{q_{\gamma(x)}\left(\frac{\overline{F}(\omega_n;x)}{\beta_n}\right)a\left(\frac{1}{\overline{F}(\omega_n;x)};x\right)}\xrightarrow{D}N\left(\lambda(x)\widetilde{\mu},\sigma^2\right),$$

where

$$\widetilde{\mu}:=\begin{cases}\frac{\gamma(x)+\widetilde{\rho}(x)-\gamma(x)\widetilde{\rho}(x)}{\gamma(x)(1-\widetilde{\rho}(x))^2}, & \gamma(x)>0\\ \frac{\rho(x)(1-\gamma(x))}{(\gamma(x)+\rho(x))(1-\gamma(x)-\rho(x))(1-2\gamma(x)-\rho(x))}, & \gamma(x)<\rho(x)\leq 0\\ -\frac{\gamma(x)(1-3\gamma^2(x))}{(1-\gamma(x))(1-2\gamma(x))(1-3\gamma(x))}, & \rho(x)<\gamma(x)\leq 0\end{cases}$$

and

$$\sigma^2 := var(\Gamma)+\gamma_-^2(x)var(\Lambda)-2\gamma_-(x)cov(\Gamma,\Lambda).$$

In the case where all the kernel functions in Theorem 4 are equal, the asymptotic variance σ^2, reduces to

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

which is, apart from $\frac{\|K\|_2^2}{g(x)}$, the same as for the classical moment estimator in the univariate framework (de Haan and Ferreira, 2006). This is also the case for the asymptotic mean.

4 A small simulation study

In the simulation experiment, we compare our estimator with the following estimator proposed by Daouia *et al.* (2013):

$$\widehat{q}_n^{RP}(\beta_n; x) := \widehat{q}_n(\alpha_n; x) + D_{\widehat{\gamma}_n^{RP}(x)}\left(\frac{\alpha_n}{\beta_n}\right)\widehat{a}_n^{RP}(x),$$

where

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

with $\widehat{\overline{F}}(y;x)$ defined in (4), and

$$\widehat{\gamma}_n^{RP}(x) := -\frac{1}{\ln 3}\ln\left(\frac{\widehat{q}_n(\alpha_n; x) - \widehat{q}_n\left(\frac{1}{3}\alpha_n; x\right)}{\widehat{q}_n\left(\frac{1}{3}\alpha_n; x\right) - \widehat{q}_n\left(\frac{1}{9}\alpha_n; x\right)}\right)$$

while

$$\widehat{a}_n^{RP}(x) := \frac{1}{D_{\widehat{\gamma}_n^{RP}(x)}\left(\frac{1}{3}\right)}\left(\frac{1}{3}\right)^{\widehat{\gamma}_n^{RP}(x)}\left(\widehat{q}_n(\alpha_n; x) - \widehat{q}_n\left(\frac{1}{3}\alpha_n; x\right)\right),$$

for $\alpha_n \in (0, 1)$. This estimator belongs to a much larger class of estimators introduced in Daouia *et al.* (2013), but we compare our estimator with this specific one since it has a good performance over a wide range of distributions.

We assume that X follows the uniform distribution on $(0, 1)$ and we let $\varphi : [0, 1] \to \mathbb{R}^+$ and $\psi : [0, 1] \to \mathbb{R}^+$ be the functions defined by

$$\begin{aligned}
\varphi(x) &= \frac{1}{2}\left(\frac{1}{10} + \sin(\pi x)\right)\left(\frac{11}{10} - \frac{1}{2}e^{-64\left(x-\frac{1}{2}\right)^2}\right)\\
\psi(x) &= \frac{1}{4}\left(1 + e^{-60\left(x-\frac{1}{4}\right)^2}\mathbf{1}\left\{x \in \left[0, \frac{1}{3}\right]\right\} + e^{-\frac{5}{12}}\mathbf{1}\left\{x \in \left(\frac{1}{3}, \frac{2}{3}\right]\right\}\right.\\
&\qquad \left. +(5 - 6x)\left(e^{-\frac{5}{12}}1\!\!1_{\{x \in (\frac{2}{3}, \frac{5}{6}]\}} - \mathbf{1}\left\{x \in \left(\frac{5}{6}, 1\right]\right\}\right)\right).
\end{aligned}$$

Now, consider the four following conditional distributions of Y given $X = x$:

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

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

  for which $\gamma(x) = -\frac{1}{\lambda(x)\tau(x)}$ and $\rho(x) = -\frac{1}{\lambda(x)}$. Here we always use $\eta(x) = 3$, $y^*(x) = 5$ and $\lambda(x) = 0.25, 0.5, 0.75, 1$ for the two functions

  $$\gamma_1(x) = -\varphi(x) \quad \text{and} \quad \gamma_2(x) = -\psi(x).$$

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

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

for which $\gamma(x) = 0$ and $\rho(x) = 0$. Note that this distribution does not fit 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 $\lambda(x) = 0.25, 0.5, 0.75, 1$ for the two functions

$$\tau_1(x) = \frac{1}{\varphi(x)} \quad \text{and} \quad \tau_2(x) = \frac{1}{\psi(x)}.$$

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

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

for which $\gamma(x) = \frac{1}{\lambda(x)\tau(x)}$ and $\rho(x) = -\frac{1}{\lambda(x)}$. We let $\eta(x) = 1$, and consider the cases $\lambda(x) = 0.25, 0.5, 0.75, 1$ for the two functions

$$\gamma_1(x) = \varphi(x) \quad \text{and} \quad \gamma_2(x) = \psi(x).$$

- The Fréchet($\alpha(x)$) distribution,

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

for which $\gamma(x) = \frac{1}{\alpha(x)}$ with again the same gamma functions as for the Burr.

For the practical implementation of our estimator, we have to select the parameters h_n and ω_n, where we for the latter choose the $(k+1)$'th largest response for which the covariate is contained in the ball $B(x, h_n)$. This threshold selection is quite usual now in the extreme value literature. In all the cases, the bi-quadratic kernel function $K(x) = \frac{15}{16}\left(1 - x^2\right)^2 \mathbf{1}\{x \in [-1,1]\}$ is used. The selection of (h_n, k) is done in two steps by a completely data driven method, which has been used several times in the recent literature dedicated to local estimation of parameters of interest in an extreme value setting (see for instance Daouia *et al.*, 2013; Goegebeur *et al.*, 2014). First the bandwidth parameter h_n is selected using a cross validation criterion, which consists of selecting h_n as

$$h_c := \underset{h_n \in \mathcal{H}}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\mathbf{1}\left\{Y_i \le Y_j\right\} - \widehat{F}_{n,-i}\left(Y_j; X_i\right) \right)^2, \tag{8}$$

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

$$\widehat{F}_{n,-i}\left(y;x\right) := \frac{\sum_{j=1, j\neq i}^{n} K_{h_n}\left(x - X_j\right) \mathbf{1}\left\{Y_j \le y\right\}}{\sum_{j=1, j\neq i}^{n} K_{h_n}\left(x - X_j\right)}.$$

Next for each point of interest in the covariate space $z_1, \ldots, z_L$, we do the following

- Compute $\widehat{U}\left(\frac{1}{\beta_n}; z_\ell\right)$ for $k = 5, \ldots, n^* - 1$, where n^* is the number of observations contained in the ball $B(x, h_n)$;

- Split the estimates $\widehat{U}\left(\frac{1}{\beta_n}; z_\ell\right)$ into blocks of size $\lfloor \sqrt{n^*} \rfloor$, and compute the standard deviation of each block;

- Select the median of the estimates $\widehat{U}\left(\frac{1}{\beta_n}; z_\ell\right)$ from the block with the smallest standard deviation.

The same approach is applied to the estimator $\widehat{q}_n^{RP}(\beta_n; x)$ by computing it for $\alpha_n = \frac{k}{n^*}$ with $k = 5, \ldots, n^* - 1$. In all cases the grid of values for h_n is chosen as $\mathcal{H} = \{0.05, 0.075, \ldots, 0.3\}$.

For all the distributions we simulate $N = 500$ samples of size $n = 1000$ and estimate the $(1 - \beta_n)$'th quantile at $z_1, \ldots, z_L$ equidistantly spaced points in the interval $[0.1, 0.9]$ using $\widehat{U}\left(\frac{1}{\beta_n}; z_\ell\right)$ and $\widehat{q}_n^{RP}(\beta_n; z_\ell)$ with β_n chosen as either $\beta_n = \frac{1}{1200}$ or $\beta_n = \frac{1}{2000}$. As a measure of efficiency, we compute the mean squared error

$$MSE\left(\widehat{Q}\left(\frac{1}{\beta_n}; \cdot\right)\right) = \frac{1}{LN} \sum_{i=1}^{N} \sum_{\ell=1}^{L} \left(\frac{\widehat{Q}_i(1 - \beta_n; z_\ell)}{U\left(\frac{1}{\beta_n}; z_\ell\right)} - 1 \right)^2,$$

where $\widehat{Q}_i(1 - \beta_n; z_\ell)$ is either $\widehat{U}\left(\frac{1}{\beta_n}; z_\ell\right)$ or $\widehat{q}_n^{RP}(\beta_n; z_\ell)$ in the i'th simulation run at position z_ℓ and $L = 41$.

The results of the simulations are summarized in Tables 1-8, from which the following conclusions can be made:

- For the reversed Burr distribution, our estimator seems to always outperform the benchmark estimator by Daouia *et al.* (2013), although in some cases the results are quite close;

- In the case of the strict Weibull distribution, the benchmark estimator appears to have a slightly better performance, except for the case where $\lambda(x) = 1$ and $\beta_n = 1/1200$. However, for such a distribution, our assumptions are violated, but our estimator still remains very competitive. In that sense, we can say that our estimator is robust with respect to a violation of the conditions;

- For the Burr and Fréchet distributions, our estimator is superior in almost all cases.

	Reversed Burr$(\eta(x), \tau(x), \lambda(x))$			
	$\gamma_1(x)$		$\gamma_2(x)$	
$\lambda(x)$	$\widehat{U}(1200; x)$	$\widehat{q}_n^{RP}(1/1200; x)$	$\widehat{U}(1200; x)$	$\widehat{q}_n^{RP}(1/1200; x)$
0.25	0.00073	0.00109	0.00056	0.00087
0.50	0.00106	0.00119	0.00094	0.00107
0.75	0.00110	0.00115	0.00102	0.00108
1.00	0.00099	0.00107	0.00093	0.00102

Table 1: Performance of $\widehat{U}(1200; x)$ and $\widehat{q}_n^{RP}(1/1200; x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

	Reversed Burr$(\eta(x), \tau(x), \lambda(x))$			
	$\gamma_1(x)$		$\gamma_2(x)$	
$\lambda(x)$	$\widehat{U}(2000; x)$	$\widehat{q}_n^{RP}(1/2000; x)$	$\widehat{U}(2000; x)$	$\widehat{q}_n^{RP}(1/2000; x)$
0.25	0.00085	0.00121	0.00067	0.00097
0.50	0.00124	0.00132	0.00111	0.00119
0.75	0.00129	0.00128	0.00120	0.00120
1.00	0.00117	0.00118	0.00112	0.00112

Table 2: Performance of $\widehat{U}(2000; x)$ and $\widehat{q}_n^{RP}(1/2000; x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

5 Real data analysis

In this section, we illustrate our methodology on a real dataset concerning the world catalogue of earthquake magnitudes. 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 *et al.*, 1981; Ekström *et al.*, 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 extreme value index and of extreme conditional quantiles, we have to take into account that earthquakes happen at a random location. Hence, this dataset is very well suited for illustration of our local estimation method. Goegebeur *et al.* (2014) already considered this dataset and they estimated the conditional extreme value index $\gamma(x)$. We refer to their paper for more details about their estimation.
Now we look at the problem of estimating an extreme quantile with $\beta_n = 1/10000$ which is locally extreme since we have a total of approximately 33000 earthquakes, which are spread around the entire earth. We classified the severity of earthquakes by converting the seismic moment to the moment magnitude scale, which is the scale that is used in public when an earthquake has happened. The seismic moment can be converted to the moment magnitude scale by the transformation

$$M_W = \frac{2}{3} \log_{10} (M_S) - \frac{32}{3},$$

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

$\lambda(x)$	$\tau_1(x)$		$\tau_2(x)$	
	$\widehat{U}(1200;x)$	$\widehat{q}_n^{RP}(1/1200;x)$	$\widehat{U}(1200;x)$	$\widehat{q}_n^{RP}(1/1200;x)$
0.25	0.06342	0.05464	0.06346	0.05202
0.50	0.06075	0.05506	0.05931	0.05272
0.75	0.05803	0.05735	0.05512	0.05429
1.00	0.05712	0.05938	0.05238	0.05490

Table 3: Performance of $\widehat{U}(1200;x)$ and $\widehat{q}_n^{RP}(1/1200;x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

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

$\lambda(x)$	$\tau_1(x)$		$\tau_2(x)$	
	$\widehat{U}(2000;x)$	$\widehat{q}_n^{RP}(1/2000;x)$	$\widehat{U}(2000;x)$	$\widehat{q}_n^{RP}(1/2000;x)$
0.25	0.07651	0.06326	0.07486	0.06059
0.50	0.07418	0.06403	0.07200	0.06170
0.75	0.07289	0.06671	0.06933	0.06350
1.00	0.07299	0.06905	0.06854	0.06430

Table 4: Performance of $\widehat{U}(2000;x)$ and $\widehat{q}_n^{RP}(1/2000;x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

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

$\lambda(x)$	$\gamma_1(x)$		$\gamma_2(x)$	
	$\widehat{U}(1200;x)$	$\widehat{q}_n^{RP}(1/1200;x)$	$\widehat{U}(1200;x)$	$\widehat{q}_n^{RP}(1/1200;x)$
0.25	0.24436	0.43920	0.20042	0.34102
0.50	0.28314	0.39717	0.23661	0.35427
0.75	0.34517	0.40045	0.28202	0.37245
1.00	0.40774	0.39967	0.33593	0.37098

Table 5: Performance of $\widehat{U}(1200;x)$ and $\widehat{q}_n^{RP}(1/1200;x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

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

$\lambda(x)$	$\gamma_1(x)$		$\gamma_2(x)$	
	$\widehat{U}(2000;x)$	$\widehat{q}_n^{RP}(1/2000;x)$	$\widehat{U}(2000;x)$	$\widehat{q}_n^{RP}(1/2000;x)$
0.25	0.31726	0.51694	0.25635	0.39776
0.50	0.37398	0.45355	0.30402	0.41380
0.75	0.46879	0.45990	0.36716	0.43212
1.00	0.56478	0.46306	0.44545	0.43051

Table 6: Performance of $\widehat{U}(2000;x)$ and $\widehat{q}_n^{RP}(1/2000;x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

Fréchet($\alpha(x)$)			
$\gamma_1(x)$		$\gamma_2(x)$	
$\widehat{U}(1200;x)$	$\widehat{q}_n^{RP}(1/1200;x)$	$\widehat{U}(1200;x)$	$\widehat{q}_n^{RP}(1/1200;x)$
0.31379	0.37742	0.26260	0.31994

Table 7: Performance of $\widehat{U}(1200;x)$ and $\widehat{q}_n^{RP}(1/1200;x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

Fréchet($\alpha(x)$)			
$\gamma_1(x)$		$\gamma_2(x)$	
$\widehat{U}(2000;x)$	$\widehat{q}_n^{RP}(1/2000;x)$	$\widehat{U}(2000;x)$	$\widehat{q}_n^{RP}(1/2000;x)$
0.41501	0.43438	0.33491	0.37438

Table 8: Performance of $\widehat{U}(2000;x)$ and $\widehat{q}_n^{RP}(1/2000;x)$. The results are averaged over $N = 500$ simulations with $n = 1000$ observations.

where M_W is the moment magnitude scale and M_S is the seismic moment. A value of M_W greater than 8 indicates a severe earthquake which may cause important damages and losses. This leads to Figure 1 for $\widehat{U}(10000;x)$. This figure is very similar to Figure 2 in Goegebeur *et al.* (2014) concerning the conditional extreme value index $\gamma(x)$. Also a similar figure can be obtained with the Daouia *et al.* (2013) estimator, even though their estimates seem to be a bit lower. Generally, for the mid-ocean ridges, the moment magnitude scale of the quantile seems to be smaller than 7 whereas a larger value is expected for the other areas, in particular at the northern part of Japan, Indonesia, southern part of Mexico, and various places along the western coast of South America. As a comparison, under the assumption that the seismic moment of the earthquakes are also identically distributed, i.e. the covariate effect of location of the earthquake is ignored, an estimate of the $U(10000)$ quantile was found to be 9.0487 using the univariate moment estimator and associated quantile estimator. To emphasize the fact that one has to differentiate the tail heaviness of the earthquake energy distribution geographically, we calculated also a bootstrap 95% confidence interval for this extreme quantile estimate, which leads to $(8.97; 9.26)$. Figure 2 shows the earthquake positions for which the extreme quantile estimate falls inside this confidence interval. As is clear from Figure 2, the confidence interval covers only very few quantiles that we estimated. This can be explained by the fact that the seismic moment of the earthquakes are not identically distributed, so this estimate has a limited practical relevance.

Acknowledgements

The authors are grateful to two anonymous referees for their helpful and constructive comments on the preliminary version of the paper. This work was supported by a research grant (VKR023480) from VILLUM FONDEN and an international project for scientific cooperation (PICS-6416).

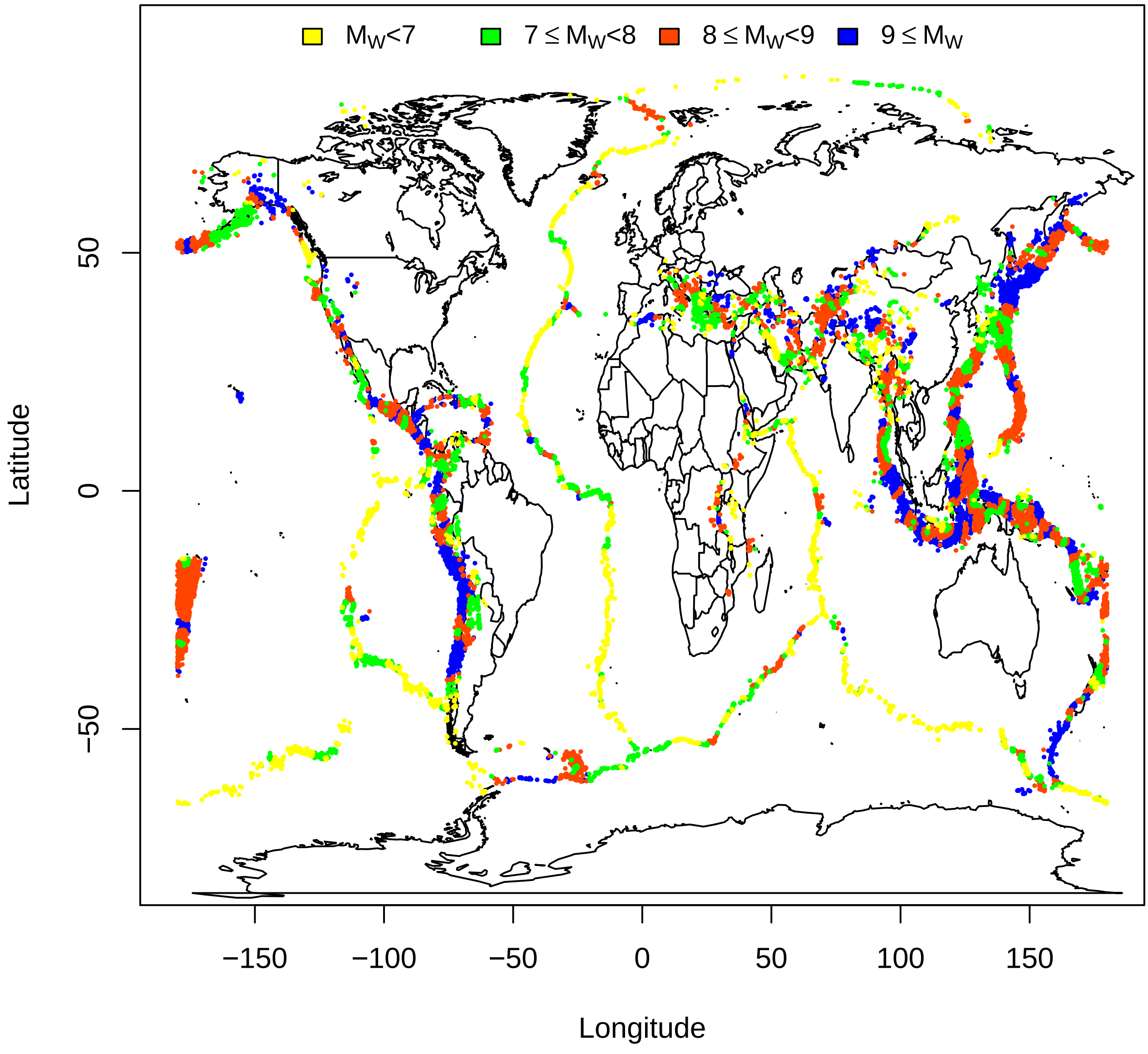

Figure 1: Local estimates of $U(10000; x)$ at locations where earthquakes have been observed, after converting the seismic moment into the moment magnitude scale M_W.

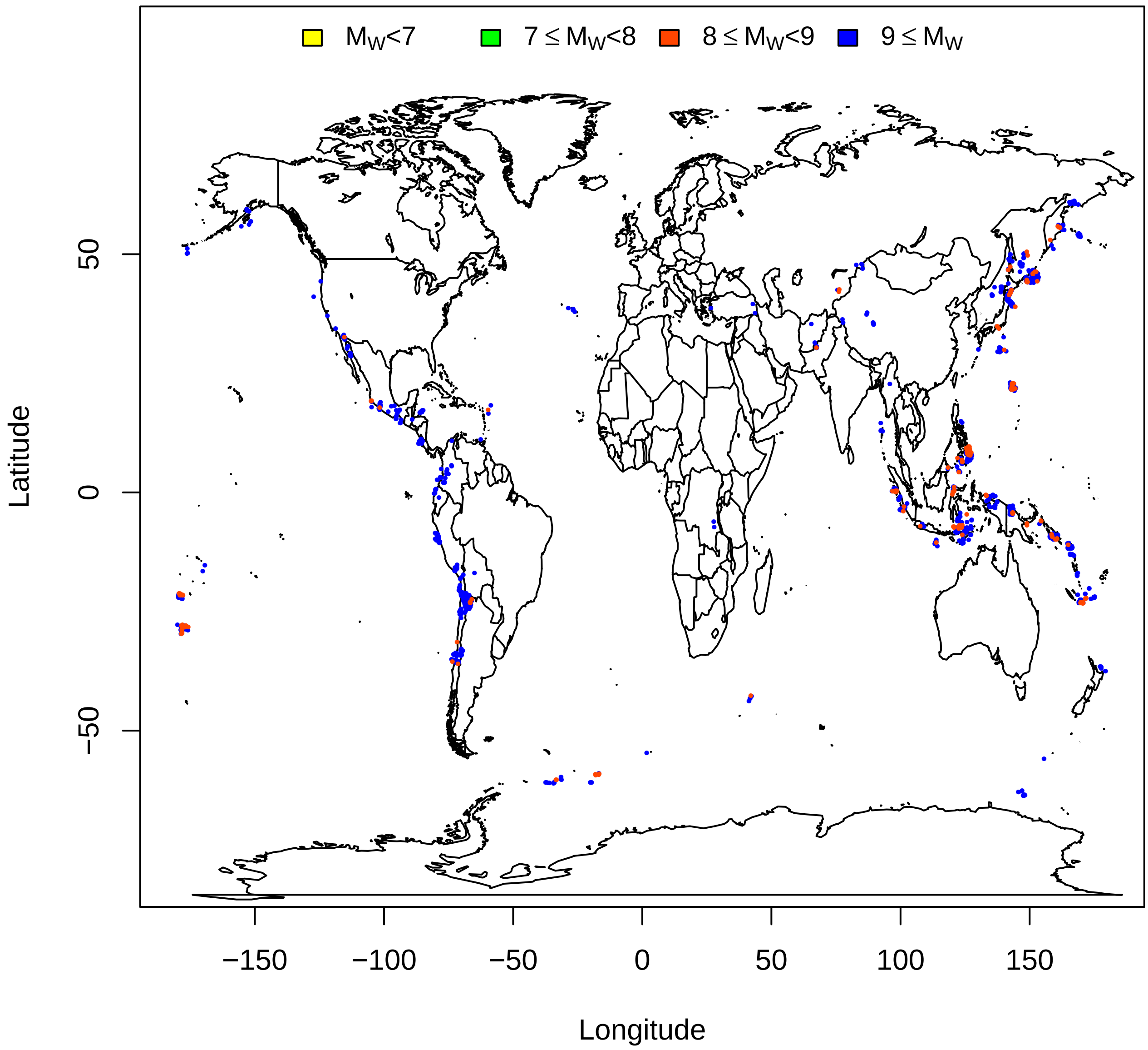

Figure 2: Earthquake positions for which the extreme quantile estimate falls inside the 95% bootstrap confidence interval (8.97; 9.26) constructed under the assumption that the seismic moment of the earthquakes are identically distributed.

References

[1] Daouia, A., Gardes, L., Girard, S., Lekina, A. (2011). Kernel estimators of extreme level curves. Test 20: 311–333.

[2] Daouia, A., Gardes, L., Girard, S. (2013). On kernel smoothing for extremal quantile regression. Bernoulli 19: 2557–2589.

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

[4] Dekkers, A.L.M., de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation. The Annals of Statistics 17: 1795–1832.

[5] 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.

[6] 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.

[7] 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.

[8] Gardes, L., Stupfler, G. (2013). Estimation of the conditional tail index using a smoothed local Hill estimator. Extremes 17: 45–75.

[9] Goegebeur, Y., Guillou, A., Osmann, M. (2014). A local moment type estimator for the extreme value index in regression with random covariates. To appear in The Canadian Journal of Statistics.

[10] de Haan, L., Ferreira, A. (2006). *Extreme Value Theory: An Introduction.* New York: Springer.

[11] Parzen, E. (1962). On estimation of a probability density function and mode. Annals of Mathematical Statistics 33: 1065–1076.

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

[13] Wang, H., Tsai, C.L. (2009). Tail index regression. Journal of the American Statistical Association 104: 1233–1240.

6 Appendix A

Proof of Lemma 1

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

$$r_n \left(\frac{\widehat{\overline{F}}(\omega_n; x)}{\overline{F}(\omega_n; x)} - 1 \right) = \frac{g(x)}{\widehat{g}_n(x)} r_n \left(\widetilde{T}_n^{(0)}(x, K) - 1 \right) + r_n \frac{g(x)}{\widehat{g}_n(x)} \frac{g(x) - \widehat{g}_n(x)}{g(x)}$$
$$=: T_1 + T_2.$$

By adapting Lemma 3 in Daouia *et al.* (2013) to our assumption ($\mathcal{G}$) and using our Theorem 1, we get that

$$T_1 \stackrel{D}{\to} P.$$

The term T_2 can be written as

$$T_2 = O_{\mathbb{P}}\left(r_n h_n^{\eta_g}\right) + O_{\mathbb{P}}\left(\sqrt{\overline{F}(\omega_n; x)}\right)$$

according to the adaptation of Lemma 3 in Daouia *et al.* (2013) to our framework. The result then easily follows.

Proof of Lemma 2

Let $\varepsilon > 0$ be arbitrary, and consider

$$\frac{U\left(\frac{1}{\overline{F}(t;x)}; x\right) - t}{a\left(\frac{1}{\overline{F}(t;x)}; x\right) A\left(\frac{1}{\overline{F}(t;x)}; x\right)} = -\frac{U\left(\frac{1}{\overline{F}(t;x)} + \frac{\varepsilon A\left(\frac{1}{\overline{F}(t;x)};x\right)}{\overline{F}(t;x)}; x\right) - U\left(\frac{1}{\overline{F}(t;x)}; x\right)}{a\left(\frac{1}{\overline{F}(t;x)}; x\right) A\left(\frac{1}{\overline{F}(t;x)}; x\right)}$$
$$+ \frac{U\left(\frac{1}{\overline{F}(t;x)} + \frac{\varepsilon A\left(\frac{1}{\overline{F}(t;x)};x\right)}{\overline{F}(t;x)}; x\right) - t}{a\left(\frac{1}{\overline{F}(t;x)}; x\right) A\left(\frac{1}{\overline{F}(t;x)}; x\right)}.$$

Since

$$\frac{U\left(\frac{1}{\overline{F}(t;x)} + \frac{\varepsilon A\left(\frac{1}{\overline{F}(t;x)};x\right)}{\overline{F}(t;x)}; x\right) - t}{a\left(\frac{1}{\overline{F}(t;x)}; x\right) A\left(\frac{1}{\overline{F}(t;x)}; x\right)} \geq 0$$

and according to the proof of Theorem B.3.19 in de Haan and Ferreira (2006),

$$\lim_{t \to y^*(x)} \frac{U\left(\frac{1}{\overline{F}(t;x)} + \frac{\varepsilon A\left(\frac{1}{\overline{F}(t;x)};x\right)}{\overline{F}(t;x)}; x\right) - U\left(\frac{1}{\overline{F}(t;x)}; x\right)}{a\left(\frac{1}{\overline{F}(t;x)}; x\right) A\left(\frac{1}{\overline{F}(t;x)}; x\right)} = \varepsilon,$$

it follows that

$$\liminf_{t\to y^*(x)} \frac{U\left(\frac{1}{\overline{F}(t;x)};x\right)-t}{a\left(\frac{1}{\overline{F}(t;x)};x\right)A\left(\frac{1}{\overline{F}(t;x)};x\right)} \geq -\varepsilon.$$

Similarly, consider

$$\frac{U\left(\frac{1}{\overline{F}(t;x)};x\right)-t}{a\left(\frac{1}{\overline{F}(t;x)};x\right)A\left(\frac{1}{\overline{F}(t;x)};x\right)} = -\frac{U\left(\frac{1}{\overline{F}(t;x)}-\frac{\varepsilon A\left(\frac{1}{\overline{F}(t;x)};x\right)}{\overline{F}(t;x)};x\right)-U\left(\frac{1}{\overline{F}(t;x)};x\right)}{a\left(\frac{1}{\overline{F}(t;x)};x\right)A\left(\frac{1}{\overline{F}(t;x)};x\right)}$$
$$+\frac{U\left(\frac{1}{\overline{F}(t;x)}-\frac{\varepsilon A\left(\frac{1}{\overline{F}(t;x)};x\right)}{\overline{F}(t;x)};x\right)-t}{a\left(\frac{1}{\overline{F}(t;x)};x\right)A\left(\frac{1}{\overline{F}(t;x)};x\right)}.$$

Then

$$\frac{U\left(\frac{1}{\overline{F}(t;x)}-\frac{\varepsilon A\left(\frac{1}{\overline{F}(t;x)};x\right)}{\overline{F}(t;x)};x\right)-t}{a\left(\frac{1}{\overline{F}(t;x)};x\right)A\left(\frac{1}{\overline{F}(t;x)};x\right)} \leq 0$$

and according to the proof of Theorem B.3.19 in de Haan and Ferreira (2006),

$$\lim_{t\to y^*(x)} \frac{U\left(\frac{1}{\overline{F}(t;x)}-\frac{\varepsilon A\left(\frac{1}{\overline{F}(t;x)};x\right)}{\overline{F}(t;x)};x\right)-U\left(\frac{1}{\overline{F}(t;x)};x\right)}{a\left(\frac{1}{\overline{F}(t;x)};x\right)A\left(\frac{1}{\overline{F}(t;x)};x\right)} = -\varepsilon,$$

so

$$\limsup_{t\to y^*(x)} \frac{U\left(\frac{1}{\overline{F}(t;x)};x\right)-t}{a\left(\frac{1}{\overline{F}(t;x)};x\right)A\left(\frac{1}{\overline{F}(t;x)};x\right)} \leq \varepsilon.$$

Consequently,

$$-\varepsilon \leq \liminf_{t\to y^*(x)} \frac{U\left(\frac{1}{\overline{F}(t;x)};x\right)-t}{a\left(\frac{1}{\overline{F}(t;x)};x\right)A\left(\frac{1}{\overline{F}(t;x)};x\right)} \leq \limsup_{t\to y^*(x)} \frac{U\left(\frac{1}{\overline{F}(t;x)};x\right)-t}{a\left(\frac{1}{\overline{F}(t;x)};x\right)A\left(\frac{1}{\overline{F}(t;x)};x\right)} \leq \varepsilon,$$

and since ε can be chosen arbitrarily small,

$$\lim_{t\to y^*(x)} \frac{U\left(\frac{1}{\overline{F}(t;x)};x\right)-t}{a\left(\frac{1}{\overline{F}(t;x)};x\right)A\left(\frac{1}{\overline{F}(t;x)};x\right)} = 0.$$

Proof of Theorem 2

Let $\alpha_n := \overline{F}(\omega_n;x)$. Define the function f as $f(u,v,w) := \frac{v}{2u}\left(1-\frac{\left(\frac{v}{u}\right)^2}{\frac{w}{u}}\right)^{-1}$. The two cases $\gamma(x)>0$ and $\gamma(x)\le 0$ will be considered separately. First assume that $\gamma(x)>0$. Then,

$$
\begin{aligned}
r_n\left[\frac{\widehat{a}_n(x)}{a\left(\frac{1}{\alpha_n};x\right)}-1\right] =& r_n\left[\frac{\omega_n}{a\left(\frac{1}{\alpha_n};x\right)}\frac{\widetilde{T}_n^{(1)}(x,K_1)}{2\widetilde{T}_n^{(0)}(x,K_0)}\left(1-\frac{\left(\frac{\widetilde{T}_n^{(1)}(x,K_1)}{\widetilde{T}_n^{(0)}(x,K_0)}\right)^2}{\frac{\widetilde{T}_n^{(2)}(x,K_2)}{\widetilde{T}_n^{(0)}(x,K_0)}}\right)^{-1}-1\right] \\
=& \frac{1}{\gamma(x)}r_n\left[f\left(\widetilde{T}_n^{(0)}(x,K_0),\widetilde{T}_n^{(1)}(x,K_1),\widetilde{T}_n^{(2)}(x,K_2)\right)-f\left(1,\gamma(x),2\gamma^2(x)\right)\right] \\
&+r_n\left(\frac{U\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)}-\frac{1}{\gamma(x)}\right) \\
&\times\left[f\left(\widetilde{T}_n^{(0)}(x,K_0),\widetilde{T}_n^{(1)}(x,K_1),\widetilde{T}_n^{(2)}(x,K_2)\right)-f\left(1,\gamma(x),2\gamma^2(x)\right)\right] \\
&+r_n\left(\frac{U\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)}-\frac{1}{\gamma(x)}\right)\gamma(x) \\
&+r_n\left(\frac{\omega_n-U\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)}\right) \\
&\times\left[f\left(\widetilde{T}_n^{(0)}(x,K_0),\widetilde{T}_n^{(1)}(x,K_1),\widetilde{T}_n^{(2)}(x,K_2)\right)-f\left(1,\gamma(x),2\gamma^2(x)\right)\right] \\
&+r_n\left(\frac{\omega_n-U\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)}\right)f\left(1,\gamma(x),2\gamma^2(x)\right) \\
=:& T_3+T_4+T_5+T_6+T_7.
\end{aligned}
$$

By a straightforward application of the delta method, it follows that

$$
T_3 \xrightarrow{D} -2P+\frac{3}{\gamma(x)}Q-\frac{1}{2\gamma^2(x)}R.
$$

Concerning T_4, we clearly have $T_4 = O_{\mathbb{P}}\left(B\left(\frac{1}{\alpha_n};x\right)\right)$, while

$$
\begin{aligned}
T_5 &= r_n\left(\frac{\gamma(x)}{\overline{A}\left(\frac{1}{\alpha_n};x\right)+\gamma(x)}-1\right) \\
&= -\frac{1}{\gamma(x)}r_nB\left(\frac{1}{\alpha_n};x\right)(1+o(1)) \\
&\to -\frac{\lambda(x)}{\gamma(x)}.
\end{aligned}
$$

Finally, Lemma 2 implies that $T_6 = o_{\mathbb{P}}\left(B\left(\frac{1}{\alpha_n};x\right)\right)$ and $T_7 = o_{\mathbb{P}}(1)$. By collecting the terms, the result follows for $\gamma(x) > 0$. Next, assume that $\gamma(x) \leq 0$. Then it follows that

$$
\begin{aligned}
r_n\left[\frac{\widehat{a}_n(x)}{a\left(\frac{1}{\alpha_n};x\right)} - 1\right] &= r_n\left[\frac{\omega_n}{U\left(\frac{1}{\alpha_n};x\right)}\frac{a^*\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)}\frac{1}{2}\frac{\widetilde{T}_n^{(1)}(x,K_1)}{\widetilde{T}_n^{(0)}(x,K_0)}\left(1 - \frac{\left(\frac{\widetilde{T}_n^{(1)}(x,K_1)}{\widetilde{T}_n^{(0)}(x,K_0)}\right)^2}{\frac{\widetilde{T}_n^{(2)}(x,K_2)}{\widetilde{T}_n^{(0)}(x,K_0)}}\right)^{-1} - 1\right] \\
&= r_n\Bigg[f\left(\widetilde{T}_n^{(0)}(x,K_0), \widetilde{T}_n^{(1)}(x,K_1), \widetilde{T}_n^{(2)}(x,K_2)\right) - f\left(1, \frac{1}{1-\gamma(x)}, \frac{2}{(1-\gamma(x))(1-2\gamma(x))}\right)\Bigg] \\
&\quad + r_n\left(\frac{a^*\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)} - 1\right)\Bigg[f\left(\widetilde{T}_n^{(0)}(x,K_0), \widetilde{T}_n^{(1)}(x,K_1), \widetilde{T}_n^{(2)}(x,K_2)\right) \\
&\qquad\qquad - f\left(1, \frac{1}{1-\gamma(x)}, \frac{2}{(1-\gamma(x))(1-2\gamma(x))}\right)\Bigg] \\
&\quad + r_n\left(\frac{\omega_n}{U\left(\frac{1}{\alpha_n};x\right)} - 1\right)\Bigg[f\left(\widetilde{T}_n^{(0)}(x,K_0), \widetilde{T}_n^{(1)}(x,K_1), \widetilde{T}_n^{(2)}(x,K_2)\right) \\
&\qquad\qquad - f\left(1, \frac{1}{1-\gamma(x)}, \frac{2}{(1-\gamma(x))(1-2\gamma(x))}\right)\Bigg] \\
&\quad + r_n\left(\frac{\omega_n}{U\left(\frac{1}{\alpha_n};x\right)} - 1\right)\left(\frac{a^*\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)} - 1\right) \\
&\quad \times\Bigg[f\left(\widetilde{T}_n^{(0)}(x,K_0), \widetilde{T}_n^{(1)}(x,K_1), \widetilde{T}_n^{(2)}(x,K_2)\right) - f\left(1, \frac{1}{1-\gamma(x)}, \frac{2}{(1-\gamma(x))(1-2\gamma(x))}\right)\Bigg] \\
&\quad + r_n\left(\frac{a^*\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)} - 1\right) \\
&\quad + r_n\left(\frac{\omega_n}{U\left(\frac{1}{\alpha_n};x\right)} - 1\right) + r_n\left(\frac{\omega_n}{U\left(\frac{1}{\alpha_n};x\right)} - 1\right)\left(\frac{a^*\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)} - 1\right) \\
&=: T_8 + T_9 + T_{10} + T_{11} + T_{12} + T_{13} + T_{14}.
\end{aligned}
$$

By the delta method,

$$T_8 \overset{D}{\to} -2(1-\gamma(x))P + (1-\gamma(x))(3-4\gamma(x))Q - \frac{1}{2}(1-\gamma(x))(1-2\gamma(x))^2 R,$$

while the term T_9, can be seen to be $O_{\mathbb{P}}\left(B\left(\frac{1}{\alpha_n};x\right)\right)$. Note that

$$\frac{\omega_n}{U\left(\frac{1}{\alpha_n};x\right)}-1=\overline{A}\left(\frac{1}{\alpha_n};x\right)B\left(\frac{1}{\alpha_n};x\right)\frac{\omega_n-U\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)B\left(\frac{1}{\alpha_n};x\right)}$$
$$=o\left(\overline{A}\left(\frac{1}{\alpha_n};x\right)B\left(\frac{1}{\alpha_n};x\right)\right),$$

according to Lemma 2, so $T_{10}=o_{\mathbb{P}}\left(\overline{A}\left(\frac{1}{\alpha_n};x\right)B\left(\frac{1}{\alpha_n};x\right)\right)$, $T_{11}=o_{\mathbb{P}}\left(\overline{A}\left(\frac{1}{\alpha_n};x\right)B^2\left(\frac{1}{\alpha_n};x\right)\right)$, $T_{13}=o\left(\overline{A}\left(\frac{1}{\alpha_n};x\right)\right)$ and $T_{14}=o\left(\overline{A}\left(\frac{1}{\alpha_n};x\right)B\left(\frac{1}{\alpha_n};x\right)\right)$. Finally, we get $T_{12}\to\lambda(x)b^{(3)}_{\gamma(x),\rho(x)}$, so the result follows by collecting the terms.

Proof of Theorem 4

Let $\widehat{\alpha}_n:=\widehat{\overline{F}}(\omega_n;x)$. First, note that as $t\to\infty$,

$$q_{\gamma(x)}(t)=\begin{cases}\frac{1}{\gamma(x)}t^{\gamma(x)}\ln(t)(1+o(1)), & \gamma(x)>0\\ \frac{1}{2}\ln^2 t, & \gamma(x)=0\\ \frac{1}{\gamma^2(x)}(1+o(1)), & \gamma(x)<0\end{cases}$$

and

$$\frac{t^{\gamma(x)}-1}{\gamma(x)}=\begin{cases}\frac{1}{\gamma(x)}t^{\gamma(x)}(1+o(1)), & \gamma(x)>0\\ \ln t, & \gamma(x)=0\\ -\frac{1}{\gamma(x)}(1+o(1)), & \gamma(x)<0\end{cases}.$$

Remark that

$$
\begin{aligned}
r_n \frac{\widehat{U}\left(\frac{1}{\beta_n};x\right) - U\left(\frac{1}{\beta_n};x\right)}{q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right) a\left(\frac{1}{\alpha_n};x\right)} =& \frac{r_n}{q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \frac{\omega_n - U\left(\frac{1}{\beta_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)} + \frac{r_n}{q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \frac{\widehat{a}_n(x)}{a\left(\frac{1}{\alpha_n};x\right)} \frac{\left(\frac{\widehat{\alpha}_n}{\beta_n}\right)^{\widehat{\gamma}_n(x)} - 1}{\widehat{\gamma}_n(x)} \\
=& \frac{r_n}{q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \frac{\widehat{a}_n(x)}{a\left(\frac{1}{\alpha_n};x\right)} \left(\frac{\left(\frac{\widehat{\alpha}_n}{\beta_n}\right)^{\widehat{\gamma}_n(x)} - 1}{\widehat{\gamma}_n(x)} - \frac{\left(\frac{\alpha_n}{\beta_n}\right)^{\gamma(x)} - 1}{\gamma(x)} \right) \\
&+ \frac{r_n}{q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \frac{\widehat{a}_n(x)}{a\left(\frac{1}{\alpha_n};x\right)} \left(\frac{\left(\frac{\widehat{\alpha}_n}{\beta_n}\right)^{\widehat{\gamma}_n(x)} - 1}{\widehat{\gamma}_n(x)} - \frac{\left(\frac{\alpha_n}{\beta_n}\right)^{\widehat{\gamma}_n(x)} - 1}{\widehat{\gamma}_n(x)} \right) \\
&+ \frac{r_n}{q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \left(\frac{\widehat{a}_n(x)}{a\left(\frac{1}{\alpha_n};x\right)} - 1 \right) \frac{\left(\frac{\alpha_n}{\beta_n}\right)^{\gamma(x)} - 1}{\gamma(x)} \\
&- \frac{r_n}{q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \left(\frac{U\left(\frac{\alpha_n}{\beta_n}\frac{1}{\alpha_n};x\right) - U\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)} - \frac{\left(\frac{\alpha_n}{\beta_n}\right)^{\gamma(x)} - 1}{\gamma(x)} \right) \\
&+ \frac{r_n}{q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \frac{\omega_n - U\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)} \\
=&: T_{15} + T_{16} + T_{17} - T_{18} + T_{19}.
\end{aligned}
$$

For T_{15}, it follows that

$$
\begin{aligned}
T_{15} =& \frac{r_n\left(\widehat{\gamma}_n(x) - \gamma(x)\right)}{q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \int_1^{\frac{\alpha_n}{\beta_n}} s^{\gamma(x)-1} \frac{e^{(\widehat{\gamma}_n(x)-\gamma(x))\ln s} - 1}{(\widehat{\gamma}_n(x) - \gamma(x))\ln s} \ln s ds\left(1 + o_{\mathbb{P}}(1)\right) \\
=& r_n\left(\widehat{\gamma}_n(x) - \gamma(x)\right)\left(1 + o_{\mathbb{P}}(1)\right) \\
&+ \frac{r_n\left(\widehat{\gamma}_n(x) - \gamma(x)\right)}{q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \int_1^{\frac{\alpha_n}{\beta_n}} s^{\gamma(x)-1} \left(\frac{e^{(\widehat{\gamma}_n(x)-\gamma(x))\ln s} - 1}{(\widehat{\gamma}_n(x) - \gamma(x))\ln s} - 1 \right) \ln s ds\left(1 + o_{\mathbb{P}}(1)\right) \\
=&: T_{15}^{(1)} + T_{15}^{(2)}.
\end{aligned}
$$

Clearly $T_{15}^{(1)} \xrightarrow{D} \Gamma$, while

$$
|T_{15}^{(2)}| \leq |r_n\left(\widehat{\gamma}_n(x) - \gamma(x)\right)| \sup_{1 \leq s \leq \frac{\alpha_n}{\beta_n}} \left| \frac{e^{(\widehat{\gamma}_n(x)-\gamma(x))\ln s} - 1}{(\widehat{\gamma}_n(x) - \gamma(x))\ln s} - 1 \right| \left(1 + o_{\mathbb{P}}(1)\right).
$$

An application of the mean value theorem shows that

$$
\sup_{1 \leq s \leq \frac{\alpha_n}{\beta_n}} \left| \frac{e^{(\widehat{\gamma}_n(x)-\gamma(x))\ln s} - 1}{(\widehat{\gamma}_n(x) - \gamma(x))\ln s} - 1 \right| \leq \sup_{1 \leq s \leq \frac{\alpha_n}{\beta_n}} \left| e^{Q} - 1 \right|,
$$

where Q is a random value between 0 and $(\widehat{\gamma}_n(x) - \gamma(x)) \ln s$. By applying the mean value theorem again, it follows that

$$\sup_{1 \le s \le \frac{\alpha_n}{\beta_n}} \left| \frac{e^{(\widehat{\gamma}_n(x)-\gamma(x))\ln s} - 1}{(\widehat{\gamma}_n(x) - \gamma(x))\ln s} - 1 \right| \le |\widehat{\gamma}_n(x) - \gamma(x)| \left| \ln\left(\frac{\alpha_n}{\beta_n}\right) \right| e^{|\widehat{\gamma}_n(x)-\gamma(x)|\left|\ln\left(\frac{\alpha_n}{\beta_n}\right)\right|}$$
$$= o_{\mathbb{P}}(1),$$

where the last equality is obtained using the assumption that $\ln\left(\frac{\overline{F}(\omega_n;x)}{\beta_n}\right) = o\left(\sqrt{nh_n^p \overline{F}(\omega_n;x)}\right)$. Hence $T_{15}^{(2)} = o_{\mathbb{P}}(1)$. Concerning T_{16}, we have

$$\begin{aligned} T_{16} =& r_n \left(\left(\frac{\widehat{\alpha}_n}{\alpha_n} \right)^{\widehat{\gamma}_n(x)} - 1 \right) \frac{\left(\frac{\alpha_n}{\beta_n}\right)^{\gamma(x)} - 1}{\gamma(x) q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} (1 + o_{\mathbb{P}}(1)) \\ &+ \frac{r_n}{q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \left(\left(\frac{\widehat{\alpha}_n}{\alpha_n} \right)^{\widehat{\gamma}_n(x)} - 1 \right) \left(\frac{\left(\frac{\alpha_n}{\beta_n}\right)^{\widehat{\gamma}_n(x)} - 1}{\widehat{\gamma}_n(x)} - \frac{\left(\frac{\alpha_n}{\beta_n}\right)^{\gamma(x)} - 1}{\gamma(x)} \right) (1 + o_{\mathbb{P}}(1)) \\ &+ \frac{r_n}{q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \frac{\left(\frac{\widehat{\alpha}_n}{\alpha_n}\right)^{\widehat{\gamma}_n(x)} - 1}{\widehat{\gamma}_n(x)} (1 + o_{\mathbb{P}}(1)) \\ =&: T_{16}^{(1)} + T_{16}^{(2)} + T_{16}^{(3)}. \end{aligned}$$

By a straightforward application of the delta method, $r_n \left(\left(\frac{\widehat{\alpha}_n}{\alpha_n} \right)^{\widehat{\gamma}_n(x)} - 1 \right) \xrightarrow{D} \gamma(x) P$, while $\frac{\left(\frac{\alpha_n}{\beta_n}\right)^{\gamma(x)} - 1}{\gamma(x) q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \to -\gamma_-(x)$, when $n \to \infty$. Hence

$$T_{16}^{(1)} \xrightarrow{D} -\gamma_-^2(x) P.$$

The term $T_{16}^{(2)}$ can be handled similarly as for T_{15} and $T_{16}^{(1)}$. This gives $T_{16}^{(2)} = O_{\mathbb{P}}\left(\frac{1}{r_n}\right)$. Next

$$T_{16}^{(3)} \xrightarrow{D} \gamma_-^2(x) P,$$

by an application of the delta method. Combining the previous convergences, we deduce that

$$T_{16} = o_{\mathbb{P}}(1).$$

From Theorem 3, it easily follows that

$$T_{17} \xrightarrow{D} -\gamma_-(x) \Lambda,$$

while

$$T_{18} = r_n B\left(\frac{1}{\alpha_n}; x\right) \frac{A\left(\frac{1}{\alpha_n}; x\right)}{B\left(\frac{1}{\alpha_n}; x\right)} \frac{\left(\frac{\alpha_n}{\beta_n}\right)^{\gamma(x)} - 1}{\gamma(x) q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \frac{\frac{U\left(\frac{\alpha_n}{\beta_n}\frac{1}{\alpha_n};x\right) - U\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)} \frac{\gamma(x)}{\left(\frac{\alpha_n}{\beta_n}\right)^{\gamma(x)} - 1} - 1}{A\left(\frac{1}{\alpha_n}; x\right)},$$

for which we know that $r_n B\left(\frac{1}{\alpha_n}; x\right) \to \lambda(x)$, $\frac{\left(\frac{\alpha_n}{\beta_n}\right)^{\gamma(x)} - 1}{\gamma(x) q_{\gamma(x)}\left(\frac{\alpha_n}{\beta_n}\right)} \to -\gamma_-(x)$,

$$\frac{\frac{U\left(\frac{\alpha_n}{\beta_n}\frac{1}{\alpha_n};x\right) - U\left(\frac{1}{\alpha_n};x\right)}{a\left(\frac{1}{\alpha_n};x\right)} \frac{\gamma(x)}{\left(\frac{\alpha_n}{\beta_n}\right)^{\gamma(x)} - 1} - 1}{A\left(\frac{1}{\alpha_n}; x\right)} \to -\frac{1}{\rho(x) + \gamma_-(x)},$$

according to Lemma 4.3.5 in de Haan and Ferreira (2006), and

$$\frac{A\left(\frac{1}{\alpha_n}; x\right)}{B\left(\frac{1}{\alpha_n}; x\right)} \to \begin{cases} 0, & c = \pm\infty \\ \frac{\gamma(x)+\rho(x)}{\gamma(x)}, & c = \frac{\gamma(x)}{\gamma(x)+\rho(x)} \\ 1, & \text{otherwise} \end{cases}.$$

Hence

$$T_{18} \to \lambda(x) \frac{\gamma_-(x) \mathbf{1}\{\gamma(x) < \rho(x) \leq 0\}}{\gamma_-(x) + \rho(x)}.$$

Finally, Lemma 2 implies that $T_{19} = o(1)$, so the result is established by combining the terms.

7 Appendix B

We formulated Assumption $(\mathcal{F})$ in terms of conditional expectations, which is convenient as it covers the three subclasses of the max-domain of attraction, corresponding to $\gamma(x) < 0$, $\gamma(x) = 0$ and $\gamma(x) > 0$, in an elegant way. For each subclass, one can verify that $(\mathcal{F})$ can be satisfied by imposing some more structure on $\overline{F}$ and by introducing additional conditions on h_n and ω_n. For instance, in case $\gamma(x) \in \mathbb{R}_0^+$, $\forall x \in \mathbb{R}^p$, assume

$$\overline{F}(y; x) = y^{-1/\gamma(x)} \ell(y; x), \quad y \geq 1,$$

where $\ell(.; x)$ is a normalised slowly varying function, i.e.

$$\ell(y; x) = c_\ell(x) \exp\left(\int_1^y \frac{\alpha(v; x)}{v} dv\right), \tag{9}$$

with $c_\ell(x) > 0$ and $\alpha(.; x)$ is a function converging to zero at infinity. Besides $(\mathcal{G})$, assume the following Hölder continuity conditions: for all $x, x' \in \mathbb{R}^p$ we have

$$\begin{aligned} |\gamma(x) - \gamma(x')| &\leq M_\gamma \|x - x'\|^{\eta_\gamma}, \\ |\log c_\ell(x) - \log c_\ell(x')| &\leq M_{c_\ell} \|x - x'\|^{\eta_{c_\ell}}, \\ \sup_{y \geq 1} |\alpha(y; x) - \alpha(y; x')| &\leq M_\alpha \|x - x'\|^{\eta_\alpha}. \end{aligned}$$

Now consider $m(\omega_n, t; z)$. Trivially, $m(\omega_n, 0; z) = \overline{F}(\omega_n; z)$. For $t > 0$, apply integration by parts to obtain

$$\begin{aligned} m(\omega_n, t; z) &= t \int_{\omega_n}^{\infty} (\ln y - \ln \omega_n)^{t-1} \frac{\overline{F}(y; z)}{y} dy \\ &= t\overline{F}(\omega_n; z) \int_1^{\infty} (\ln u)^{t-1} \frac{\overline{F}(u\omega_n; z)}{u\overline{F}(\omega_n; z)} du \\ &= t\overline{F}(\omega_n; z) \int_1^{\infty} (\ln u)^{t-1} u^{-1/\gamma(z)-1} du \\ &\quad + t\overline{F}(\omega_n; z) \int_1^{\infty} (\ln u)^{t-1} u^{-1/\gamma(z)-1} \left(\frac{\ell(u\omega_n; z)}{\ell(\omega_n; z)} - 1 \right) du \\ &=: \breve{T}_1 + \breve{T}_2. \end{aligned}$$

By straightforward integration, $\breve{T}_1 = \gamma^t(z)\Gamma(t+1)\overline{F}(\omega_n; z)$. Concerning $\breve{T}_2$, let $\widetilde{T}_2$ denote the integral, and apply (9) to obtain

$$|\widetilde{T}_2| \leq \int_1^{\infty} (\ln u)^{t-1} u^{-1/\gamma(z)-1} \left| \exp\left(\int_{\omega_n}^{u\omega_n} \frac{\alpha(v; z)}{v} dv \right) - 1 \right| du.$$

Whence, by the well-known inequality $|e^u - 1| \leq e^{|u|}|u|$ and denoting $\overline{\alpha}(y; z) := \sup_{v \geq y} |\alpha(v; z)|$,

$$|\widetilde{T}_2| \leq \overline{\alpha}(\omega_n; z) \int_1^{\infty} (\ln u)^t u^{-1/\gamma(z)-1+\overline{\alpha}(\omega_n; z)} du.$$

Let $B(x, h_n)$ denote the closed ball with center x and radius h_n. Then, using the uniform Hölder-type condition on α, we have

$$\overline{\alpha}(\omega_n; z) \leq \overline{\alpha}(\omega_n; x) + \sup_{y \geq 1} |\alpha(y; z) - \alpha(y; x)| \leq \overline{\alpha}(\omega_n; x) + M_\alpha h_n^{\eta_\alpha}, \tag{10}$$

for all $z \in B(x, h_n)$, and thus $\overline{\alpha}(\omega_n; z) \to 0$, uniformly in $z \in B(x, h_n)$. Now take c such that $0 < c < 1/\bar{\gamma}$, where $\bar{\gamma} := \sup_{z \in B(x, h_n)} \gamma(z)$. There exists an n_0 such that for $n \geq n_0$ we have $\overline{\alpha}(\omega_n; z) < c$ for all $z \in B(x, h_n)$, and hence for $n \geq n_0$

$$|\widetilde{T}_2| \leq \frac{\overline{\alpha}(\omega_n; z)\Gamma(t+1)}{\left(\frac{1}{\bar{\gamma}} - c\right)^{t+1}} =: \tilde{c}\, \overline{\alpha}(\omega_n; z)\Gamma(t+1),$$

for all $z \in B(x, h_n)$. Combining the above, we have the following bound for $n \geq n_0$:

$$\begin{aligned} \left| \frac{m(\omega_n, t; z)}{\gamma^t(z)\Gamma(t+1)\overline{F}(\omega_n; z)} - 1 \right| &\leq \frac{\tilde{c}t\overline{\alpha}(\omega_n; z)}{\gamma^t(z)} \\ &\leq \frac{\tilde{c}t\overline{\alpha}(\omega_n; z)}{\inf_{z \in B(x, h_n)} \gamma^t(z)} =: \check{c}t\overline{\alpha}(\omega_n; z), \end{aligned} \tag{11}$$

for all $z \in B(x, h_n)$. Note that this inequality applies in fact also to the case $t = 0$.

Now, for the ratio of conditional expectations we can obtain the inequality

$$
\begin{aligned}
&\left|\frac{m(\omega_n,t;x-zh_n)}{m(\omega_n,t;x)}-1\right| \\
&\quad\le\ \left|\frac{\overline{F}(\omega_n;x-zh_n)}{\overline{F}(\omega_n;x)}-1\right|+\left|\frac{\gamma^t(x-zh_n)}{\gamma^t(x)}-1\right|\frac{\overline{F}(\omega_n;x-zh_n)}{\overline{F}(\omega_n;x)} \\
&\qquad+\left|\frac{\gamma^t(x)\Gamma(t+1)\overline{F}(\omega_n;x)}{m(\omega_n,t;x)}-1\right|\frac{\gamma^t(x-zh_n)\overline{F}(\omega_n;x-zh_n)}{\gamma^t(x)\overline{F}(\omega_n;x)} \\
&\qquad+\left|\frac{m(\omega_n,t;x-zh_n)}{\gamma^t(x-zh_n)\Gamma(t+1)\overline{F}(\omega_n;x-zh_n)}-1\right|\frac{\gamma^t(x)\Gamma(t+1)\overline{F}(\omega_n;x)}{m(\omega_n,t;x)}\frac{\gamma^t(x-zh_n)\overline{F}(\omega_n;x-zh_n)}{\gamma^t(x)\overline{F}(\omega_n;x)} \\
&\quad=:\ \breve{T}_3+\breve{T}_4+\breve{T}_5+\breve{T}_6.
\end{aligned}
$$

Concerning $\breve{T}_3$, note that for $z\in\Omega$, $\omega_n\ge e$ and $h_n\le 1$ we have for some constant $M_{\overline{F}}$ that

$$\left|\ln\frac{\overline{F}(\omega_n;x-zh_n)}{\overline{F}(\omega_n;x)}\right|\le M_{\overline{F}}h_n^{\eta}\ln\omega_n,$$

where $\eta:=\eta_\gamma\wedge\eta_{c_\ell}\wedge\eta_\alpha$, and thus, after application of Taylor's theorem,

$$\sup_{t\in\{0,1,\ldots,6\}}\sup_{z\in\Omega}\left|\frac{\overline{F}(\omega_n;x-zh_n)}{\overline{F}(\omega_n;x)}-1\right|=O(h_n^{\eta}\ln\omega_n).$$

For $\breve{T}_4$, apply the mean value theorem and obtain, with $\tilde{\gamma}$ some value between $\gamma(x)$ and $\gamma(x-zh_n)$,

$$
\begin{aligned}
\left|\frac{\gamma^t(x-zh_n)}{\gamma^t(x)}-1\right| &= \frac{t\tilde{\gamma}^{t-1}}{\gamma^t(x)}|\gamma(x-zh_n)-\gamma(x)| \\
&\le \widetilde{M}_\gamma h_n^{\eta_\gamma},
\end{aligned}
$$

for some positive constant $\widetilde{M}_\gamma$. Thus

$$\sup_{t\in\{0,1,\ldots,6\}}\sup_{z\in\Omega}\breve{T}_4=O(h_n^{\eta_\gamma}).$$

The term $\breve{T}_5$ can be analysed with Taylor's theorem and inequality (11), leading to

$$\sup_{t\in\{0,1,\ldots,6\}}\sup_{z\in\Omega}\breve{T}_5=O(\overline{\alpha}(\omega_n;x)).$$

For $\breve{T}_6$, apply inequality (11) to obtain

$$\sup_{t\in\{0,1,\ldots,6\}}\sup_{z\in\Omega}\breve{T}_6=O(\sup_{z\in\Omega}\overline{\alpha}(\omega_n;x-zh_n)).$$

By combining the above results we have that

$$\sup_{t\in\{0,1,\ldots,6\}}\sup_{z\in\Omega}\left|\frac{m(\omega_n,t;x-zh_n)}{m(\omega_n,t;x)}-1\right|=O(h_n^{\eta}\ln\omega_n\vee\sup_{z\in\Omega}\overline{\alpha}(\omega_n;x-zh_n)).$$

Using a bound similar to (10), Assumption ($\mathcal{F}$) is thus satisfied if

$$h_n^{\eta}\ln\omega_n\to 0.$$

Similar arguments hold for $\gamma(x)=0$ and $\gamma(x)<0$ but for brevity they are omitted.