Charles University in Prague
Faculty of Mathematics and Physics
Department of probability and mathematical statistics

Stochastical inference in the model of extreme events

DOCTORAL THESIS

September 2011

Mgr. et Mgr. Jan Dienstbier

Charles University in Prague
Faculty of Mathematics and Physics

DOCTORAL THESIS

Jan Dienstbier

Stochastical inference in the model of extreme events

Department of probability and mathematical statistics

Supervisor of the doctoral thesis: Doc. RNDr. Jan Picek, CSc.

Study programme:
Probability and Mathematical Statistics (M4)

Prague 2011

Název práce: Stochastická inference v modelu extrémních událostí

Autor: Jan Dienstbier

Katedra/Ústav: Katedra pravděpodobnosti a matematické statistiky

Vedoucí doktorské práce: Doc. RNDr. Jan Picek, CSc., Technická Univerzita v Liberci

Abstrakt: Práce se věnuje extremálním aspektům lineárních modelů. Obsahuje stručný výklad teorie extremálních hodnot a uvádí do problému lineárních modelů $\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \mathbf{E}_{n\times 1}$ s chybami $E_i \sim F, i = 1, \ldots, n$, kde distribuční funkce F náleží do některé sféry extremální přitažlivosti. Pracujeme s regresními kvantily odvozenými v článku Koenker and Basset (1978) a ukazujeme jejich extremální vlastnosti. V rámci odvození nových metod je v práci podán důkaz aproximace regresních kvantilů založený na na starších výsledcích Gutenbrunner et al. (1993). Náš výsledek platí na intervalu $[\alpha_n^*, 1 - \alpha_n^*]$ s lepším řádem konvergence $\alpha_n^* \to 0$, než byl dosud odvozen ve starší literatuře. Tato aproximace umožňuje vybudovat aproximaci chvostů regresních kvantilů, na které je potom založena teorie hladkých funkcionálů procesu regresních kvantilů. Pomocí této teorie pak lze odvodit novou třídu odhadů Paretova indexu vhodnou pro regresní situaci. V práci probíráme vlastnosti této třídy odhadů a demonstrujeme jejich vlastnosti na simulační studii.

Klíčová slova: regresní kvantily, Paretův index, Bahadurova reprezentace

Title: Stochastical inference in the model of extreme events

Author: Jan Dienstbier

Department/Institute: Department of probability and mathematical statistics
Supervisor of the doctoral thesis: Doc. RNDr. Jan Picek, CSc.

Abstract: The thesis deals with extremal aspects of linear models. We provide a brief explanation of extreme value theory. The attention is then turned to linear models $\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \mathbf{E}_{n\times 1}$ with the errors $E_i \sim F, i = 1, \ldots, n$ fulfilling the domain of attraction condition. We examine the properties of the regression quantiles of Koenker and Basset (1978) under this setting we develop theory dealing with extremal characteristics of linear models. Our methods are based on an approximation of the regression quantile process for $\alpha \in [0, 1]$ expanding older results of Gutenbrunner et al. (1993). Our result holds in $[\alpha_n^*, 1 - \alpha_n^*]$ with a better rate of $\alpha_n^* \to 0$ than the other approximations described previously in the literature. Consecutively we provide an approximation of the tails of regression quantile. The approximations of the tails enable to develop theory of the smooth functionals, which are used to establish a new class of estimates of extreme value index. We prove $T(F_n^{-1}(1 - k_n t/n))$ is consistent and asymptotically normal estimate of extreme for any T member of the class. Various possible estimates of the empirical tail quantile functions are considered, discussed and illustrated on a simulation study.

Keywords: regression quantiles, extreme value index, Bahadur representation

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources. I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague 14.9.2011 signature

To my family & Kamila.

Contents

List of Figures

List of Abbreviations and Symbols

Symbol	**Meaning**
$A \subset B$	Means A is the proper subset of B
$I[\,\cdot\,]$	The indicator function
$\mathbb{N}$	The set of natural numbers
$\mathbb{R}$	The set of real numbers
$\mathbb{R}^d$	The real d-dimensional vector space ($d \in \mathbb{N}$)
$C[0,1]$	The space of continuous functions on $[0,1]$ equipped with the supremum norm
$D[0,1]$	The space of left continuous functions on $[0,1]$ equipped with Skorokhod metric
$\|\cdot\|$	The supremum norm
$\|\cdot\|_{\gamma,h}$	The weighted metric seminorm of real functions on unit interval, see (1.55)
$a \vee b$	$\max(a,b)$
$a \wedge b$	$\min(a,b)$
$[a]$	Means the largest integer less than or equal to a
$X \stackrel{\mathcal{D}}{=} Y$	Means X and Y have the same distribution
$X \sim F$	Means X have distribution function F
$\xrightarrow{\text{a.s.}}$	Almost sure convergence
$\xrightarrow{\text{P}}$	Convergence in probability
$\xrightarrow{\mathcal{D}}$	Convergence in distribution
$\xrightarrow{\mathcal{W}}$	Weak convergence of stochastic processes
F^{-1}	Left-continuous quantile function
x^*	The upper right endpoint of the distribution, $\inf\{x : F(x) = 1\}$

x_*	The lower left endpoint of the distribution, $\sup\{x : F(x) = 0\}$
$X \approx Y$	Means $X = \mathcal{O}(Y)$ and $Y = \mathcal{O}(X)$
γ	Extreme value index (EVI)
G_γ	Generalized extreme value distribution
$\mathrm{MDA}(G_\gamma)$	Maximum domain of attraction
ρ	Second order parameter
ℓ	Slowly varying function
a.s.	Almost surely
CLT	Central limit theorem
LIL	Law of the iterated logarithm
EVT	Extreme value theory
PWM	Probability weighted moments
ML	Maximum likelihood
i.i.d.	Independent identically distributed

Preface

The thesis deals with stochastic inference oriented on extremal properties of random samples. However, the topic would be in its full generality obviously beyond the possibility of any dissertation thesis. An introduction to the general methodology how to deal with extremes in samples, extreme value theory, is provided in the first chapter. Extreme value theory is now an independent discipline of statistics and has been treated in many books and in a vast number of articles (cf. de Haan and Ferreira (2006), Beirlant et al. (2004), Resnick (2007) and others mentioned in the first chapter). Therefore, the topic of this thesis is much more modest. We shall concentrate on the inference of extremal properties of linear models. While the univariate and multivariate branches of extreme value theory have been examined extensively during past years, the special case of linear models attracted far less attention. Moreover, many findings and approaches of the current theory have not been based on firm theoretical grounds. The thesis should be a humble contribution to the discussion about this narrowly defined topic. My wish is to shed light on the theoretical backgrounds of extreme value inference in linear models. The main tools in this task are the quantile sensitive regression methods such as regression quantiles or two-step quantiles (cf. Koenker and Basset (1978), Jurečková and Picek (2005)) whose properties in connection with the theory of empirical quantiles are described in the second chapter. The asymptotic approximation of regression quantiles provided therein is an important conclusion of the thesis. The techniques developed in last chapters are all based on this approximation.

Extreme value theory is a theory of assumptions. As long as one can hope that these assumptions hold it is possible to provide answers. For real life datasets the standing point of a statistician is much more complicated, as it is no longer clear which assumptions hold. Extreme value theory with its basic purpose to provide at least some conclusions even on a very weak basis of facts, cannot often provide unambiguous solutions. It has turned out in many situations that an expert guess is a valuable addition to the theoretically grounded procedures. This is certainly nothing against the theory, as without it one would be even blinder facing the facts, but it means that the theory is just a tool in hands of a statistician who has the responsibility for the decision. This belief was crucial in formulating the thesis conclusions. Rather than provide just answers concerning extremes in linear models, I sought to provide questions which should be asked, as only questioning the very nature of investigations can guarantee that our description of the reality would not be flawed from the very beginning.

I could not complete the thesis without the help of many among my teachers, colleagues and friends. First of all I would like to express my gratitude to doc. RNDr. Jan Picek, PhD., who fairly exceeded his duty as a supervisor and I would have never finished this work without his friendly support. My thanks also belong to my teachers and colleagues from the Department of Probability and Statistics on Faculty of mathematics and physics of Charles University. Let me mention at least four of them I have been obliged to over the years: prof. RNDr. Jana Jurečková, DrSc., prof. RNDr. Marie Hušková, DrSc., Doc. RNDr. Daniel Hlubinka, PhD and Ing. Marek Omelka, PhD. Finally my deepest obligation comes to my family and Kamila as it was also a heavy burden for them.

Concerning the financial arrangements, this work was supported by the project of Czech Ministry of Education LC06024 and partly also by the grant of Charles University in Prague GAUK 153407-B.

Jan Dienstbier

Charles University in Prague
Prague, September 2011

Chapter 1

Extreme value theory

The purpose of this chapter is to provide an introduction to *extreme value theory* (EVT). The following survey of mathematical, probabilistic, and statistical tools of EVT is just a brief one – a reader can find a suitable treatment of the topic in any modern book covering EVT. Let me mention some of them.

Firstly I would like to point to the recent book by de Haan and Ferreira (2006) that offers a thorough introduction to the theory from the mathematical point of view. The authors present the subject in its full generality as well as the precise details. To understand the inner nature of the theory, it is useful to look also at the treatment of heavy-tailed phenomena by Sidney L. Resnick. The book Resnick (2007) goes pretty much to the core of the regular variation problematic from the point process point of view. It can be recommended for a clear and lively explanation of the basic concepts hidden behind EVT. Finally, I should mention Beirlant et al. (2004), which is valuable for the wide range of covered statistical and computations techniques. A wide range of real data cases illuminates these techniques. These three books have been the most influential for the conception of this chapter, but the chapter also contains a lot from other sources for example from the articles Drees (1998*b*), Dietrich et al. (2002), Buch-Kromann (2006) and others.

1.1 Why extreme value theory?

Extreme value theory is a branch of statistics devoted to the study of rare phenomena occurring in sample observations. The theory itself represents an independent statistical discipline which has been developing steadily during the past years.

Uniformed observer may ask, why one should pay special attention to the problem of an inference of extreme phenomena at all. Are not they covered by the classical theory of statistic? Is it desirable to concentrate on rarely observable events at all? While rare events do not play a principal role in the "classical" methods concerned by average values, they can be crucial for a significant number of real life problems. A flood is a perfect example of a rare event, but extremes appear in many other fields. Extreme losses on financial markets, enormous insurance claims, or peaks of teletraffic communication in computer nets are just few cases that can be dealt with the theory. In all these examples there exists a considerable demand for any information about the rare events, their inner structure and probability distributions.

Nevertheless, it should also be clarified why to demand a special theory concerning the extremes only. Simply, "traditional" techniques of mathematical statistics are not suitable for the the task. We shall demonstrate it on the following example:

> Have a random sample $\mathbf{X} = \{X_1, \ldots, X_n\}$, i.e. independent identically distributed observations generated by an unknown random variable with a distribution function F. The task is to estimate the probability of exceedance of quantiles of F, $p = P(X > x)$ on the basis of the data.

The simplest and in some sense also the most efficient way to estimate the probability is by *the empirical distribution function*

$$F_n := \frac{1}{n} \sum_{i=1}^{n} I[x \geq X_i],$$

where $I[\,\cdot\,]$ stands for *the indicator function*. The estimate is $\hat{p} = 1 - F_n(x)$. However, this approach is clearly inadequate if $x > X_{n:n}$, where $X_{1:n} \leq \ldots \leq X_{n:n}$ denotes *the order statistics* of the sample $X_1, \ldots, X_n$, as in that case $\hat{p} = 0$ regardless of the exact distribution function of the sample. It is obvious that underestimate severely the risk if the distribution is heavy-tailed. Similarly one can expect that even the estimate of p for $X_{n-1:n} < x < X_{n:n}$ is not the fully accurate one. The conclusion of such an observation holds to some extent also for other high quantiles.

The conclusion is that the reliability and usefulness of the estimates based on empirical distribution functions are questionable considering high quantiles of the distribution. The events corresponding to the distribution tails are commonly called *extreme events*. From similar arguments one could also derive that also other common strategies to estimate the density or the distribution function of the sample distribution are inefficient for "extreme" areas. Namely kernel estimators typically work with the preselected "heaviness" of the tail (i.e. the rate of the decreasing tail probability) which again results in an underestimation (or overestimation) of the risk of extreme events.

On the other hand, the motivation to estimate high quantiles of sample distribution is obvious and demanded in various areas of human activity. As an example consider the situation appearing in actuarial practice, especially in reinsurance. The success of insurance companies depends on the estimation of this probability or indices derived from it. *The mean excess function* or *mean residual life function* $e(\cdot)$ is used as a common tool. It is defined as

$$e(t) := \mathrm{E}(X - t|X > t),$$

provided that $\mathrm{E}(X) < \infty$. Estimating $e(t)$ for a high threshold t brings the same type of problems as in our introductory example and it is clearly desirable to have a solution adaptable to the extreme phenomena (for some possibilities and their application on the real data cases see Beirlant et al. (2004), pp. 14–19).

The actuarial practice resembles the situation in hydrology. Hydrologists are often occupied with an estimation of the *T-year flood discharge*, which is a level exceeded every T years on average. A time span of 100 years is usually taken, but the estimation must be often carried out on the basis of water discharges for a much shorter period. Moreover, it is advantageous to provide even a more detailed high-quantile estimation of the water level. The knowledge of high-quantile properties of flood discharges is important for modelling of the effective water course systems capacity, urban drainage and water runoff.

Similar motivation stands behind the urge to estimate the extreme rainfall or wind speed. The real case data studies demand even more complicated models adaptive to dependency in the data set. As an example consider data from two different locations which have something in common (they are on the same river separated by a longer distance or in the same location and in a close vicinity for the case of rainfall).

Even the decision which observations are rare or extreme is not an easy one. Deriving from a (so far only vague) definition of rare or extreme events, we can expect that in a dataset there are none or at best just a few observations appllying to an area of extreme events. However, the number of really extreme observations should also depend on a sample size. But how? To come up with at least partial conclusion we must assume stricter rules governing the sample distribution. At least, one should assume that the really extreme observations of the tails share the same pattern regardless of given high threshold (at least asymptotically). As events $\{X > t\}$ and $\{X < t\}$ are independent for any single $t \in \mathbb{R}$, we can require that such an assumption holds only for the tail data greater than any given fixed threshold. Hence we can readily restrict ourselves on areas $\{X > t_n\}$ for a suitable sequence of thresholds $t_n \in \mathbb{R}$ according to the sample size $n \in \mathbb{N}$. We shall see later on that this type of approach leads to a general theory of extreme events.

However, at least a bit should be also said about other possible solutions. In some situations it is reasonable to approximate the tail by some *ad hoc* chosen distribution. For example in insurance practice so called Champernowne distribution is used, see Buch-Kromann (2006). The distribution function of the modified Champernowne distribution is defined

$$F_{\alpha,M,c} = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha}, \qquad \text{for } x \in \mathbb{R}^+, \tag{1.1}$$

where $\alpha > 0$, $M > 0$, and $c \geq 0$ are parameters of the distribution. The parameters can be estimated from the data using the maximum likelihood method. Buch-Kromann proposes (in the cited article) further corrections of estimates by the means of kernel density estimator. Finally the tail is approximated by the suitably transformed kernel density estimator to obtain exceedance probabilities etc.

The treatment by EVT represents a general alternative to such more or less *ad hoc* solutions. Again, have a sample $X_1, \ldots, X_n$ generated by an unknown distribution F. We are interested in an approximation of the distribution of the maxima $X_{n:n} := \max_{i=1,\ldots,n}\{X_i\}$ for $n \to \infty$. The distribution itself tends to be degenerate in infinity, which is not too much informative − by the strong law of large numbers it holds $P(X_{n:n} \to F(x_*)) = 1$, where x_* is the right upper end-point of the distribution F.

However, much more information can be obtained by an approach similar to the concept of *the central limit theorem* (CLT) or the stable laws. The main idea is worth of more detailed explanation. Let $S_n := X_1 + \ldots + X_n$ by the CLT holds $\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow[n\to\infty]{\mathrm{P}} Z$, where Z has the standard normal distribution $N(0,1)$, provided that $\mathrm{E}X_1 = \mu$ and $\mathsf{var}\, X_1^2 = \sigma^2 < \infty$. Similar idea stands behind the concept of *stable domain of attraction*. Suppose there exist some $A_n > 0$ and B_n such that,

$$\frac{S_n - A_n}{B_n} \xrightarrow[n\to\infty]{\mathcal{D}} Y \sim \Phi. \tag{1.2}$$

We say in that case that F belongs to the stable domain of attraction of Φ or, in other words, Φ possesses the stable domain of attraction. It can be proven that a set of the stable distributions is closed, as any distribution F posseses a stable domain of attraction if and only if F has a stable distribution, i.e. it holds for some constants a_n and $b_n > 0$ that $S_n \sim b_n X_1 + a_n$. This is just an introductory idea of the more extensive theory of stable laws, see Shorack (2000), pp. 407-411, for more.

Similar ideas can be used even if the maxima are on the spot of the sums in (1.2). Define for the sample $X_1, \ldots, X_n$ order statistics $X_{1:n} \leq \cdots \leq X_{n:n}$. Suppose that there exist constants $a_n \in \mathbb{R}$ and $b_n > 0$ that it holds

$$\frac{X_{n:n} - a_n}{b_n} \xrightarrow[n\to\infty]{\mathcal{D}} Y \sim G, \tag{1.3}$$

where G is some nondegenerate distribution of Y. The question is what are the possible forms of the distribution G.

1.2 Domains of attraction

Suppose $X_1, \ldots, X_n$ are independent random variables (i.i.d.) with a common distribution function F. Let $X_{1:n} \leq \cdots \leq X_{n:n}$ be the order statistics of $X_1, \ldots, X_n$. Extreme value theory is based on the following assumption:

(EVT.1) Suppose there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$P\left(X_{n:n} \le a_n x + b_n\right) = F^n(a_n x + b_n) \to G(x) \tag{1.4}$$

for all $x \in \mathbb{R}$, where G is a non-degenerate function.

Assuming (EVT.1) holds we say that G is *the extreme value distribution* and F is in *the domain of attraction* of G (notation $F \in \mathrm{MDA}(G)$). The following theorem is fundamental result of EVT. It provides a characterization of the possible non-degenerate distributions that can occur in (1.4).

Theorem 1.2.1 (Fisher-Tippet-Gnedenko)**.** *The class of non-degenerate limit distributions $G(x)$ which can fulfill (1.4) is $G(x) = G_\gamma(ax + b)$ with $a > 0$, $b \in \mathbb{R}$, where*

$$G_\gamma(x) = \begin{cases} \exp\left(-(1-\gamma x)^{-1/\gamma}\right), & 1 + \gamma x > 0, \\ \exp(-e^{-x}) & \gamma = 0, \end{cases} \tag{1.5}$$

and $\gamma \in \mathbb{R}$.

Proof. See de Haan and Ferreira (2006), Theorem 1.1.3. □

The parameter γ in (1.5) is called *extreme value index*. This parameter has a central role in the theory. We divide the distributions fulfilling (1.4) into groups of domains of attraction according to $\gamma \in \mathbb{R}$ and its sign. The sign itself has an important meaning as a delimiter between three basic classes of distributions according to the mass concentrated on the tail (see below). Note that a similar condition can be formulated for the minima $X_{1:n}$ of the sample as well. In that case similar version of Theorem 1.2.1 holds and the characterizations remains principally the same (consider the properties of the maxima of $\{-X_1, \dots, -X_n\}$).

The versatility of the condition (1.4) is demonstrated by the fact that the different criteria on F lead to the same domains of attraction.

Theorem 1.2.2. *For $\gamma \in \mathbb{R}$ the following statements are equivalent:*

(i) There exist real constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right),$$

for all x with $1 + \gamma x > 0$ and $G_0(x)$ defined by the limit as in (1.5).

(ii) There is a positive function $a(\cdot)$ such that for $x > 0$,

$$\lim_{t\downarrow 0} \frac{F^{-1}(1-tx) - F^{-1}(1-t)}{a(t)} = \frac{x^{-\gamma} - 1}{\gamma}, \tag{1.6}$$

where for $\gamma = 0$ the right-hand side is interpreted as $\log x$.

(iii) There exists a positive function h such that

$$\lim_{t \uparrow x^*} \frac{1 - F(t + xh(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma} \tag{1.7}$$

for all x such that $1 + \gamma x > 0$, where $x^* = \sup\{x : F(x) < 1\}$.

Proof. See de Haan and Ferreira (2006), Theorem 1.1.6. □

The condition (EVT.1) can be simplified for smooth and differentiable distributions. In that case a sufficient condition for belonging to a domain of attraction is given by so called von Mises condition.

Theorem 1.2.3 (von Mises condition)**.** *Let F be a distribution function and x^* its right endpoint. Suppose F has a density f that is positive and differentiable for all x in some left neighbourhood of x^*. If*

$$\lim_{t \uparrow x^*} \frac{(1 - F(t))f'(t)}{f^2(t)} = -\gamma - 1 \tag{1.8}$$

then F is in the domain of attraction of G_γ, defined in (1.5).

Proof. See de Haan and Ferreira (2006), Theorem 1.1.8. □

Furthermore, if γ is positive or negative, the notation (1.8) can be simplified even a little bit more.

Theorem 1.2.4 (von Mises condition for $\gamma > 0$)**.** *Suppose $x^* = \infty$ and there exists f, the positive density of F. If for some $\gamma > 0$ holds*

$$\lim_{t \to \infty} \frac{tf(t)}{1 - F(t)} = \frac{1}{\gamma}, \tag{1.9}$$

then F is in the domain of attraction of G_γ defined in (1.5).

Proof. See de Haan and Ferreira (2006), Theorem 1.1.11. □

An analogue assertion holds also in the case $\gamma < 0$.

Theorem 1.2.5 (von Mises condition for $\gamma < 0$)**.** *Suppose $x^* < \infty$ and for $x < x^*$ there exists f, the density of F. If*

$$\lim_{t \uparrow x^*} \frac{(x^* - t)f(t)}{1 - F(t)} = -\frac{1}{\gamma} \tag{1.10}$$

for some $\gamma < 0$, then F is in the domain of attractions of G_γ defined in (1.5).

Proof. See de Haan and Ferreira (2006), Theorem 1.1.13. □

Theorems 1.2.4 and 1.2.5 show that the sign of γ really matters. By Theorem 1.2.4 one gets that smooth distribution functions appertaining to $\{G_\gamma, \gamma \in \mathbb{R}^+\}$ have an infinite upper endpoint x^*. Similarly for the distributions in $\{G_\gamma, \gamma \in \mathbb{R}^-\}$ it holds that their left tails are restricted by finite x^*. The third important class arises from the case $\gamma = 0$.

In the following part we will also use the word "domain" in a wider sense when referring to the three larger classes respective to the sign of γ. The above described property concerning finiteness of their tails holds also for the distributions fulfilling only the domain of attraction condition and not von Mises condition, for details see de Haan and Ferreira (2006), Theorem 1.2.1. The characteristics of these three classes are summed up in the following subsections.

1.2.1 Fréchet-Pareto domain, $\gamma > 0$

Fréchet-Pareto (or just Fréchet) domain traditionally receives the largest attention of researchers. The distributions, which lie in the domain are called *heavy tailed.* They have an infinite endpoint $x^* = \infty$ and their higher moments are not finite depending on the value of γ. Relation (1.7) can be simplified to

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma}, \qquad \text{for} \qquad x > 0 \tag{1.11}$$

or in terms of quantile function

$$\lim_{t\downarrow 0} \frac{F^{-1}(1-tx)}{F^{-1}(1-t)} = x^{-\gamma}, \qquad \text{for} \qquad x > 0 \tag{1.12}$$

see de Haan and Ferreira (2006), Theorem 1.2.1., which in terms of regular variation means that $1 - F$ is regularly varying in infinity with index $-1/\gamma$, provided that $F \in \mathrm{MDA}(G_\gamma), \gamma > 0$. Hence,

$$1 - F(x) = x^{-\frac{1}{\gamma}} \ell_F(x) \qquad \text{and} \qquad F^{-1}(1-x) = x^{-\gamma}\ell_Q(x), \tag{1.13}$$

where ℓ_F and ℓ_Q are two slowly varying functions (ℓ_Q is slowly varying at zero) linked together via de Bruyn conjugation, for the definition of de Bruyn conjugation and more details see Beirlant et al. (2004), pp. 79–82.

Representation (1.13) reveals that with the increasing value of γ the tail becomes heavier, or in other words, it is more likely that large outliers appear in the sample. This also reflects the dispersion of the data (its structure not the variance) and finiteness of the moments. For $\gamma > 0.5$ the variance of the distribution does not exist. For $\gamma > 1$ there does not exist even expected value $\mathrm{E}X$. More precisely,

$$\mathrm{E}(X_+^a) \begin{cases} = \infty, & a\gamma > 1. \\ < \infty, & a\gamma < 1, \end{cases}$$

where X_+ denotes the positive part of X, see Beirlant et al. (2004), pp. 58.

The distributions of Fréchet–Pareto domain have an important role in practical applications and have been successfully used to model many real life situations. To give some examples, the financial data such as insurance claims, log returns of the stocks values, teletraffic data or high water levels are often heavy tailed. The importance of studying the theoretical properties of the class is even stronger considering the fact that many traditional statistical methods can be severely affected by the tail properties of the sample (consider e.g. least-square linear regression). EVT provides a general framework to treat the extremal aspects of the inference. Thorough attention gained by the class has resulted in specific inferential methods usable for Fréchet class, which are not consistent if $\gamma \leq 0$. The most prominent example of this kind of estimators is Hill's estimator, see page 14. The following table shows some examples of the distributions belonging to the Fréchet domain.

Distribution	**Notation**	$F(x)$	γ
Pareto	$\mathrm{Pa}(\alpha)$	$1-x^{-\alpha}$ $x>1;\gamma>0$	$\frac{1}{\alpha}$
Generalized Pareto	$\mathrm{GP}(\sigma,\gamma)$	$1-\left(1+\frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}}$ $x>0;\sigma,\gamma>0$	γ
Fréchet	$\mathrm{Fr}(\alpha)$	$1-\exp(-x^{-\alpha})$ $x>1;\alpha>0$	$\frac{1}{\alpha}$
Student's t-distribution (absolute values)	$\lvert t_n\rvert$	$1-\int_x^\infty \frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{\omega^2}{n}\right)^{-\frac{n+1}{2}}\mathrm{d}\omega,$ $x>0;n>0$	$\frac{1}{n}$
Cauchy		$\frac{1}{\pi}\arctan(x)$ $x\in\mathbb{R}$	1
Burrleigh	$\mathrm{Burr}(\eta,\tau,\lambda)$	$1-\left(\frac{\eta}{\eta+x^\tau}\right)^\lambda$ $x>0;\eta,\tau,\lambda>0$	$\frac{1}{\lambda\tau}$

1.2.2 Weibull domain, $\gamma < 0$

The distributions in this domain are chiefly characterized by the finite upper endpoint x^*. As such, they represent the opposite of the previously described heavy-tailed distributions of Fréchet domain. The extremes of the distributions in the class are bounded and $\mathrm{E}X^a < \infty, a \in \mathbb{R}^+$. The basic characterization of the distributions in the domain can be written as

$$\lim_{t\downarrow 0}\frac{1-F(x^*-tx)}{1-F(x^*-t)} = x^{-1/\gamma} \tag{1.14}$$

The relation (1.14) can be rewritten in terms of regular variation

$$1-F(x^*-1/x) = x^{-1/\gamma}\ell_F(x) \qquad \text{and} \qquad F^{-1}(x) = x_* - x^{-\gamma}\ell_Q(x) \tag{1.15}$$

where ℓ_F and ℓ_U are slowly varying functions linked together by de Bruyn conjugation similarly as in the case $\gamma > 0$, again see Beirlant et al. (2004), pp. 79–82.

The problem of an inference of γ in the Weibull domain has attracted far less attention than that of Fréchet and Gumbel domains (see next section). In fact for $\gamma < -1$ one gets better estimates of the distribution tails using standard statistical methods than the methods of EVT, see section 1.6 and the references cited therein.

Nevertheless, techniques dealing specially with the Weibull case have been developing during the past years, e.g. Negative Hill's estimator of γ is estimator of the extreme value index working only for the Weibull class. This estimator can be seen as a counterpart of Hill's estimator, whose consistency is limited to the case of $\gamma > 0$.

Distribution	**Notation**	$F(x)$	γ
Uniform	U[0, 1]	$x, x \in [0, 1]$	-1
Beta	Beta(p, q)	$\int_0^x \frac{1}{B(p,q)} t^{p-1}(1-t)^{q-1}\mathrm{d}t$, $x \in (0, 1)$	$-\frac{1}{q}$
Extreme value Weibull		$\exp\left((x^* - x)^\alpha\right)$ $x \in (0, x^*)$	$-\frac{1}{\alpha}$

The distribution which have finite extremes may seem easy to deal with. Nevertheless, these distribution requires special treatment and normalizations in many applications, especially in the case $\gamma < -1/2$. Note, that some of the distributions in Weibull domain have an infinite density near the upper endpoint. As an example consider the distribution given by the distribution function

$$F(x) = 1 - \sqrt{1-x}, \qquad x \in [0, 1].$$

It has the positive density $f(x) = \frac{1}{\sqrt{1-x}}$ for which holds

$$\lim_{x \to 1} \frac{1}{\sqrt{1-x}} = \infty, \tag{1.16}$$

i.e. the density f tends to infinity when approaching the upper endpoint. Moreover $f'(x) = \frac{1}{2}(1-x)^{-3/2}$ and

$$\lim_{x \to 1} \frac{\sqrt{1-x} \cdot \frac{1}{2}(1-x)^{-3/2}}{(1-x)^{-1}} = \frac{1}{2}. \tag{1.17}$$

Hence F fulfills von Mises condition (1.8) with $\gamma = -3/2$ and belongs to Weibull domain of attraction. This is a reason why we will restrict ourselves mainly to the assumption $\gamma > -1/2$.

1.2.3 Gumbel domain, $\gamma = 0$

The Gumbel domain of attraction is more diverse than the previous ones. The upper right endpoint x^* of the distributions within the domain can be finite as well as infinite and also the basic classification of the domain is more complex than in the case of Fréchet or Weibull domains: For any distribution function F appertaining to Gumbel domain there exists a positive function h such that

$$\lim_{t \uparrow x^*} \frac{1 - F(t + xh(t))}{1 - F(t)} = \exp(-x), \tag{1.18}$$

for $x \in \mathbb{R}$. Function h can be defined as

$$h(t) := \frac{\int_t^{x^*} (1 - F(s)) \mathrm{d}s}{1 - F(t)},$$

see de Haan and Ferreira (2006), Theorem 1.2.1. The distributions in the domain also fulfill relation $\int_t^{x^*} (1 - F(s)) \mathrm{d}s < \infty$ for $t < x^*$. Such distributions are often called exponentially light-tailed. Indeed, there exists $\mathrm{E}X^a$ for any $a \in \mathbb{R}^+$. Exponential distribution as well as Normal distribution are typical example of distribution appertaining to Gumbel domain of attraction.

Distribution	**Notation**	$F(x)$
Exponential	$\mathrm{Exp}(\lambda)$	$1 - \exp(-\lambda x)$m $x > 0; \lambda > 0$
Normal	$\mathcal{N}(\mu, \sigma^2)$	$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) \mathrm{d}u = \sigma(\Phi(x) + \mu)$, $x \in \mathbb{R}; \mu \in \mathbb{R}, \sigma > 0$
Gamma	$\Gamma(\lambda, m)$	$\frac{\lambda^m}{\Gamma(m)} \int_{-\infty}^{x} u^{m-1} \exp(-\lambda u) \mathrm{d}u$, $x > 0; \lambda, m > 0$
Logistic	Log	$1 - 1/(1 + \exp(x))$, $x \in \mathbb{R}$
Log-normal	$\mathcal{LN}(\mu, \sigma^2)$	$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2 x}} \exp\left(-\frac{(\log(u)-\mu)^2}{2\sigma^2}\right) \mathrm{d}u$, $x > 0; \mu \in \mathbb{R}, \sigma > 0$
Weibull		$1 - \exp(-\lambda x^\tau)$, $x > 0; \lambda, \tau > 0$

1.2.4 Distributions which do not lie in any domain

We have seen that the distributions meeting the extreme value condition can be divided into three classes distinguished by the sign of γ which reflects their properties. At least something should be written also about the distributions, which do not fulfill condition (1.4). Not much attention has been paid to them in EVT as such distributions can be seen to be contradicting to the philosophy of the theory. Nevertheless, one should be aware of their existence. In fact they actually appear in a limited field of applications. To meet these purposes and to distinguish these distributions from the distributions fulfilling the extreme value condition, some tools have been recently developed.

It has been shown on previous pages that it is the unifying concept of regular variation, which characterizes the properties of distribution tails. The distributions falling outside any domain of attraction simply do not have regularly varying tails. In few examples we shall try to summarize some sources of such an irregularity.

Rather surprisingly, one reason, why some distributions do not fall in any domain of attraction, is that they are "too heavy". Consider the distribution given by distribution function

$$F(x) = 1 - \frac{1}{\log ex}, \qquad x \in [1, \infty]. \tag{1.19}$$

For any $a \in \mathbb{R}^+$ it holds that $\mathrm{E}X^a = \infty$, therefore the distribution can lie only in the Fréchet sphere. On the other hand $1 - F(x) = 1/\log ex$ is not regularly varying and cannot be represented as in (1.13) – the tails of the distribution F are "too heavy" even for Fréchet sphere.

In some cases, the problem lies in the smoothness of the distribution function. Discrete distributions are not often in any domain of attraction, e.g. it is such in case of the Poisson distribution. Another example is the truncated exponential distribution

$$X \sim \exp[E], \tag{1.20}$$

where E is a random variable with the standard exponential distribution and $[\cdot]$ denotes the function returning the whole number of the input value. Again distribution (1.20) does not fulfill the domain of attraction condition, see Hüsler and Li (2006).

The changes in tails causing that asymptotic of (1.20) can not be described by regular variation can be also smooth. The following classical example have been introduced by R. von Mises:

$$F(x) = 1 - \exp(-x - \sin x), \qquad x > 0. \tag{1.21}$$

This distribution does not belong to any domain of attraction, see de Haan and Ferreira (2006), pp. 36. In the example (1.21) the crucial element causing that the distribution does not lie in any domain of attraction is the function $\sin x$.

Nevertheless, the assumption that the sample distribution belongs to some domain of attraction is still reasonable in majority of real life situations. The opposite case can be tested by means of recently developed tests, see below. However, the falsehood of the domain of attraction condition is also indicated by the results of the estimations and other simple observations. Simulations results in Dienstbier (2009) indicate that the problems with assumption (1.4) can also detected by a suspicious behaviour of the common estimators of γ being compared together. Calculation of more than one estimator is highly recommended as the estimation is often based on a different property of the limiting distribution G_γ.

The tests are testing the hypothesis

$$H_0: \qquad F \in \mathrm{MDA}(G_\gamma) \qquad \text{for some } \gamma \in \mathbb{R} \tag{1.22}$$

against various alternatives. These tests were derived by Dietrich et al. (2002) and Drees et al. (2006), while Hüsler and Li (2006) provided extensive simulation study comparing the both methods. A reader can find an extensive overview of these methods and other alternatives such as a modification of goodness of fit tests in Neves and Fraga-Alves (2008).

1.3 Assumptions of the second order

Suppose that condition (EVT.1) holds, hence the tail of the distribution can be described by its regularly varying component. A natural question arises what is the difference between the exact tails of the distribution and their approximation by the regularly varying function, i.e. the remains after the tail approximation (1.6). We shall define

$$R(t,x) := \frac{F^{-1}(1-tx) - F^{-1}(1-t)}{a(t)} - z_\gamma(x). \tag{1.23}$$

where

$$z_\gamma(x) := \begin{cases} \frac{x^{-\gamma}-1}{\gamma} & \gamma \neq 0, \\ -\log(x) & \gamma = 0. \end{cases} \tag{1.24}$$

We shall see that the properties of $R(t,x)$ are important when considering the rate of convergence for various estimators of γ. It is a natural idea to assume that the remainder can also be described in terms of regular variation. The following assumption is was formulated in de Haan and Resnick (1996), but equivalent formulations can be found also by other authors.

(EVT.2) Suppose there exists a function $A(t) \to 0$ (as $t \to \infty$) with constant sign near infinity such that

$$\lim_{t \downarrow 0} \frac{\frac{F^{-1}(1-tx)-F^{-1}(1-t)}{a(t)} - z_\gamma(x)}{A(t)} = \lim_{t\downarrow 0} \frac{R(t,x)}{A(t)} = K(x) \tag{1.25}$$

for all $x > 0$, where $K(x)$ is assumed not to be a multiple of $z_\gamma(x)$.

Condition (EVT.2) is called *the second order condition*. It describes the variation of the remainder between the first order approximation and the real distribution. Functions $K(x)$ and $A(x)$ can be obtained in exact forms, see de Haan and Ferreira (2006), pp. 43–49. Chiefly there exist constants $c_1, c_2 \in \mathbb{R}$ and a parameter $\rho \leq 0$ such that

$$K(x) = c_1 \frac{1}{\rho} \left(z_{\gamma+\rho}(x) - z_\gamma(x) \right) + c_2 z_{\gamma+\rho}(x) \tag{1.26}$$

if $\gamma \neq 0 \neq \rho$,

$$K(x) = c_1 \frac{1}{\rho} \left(z_{\gamma+\rho}(x) - z_\gamma(x) \right) + c_2 z_\gamma(x) \tag{1.27}$$

if $\gamma \neq 0 = \rho$, and finally

$$K(x) = c_1 \frac{1}{2} (\log x)^2 - c_2 \log x, \tag{1.28}$$

if $\gamma = 0 = \rho$, with z_γ as in (1.24).

Moreover, the normalising function A is regularly varying at zero, i.e.

$$\lim_{t \downarrow 0} \frac{A(tx)}{A(t)} = x^{-\rho} \tag{1.29}$$

Relations (1.26), (1.27), (1.28), and (1.29) show that the parameter ρ has an analogous role in this second order approximation as γ has in the approximation of the first order. We call ρ *the second order index* of regular variation. We shall see that the asymptotic bias of an γ estimators is tied to ρ in next section.

For the sake of simplicity, some authors introduce a less general assumption than (EVT.2):

(**H.1**) Assume $\gamma > 0$, then F belong to the so called *Hall class* if

$$F^{-1}(1-t) = ct^{-\gamma}(1 + dt^{-\rho} + o(t^{-\rho})), \qquad t \in (0,1) \tag{1.30}$$

for some $c > 0$, $\rho \leq 0$, and $d \in \mathbb{R}$.

Note that if F is a member of the Hall class (1.30), it also fulfills the second order condition (1.25) with the parameter ρ, which is identical with that in the definition (1.30). While the assumption (1.30) is stricter than (1.25), many properties associated with the second order approximation can be readily described in somewhat simpler version. Also the definition of the Hall class itself can be expanded to more general version, see Beirlant et al. (2004), pp. 93 (cf. also the literature cited therein).

Introducing the second order condition can be seen problematic as in the real data analysis one does not know that at least the first order condition, i.e. domain of attraction, holds. Indeed, following the steps of von Mises example (1.21) we can derive distributions, which fulfill the domain of attraction condition (EVT.1), but do not fulfill (EVT.2). An example is

$$F(x) = 1 - x^{-1} \left(1 + \frac{1}{x} \exp\left(\sin \log x\right) \right), \tag{1.31}$$

see de Haan and Ferreira (2006), pp. 61. On the other hand at least the consistency of the estimation of γ is usually assured by the domain of attraction condition alone.

Generally, the second order condition is assumed in the theory. However, at least some violations of the condition are can be seen as favourable for the usual estimation procedures: the Pareto distribution $F(x) = 1 - x^{-\alpha}$, $\alpha > 0$ naturally does not fulfill the second order condition as the remainder $R(t, x)$ is equal to zero and thus not regularly varying. On the other hand the estimation methods are based on the fact that the properly standardized tail has asymptotically the Pareto distribution and hence they perform very well on a Pareto distributed sample.

In the following lines we will generally assume that the second order condition (EVT.2) holds, but similarly as in Drees (1998*b*) we shall also cover these cases where the remainder (1.23) converge to 0 fast enough and the second order condition does not hold. We will see that in this case the bias of the estimation tends to be zero, which indicated the example with the Pareto distribution.

1.4 Estimating γ and intermediate sequences

In the previous section we have described the basic ideas of EVT. We have shown that an extremal part of the distribution can be approximated by one-parameter extreme value distribution. For a real data sample we naturally do not know the exact value of the parameter, even if we assume that the domain of attraction condition holds. As it was stated in the beginning, the strategy shall be to approximate the tails of the distribution by its associated extreme value distribution. Hence γ must be estimated on the basis of the data sample. We shall see later that a good implementation of this task is usually a sufficient step forward (at least from the asymptotic point of view) to the estimation of conditional extreme quantiles and other goals of the theory, see section 1.6. All in all, a good estimate of γ matters.

There are many known estimators of γ and the literature covering the problem is very rich. On the following pages we thus cover only a brief selection of the possible estimators of γ. For a broader overview of possible methods we refer the reader to Beirlant et al. (2004), where the additional literature can be also found.

1.4.1 Scale invariant estimators

Rather surprisingly (at least at the first sight) many popular estimators used in EVT practice are only scale but not location invariant. This is also true for the most popular one – *Hill's estimator*. It is defined as

$$\hat{\gamma}^{H}_{n,k} = \frac{1}{k} \sum_{i=1}^{n} \log X_{n-k+1:n} - \log X_{n-k:n}, \tag{1.32}$$

where $k = k_n \in \mathbb{N}$ is *an intermediate integer sequence*, i.e., $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. Hill's estimator is traditionally one of the most popular estimators and probably the one which has received the greatest attention among all estimators proposed in the literature. It was introduced during the early stage of the development of EVT by Hill (1975), but its consistency and asymptotic normality were not been fully established until Mason (1982). The estimator has a lot of notable properties which it shares with many other estimators of γ. This is the reason we shall speak about it in a more detail.

Firstly, note the special role of k in the definition (1.32). The selection of k determines the number of $k = k_n$ highest order statistics out of a total size n which are used for the estimation. The estimator is consistent if $\gamma > 0$ and k is an intermediate sequence or goes to the infinity with even faster rate. The rate of $k = k_n$ has an important consequences for the form of the asymptotic distribution of the estimator. Suppose that the second order condition (EVT.2) holds and for A as in (1.25) holds

$$\lim_{n\to\infty} \sqrt{k} A\left(\frac{k}{n}\right) = \lambda \in [0, \infty) \tag{1.33}$$

Then

$$\sqrt{k}(\hat{\gamma}^H_{n,k} - \gamma) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\right) \tag{1.34}$$

provided that $k = k_n \to \infty$, $k/n \to 0$, $n \to \infty$, see de Haan and Ferreira (2006), Theorem 3.2.5. The optimal rate of k in $\hat{\gamma}^H_{n,k}$ coincides through the relations (1.33) and (1.34) with the second order index ρ. Assume for simplicity, that F is in the Hall class defined by (1.30) for some $\rho < 0$ and $\gamma > 0$ (hence also $F \in \mathrm{MDA}(G_\gamma)$). The optimal rate of $\hat{\gamma}^H_{n,k}$ is attained for $k \sim n^{\frac{2\rho}{2\rho-1}}$, while for $\sqrt{k}A(k/n) \to 0$ and $k = o\left(n^{\frac{2\rho}{2\rho-1}}\right)$ the rate is also slower in the situation $\sqrt{k}|A(k/n)| \to \infty$ when

$$\frac{\hat{\gamma}_{n,k} - \gamma}{A\left(\frac{k}{n}\right)} \xrightarrow[n\to\infty]{\mathrm{P}} \frac{1}{1-\rho}. \tag{1.35}$$

see de Haan and Ferreira (2006), pp. 77. An interesting coverage of the problematic under more general condition is provided in Resnick (2007), pp. 291–303. The condensed conclusion of Resnick's observations and reviews of older results is that the asymptotic normality of Hill's estimator is equivalent to some kind of the second order regular variation.

The relation (1.34) leads to the following conclusion: When k_n is small, the variance of the estimator is large and on the contrary, large k_n results in a serious bias of the estimation. Note that the optimal rate of k_n depends on an unknown parameter ρ linked together with the second order condition, cf. with (1.30) – it holds also in the general case when we do not assume that F belongs to the Hall class but just fulfills the second order condition. As we do not know ρ, there is not any golden rule how to achieve the optimal proportion between k_n and n for a given sample of observations. The only possibility is to learn more by the further inference and assume stricter conditions on the underlying distribution of the sample.

The problem with the selection of "optimal" k and the reduction of the bias of the estimation has gained a considerable attention in the literature, see a brief overview in

Beirlant et al. (2004), pp. 113–129. However, these refined and sometimes a little bit complicated methods usually cause a greater asymptotic variance than the one attain by the Hill's estimator. In fact it can be proven, that the Hill's estimator attain the lowest asymptotic variance, see Drees (1998*a*), while for all the estimators for $\gamma > 0$ achieve the best rate of convergence, see Drees (1998*c*).

Asymptotic variance in a combination with an ambiguous selection of k_n poses a practical problem with the Hill's estimator (and also with other estimators), the estimate usually differs considerably for a different selection of k_n. Which estimate should be used? The mentioned methods calculating "the best" k on the basis of the data cannot be seen as a final and all-time dependable solutions of the problem – naturally, each of these methods is based on a different strategies and leads to a different value of k. It is questionable, whether to depend on any of these methods in a real data analysis. Any researcher using EVT should double check any assumptions of his analysis. Hence blindly choosing just a one out of different possible k_n to obtain the estimate of γ seems to be against the philosophy of the theory. An expert guess or a combination of various methods based on various aspects of the tail distribution seem to be a wiser solution.

In the real data analysis so called *Hill plot* is often used as an auxiliary tool in decision which k should be used

$$\left\{(k, \hat{\gamma}^H_{n,k}), 1 \leq k \leq n\right\}. \tag{1.36}$$

An example of the practical realisation of the plot is shown in Figure 1.1. As an estimate of γ we accept such a value of $\hat{\gamma}^H_{n,k}$ that lies on the plot in a region which seems to be stable. Sometimes this works satisfactory (as in Figure 1.1), sometimes less (as in Figure 1.2) and sometimes the plot is not very revealing. This happens for example for the sample generated by the perturbed Pareto distribution given by its quantile function $F^{-1}(1-t) = (t|\log(t)|^2)^{-1}$, see Figure 1.3. This distribution has by (1.12) regularly varying tails with $\gamma = 1$. Nevertheless, its slowly varying factor $\|\log(t)\|^2$, which does not play any role from the asymptotic point of view in (1.12) as $\log^2(tx)/\log^2(t) \to 1$ with $t \to \infty$, stands behind the severe bias of the estimation as indicated on Figure 1.3. The region of stability does not coincide with real value of γ for any $k \in \mathbb{N}$. In fact slowly varying factor in perturbed Pareto distribution dies out very slowly and even considerable amount of observations does not solve the problem as is indicated on Figure 1.4.

A different scaling of the Hill plot leads in some cases to the wider region of the stability, e.g. this is the case of Starica plot, see Resnick (2007), pp. 314–321. However, in some cases we cannot decide without a more complex methodology aiming at the reduction of the bias itself through some estimation, see Resnick (2007), pp. 88–89 and Beirlant et al. (2004), pp. 113–129 for some possibilities. We do not stick to details, but the conclusion is clear – the estimation of γ is a complex task whose difficulty should not be underestimated by no means.

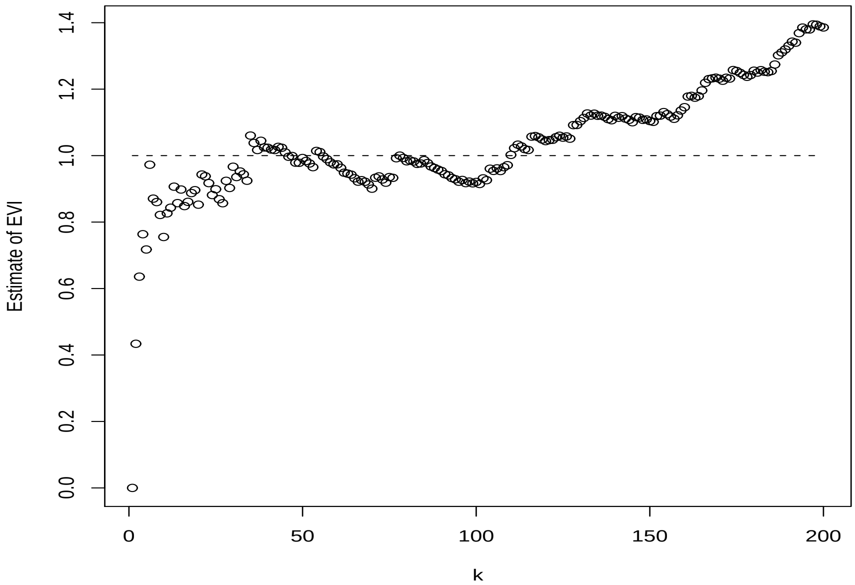

Figure 1.1: Different estimates of γ according to the Hill's estimator and a different number of k. The data consists of 250 random values generated from Fréchet distribution with $\gamma = 1$ (indicated by the dashed line).

Hill's estimator is suitable only for the Fréchet domain, i.e. $\gamma > 0$, however the estimation can be extended to any domain, an example is *the moment estimator* introduced by Dekkers et al. (1989)

$$\hat{\gamma}_{n,k}^{M} := M_{n,k}^{(1)} + 1 - \frac{1}{2}\left(1 - 2\frac{M_{n,k}^{(1)}}{M_{n,k}^{(2)}}\right)^{-1} \tag{1.37}$$

where

$$M_{n,k}^{(j)} = \frac{1}{k}\sum_{i=1}^{k}\left(\log X_{n-i+1:n} - \log X_{n-k:n}\right)^{j}, \qquad j = 1, 2. \tag{1.38}$$

This estimator which can be seen as a correction of Hill's estimator (note that $M_{n,k}^{(1)} = \hat{\gamma}_{n,k}^{H}$) is consistent for $\gamma \in \mathbb{R}$, see de Haan and Ferreira (2006), pp. 100–109.

Both estimators introduced in this section are scale invariant only. This can pose a serious problem during the analysis of a real dataset as even adding a constant to all values of the sample can affect the estimation. Even worse, Hill's estimator is actually surprisingly sensitive to such changes in location, some examples are provided by Resnick (2007), pp. 88. A natural question arises at this point: Why to use the Hill's estimator at all? In fact, if using carefully, Hill estimator can be a very efficient estimator, which can be also seen as the best solution from one point of view. For $\gamma > 0$ it has the lowest

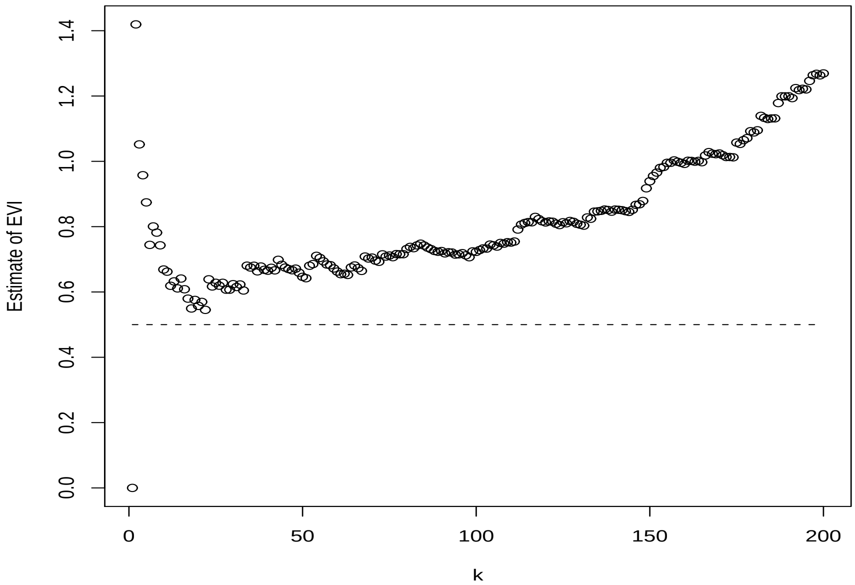

Figure 1.2: Different estimates of γ according to the Hill's estimator and a different number of k. The data consists of 250 absolute values of random values generated from Student's t-distribution with 2 degrees of freedom ($\gamma = 0.5$ is indicated by the dashed line). Notice that the bias is larger and faster growing than in Figure 1.1.

asymptotic variance among all location and scale invariant estimators of γ. Moreover, if $\gamma > 0$ and $\gamma > \rho$ the asymptotic variance of the Hill's estimator cannot be attained by estimators which are also location and not only scale invariant, see Drees (1998a). Taking also the bias into account, it turns out that no estimator having minimal mean-squared error simultaneously for all underlying distribution functions satisfying the second order condition exits, see Drees (1998b). Each of the estimators aims at a slightly different aspect of the underlying extreme value distribution of the sample and none of them can hold the title "the best one". It is therefore strongly advisable to estimate γ using more procedures if possible.

1.4.2 Location and scale invariant estimators

In this thesis we would rather deal with the estimators which are not only scale but also location invariant. As in the previous section we provide only very basic ideas and refer to a survey literature such as Beirlant et al. (2004) for further details.

One of the simplest, oldest, but also one of the most commonly used estimator is

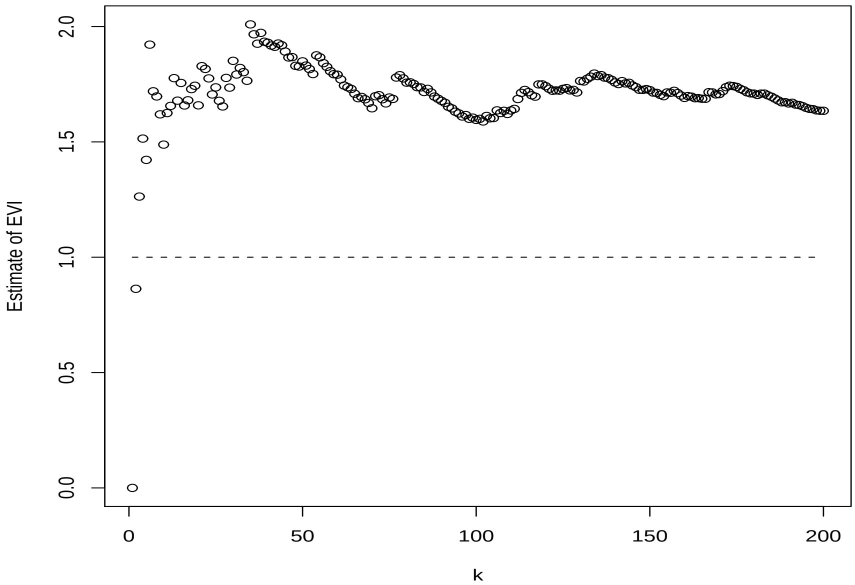

Figure 1.3: Different estimates of γ according to the Hill estimator and a different number of k. The data consists of 250 random values generated from the perturbed Pareto distribution given by $F^{-1}(1-t) = (t|\log(t)|^2)^{-1}$ (real value of $\gamma = 1$ is indicated by the dashed line).

Pickands estimator,

$$\hat{\gamma}^P_{n,k} := \frac{1}{\log 2} \log \frac{X_{n-[k/4]:n} - X_{n-[k/2]:n}}{X_{n-[k/2]:n} - X_{n-k:n}}. \tag{1.39}$$

The estimator is consistent if $k = k_n$ is an intermediate sequence, asymptotic normality can be proven under the second order condition provided that

$$\lim_{n\to\infty} \sqrt{k_n} A\left(\frac{k_n}{n}\right) = \lambda \tag{1.40}$$

see de Haan and Ferreira (2006), pp. 83–89. Note that the Pickands estimator is calculated from just four values of higher order statistics, the other values of the sample are just determining the order of observations. This is a slightly different approach from that of Hill's estimator and also other estimators mentioned so far, which are based on k_n values of observations. An inauspicious consequence is that the volatility of the Pickands plot $\left\{(k, \hat{\gamma}^P_{n,k}), 1 \le k \le n\right\}$ is usually considerably higher than that of estimators based on infinitely growing number of high order statistics like the Hill estimator. It often requires to have several thousands of observations to guess a stability region of the Pickands plot more accurately, but even in such case one needs to solve similar problems as in the case of the Hill's estimator explained previously.

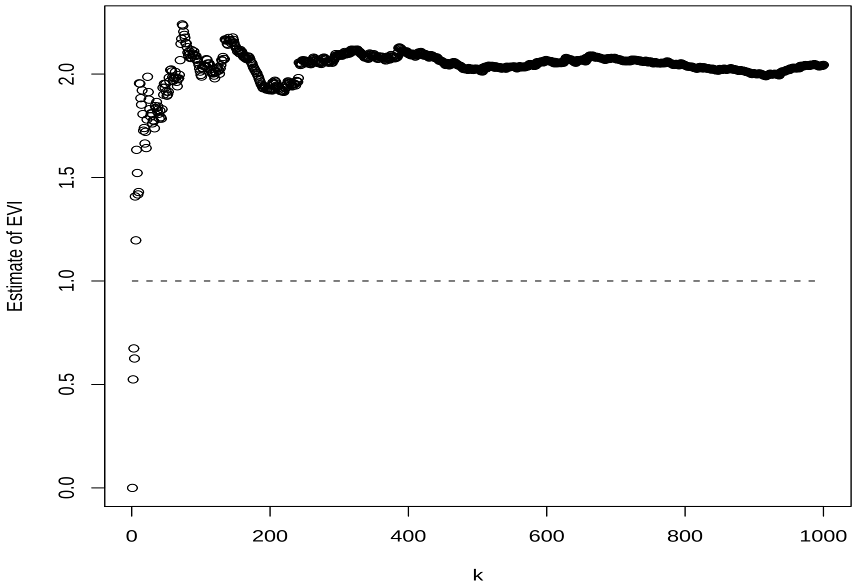

Figure 1.4: Different estimates of γ according to the Hill estimator and a different number of k. The data consists of 10^5 random values generated from perturbed Pareto distribution with given by $F^{-1}(1-t) = (t|\log(t)|^2)^{-1}$ (real value of $\gamma = 1$ is indicated by the dashed line).

Another popular estimator, which is widely used, is *the maximum likelihood* estimator. The idea is as follows: Using relation (1.7) one gets that for some positive function h

$$\lim_{t \uparrow x^*} P\left(\frac{X-t}{h(t)} > x \middle| X > t\right) = 1 - H_\gamma(x) := (1+\gamma x)^{-1/\gamma}. \tag{1.41}$$

In other words, the largest observations approximately follow a generalized Pareto (GP) distribution which opens a possibility of applying the maximum likelihood procedure to the largest observations and using GP distribution as a model. The careful treatment of the threshold t in (1.41) is necessary. It shows up that it is equivalent to consider excesses over $X_{n-k:n}$ where k is an intermediate sequence, see de Haan and Ferreira (2006), pp. 89–91. Using the usual procedure we are looking for a maximum of the density of an approximative likelihood $\prod_{i=1}^k h_\gamma(Z_i)$ where $h_{\gamma,\sigma}(x) = \partial H_\gamma(x/\sigma)/\partial x$ is the density of $H_\gamma(x/\sigma)$ (it is also necessary to estimate the scale parameter $\sigma \in (0,\infty)$) and $Z_i = X_{n-i+1:n} - X_{n-k:n}$. The likelihood equations are given by the partial derivatives

$$\begin{aligned}
\frac{\partial \log h_{\gamma,\sigma}(z)}{\partial \gamma} &= \frac{1}{\gamma}\log\left(1+\frac{\gamma}{\sigma}z\right) - \left(\frac{1}{\gamma}\right)\frac{\frac{z}{\sigma}}{1+\frac{\gamma}{\sigma}z} \\
\frac{\partial \log h_{\gamma,\sigma}(z)}{\partial \sigma} &= -\frac{1}{\sigma} - \left(\frac{1}{\gamma}+1\right)\frac{-\frac{\gamma}{\sigma^2}z}{1+\frac{\gamma}{\sigma}z},
\end{aligned} \tag{1.42}$$

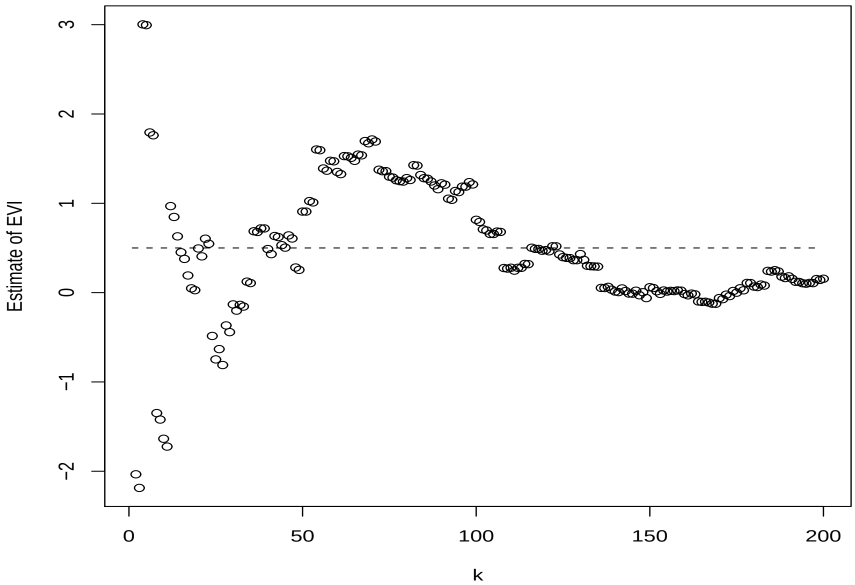

Figure 1.5: Different estimates of γ according to the Pickands estimator and a different number of k. The data consists of 250 absolute values from the sample generated from Student t-distibution with 2 degrees of freedom (real value of $\gamma = 0.5$ is indicated by the dashed line). Note a considerable volatility of the plot in comparison with the Hill estimate on Figure 1.2.

which are for $\gamma = 0$ interpreted as their limits $\gamma \to 0$. From (1.42) follows likelihood equations

$$
\begin{aligned}
\sum_{i=1}^{k} \frac{1}{\gamma^2} \log\Big(1 + \frac{\gamma}{\sigma}(X_{n-i+1:n} - X_{n-k:n})\Big) & \\
- \left(\frac{1}{\gamma} + 1\right) \frac{\frac{1}{\sigma}(X_{n-i+1:n} - X_{n-k:n})}{1 + \frac{\gamma}{\sigma}(X_{n-i+1:n} - X_{n-k:n})} &= 0, \\
\sum_{i=1}^{k} \left(\frac{1}{\gamma} + 1\right) \frac{\frac{\gamma}{\sigma}(X_{n-i+1:n} - X_{n-k:n})}{1 + \frac{\gamma}{\sigma}(X_{n-i+1:n} - X_{n-k:n})} &= k.
\end{aligned} \tag{1.43}
$$

which for $\gamma \neq 0$ can be finally simplified to

$$
\begin{aligned}
\frac{1}{k} \sum_{i=1}^{k} \log\Big(1 + \frac{\gamma}{\sigma}\left(X_{n-i+i:n} - X_{n-k:n}\right)\Big) &= \gamma, \\
\frac{1}{k} \sum_{i=1}^{k} \frac{1}{1 + \frac{\gamma}{\sigma}\left(X_{n-i+1:n} - X_{n-k:n}\right)} &= \frac{1}{\gamma + 1},
\end{aligned} \tag{1.44}
$$

see de Haan and Ferreira (2006), p. 91. The equations (1.44) can be solved only if

$\gamma \geq -1$ as otherwise (1.42) cannot be correctly defined if $\gamma/\sigma \downarrow -(X_{n:n} - X_{n-k:n})^{-1}$. It has been proven that the estimator $\hat{\gamma}_{n,k}^{ML} = \gamma_0$, where $\gamma+$ is the solution of (1.44), is consistent for $\gamma > -1$ if k_n is an intermediate sequence and under the second order condition with an intermediate sequence satisfying (1.40) also asymptotically normal, see de Haan and Ferreira (2006), Theorem 3.4.2. for the region $-1/2 < \gamma < \infty$ and Zhou (2010) for the remaining case $-1 < \gamma \leq -1/2$.

The last estimator we would like to mention in this section is *the probability weighted moments estimator*. It is defined as

$$\hat{\gamma}_{n,k}^{PWM} := \frac{M_n - 4J_n}{M_n - 2J_n} = 1 - \left(\frac{M_n}{2J_n} - 1\right)^{-1}, \tag{1.45}$$

where

$$J_n := \frac{1}{k}\sum_{i=1}^{k} X_{n-i+1:n} - X_{n-k:n} \tag{1.46}$$

and

$$J_n := \frac{1}{k}\sum_{i=1}^{k} \frac{i}{k}\left(X_{n-i+1:n} - X_{n-k:n}\right). \tag{1.47}$$

As the estimator si based on moments of approximate Pareto distribution of the tails and these moments exist only if $\gamma < 1$, the estimator is consistent for intermediate sequences k_n and $\gamma < 1$. The asymptotic normality can be obtained under the second order condition only if $\gamma < 1/2$ with k_n fulfilling (1.40), see de Haan and Ferreira (2006), Theorem 3.6.1.

1.5 Smooth statistical tail functionals

In the previous sections we introduced some popular estimators of γ out of the vast number of estimators suggested by the literature. Each of them has its own asymptotic theory, a different proof of consistency and asymptotic normality, but the conditions under which the properties hold are very similar (they have already shown up at this point, it is some sort of the second order condition must with appropriate rate of intermediate sequence fulfilling (1.40)). The question arises, whether can be the theoretical properties of the estimators dealt on one common ground. The positive answer to this question is given by the theory of smooth functionals of the empirical tail quantile function. The theory was established in Drees (1998*b*) and Drees (1998*a*) and we shall give a brief introduction to it on the following lines.

The central idea of H. Drees' approach is that all the estimators considered previously (and many more) can be (at least approximately) represented as smooth functionals $T(\cdot)$

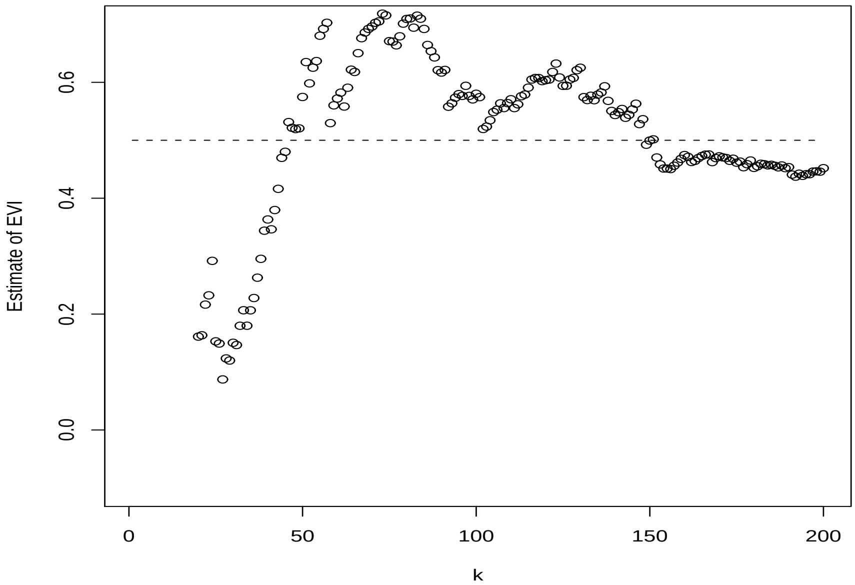

Figure 1.6: Different estimates of γ according to the ML-estimator and a different number of k. The data consists of 250 absolute values of sample generated from Student t-distribution with 2 degrees of freedom ($\gamma = 1/2$ is indicated by the dashed line).

of the empirical tail quantile function which is defined by

$$Q_{n,k}(t) := F_n^{-1}\left(1 - \frac{k_n}{n}t\right) = X_{n-[k_n t]:n}, \qquad t \in [0,1]. \tag{1.48}$$

Note that the distribution of $Q_{n,k}$ is fully determined by the distribution of $k_n + 1$ largest observations. $Q_{n,k}$ can be approximated by

$$F^{-1}\left(1 - \frac{k_n}{n}t\right), \qquad t \in [0,1].,\tag{1.49}$$

which in turn can be approximated (for a suitable standardization and with k_n being an intermediate sequence) by a generalized extreme value distribution as shown in Theorem 1.2.2. In order to get asymptotic results for $T(Q_{n,k})$, it is reasonable to assume that T is both smooth and location and scale invariant, i.e.

$$T(az + b) = T(z) \tag{1.50}$$

for all $a > 0, b \in \mathbb{R}$ and all z belonging to an appropriate function space over which T is smooth. The case where T is only scale invariant, i.e. $T(az) = T(z)$, is analogical to $T(az + b) = T(z)$, but we shall not discuss its details, see Drees (1998a) for the full explanation of the problem. The reason why we largely omit this class of estimators is

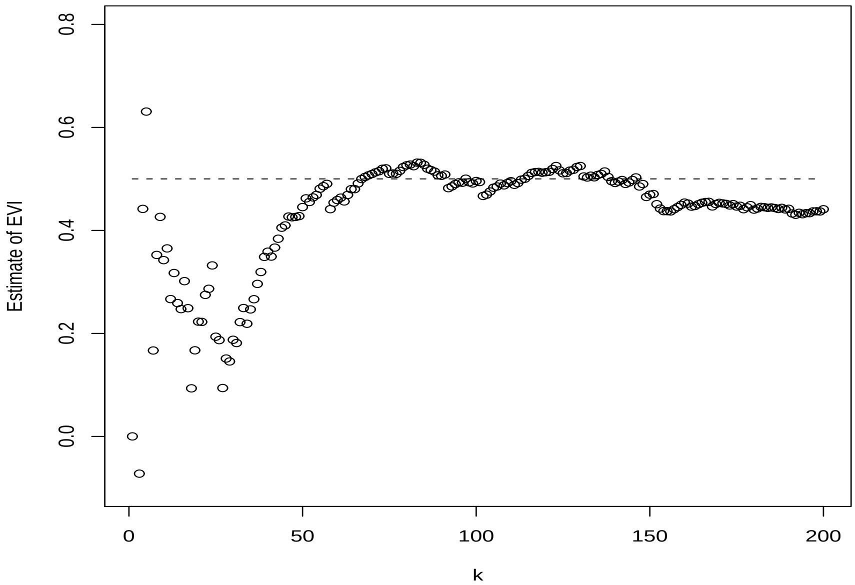

Figure 1.7: Different estimates of γ according to the PWM-estimator and a different number of k. The data consists of 250 absolute values of sample generated from Student t-distribution with 2 degrees of freedom ($\gamma = 1/2$ is indicated by the dashed line).

that they have usually a poor performance when applied to the case of linear models as we shall seen in final sections on simulation of our methods.

The estimator definition as a functional is a natural approach for the both cases (location and scale invariant estimators as well as only scale invariant estimators), one can write all important estimators in this way – e.g. the Hill's estimator is given simply by the functional

$$T_H(z) := \int_0^1 \log^+(z(t)/z(1))\mathrm{d}t, \tag{1.51}$$

if the right-hand side is positive and finite and with $T_H(z) := 0$ otherwise; $\log^+$ means the positive value of log function. Similarly Pickands estimator can be defined as

$$T_{\text{Pick}}(z) := \frac{1}{\log 2} \log\left(\frac{z(1/4) - z(1/2)}{z(1/2) - z(1)}\right) I\left[\frac{z(1/4) - z(1/2)}{z(1/2) - z(1)} > 0\right]. \tag{1.52}$$

It can be easily verified that (1.51) and (1.52) applied to empirical tail quantile function $Q_{n,k}(t)$ created by the largest observations of X_i with $k = k_n$ being an intermediate sequence generate the estimators (1.32) and (1.39) respectively.

More interesting is the reverse idea. If the functional T is smooth in a sequence of neighbourhoods of $Q_{n,k}(t)$ a consistency of the estimator $T(Q_{n,k}(t))$ is assured under the condition that T returns γ when applied to the theoretical counterpart of $Q_{n,k}(t)$.

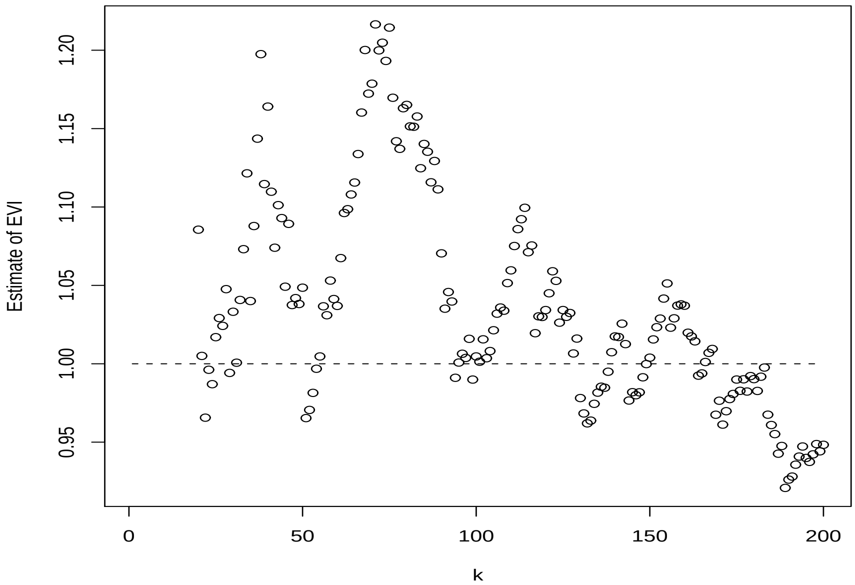

Figure 1.8: Different estimates of γ according to the ML-estimator and a different number of k. The data (absolute values of sample generated from Cauchy distribution) are the same set as in Figure 1.7 ($\gamma = 1$ is indicated by the dashed line). Note large volatility of the plot.

The natural counterpart of $Q_{n,k}(t)$ is the tail quantile function of the associated extreme value distribution z_γ (or $z_\gamma + 1/\gamma$ in the case of estimators which are only scale invariant).

The suitable uniform approximation of $Q_{n,k}(t)$ to z_γ, or, more precisely, that of standardization $(Q_{n,k}(t) - F^{-1}(1 - k_n/n))/a(k_n/n)$ can be established only in a suitable metric space of functions similar to $Q_{n,k}(t)$ living on $[0,1]$. Moreover, as $F^{-1}(1-t)$ and its theoretical counterpart $z_\gamma(t)$ diverges as $t \downarrow 0$ if the right endpoint is infinite (which is true in the case of heavy tailed distribution with $\gamma > 0$ as well as for many distributions with $\gamma = 0$), it is plausible to built the metric in a way which would reflect this fact. A natural way is to introduce weight functions to tie down the empirical tail quantile function $Q_{n,k}(t)$ at $t = 0$. Let

$$\tilde{h}(t) := (t/\log\log(3/t))^{1/2}, \qquad \text{for } t \in [0,1] \tag{1.53}$$

and define accordingly an appropriate space of auxiliary weight functions

$$\mathcal{H} := \left\{ h : [0,1] \mapsto [0,\infty) \,\middle|\, h \in C[0,1],\, \lim_{t\downarrow 0} h(t)/\tilde{h}(t) = 0 \right\}, \tag{1.54}$$

where $C[0,1]$ is the space of continuous functions on the closed interval $[0,1]$. For each $\gamma \in \mathbb{R}$ and $h \in \mathcal{H}$ we define a weighted seminorm $\|\cdot\|_{\gamma,h}$ on the space of real functions

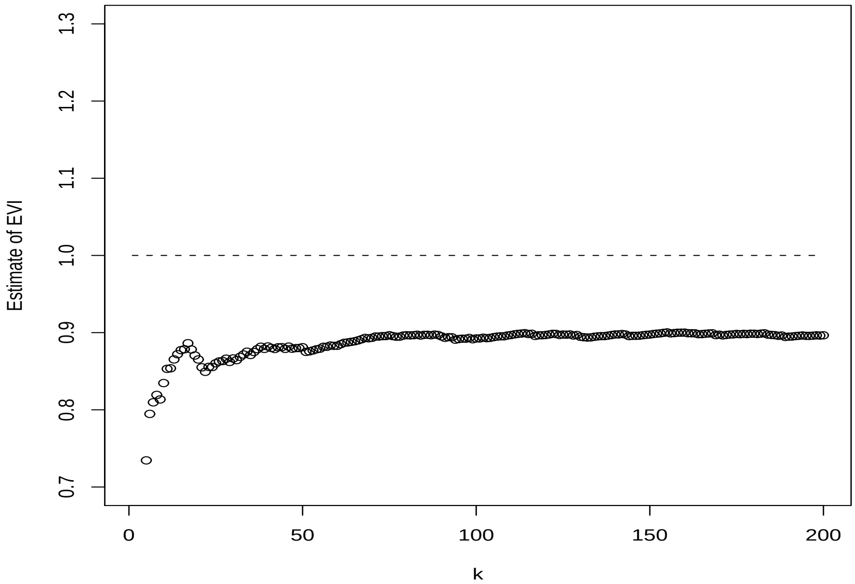

Figure 1.9: Different estimates of γ according to the PWM-estimator and a different number of k. The data, the data are the same set of absolute values from Cauchy distribution as used in Figure 1.8 ($\gamma = 1$ is indicated by the dashed line). The estimator is not consistent for this value of γ, which is in fact indicated by the plot showing that the real value of γ lies is 1 or larger.

on the unit interval by

$$\|z\|_{\gamma,h} := \sup_{t\in[0,1]} t^{\gamma} h(t)|z(t)|. \tag{1.55}$$

(We suppose the convention $0 \cdot \infty = 0$). Note that $\|t^{-\gamma-1}W(t)\|_{\gamma,h} < \infty$ with $W(t)$ being a Wiener process. Finally the space containing all possible empirical tail quantile functions and their theoretical counterparts is the space of real functions on the unit interval equipped with seminorm $\|\cdot\|_{\gamma,h}$ is

$$\mathcal{D}_{\gamma,h} := \left\{ z : [0,1] \to \mathbb{R} \,\middle|\, \lim_{t\downarrow 0} t^{\gamma} h(t) z(t) = 0,\ (t^{\gamma} h(t) z(t))_{t\in[0,1]} \in D[0,1] \right\}. \tag{1.56}$$

Now we can finally establish the approximation of $Q_{n,k}$.

Theorem 1.5.1. *Under the second order condition (EVT.2), there exists a sequence of Wiener processes $\{W_n(t)\}_{n\in\mathbb{N}}$ such that for all $h \in \mathcal{H}$ and functions a, A and K as in (1.25) it holds*

$$\left\| \frac{Q_{n,k}(t) - F^{-1}\left(1-\frac{k}{n}\right)}{a\left(\frac{k}{n}\right)} - \left(z_\gamma(t) - k^{-1/2}t^{-(\gamma+1)}W(t) + A\left(\frac{k}{n}\right)K(t)\right)\right\|_{\gamma,h} = o_P\left(k^{-1/2} + |A(k/n)|\right), \quad (1.57)$$

where $k = k_n$ is an intermediate sequence $k \to \infty$, $k/n \to 0$, $n \to \infty$.

Proof. See Theorem 2.1 in Drees (1998*b*). A slightly different formulation can be found in de Haan and Ferreira (2006), Theorem 2.4.2. □

An analogue to Theorem 1.5.1 can be also written if we can assume only the domain of attraction condition (EVT.1). Nevertheless, to write an approximation similar to (1.57) we need to specify the rate of convergence of $R(t,x)$ in (1.25) to zero.

Theorem 1.5.2. *Suppose that it holds (EVT.1). Then there exists a sequence of Wiener processes $\{W_n(t)\}_{n\in\mathbb{N}}$ such that for all $h \in \mathcal{H}$ and $\varepsilon > 0$ and the function a as in (1.6) it holds*

$$\left\| \frac{Q_{n,k}(t) - F^{-1}\left(1-\frac{k}{n}\right)}{a\left(\frac{k}{n}\right)} - \left(z_\gamma(t) - k^{-1/2}t^{-(\gamma+1)}W_n(t)\right)\right\|_{\gamma,h} = o_P\left(k^{-1/2}\right) + \mathcal{O}\left(\sup_{x\in(0,1+\varepsilon)} x^{\gamma+\frac{1}{2}}\left|R\left(\frac{k}{n},x\right)\right|\right), \quad (1.58)$$

where $k = k_n$ is an intermediate sequence $k \to \infty$, $k/n \to 0$, $n \to \infty$.

Proof. See Theorem 2.1 in Drees (1998*b*). □

Notably, if the convergence of $(F^{-1}(1-tx)-F^{-1}(t))/a(t)$ to z_γ is faster than any negative power of t, i.e.

$$\lim_{t\to 0} t^\alpha \left(\frac{F^{-1}(1-tx)-F^{-1}(t)}{a(t)} - \frac{x^{-\gamma}-1)}{\gamma}\right) = 0 \quad (1.59)$$

for all $x \in [0,1]$ and $\alpha < 0$, then the right side of (1.58) is obviously $o(k_n^{-1/2})$.

Theorems 1.5.1 and 1.5.2 describe the remainder of $Q_{n,k}(t)$ or the joint distribution of the upper order statistics. While any reasonable estimator of γ is supposed to be a function of the upper order statistics it can be also written in terms of $Q_{n,k}(t)$. Hence any such an estimator can be written as a functional applied on the empirical tail quantile function: $T(Q_{n,k}(t))$. If T is smooth and it returns γ for z_γ, it returns similar values for functions nearby the associated Pareto distribution of the sample or in other words $T(Q_{n,k}(t))$ is consistent. We shall consider a class of statistical functionals fulfilling the following properties:

Assume that for $\gamma \in \mathbb{R}$ and some $h \in \mathcal{H}$ the functional $T : \operatorname{span}(\mathcal{D}_{\gamma,h}, 1) \to \mathbb{R}$ satisfies

(T.1) $T_{|\mathcal{D}_{\gamma,h}}$ is $\mathcal{B}(\mathcal{D}_{\gamma,h})$, $\mathcal{B}(\mathbb{R})$-measurable (with $\mathcal{B}(\cdot)$ denoting the appertaining Borel-σ-field)

(T.2) $T(az+b) = T(z) \quad$ for all $z \in \mathcal{D}_{\gamma,h}$, $a > 0$, $b \in \mathbb{R}$,

(T.3) $T(z_\gamma) = \gamma$,

where the z_γ is defined same as in (1.24).

Theorem 1.5.3. *If $F \in MDA(G_\gamma)$, and k_n is an intermediate sequence, T satisfies conditions (T.1–3) and, in addition, $T_{|\mathcal{D}_{\gamma,h}}$ is continuous in z_γ, then*

$$T\left(Q_{n,k}\right) \xrightarrow[n\to\infty]{\mathrm{P}} \gamma. \tag{1.60}$$

Proof. See Drees (1998*b*), Theorem 3.1. □

If functional $T_{|\mathcal{D}_{\gamma,h}}$ is also differentiable, one can obtain an asymptotic normality as well.

(T.4) $T_{|\mathcal{D}_{\gamma,h}}$ is Hadamard differentiable tangentially to $C_{\gamma,h}$ at z_γ with derivative T'_γ, where

$$C_{\gamma,h} := \left\{ z \in \mathcal{D}_{\gamma,h} \,\middle|\, z_{|(0,1]} \in C(0,1] \right\},$$

i.e. there exists signed measure $\nu_{T,\gamma}$ and continuous linear map $T'_\gamma : C_{\gamma,h} \to \mathbb{R}$ such that

$$\frac{T(z_\gamma + \varepsilon_n y_n) - T(z_\gamma)}{\varepsilon_n} \xrightarrow{\varepsilon\to 0} T'_\gamma(y) = \int_0^1 y \mathrm{d}\, \nu_{T,\gamma} \tag{1.61}$$

for any $y_n \in C_{\gamma,h}$ converging to $y \in C_{\gamma,h}$ and for all sequences $\varepsilon_n \downarrow 0$.

The asymptotic normality of functional $T_{|\mathcal{D}_{\gamma,h}}$ can be then written in terms of Riesz representation of its Hadamard derivative defined by (1.61).

Theorem 1.5.4. *Assume that $T : span(\mathcal{D}_{\gamma,h}, 1) \to \mathbb{R}$ satisfies conditions (T.1)–(T.4). Let*

$$\sigma_{T,\gamma} := \int_{[0,1]\times[0,1]} (st)^{-\gamma+1} \min(s,t) \nu^2_{T,\gamma}(\mathrm{d}\, s, \mathrm{d}\, t) \tag{1.62}$$

and

$$\mu_{T,\gamma,K_{\gamma,\rho}} := \int_0^1 K_{\gamma,\rho} \mathrm{d}\, \nu_{T,\gamma} \tag{1.63}$$

Then under the condition (EVT.2) and $\lim_{n\to\infty} k_n A\left(k/n\right) = \lambda \in [0,\infty]$ *follows*

(i) $\lambda \in (0,\infty)$

$$k^{1/2}\left(T(Q_n) - \gamma\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(\lambda\mu_{T,\gamma,\rho}, \sigma^2_{T,\gamma}). \tag{1.64}$$

(ii) $\lambda = \infty$

$$k^{1/2}\left(T(Q_n) - \gamma\right) \xrightarrow[n\to\infty]{\mathrm{P}} \mu_{T,\gamma,\rho} \tag{1.65}$$

Moreover, if only (EVT.1) holds and

$$\sup_{x\in(0,1+\varepsilon]} x^{\gamma+1/2}\left|R\left(k_n/n, x\right)\right| = o\left(k_n^{-1/2}\right) \qquad \textit{for some } \varepsilon > 0, \tag{1.66}$$

then

$$k^{1/2}\left(T(Q_n) - \gamma\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2_{T,\gamma}). \tag{1.67}$$

Proof. See Drees (1998*b*), Theorem 3.2. □

An analogue to Theorems 1.5.3 and 1.5.4 can be derived for scale invariant estimators, which are not location invariant as well (which is the case of (1.51). In this case it is plausible to assume, that instead of (T.2)–(T.3) it holds

(T.5) $T(az) = T(z)$ for all $z \in \mathcal{D}_{\gamma,h}$, $a > 0$,

(T.5) $T(\overline{z}_\gamma) = \gamma$,

where

$$\overline{z}_\gamma(x) := \begin{cases} x^{-\gamma} & \gamma > 0, \\ -\log(x) \quad \text{if} & \gamma = 0, \\ -x^{-\gamma} & \gamma < 0. \end{cases} \tag{1.68}$$

If it holds (T.1) and (T.5)–(T.6), the estimator $T(Q_{n,k}(t)$ is consistent. Moreover if the functional is also Hadamard differentiable tangentially to $C_{\gamma,h}$ at $\overline{z}_\gamma$, $T(Q_{n,k}(t)$ is also asymptotically normal, see Drees (1998*a*) for details.

Virtually almost any location and scale invariant estimator can be written in terms of functionals fulfilling (T.1)–(T.4) (resp. (T.5)–(T.6)). We have already introduced the functional generating Pickands estimator

$$T_{\text{Pick}}(z) = \frac{1}{\log 2}\log\left(\frac{z(1/4) - z(1/2)}{z(1/2) - z(1)}\right) I\left[\frac{z(1/4) - z(1/2)}{z(1/2) - z(1)} > 0\right] \tag{1.69}$$

Functional (1.52) is location and scale invariant and also Hadamadard differentiable in z_γ with derivative

$$T'_{\text{Pick},\gamma}(z) := \frac{\gamma}{(2^\gamma - 1)\log 2}\left(z(1) - (1 + 2^{-\gamma})z(1/2) + 2^{-\gamma}z(1/4)\right) \tag{1.70}$$

and appertaining signed measure

$$\nu_{\text{Pick},\gamma} = \frac{\gamma}{(2^\gamma - 1)\log 2}\left(\delta_1 - (1 + 2^{-\gamma})\delta_{1/2} + 2^{-\gamma}\delta_{1/4}\right), \tag{1.71}$$

where δ_x denotes the Dirac measure with mass 1 at x. If the second order condition (EVT.2) holds and $\lambda = \lim_{n\to\infty} \sqrt{k_n} A(k_n/n)$ then

$$k_n^{1/2} \left(T_{\mathrm{Pick}}(Q_{n,k}) - \gamma\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(\lambda \mu_{\mathrm{Pick},\gamma,\rho}, \sigma^2_{\mathrm{Pick},\gamma}\right) \tag{1.72}$$

with

$$\mu_{\mathrm{Pick},\gamma,\rho} = \frac{\gamma}{\log(2)\left(2^\gamma - 1\right)} \left(-1\left(1 + 2^{-\gamma}\right) A(1/2) + 2^{-\gamma} A(1/4)\right) \tag{1.73}$$

and

$$\sigma^2_{\mathrm{Pick},\gamma} = \left(\frac{\gamma}{\log(2)\left(2^\gamma - 1\right)}\right)^2 \left(1 + 2^{2\gamma+1}\right). \tag{1.74}$$

A generalized version of PWM-estimator introduced in (1.45) can be generated by

$$T_{\mathrm{GPWM},\nu_1,\nu_2}(z) := \frac{\int_0^1 z \mathrm{d}\nu_1}{\int_0^1 z \mathrm{d}\nu_2} I\left[\int_0^1 z \mathrm{d}\nu_2 \neq 0\right] \tag{1.75}$$

for suitable finite signed Borel measures ν_1 and ν_2 on $[0,1]$, e.g. the basic version of PWM-estimator defined in (1.45) is generated by

$$T_{\mathrm{PWM}}(z) := \frac{\int_0^1 (z(t) - z(1))(1-4t)\mathrm{d}t}{\int_0^1 (z(t) - z(1))(1-2t)\mathrm{d}t} I\left[\int_0^1 (z(t) - z(1))(1-2t)\mathrm{d}t > 0\right] \tag{1.76}$$

By simple arithmetic one immediately gets that $T_{\mathrm{PWM}}(z_\gamma) = \gamma$. As that the estimator is continuous and Hadamard differentiable in z_γ, $T_{\mathrm{PWM}}(Q_{n,k})$ is consistent and asymptotically normal in the region described on page 22.

The concept of smooth tail functionals is broad enough to cover the estimators, which can be defined only in an implicit form. The most important example of such estimators is the ML-estimator defined by the equations (1.44). In terms of functions we define the estimator as the first coordinate of a solution of the equations

$$\eta(Q_{n,k}, \gamma, \sigma) = \left(\eta_1\left(Q_{n,k}, \gamma, \sigma\right), \eta_2\left(Q_{n,k}, \gamma, \sigma\right)\right) = (0,0), \tag{1.77}$$

where

$$\eta_1(z,\gamma,\sigma) := \int_0^1 \frac{\mathrm{d}t}{1 + \frac{\gamma}{\sigma}\left(z(t) - z(1)\right)} - \frac{1}{\gamma+1}, \tag{1.78}$$

$$\eta_2(z,\gamma,\sigma) := \int_0^1 \log\left(1 + \frac{\gamma}{\sigma}\left(z(t) - z(1)\right)\right) \mathrm{d}t - \gamma, \tag{1.79}$$

see relations (1.44). Have for simplicity $\gamma_0 > 0$, then consider $\eta\,(z_\gamma,\cdot,\cdot)$ as an element of two dimensional continuous functions $C^2(\Theta)$, where Θ is some neighbourhood of $(\gamma_0, 1)$. Then $\eta(z_\gamma,\cdot,\cdot)$ is differentiable at $(\gamma_0, 1)$ with an invertible Jacobian matrix

$$J_{\eta,\gamma_0} = \frac{\gamma_0}{1+\gamma_0}\begin{pmatrix} \frac{1}{(\gamma_0+1)(2\gamma_0+1)} & \frac{1}{(2\gamma_0+1)} \\ -1 & -1 \end{pmatrix}. \tag{1.80}$$

Hence using implicit function theorems 1.4.7. and 1.4.2 in Rieder (1994) there exist a neighbourhood U of function $\eta(z_\gamma,\cdot,\cdot)$ and a map $M : U \to \Theta$ such that $\xi(M(\xi)) = 0$ for all functions $\xi \in U$. Moreover, M is uniquely determined on the subset of continuously differentiable functions in U and it is Hadamard differentiable at $\eta(z_\gamma,\cdot,\cdot)$ with derivative

$$M'_{\gamma_0} = -J^{-1}_{\eta,\gamma_0}\xi(\gamma_0, 1).$$

Define η as continuous mapping from $\mathcal{D}^*_{\gamma_0,h} \to C^2(\Theta)$, where

$$\mathcal{D}^*_{\gamma_0,h} := \{z \in \mathcal{D}_{\gamma_0,h} | z \text{ non-increasing}\}$$

for some suitable weight function $h \in \mathcal{H}$. Drees (1998*b*) have shown that η is Hadamard differentiable at z_γ tangentially to $C_{\gamma_0,h}$ with derivative

$$\eta'_{\gamma_0}(z)(\gamma,\sigma) := \left(-\int_0^1 \frac{\gamma(z(t)-z(1))}{\sigma\left(1+\frac{\gamma}{\gamma_0\sigma(t^{-\gamma_0}-1)}\right)^2}\mathrm{d}t, \int_0^1 \frac{\gamma(z(t)-z(1))}{\sigma\left(1+\frac{\gamma}{\gamma_0\sigma(t^{-\gamma_0}-1)}\right)}\mathrm{d}t \right). \tag{1.81}$$

Finally we can define the functional for ML-estimate as the first coordinate of $M \circ \eta$, i.e. map $\mathcal{D}_{\gamma_0,h} \to \mathbb{R}$. One gets functional T which is unique on a neighbourhood of z_γ and Hadamard differentiable in z_γ tangentially to $C_{\gamma_0,h}$. Its derivative is given by the chain rule from (1.81) and (1.80), which gives after some calculations

$$T'_{\gamma_0}(z) = \frac{(\gamma_0+1)^2}{\gamma_0}\int_0^1 \left(t^{\gamma_0} - (2\gamma_0+1)t^{2\gamma_0}\right)(z(t)-z(1))\,\mathrm{d}t,$$

see Drees (1998*b*) for more details.

1.6 High quantiles and other characteristics

Recall again the introductory example of this chapter on page 2. We shall estimate $F^{-1}(1-p)$ for a small probability $p = p_n \to 0$ as $n \to \infty$. If $np_n \to \infty$, then the estimator $F_n^{-1}(1-p) = X_{n-[np]:n}$ is reasonable and EVT is not needed. On the other hand if $np_n = \mathcal{O}(1)$ as $n \to \infty$, then the empirical quantile $X_{n:n}$ is usually a poor estimate and some approximation by EVT could be a better match.

We shall restrict ourselves to the heavy tailed case, i.e. $F \in \mathrm{MDA}(G_\gamma)$, $\gamma > 0$. Using (1.11) we get also

$$\lim_{t\to 0} \frac{F^{-1}(1-tx)}{F^{-1}(1-t)} = x^{-\gamma}. \tag{1.82}$$

Replacing t with k/n, where $k = k_n$ is an intermediate sequence, i.e. $k \to \infty$ and $k/n \to 0$, and x with np/k one gets an approximation

$$F^{-1}(1-p) \quad \approx \quad F^{-1}\left(1-\frac{k}{n}\right)\left(\frac{np}{k}\right)^{-\gamma} \approx X_{n-k:n}\left(\frac{np}{k}\right)^{-\hat{\gamma}_{n,k}} \tag{1.83}$$

where $\hat{\gamma}_{n,k}$ is some consistent estimate of γ. Hence we shall define the estimate of high quantile $F^{-1}(1-p)$ as

$$\hat{X}_{p,n,k} := X_{n-k:n}\left(\frac{np}{k}\right)^{-\hat{\gamma}_{n,k}} \tag{1.84}$$

It remains to solve, how far the estimate from its theoretical value is. We get

$$\begin{aligned}
\log \frac{\hat{X}_{p,n,k}}{F^{-1}(1-p)} &= \log \frac{X_{n-k:n}}{F^{-1}(1-k/n)} + \log\left(\frac{F^{-1}(1-k/n)}{F^{-1}(1-p)}\left(\frac{np}{k}\right)^{-\hat{\gamma}_{n,k}}\right) \\
&= \log \frac{X_{n-k:n}}{F^{-1}(1-k/n)} + \log\left(\frac{F^{-1}(1-k/n)}{F^{-1}(1-p)}\left(\frac{np}{k}\right)^{-\gamma}\right) \\
&\qquad -(\hat{\gamma}_{n,k}-\gamma)\log\left(\frac{np}{k}\right).
\end{aligned} \tag{1.85}$$

For the sake of simplicity suppose that F is a member of the Hall class, i.e. $F^{-1}(1-t) = ct^{-\gamma}(1 + dt^{-\rho} + o(t^{-\rho}))$ for some $c > 0$, $\rho \le 0$, and $d \in \mathbb{R}$. Moreover we shall suppose that k is optimal, i.e. $k = \mathcal{O}(n^{2\rho/(2\rho-1)})$. Let $\{U_{1:n}, \dots, U_{n:n}\}$ be an ordered sample of the uniform $U[0,1]$ distribution then

$$\log\left(\frac{n}{k}U_{k:n}\right)^{-\gamma} \stackrel{\mathcal{D}}{=} \gamma\left(E_{n-k:n} - \log\frac{n}{k}\right), \tag{1.86}$$

where $\{E_{k:n}\}_{k=1}^n$ is an ordered sample generated by the standard exponential distribution. Moreover as $\sqrt{k}\left(E_{n-k:n} - \log\frac{k}{n}\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$ by the central limit theorem as $k \to \infty$ and $k/n \to 0$, it holds that (1.86) is $\mathcal{O}(k^{-1/2})$. For the first term on the right side of (1.85) it holds

$$\begin{aligned}
\log \frac{X_{n-k:n}}{F^{-1}(1-k/n)} &\stackrel{\mathcal{D}}{=} \log \frac{F^{-1}(1-U_{k+1:n})}{F^{-1}(1-k/n)} \\
&= \log\left(\left(\frac{n}{k}U_{k+1:n}\right)^{-\gamma}\left(1+\mathcal{O}\left(\left(\frac{k}{n}\right)^{-\rho}\right)\right)\right) \\
&= \mathcal{O}(k^{-1/2} + (k/n)^{-\rho}).
\end{aligned} \tag{1.87}$$

The second term of (1.85) can be approximated using the property of Hall class by

$$\log\left(\frac{F^{-1}(1-k/n)}{F^{-1}(1-p)}\left(\frac{np}{k}\right)^{-\gamma}\right) = \mathcal{O}\left((k/n)^{-\rho}\right). \tag{1.88}$$

For the rate of the estimation it follows from Theorem 1.5.4 that $(\hat{\gamma}_{n,k}-\gamma) = \mathcal{O}_P(k^{1/2})$, see also Drees (1998*c*) for detailed discussion on the possible rates of convergence of the γ-estimators. Have now p such that $\log(np) = o(k^{1/2})$. It follows

$$\begin{aligned} \frac{k^{1/2}}{\log(np/k)}\left(\frac{\hat{X}_{p,n,k}}{F^{-1}(1-p)}-1\right) &= \mathcal{O}\left(\frac{k^{1/2}}{\log(np/k)}\log\frac{\hat{X}_{p,n,k}}{F^{-1}(1-p)}\right) \\ &= \mathcal{O}\left(-k^{-1/2}(\hat{\gamma}_{n,k}-\gamma)\right) \end{aligned} \tag{1.89}$$

where the first approximation is due to the property of the logarithm and the second by (1.85), (1.87), and (1.88).

It follows from (1.89) that from the asymptotic point of view only the asymptotic error of $\hat{\gamma}_{n,k}$ matters if the optimal fraction of k is attained. Only in the case of the moderate sample size (i.e. $k^{1/2} = o(\log(np))$), next term which could not be omitted in the approximations leading to (1.89) would have been of order $\log(np/k)$, cf. with (1.88).

The previous can be seen as a justification why we shall concentrate solely on the development of the asymptotic theory of γ estimation. The question, how to deal with high quantile estimates and a moderate sample size, is neither trivial nor automatically solved, but he problem how to develop asymptotically consistent and normal estimates of γ seems to be more crucial.

The situation slightly differs if $\gamma \le 0$. Relation (1.82) does not hold, hence also the estimate introduced in (1.83) needs some correction. By the domain of attraction condition and (1.6) one gets by some approximations the general version of the estimate for highest quantiles as

$$\begin{aligned} F^{-1}(1-p) &\approx F^{-1}(1-k/n) + a(k/n)\frac{(np/k)^{-\gamma}-1}{\gamma} \\ &\approx X_{n-k:n} + \hat{a}(k/n)\frac{(np/k)^{-\hat{\gamma}_{n,k}}-1}{\hat{\gamma}_{n,k}} =: \hat{X}_{p,n,k}, \end{aligned} \tag{1.90}$$

where an estimate of of the scale parameter $\hat{a}(k/n)$ and its asymptotic properties have to be considered, see some clues in Beirlant et al. (2004), pp. 156–159, and Drees (2003). The estimator suggested in (1.90) is profitable only if $\gamma \ge -1/2$, as in the other case it turns out that it is better to use rather the empirical quantile alone then any approximation by EVT. Suppose $p = o(1/n)$, then with the probability tending to 1 it follows

$$\left|F^{-1}(1-p) - X_{n:n}\right| \le \left|F^{-1}(1) - X_{n:n}\right| = \mathcal{O}(n^{\gamma}) \ll n^{\frac{\rho}{1-2\rho}},$$

for any $\rho \leq 0$. In other words, $X_{n:n}$ is itself tending to $F^{-1}(1-p)$ faster than any estimator using EVT correction such as (1.90).

.

Chapter 2

Regression quantiles and their approximations

This chapter deals with linear models and the concept of quantile regression. We shall describe the properties of *regression quantiles* aiming at their asymptotic representations and approximations by Brownian bridges. The key result is the extension of representations of regression quantiles established by Gutenbrunner et al. (1993) and Jurečková (1999). Under somewhat stronger conditions we get a linear representation of regression quantiles holding uniformly on some interval $[\alpha_n^*, 1-\alpha_n^*] \subset [0,1]$, where of $\alpha_n^* \to 0$ with a rate almost $1/n$. It shows up that important condition is linked with von Mises condition. Finally we shall formulate the approximation of high regression quantiles which provides an analogy to the approximation of the tail quantile function, c.f. with the previous chapter and articles Drees (1998*b*) and Drees (1998*a*).

2.1 Basic model of quantile regression

Consider the linear regression model

$$\mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \mathbf{E} \tag{2.1}$$

where $\mathbf{Y}_n = \mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ is a vector of n-observations, $\mathbf{X}_n = \mathbf{X}$ is an $(n \times p)$ known design matrix, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ is an unknown p-dimensional parameter and $\mathbf{E}_n = \mathbf{E} = (E_1, \ldots, E_n)^\top$ is the vector of i.i.d. errors with a (generally unknown) distribution function F. We assume that $0 < F(x) < 1$ for $x \in (x_*, x^*)$. This define *the lower left* and *the upper right endpoint* x_* respectively x^* as

$$-\infty \le x_* := \sup\{x : F(x) = 0\} \qquad \text{and} \qquad +\infty \ge x^* := \inf\{x : F(x) = 1\}\,. \tag{2.2}$$

For the design matrix assume throughout this text that β_1 is an intercept, i.e. the first column of $\mathbf{X}_n$ equal to $\mathbf{1}_n = (1, \ldots, 1)^\top$. For the notational purpose we shall denote $\mathbf{x}_i \in \mathbb{R}^p$ the i-th row of matrix $\mathbf{X}_n$ ($\mathbf{x}_i \in \mathbb{R}^p, i = 1, \ldots, n$, is treated as a column vector). We define *the regression quantiles* of model (2.1) as the solution of the minimization problem

$$\widehat{\boldsymbol{\beta}}(\alpha) = \widehat{\boldsymbol{\beta}}_n\left(\alpha \middle| Y, \mathbf{x}\right) := \arg\min_{\mathbf{b}\in\mathbb{R}^p} \sum_{i=1}^{n} \rho_\alpha\left(Y_i - \mathbf{x}_i^\top \mathbf{b}\right), \tag{2.3}$$

where ρ_α denotes *the loss function*

$$\rho_\alpha(u) := u\psi_\alpha(u), \qquad \psi_\alpha(u) := (\alpha - I[\, u < 0\,])\,, \qquad u \in \mathbb{R}^1. \tag{2.4}$$

Regression quantiles were introduced by Koenker and Basset (1978) to provide a generalization of quantile idea to linear models. The concept has been widely accepted and has gained a steady popularity. Due to the interest of various authors the theory of quantile regression has undergone fast development during the past years. Currently, the quantile regression is successfully used in various fields of statistics and brought a vast number of interesting applications, for an overview of the recent developments it the theory see Koenker (2005).

By their definition (2.3) regression quantiles are members of the M-estimators class. An M-estimator $\mathbf{M}_n$ is defined for an appropriate loss function ρ as

$$\mathbf{M}_n := \arg\min_{\mathbf{b}\in\mathbb{R}^p} \sum_{i=1}^{n} \rho(Y_i - \mathbf{x}_i^\top \mathbf{b}). \tag{2.5}$$

Thus regression quantiles are special cases of M-estimators with $\rho = \rho_\alpha(\cdot)$ while least square regression is also a member of the class with $\rho = (\cdot)^2$. The class covers various important estimators of linear regression coefficients from the mentioned non-robust least squares estimators to fairly robust Huber's M-estimate. The class was proposed by Peter J. Huber in the sixties in his endeavour to stimulate the development of the robust statistical analysis and derive new robust methods. A detailed description of the M-estimators class can be find in Huber (1981) and in various newer books, e.g. Dodge and Jurečková (2000), pp. 21–25. The asymptotic theory of M-estimators was thoroughly investigated in Jurečková and Sen (1996). We refer to M-estimators class in our context as there are many properties which are common for all members of the class.

The regression quantiles are a little bit more than just arbitrary chosen M-estimators. They also represent a natural generalization of the notion of sample quantiles and order statistics to linear regression models. Note, that the loss $\rho_{1/2} = |u|/2$ leads directly to Laplace's median regression. Thus regression quantiles generalize the ℓ_1-median regression for other quantiles depending on α in ρ_α. The parameter α of the loss function thus determines the probability level of the estimated regression quantile.

Consider the residuals $Y_i - \mathbf{x}_i^\top \mathbf{b}$, $i = 1, \dots, n$ for some $\mathbf{b} \in \mathbb{R}^p$ and plug it to $\rho_\alpha(\cdot)$. The loss is $(1-\alpha)$ times the size of the residual if the observation Y_i is below the plane $\mathbf{x}^\top \mathbf{b}$, where $\mathbf{x} \in \mathbb{R}^p$ is inside a space generated by rows of covariate matrix $\mathbf{X}$. On the other hand the loss is α-times the residual, if the observation Y_i is above the plane. Accordingly, in an univariate one sample location model regression quantiles coincide with *the empirical quantile function*

$$F_n^{-1}(\alpha) = Q_n(\alpha) = X_{k:n} \qquad \text{if } \frac{k-1}{n} < \alpha \le \frac{k}{n}, \quad 0 < \alpha < 1, \quad k = 1, \dots, n. \tag{2.6}$$

This special case is worth of a further attention. It corresponds to $\mathbf{X} = \mathbf{1}_n$, cf. the definition in (2.3). $\widehat{\boldsymbol{\beta}}_n$ is one dimensional for any $n \in \mathbb{N}$ in this case and if we suppose that the errors are not shifted by a location parameter, i.e. $Y_i = E_i$, for $i = 1, \dots, n$, then $\widehat{\boldsymbol{\beta}}_n(\alpha) = F_n^{-1}(\alpha)$ for $\alpha \in (0, 1)$. In the course of this chapter we shall formulate various approximations of regression quantiles of different dimensions and we shall also discuss the possibility of strengthening such assertions. Yet, one must keep in mind that all approximations of regression quantiles are limited by the properties of the empirical quantile function. The multivariate theory cannot be expanded farther then is possible with the univariate quantile case.

A few lines should be written about the computational aspects of the theory. We shall important that the minimization (2.3) problem can be written in the parametric linear programming form

$$\begin{aligned} \alpha \mathbf{1}_n' \mathbf{r}^+ + (1-\alpha) \mathbf{1}_n' \mathbf{r}^- &:= \min \\ \mathbf{X}\boldsymbol{\beta} + \mathbf{r}^+ - \mathbf{r}^- &= \mathbf{Y} \\ (\boldsymbol{\beta}, \mathbf{r}^+, \mathbf{r}^-) \in \mathbb{R}^p \times \mathbb{R}_+^n \times \mathbb{R}_-^n, &\qquad 0 < \alpha < 1, \end{aligned} \tag{2.7}$$

where the regression quantile $\widehat{\boldsymbol{\beta}}_n \left(\alpha \right| \mathbf{Y}, \mathbf{x})$ coincides with the component $\boldsymbol{\beta}$ of the optimal solution. This representation is important both from theoretical and practical point of view. We can formulate also the dual linear programming problem to (2.7) thus obtaining

$$\begin{aligned} \mathbf{Y}^\top \hat{\mathbf{a}}(\alpha) &:= \max \\ \mathbf{X}^\top \left(\hat{\mathbf{a}}(\alpha) - (1-\alpha) \right) \mathbf{1}_n &= \mathbf{0} \\ \hat{\mathbf{a}}(\alpha) \in [0,1]^n, &\qquad 0 < \alpha < 0. \end{aligned} \tag{2.8}$$

The optimal solution $\hat{\mathbf{a}}(\alpha) = (\hat{a}_{n,1}(\alpha), \dots, \hat{a}_{nn}(\alpha))$ forms the vector of so-called *regression rank scores*. Each $\hat{a}_{ni}$ is a continuous, piecewise linear function of $\alpha \in [0,1]$, $\hat{a}_{ni} = 0$ for $i = 1, \dots, n$. The theory of regression rank scores was developed by Gutenbrunner and Jurečková (1992). They proposed tests based on regression rank scores generated by different score functions, which are parallel to classical rank tests.

The theory was later expanded to cover wider class of tests, see Gutenbrunner et al. (1993).

Assertion (2.7) allows to compute regression quantiles quite easily using well known routines of parametric linear programming, which are usually included in statistical software. Computations in this thesis have been obtained using free statistical software `R`. In `R`, quantile regression is widely supported through the library `quantreg`. Roger Koenker himself collaborated on the development of `quantreg` library and also wrote its documentation, see Koenker (2006).

If the idea of quantile regression is to provide a generalization of the univariate quantile idea, we require some kind of consistency when $n \to \infty$. We are interested what is the theoretical counterpart of the α-regression quantile considering conditional (under $X = \mathbf{x}$) quantile function of Y,

$$F_Y^{-1}(\alpha|X = \mathbf{x}) = \mathbf{x}^\top \boldsymbol{\beta}(\alpha). \tag{2.9}$$

Consider the simplest setting, the case of univariate one sample quantiles without location parameter. In that case $\mathbf{x} = 1$ and the theoretical counterpart $\boldsymbol{\beta}(\alpha)$ of $\widehat{\boldsymbol{\beta}}_n(\alpha)$ is simply $\boldsymbol{\beta}(\alpha) = F^{-1}(\alpha)$. In this case, the consistency for any $\alpha \in (0, 1)$ is a direct consequence of Glivenko-Cantelli theorem. Rates of convergence depend on the behaviour of $F^{-1}(\cdot)$ in a neighbourhood of α. It is easy to show that if F has a continuous density f and $f(F^{-1}(\alpha))$ is bounded away from 0 and ∞ at α, one gets also

$$\sqrt{n}(\hat{\beta}_n(\alpha) - F^{-1}(\alpha)) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \sigma_\alpha^2),$$

where

$$\sigma_\alpha := \frac{(\alpha(1-\alpha))^{1/2}}{f^2(F^{-1}(\alpha))}, \tag{2.10}$$

see e.g. Koenker (2005), pp. 71–72. When $f(F^{-1}(\cdot))$ tends to infinity at α, improvements on the rate of convergence beyond $\mathcal{O}(n^{-1/2})$ are possible, while slower rates may prevail in the case $f(F^{-1}(\cdot))$ tends to zero at α. If $F(\cdot)$ is constant on some neighbourhood of $F^{-1}(\alpha)$ it is not possible to distinguish between different estimates in the neighbourhood, i.e. suppose $F(\alpha)$ is constant for $\alpha \in [\alpha_*, \alpha^*]$, for some $0 < \alpha_* < \alpha^* < 1$. Then we can only say that the sum of probabilities that $\hat{\beta}_n(\alpha)$ falls near $\beta(\alpha)$ tends to 1 as $n \to \infty$. A necessary and sufficient condition that $\hat{\beta}_n(\alpha) \to F^{-1}(\alpha)$ in the model of univariate i.i.d. sample is

$$F(F^{-1}(\alpha) - \varepsilon) < \alpha < F(F^{-1}(\alpha) + \varepsilon)$$

for all $\varepsilon > 0$. A reader can find more details about the topic, particularly about the establishment of the consistency for non-i.i.d. cases in Koenker (2005), pp. 117–118.

Consider a slight modification of the univariate model above $Y_i = \beta_1 + E_i, i = 1, \ldots, n$. The theoretical counterpart of the sample (regression) quantile

$$\hat{\beta}(\alpha) = \arg\min_{t\in\mathbb{R}} \sum_{i=1}^{n} \rho_\alpha(Y_i - \beta_1 + t), \qquad \alpha \in (0,1)$$

is $\beta(\alpha) = \beta_1 + F^{-1}(\alpha)$. This observation can be generalised even for higher dimensions. In the general linear regression model with i.i.d. errors we get that $\boldsymbol{\beta}(\alpha)$ in (2.9) should take the form

$$\boldsymbol{\beta}(\alpha) = (\beta_1 + F^{-1}(\alpha), \beta_2, \ldots, \beta_p)^\top \tag{2.11}$$

If the dimension is higher then one, the consistency of $\widehat{\boldsymbol{\beta}}(\alpha)$ also depends on the properties of the sequence of design matrices $\mathbf{X}_n$. In fact, different consistency laws can be established depending on the delicate interplay between conditions imposed on conditional distribution function $F = P(Y_i < y|\mathbf{x})$ (with a suitable set of assumptions the consistency can be achieved also in the case of heteroscedasticity) and the conditions on $\mathbf{X}_n$. We will not go to details here as different rates of asymptotic convergence are possible, see Koenker (2005), pp. 118–119, and we shall aim our attention to the problem of establishing asymptotic normality, where many can be gained from a genaral theory of M-estimators. In fact the following assertions hold for the whole M-estimators class, see Jurečková and Sen (1996), pp. 80–88 and Dodge and Jurečková (2000), pp. 23–24 for a more condensed version.

We shall assume that the distribution function F has a density.

(F.A) There exists a density f of distribution function F and it is bounded away from 0 and ∞ in $F^{-1}(\alpha)$.

Then under assumption that

$$\mathbf{D} := \lim_{n\to\infty} \frac{1}{n}\left(\sum_{j=1}^{n} \mathbf{x}_j\mathbf{x}_j^\top\right)$$

is positive definite matrix and Noether condition on the sequence of design matrices $\mathbf{X}_n$

(X.N)

$$\max\left\{\mathbf{x}_i^\top \left(\sum_{j=1}^{n} \mathbf{x}_j\mathbf{x}_j^\top\right) \mathbf{x}_i \;:\; 1 \le i \le n\right\} \to 0 \qquad \text{as } n\to\infty$$

it holds

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(\mathbf{0},\, \sigma_\alpha^2\mathbf{D}^{-1}), \tag{2.12}$$

where σ_α^2 is as in (2.10) and $\boldsymbol{\beta}(\alpha)$ as in (2.11), see Jurečková and Sen (1996), pp. 85.

2.2 Asymptotic of regression quantile process

Having established the asymptotics of any single regression quantile for $\alpha \in (0,1)$ we may ask whether $\widehat{\boldsymbol{\beta}}(\alpha)$ could be asymptotically approximated uniformly with respect to $\alpha \in (0,1)$ at least in some subset $A \subset (0,1)$. We shall thus seek an asymptotic theory of the (empirical) *regression quantiles process* $\left\{\hat{\boldsymbol{\beta}}_n(\alpha) : \alpha \in A\right\}$ or equivalently some of its standardized variant. In relation to point-wise consistency (2.12) of $\widehat{\boldsymbol{\beta}}(\alpha)$ we define the process

$$\mathbf{q}_n(\alpha) := n^{1/2} f(F^{-1}(\alpha))(\widehat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)). \tag{2.13}$$

Observe the special univariate one-sample case at first, i.e. $\mathbf{X}_n = \mathbf{1}_n$ and $\mathbf{Y}_n = \mathbf{E}_n$. It turns up that this case was already studied thoroughly studied, as (2.13) is identical with *the empirical quantile process*

$$q_n(\alpha) = n^{1/2} f(F^{-1}(\alpha))(F_n^{-1}(\alpha) - F^{-1}(\alpha)), \qquad 0 < \alpha < 1, \tag{2.14}$$

where $F_n^{-1}(\alpha)$ is *the empirical quantile function* defined as

$$F_n^{-1}(\alpha) = \begin{cases} X_{1:n}, & \text{if } t = 0, \\ X_{k:n}, & \text{if } \frac{k-1}{n} < t \leq \frac{k}{n}, \qquad 1 \leq k \leq n. \end{cases} \tag{2.15}$$

Therefore in this special case we can naturally extend the process $\{q_n(\alpha), \alpha \in (0,1)\}$ to $\{q_n(\alpha), \alpha \in [0,1]\}$ by redefining the missing regression quantiles in 0 and 1 using the definition (2.15). In fact this approach corresponds with the definition of *the largest regression quantile*

$$\widehat{\boldsymbol{\beta}}(1) := \arg\min_{\mathbf{b} \in \mathbb{R}^p} \left\{ \sum_{i=1}^{n} \mathbf{x}_i^\top \mathbf{b} \,\middle|\, Y_i \leq \mathbf{x}_i^\top \mathbf{b}, \quad i = 1, \dots, n \right\} \tag{2.16}$$

and its counterpart, *the minimum regression quantile*

$$\widehat{\boldsymbol{\beta}}(0) := \arg\max_{\mathbf{b} \in \mathbb{R}^p} \left\{ \sum_{i=1}^{n} \mathbf{x}_i^\top \mathbf{b} \,\middle|\, Y_i \geq \mathbf{x}_i^\top \mathbf{b}, \quad i = 1, \dots, n \right\}, \tag{2.17}$$

Note that this definition corresponds also with the definition of the regression quantiles we introduced in (2.3). It is

$$\widehat{\boldsymbol{\beta}}(1) = \arg\min_{\mathbf{b} \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_1 \left(Y_i - \mathbf{x}_i^\top \mathbf{b}\right), \tag{2.18}$$

$$\widehat{\boldsymbol{\beta}}(0) = \arg\min_{\mathbf{b} \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_0 \left(Y_i - \mathbf{x}_i^\top \mathbf{b}\right), \tag{2.19}$$

where $\rho_1(u) = u(1 - I[u < 0])$ and $\rho_1(u) = u(-I[u < 0])$, i.e. $\rho_1(u)$ and $\rho_0(u)$ is a left side or right side limit of $\rho_\alpha(u)$ with $\alpha \to 1$ or $\alpha \to 0$. Nevertheless the task to establish an asymptotic theory of the largest and smallest regression quantile is a little bit different from the case $\alpha \in [\varepsilon, 1-\varepsilon]$, see Portnoy and Jurečková (2000), Smith (1994).

The theory of asymptotic representations of $q_n(\alpha)$ is quite elaborate and we can present just a few of the most important results. Our aim is to develop at least the very basic analogies of these results for $\mathbf{q}_n(\alpha)$, for dimensions $p \geq 2$. For more thorough explanation of the asymptotic theory of regression quantiles we refer to Csörgő and Révész (1981) and Csörgő and Horváth (1993) written by the authors of famous “Hungarian construction”. While the first one Csörgő and Révész (1981) represents the most comprehensive explanation of the theory and its goals, some more advanced results can be found only in Csörgő and Horváth (1993), which in fact replenishes and develop the ideas of the first book.

One of the central ideas of the theory is that q_n is by sup-norm distance in a close vicinity of *the uniform quantile function* defined as

$$u_n(\alpha) = n^{1/2}(\alpha - F(F_n^{-1}(\alpha))). \tag{2.20}$$

The process (2.20) is in turn equivalent (in distribution) with the process generated by sample $U_1, \ldots, U_n$, the uniformly distributed variables on unite interval $\mathcal{U}[0,1]$, i.e.

$$\tilde{u}_n(\alpha) = n^{1/2}(\alpha - U_{[\alpha n]:n}). \tag{2.21}$$

It is the celebrated Komlós-Major-Tunsnády result, see e.g. Csörgő and Révész (1981), Theorem 4.4.1., which opened the possibilities how to establish the strong approximation of $u_n(\alpha)$. The consequences proposed for the asymptotic behaviour of the empirical quantile process, see Csörgő and Révész (1981), Theorem 4.5.5., were refined even more in Csörgő and Horváth (1993).

Theorem 2.2.1. *Assume following conditions. Let F be a continuous distribution function F and*

(F.U.1) *F is twice differentiable on (x_*, x^*), where x_* and x^* are given by (2.2).*

(F.U.2) $F'(x) = f(x) > 0, \qquad x \in (x_*, x^*)$

(F.U.3) *for some K_γ we have*

$$\sup_{0<\alpha<1} \alpha(1-\alpha)\frac{|f'(F^{-1}(\alpha))|}{f^2(\alpha)} \leq K_\gamma$$

Then we have

$$\sup_{\frac{1}{n+1} \leq \alpha \leq \frac{n}{n+1}} |q_n(\alpha) - u_n(\alpha)|$$

$$\stackrel{\text{a.s.}}{=} \begin{cases} \mathcal{O}\left(n^{-1/2}(\log\log n)^{1+K_\gamma}\right), & \text{if } K_\gamma \le 1 \\ O\left(n^{-1/2}(\log\log n)^{K_\gamma}(\log n)^{(1+\varepsilon)(K_\gamma-1)}\right), & \text{if } K_\gamma > 1 \end{cases} \tag{2.22}$$

for all $\varepsilon > 0$.

Proof. See Csörgő and Horváth (1993), Theorem 6.1.3., pp. 372–375. □

If we want to estimate the distance between u_n and q_n only in probability (and that is enough for the applications proposed in the next chapter), then we can get somewhat better rates.

Theorem 2.2.2. *Assume that it holds the conditions of of Theorem 2.2.1. Then for the version of uniform quantile process defined as in (2.20) it holds*

$$\sup_{\frac{1}{n+1}\le\alpha\le\frac{n}{n+1}} |q_n(\alpha) - u_n(\alpha)| = \mathcal{O}_P(n^{-1/2}\log\log n) \tag{2.23}$$

Proof. See Csörgő and Horváth (1993), Theorem 6.1.5, pp. 377–378. □

The consequence of the Komlós-Major-Tusnády construction states that for $\alpha \in (0,1)$ the suitably standardized uniform quantile process is from the asymptotic point of view near Brownian bridges. The relation holds also for the weighted uniform quantile process.

Theorem 2.2.3. *Have any version of the uniform quantile process (cf. (2.20) and (2.21)). Then we can define a sequence of Brownian bridges* $\{B_n(\alpha), 0 \le \alpha \le 1\}$ *such that*

$$\sup_{0\le\alpha\le1} |u_n(\alpha) - B_n(\alpha)| \stackrel{\text{a.s.}}{=} O(n^{-1/2}\log n) \tag{2.24}$$

and

$$n^{\frac{1}{2}-\nu} \sup_{\lambda/n\le\alpha\le1-\lambda/n} \frac{|u_n(\alpha) - B_n(\alpha)|}{(\alpha(1-\alpha))^\nu} = O_P(1) \tag{2.25}$$

for all $0 < \nu \le 1/2$ *and* $0 < \lambda < \infty$.

Proof. See Csörgő and Horváth (1993), Theorem 4.2.1, pp. 195–202. □

In turn the weighted empirical quantile process can be approximated by Brownian bridges as is clarified by the following theorem,

Theorem 2.2.4. *Assume that* F *is a continuous distribution function which fulfills*

(i) F *is twice differentiable on* (x_*, x^*),

(ii) $F'(x) = f(x) > 0, x \in (x_*, x^*)$,

(iii) for some $K_\gamma > 0$ it holds

$$\sup_{0<t<1} t(1-t)\frac{\left|f'(F^{-1}(t))\right|}{f^2(F^{-1}(t))} \leq K_\gamma. \tag{2.26}$$

Then for $0 \leq \nu \leq \frac{1}{2}$ we can define a sequence of Brownian bridges $\{B_n(t), 0 \leq t \leq 1\}$ such that

$$n^{\frac{1}{2}-\nu} \sup_{\frac{1}{n+1}\leq\alpha\leq\frac{n}{n+1}} \frac{|q_n(\alpha) - B_n(\alpha)|}{(\alpha(1-\alpha))^\nu} = \begin{cases} O_P(\log n), & \text{if} \quad \nu = 0 \\ O_P(1), & \text{if} \quad 0 < \nu \leq \frac{1}{2} \end{cases} \tag{2.27}$$

for all $0 < \nu \leq 1/2$ and $0 < \lambda < \infty$.

Proof. See Csörgő and Horváth (1993), Theorem 4.2.1, pp. 195–202. □

The previous asymptotic approximations can be also used to obtain an analogy of *the law of iterated logarithm* (LIL) for the empirical quantile process. This follows from Theorem 2.2.1 and the LIL established for the uniform quantile process.

Theorem 2.2.5. *Assume that conditions of Theorem 2.2.1 hold with $K_\gamma < 1$. Then there exists a $C > 0$ such that*

$$\limsup_{n\to\infty} \sup_{\frac{1}{n+1}\leq\alpha\leq\frac{n}{n+1}} (\alpha(1-\alpha)\log\log n)^{-1/2}\left|f(F^{-1}(\alpha)q_n(\alpha)\right| \leq C \qquad a.s. \tag{2.28}$$

Proof. Follows from Theorem 2.2.1 similarly as Theorem 5.3.1 of Csörgő and Révész (1981) follows from Theorem 4.5.6 ibid. See Csörgő and Révész (1981), pp. 162. □

Note that the condition $K_\gamma < 1$ in the previous theorem can be relaxed at the cost of a different rate of convegence in (2.28). However, it is a terminological question whether such a relation could be called "the law of iterated logarithm".

While regression quantiles can be seen as a generalization of the quantile idea suitable for linear moderl (2.1) a natural question arises, which of the interesting properties introduced on the previous pages hold also for the process based on regression quantiles instead of order statistics. We seek a generalization of the univariate (or location) case, or in other words, an approximations getting as close as possible to the rates and boundaries imposed on $\alpha \in [0, 1]$ in the univariate case. Actually, the boundaries for α derived for the univariate case are the best, which are possible. This fact is clarified by Mogul'skiĭ theorem, see Csörgő and Révész (1981), pp. 159–160. Therefore, the rates and boundaries on α in the multivariate case cannot be better.

A general theory of the regression quantile process requires an appropriate definition of the multivariate uniform quantile process. The analogies of the approximations thus leads to a p-variate processes of independent Brownian bridges, which have been considered in Koenker and Machado (1999) and Koenker and Xiao (2002). The goal of this thesis is more modest. We shall establish LIL for regression quantile process as an analogy to (2.28). However, such theorem differs a lot from its univariate location counterpart (2.2.5). The proof in i.i.d. univariate case can be based on the fact that for any random variable X and its distribution function F we have

$$F(X) \stackrel{\mathcal{D}}{=} U, \qquad U \sim \mathcal{U}[0,1]. \tag{2.29}$$

Consequently $U_1 = F(X_1), U_2 = F(X_2), \ldots$ are independent $\mathcal{U}[0,1]$ random variables and the order statistics $X_{1:n} \leq \ldots \leq X_{n:n}$ of the random sample $X_1, \ldots, X_n$ induce the order statistics $U_{1:n} = F(X_{1:n}) \leq \ldots \leq U_{n:n} = F(X_{n:n})$ of the uniform-$[0,1]$ random sample $U_1, \ldots, U_n$. Therefore, if one one has an approximation of the uniform quantile process, the relation (2.29) leads to quite straightforward modifications for a general distribution.

This is not true in the more complicated case of regression quantiles, which are defined as a solution of minimization (2.3). This definition is not easy to handle with and many properties which hold for univariate order statistics hold for regression quantiles only asymptotically. There is no direct correspondence between the regression quantiles and the order statistics of the error. The information about the distribution of the errors (and its tails) is mixed in regression quantiles with the information about the covariate matrix $\mathbf{X}$. In fact, the number of regression quantiles calculated from n observations of Y is not n but it depends on the exact form of $\mathbf{X}$. Usually the number of regression quantiles is much less than n^p if $\widehat{\boldsymbol{\beta}}(\alpha) \in \mathbb{R}^p$; it is approximately $\mathcal{O}_P(n \log(n))$ if the matrix is random and fulfills some reasonable assumptions, see Portnoy (1991).

A tool that can overcome this lack of correspondence as well as the implicit definition of regression quantiles is the Bahadur representation. Well know for the empirical quantile process, see Bahadur (1966), this representation can be established for regression quantile process as well, for more possibilities of construction see Koenker (2005), pp. 122 and the literature cited therein. The basic idea is to establish a representation of the regression quantile process by a sum of weighted random variables

$$n^{1/2}\left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right) = \mathbf{D}^{-1} \frac{1}{n^{1/2}} \sum_{i=1}^{n} \mathbf{x}_i \psi_\alpha \left(Y_i - F_Y^{-1}(\alpha|\mathbf{x}_i)\right) + R_n \tag{2.30}$$

where ψ_α is as in (2.4) and R_n is suitably small (i.e. $R_n = o_P(1)$). The relation (2.30) stands behind many useful approximations of regression quantile process established so far.

Considering assumptions (F.A) and (X.N), c.f. (2.12), we introduce the following set of assumptions on the distribution function F:

(F.B.1) F has a continuous density f, which is positive and finite on $\{t : 0 < F(t) < 1\}$,

and the covariate matrix $\mathbf{X}$:

(X.B.1) $x_{i1} = 1$, $i = 1, \dots, n$ and the other columns of $\mathbf{X}_n$ are orthogonal to the first one,

(X.B.2) $\max_{1 \le i \le n} \|\mathbf{x}_i\| = n^{1/2}$,

(X.B.3) $\mathbf{D}_n = \frac{1}{n}\mathbf{X}_n^\top \mathbf{X}_n \to \mathbf{D}$, where $\mathbf{D}$ is a positive definite $p \times p$ matrix.

Under these assumptions the Bahadur representation of the regression quantile processs holds uniformly on $[\varepsilon, 1-\varepsilon]$ for any $\varepsilon \in (0, 1/2)$.

Theorem 2.2.6. *Suppose that it holds conditions (F.B.1) and (X.B.1-3), then for empirical regression quantile process* $\mathbf{q}_n(\cdot)$ *and any* $\varepsilon \in (0, 1/2)$ *holds*

$$\sup_{\alpha \in [\varepsilon, 1-\varepsilon]} (f(F^{-1}(\alpha)))^{-1} \left\| \mathbf{q}_n(\alpha) - \mathbf{D}_n^{-1}\mathbf{B}_n(\alpha) \right\| = o_P(1) \tag{2.31}$$

where $\mathbf{B}_n(\alpha)$ *is a vector of* p *independent Brownian bridges on* $(0, 1)$.

Proof. See Gutenbrunner and Jurečková (1992), Theorem 1. □

Theorem 2.2.6 provide uniform asymptotic representation of the process $\mathbf{q}_n$. This extension expands the potential scope of both estimation and inference methods based on $\hat{\boldsymbol{\beta}}_n(\alpha)$ albeit at the cost of somewhat stronger conditions than (F.A) and (X.N), which were sufficient to assure consistency of a single regression quantile. The result holds on a compact subset $[\varepsilon, 1-\varepsilon]$ of the unite interval.

Similarly as in the univariate approximations we shall strenghten the result by extending its validity to the edge of unit interval as much as possible. In the ideal case we would get the approxination of the regression quantile process over whole unit interval as in the univariate case. However, this goal has not been achieved so far. Nevertheless, various analogies to Theorem 2.2.6 holding on $[\alpha_n, 1-\alpha_n]$, where $\alpha_n \to 0$ with a suitable order have been already introduced in the literature.

The relation was studied by Gutenbrunner et al. (1993). They assume the following set of assumptions on the distribution function of errors F:

(F.G.1) $|F^{-1}(\alpha)| \le (\alpha(1-\alpha)^{-a}$ for $0 < \alpha \le \alpha_0$, $1 - \alpha_0 \le \alpha < 1$, where $0 < a \le \frac{1}{4} - \varepsilon$, $\varepsilon > 0$ and $c > 0$,

(F.G.2) $1/f(F^{-1}(\alpha)) \le c(\alpha(1-\alpha))^{-1-a}$ for $0 < \alpha \le \alpha_0$ and $1 - \alpha_0 \le \alpha < 1$, $c > 0$.

(F.G.3) $f(x) > 0$ is absolutely continuous, bounded and monotonically decreasing as $x \to x_*$ and $x \to x^*$, where x_* and x^* are given by (2.2). Moreover the derivative f' is bounded almost everywhere.

(F.G.4)

$$\frac{|f'(x)|}{|f(x)|} \le c|x| \qquad \text{for } |x| \ge K \ge 0, \qquad c > 0.$$

Another set of assumptions is imposed on the sequence of covariate matrices $\mathbf{X}_n$

(X.G.1) $n^{-1}\sum_{i=1}^{n}\|\mathbf{x}_i\|^4 = O(1)$ as $n \to \infty$.

(X.G.2) $\max_{1\le i\le n}\|\mathbf{x}_i\| = O(n^{(2(b-a)-\delta)/(1+4b)})$ for some $b > 0$ and $\delta > 0$ such that $0 < b - a < \varepsilon/2$ (hence $0 < b < \frac{1}{4} - \varepsilon/2$).

For the parameter b, which is related to a from (F.G.1–4) by (X.G.2), define

$$\alpha_n^* := n^{-\frac{1}{1+4b}}. \tag{2.32}$$

Following theorem establishes the uniform asymptotic representation of regression quantiles on $[\alpha_n^*, 1-\alpha_n^*]$.

Theorem 2.2.7. *Under the conditions (F.G.1–4), (X.B.1–2), and (X.G.1–2) it holds*

$$\frac{n^{1/2}f(F^{-1}(\alpha))}{(\alpha(1-\alpha))^{1/2}}\left(\widehat{\boldsymbol{\beta}}_n\left(\alpha|\,\mathbf{Y},\mathbf{x}\right)-\boldsymbol{\beta}(\alpha)\right)$$
$$= n^{-1/2}(\alpha(1-\alpha))^{-1/2}\mathbf{D}_n^{-1}\sum_{i=1}^{n}\mathbf{x}_{ni}\psi_\alpha(E_{i\alpha}) + o_P(1) \tag{2.33}$$

uniformly in $\alpha_n^* \le \alpha \le 1-\alpha_n^*$, where $E_{i\alpha} = E_i - F^{-1}(\alpha)$. Consequently,

$$\begin{aligned}\sup_{\alpha_n^*\le\alpha\le1-\alpha_n^*}\left\|\frac{f(F^{-1}(\alpha))}{(\alpha(1-\alpha))^{1/2}}(\widehat{\boldsymbol{\beta}}_n\left(\alpha|\,\mathbf{Y},\mathbf{x}\right)-\boldsymbol{\beta}(\alpha))\right\| &= \\ \sup_{\alpha_n^*\le\alpha\le1-\alpha_n^*}\left\|\frac{\mathbf{q}_n(\alpha)}{(n\alpha(1-\alpha))^{1/2}}\right\| &= \mathcal{O}_P(n^{-1/2}C_n)\end{aligned} \tag{2.34}$$

Proof. See Gutenbrunner et al. (1993), Theorem 3.1. $\square$

Though the preceding theorem is a big leap forward, its applications are strictly limited by its assumptions. While the conditions (F.G.1)–(F.G.4) are satisfied by some common distributions, e.g. by the normal, logistic, double exponential and even t distributions with 5, or more, degrees of freedom, a lot of distributions is excluded. For example, the heavy tailed distributions with $\gamma \ge 1/4$ and even the uniform distribution on $[0, 1]$ do not fulfill these conditions. Moreover, by the definition of α_n^* in (2.32) we cannot get to the edge of the unit interval with a better rate than $O(n^{-1/2})$, c.f. with the rate $(n+1)^{-1}$ in Theorem 2.2.1, and the rate of α_n depends on the heaviness of the tails of F (by condition (F.G.1)).

These constraints was partly overcome by Jurečková (1999). She extended the Theorem 2.2.7 to a broader class of distributions covered by the conditions

(F. J.1) F is absolutely continuous with absolutely continuous, positive and bounded density $f(x)$. Derivatives f', f'' of f are bounded almost everywhere in $x \in \mathbb{R}$.

(F. J.2) f is monotonically decreasing to 0 as $x \to -\infty$ and $x \to +\infty$,

$$\lim_{x\to-\infty} \frac{-a \log F(x)}{\log |x|} = 1, \qquad \lim_{x\to\infty} \frac{-a \log(1 - F(x))}{\log x} = 1, \tag{2.35}$$

for some a (the same in each tail), $0 < a < \infty$.

For some fixed b satisfying $0 < a + \delta \le b \le 2a + \delta$ for some $\delta > 0$, α_n^* given by

$$\alpha_n^* := n^{-1/(1+2b)}, \tag{2.36}$$

and a set of design matrix conditions equivalent with (X.G.1)–(X.G.4). She concluded that under these conditions (2.34) and (2.33) hold uniformly on $[\alpha_n^*, 1-\alpha_n^*]$ for α_n^* given by (2.36), c.f. Theorem 2.1 in Jurečková (1999).

The conditions (F.J.1) and (F.J.2) are more natural than (F.G.1)–(F.G.4) as they are equivalent with von Mises conditions for the distributions from Fréchet domain of attraction, c.f. Theorem 1.2.4. Nevertheless, the conditions proposed by Jurečková (1999) do not admit some heavy-tailed distributions with light tails, as the t-distribution with less than 5 degrees of freedom due to the conditions on covariate matrix (X.G.1)–(X.G.4); Jurečková (1999) recommends to use Theorem 2.2.7 in such a case. What is worse, through the assumptions and the definition of (2.36) follows that heavier tails of the distribution means the more restricted interval $[\alpha_n^*, 1 - \alpha_n^*]$ similarly as in Theorem 2.2.7.

In the following section we shall overcome these boundaries by introducing an improved version of the Bahadur representation approximation of regression quantile process based on the previous results of Gutenbrunner et al. (1993) and Jurečková (1999). This result hold under assumption that von Mises type of condition is fulfilled for $\gamma > -1/2$. Our result not only covers broader class of distribution functions but also extends the boundaries $[\alpha_n^*, 1-\alpha_n^*]$ closer to the edge of the unit interval with the rate of α_n, which does not depend on the heaviness of the tails of F. From the point of assumptions on the distribution functions this is a generalization of the Gutenbrunner et al. (1993) and Jurečková (1999) results and it enables to establish LIL representation more suitable for EVT applications.

2.3 Extending approximations of r.q. process

Consider again the linear model (2.1) with i.i.d. errors $E_1, \ldots, E_n$ with a common distribution function F. We shall assume that F fulfills the following conditions

(F.1) F is absolutely continuous with positive density f on (x_*, x^*). There exists f', the derivative of density f.

(F.2) There exists some $0 < K_\gamma < \infty$ such that

$$\sup_{x_* < x < x^*} F(x)(1 - F(x)) \left| \frac{f'(x)}{f^2(x)} \right| \leq K_\gamma \tag{2.37}$$

(F.3) There exists limits

$$\lim_{x \uparrow x^*} \frac{(1 - F(x))f'(x)}{f^2(x))} = -1 - \gamma^* \tag{2.38}$$

and

$$\lim_{x \downarrow x_*} \frac{F(x)f'(x)}{f^2(x)} = -1 - \gamma_*. \tag{2.39}$$

for some $\gamma_*, \gamma^* \in \mathbb{R}$.

(F.4) It holds that $\gamma := \min\{\gamma_*, \gamma^*\} > -1/2$.

Another set of regularity conditions will be imposed on the design matrix $\mathbf{X}$:

(X.1) $x_{i1} = 1, \quad$ i = 1, . . . , n.

(X.2) $\lim_{n\to\infty} \mathbf{D}_n = \mathbf{D}$, where $\mathbf{D}_n = n^{-1}\mathbf{X}_n^\top \mathbf{X}_n$ and $\mathbf{D}$ is a positive definite $(p \times p)$ matrix.

(X.3) $n^{-1}\sum_{i=1}^n \|\mathbf{x}_i\|^4 = \mathcal{O}(1)$ as $n \to \infty$.

(X.4) $\max_{1\leq i \leq n} \|\mathbf{x}_i\| = \mathcal{O}(n^\Delta)$ as $n \to \infty$ for some $\Delta \leq 1/6$

Assumptions on F coincide with the concept of regular variation in tails, which is tied with the domain of attraction condition introduced in section 1.2. The concept has its meaning also in other fields of statistics. Particularly, the term inside (2.37) plays an important role in non-parametric and robust statistical analysis. Hájek et al. (1999) (among others) introduced the score function J, which is defined as

$$J(t) = -\frac{f'(F^{-1}(\alpha))}{f(F^{-1}(\alpha))} = \frac{\mathrm{d}}{\mathrm{d}\alpha} f(F^{-1}(\alpha)). \tag{2.40}$$

Thus, the condition (F.2) can be rewritten as

$$\sup_{0<\alpha<1} (\alpha(1-\alpha)) \frac{|J(\alpha)|}{f(F^{-1}(\alpha))} \leq K_\gamma. \tag{2.41}$$

Parzen (1979) studied the relation of the score function to the tails of the distribution. He provides an asymptotic theory based on the behaviour of *the density quantile function* $f(F^{-1}(\alpha))$ for $\alpha \to 1$ and establishes the domains of attraction we know from EVT (with a different parametrization) based on $J(\alpha)$.

Actually, the conditions (F.1)–(F.4) imply the domain of attraction condition for some $\gamma > -1/2$ for both tails of F. The properties of the distribution inside the interval (x_*, x^*)

are controlled by relation (2.37), which together with (F.1) implies the finite second derivatives of F. If the density function f of F is bounded, positive and differentiable inside (x_*, x^*), the conditions are (F.1)–(F.4) are fulfilled if F satisfies the domain of attraction conditions for both of its tails with tail indices greater than $-1/2$. Note that the boundary $\gamma > -1/2$ is important also in other applications of EVT, c.f. with the role of normalizing weights in the tail approximations of Drees (1998*b*) for $\gamma \leq -1/2$ and $\gamma > -1/2$ or the properties of ML-estimator in Zhou (2010) as referred on page 21.

If the distribution have a density, then its regularly varying tails are reflected in the properties of the density function $f(x)$ and its derivative $f'(x)$ as is shown by the following Lemma.

Lemma 2.3.1. *Suppose that $F(x)$ have a density $f(x)$ which is positive on some left neighbourhood of x^*. Moreover assume that $F(x)$ fulfills von Mises condition (1.8) for its upper tail with some $\gamma \in \mathbb{R}$. Moreover suppose that also $f'(x)$, the derivative of $f(x)$, exists. Then it holds*

(i) If $x^* = \infty$ then $\lim_{t\to\infty} f(t) = 0$.

(ii) If $\gamma > 0$ (and thus $x^* = \infty$ and $\lim_{t\to\infty} f(t) = 0$) then $\lim_{t\to\infty} f'(t) = 0$

(iii) If $\gamma < 0$ then $x^* < \infty$ and there exists $\lim_{t\to x^*} f(t) = K \in [0, +\infty]$

Proof. (i) Suppose $x^* = \infty$. As $f(t)$ is positive on some left neighbourhood of $+\infty$ due to the finiteness of distribution function and its relation to density as $\int_{-\infty}^{x} f(t)\mathrm{d}t$ it holds $\lim_{t\to\infty} f(t) = 0$.

(ii) As for $\gamma > 0$ holds simplified version of von Mises condition (1.9) also

$$\lim_{t\to\infty} \frac{1-F(t)}{f^2(t)} = \lim_{t\to\infty} \frac{1-F(t)}{tf(t)} \cdot \frac{t}{f(t)} = \gamma \lim_{t\to\infty} \frac{t}{f(t)} = \infty$$

one gets by the general form of von Mises condition (1.8) that $\lim_{t\to\infty} f'(t) = 0$.

(iii) Suppose that $\gamma < 0$. Then by (1.8) holds

$$\lim_{t\to\infty} \frac{1-F(t)}{(x^*-t)f(t)} = -\gamma < \infty.$$

Moreover as $F(t)$ is monotone on some left neighbourhood of x^* it exists $\lim_{t\to x^*} \frac{1-F(t)}{x^*-t} = K \in [0, +\infty]$. If $K = \infty$ then $\lim_{t\to\infty} f(t) = 0$ as $0 < -\gamma < \infty$. If $0 < K < \infty$ then must be $\lim_{t\to\infty} f(t) = -\gamma/K$ and finally if $K = 0$ then $\lim_{t\to\infty} f(t) = \infty$.

$\square$

Similarly as in one sample i.i.d. case (see Csörgő and Révész (1981), pp. 143–155 and Csörgő and Horváth (1993), pp. 369–375) we can hope in an asymptotic theory of regression quantile process only if we can "regulate" the ratio $f(F^{-1}(\alpha))/f(F^{-1}(\alpha+\xi))$ with some suitably small $\xi > 0$. To accomplish this task we will use the following lemma.

Lemma 2.3.2. *Under conditions (F.1) and (F.2) we have*

$$\frac{f(F^{-1}(\alpha_1))}{f(F^{-1}(\alpha_2))} \leq \left(\frac{\alpha_1 \vee \alpha_2}{\alpha_1 \wedge \alpha_2} \cdot \frac{1-(\alpha_1 \wedge \alpha_2)}{1-(\alpha_1 \vee \alpha_2)}\right)^{K_\gamma}, \tag{2.42}$$

for every pair $\alpha_1, \alpha_2 \in (0,1)$.

Proof. Lemma 4.5.2. in Csörgő and Révész (1981). □

Denote

$$\sigma_\alpha := \frac{(\alpha(1-\alpha))^{1/2}}{f(F^{-1}(\alpha))}. \tag{2.43}$$

and set

$$\alpha_n^* := n^{-1+\delta}, \qquad 1 > \delta > 6\Delta. \tag{2.44}$$

Before stating our main result we shall consider a crucial approximation of the function minimized in (2.3) by a quadratic function of $\mathbf{t}$. The approximation holds uniformly in an appropriate neighbourhood of $\boldsymbol{\beta}$ for $\alpha \in [\alpha_n^*, 1-\alpha_n^*]$.

Lemma 2.3.3. *Assume that the distribution function $F(x)$ fulfills conditions (F.1)-(F.4) and for the design matrix $\mathbf{X}$ holds (X.1)–(X.4). Have α_n^* defined as in (2.44). For $t \in \mathbb{R}^p$ and $\alpha \in (0,1)$, denote*

$$\begin{aligned} r_n(\mathbf{t}, \alpha) := (\alpha(1-\alpha))^{-1/2}\sigma_\alpha^{-1}\sum_{i=1}^{n}\left[\rho_\alpha\left(E_{i\alpha} - n^{-1/2}\sigma_\alpha \mathbf{x}_i^\top \mathbf{t}\right) - \rho_\alpha(E_{i\alpha})\right] \\ + n^{-1/2}(\alpha(1-\alpha))^{-1/2}\mathbf{t}^\top \sum_{i=1}^{n}\mathbf{x}_i\psi_\alpha(E_{i\alpha}) - \frac{1}{2}\mathbf{t}^\top \mathbf{D}_n \mathbf{t} \end{aligned} \tag{2.45}$$

with

$$\psi_\alpha(z) := \alpha - I[z<0], \qquad z \in \mathbb{R}^1 \tag{2.46}$$

$$E_{i\alpha} := E_i - F^{-1}(\alpha), \qquad i = 1,\ldots,n, \quad 0 < \alpha < 1. \tag{2.47}$$

Then

$$\sup\left\{|r_n(\mathbf{t},\alpha)| : \alpha_n^* \leq \alpha \leq 1-\alpha_n^*, \|\mathbf{t}\| \leq C_n\right\} = \mathcal{O}_P\left(n^{-\frac{1}{4}\delta+\frac{3}{2}\Delta}C_n^{\frac{3}{2}}(\log(n))^{\frac{1}{2}}\right), \tag{2.48}$$

where $C_n = C(\log\log n)^{1/2}$, $0 < C < \infty$.

Proof. For a fixed $\mathbf{t} \in \mathbb{R}^p$ denote

$$\varepsilon_{i\mathbf{t}\alpha} = \varepsilon_i := n^{-1/2}\sigma_\alpha \mathbf{x}_i^\top \mathbf{t}, \qquad i = 1, \dots, n. \tag{2.49}$$

By (X.4) it holds that $\max_{1\le i\le n} |\varepsilon_i \sigma_\alpha^{-1}| = \mathcal{O}(n^{-\frac{1}{2}+\Delta} C_n)$. Denote, for $i = 1, \dots, n$,

$$Q_i(\mathbf{t}, \alpha) = Q_i := (\alpha(1-\alpha))^{-1/2} \sigma_\alpha^{-1} \left[\rho_\alpha(E_{i\alpha} - \varepsilon_i) - \rho_\alpha(E_{i\alpha}) + \varepsilon_i \psi_\alpha\left(E_{i\alpha}\right)\right] \tag{2.50}$$

with ψ_α as in (2.46) and $E_{i\alpha}$ as in (2.47). Then we obtain by simple arithmetic that

$$\begin{aligned} Q_i &= (\alpha(1-\alpha))^{-1/2} \sigma_\alpha^{-1} \left\{(E_{i\alpha} - \varepsilon_i)(\psi_\alpha(E_{i\alpha} - \varepsilon_i) - \psi_\alpha(E_{i\alpha}))\right\} \\ &= \frac{\sigma_\alpha^{-1}}{(\alpha(1-\alpha))^{1/2}} \left\{(E_{i\alpha} - \varepsilon_i) I[\varepsilon_i < E_{i\alpha} < 0] + (\varepsilon_i - E_{i\alpha}) I[0 < E_{i\alpha} < \varepsilon_i]\right\}. \end{aligned} \tag{2.51}$$

By (2.49), (2.50), and (2.51), for $\varepsilon_i > 0$

$$\begin{aligned} (\alpha(1-\alpha))^{1/2}\sigma_\alpha \mathrm{E} Q_i &= \int_{F^{-1}(\alpha)}^{F^{-1}(\alpha)+\varepsilon_i} \left(\varepsilon_i - x + F^{-1}(\alpha)\right) \mathrm{d}F(x) \\ &= \int_{F^{-1}(\alpha)}^{F^{-1}(\alpha)+\varepsilon_i} \left(F^{-1}(\alpha) + \varepsilon_i - x\right) f\left(F^{-1}(\alpha)\right) \mathrm{d}x \\ &\quad + \int_{F^{-1}(\alpha)}^{F^{-1}(\alpha)+\varepsilon_i} \left(F^{-1}(\alpha) + \varepsilon_i - x\right)\left(f(x) - f\left(F^{-1}(\alpha)\right)\right) \mathrm{d}x. \end{aligned} \tag{2.52}$$

For the first term on the right side of (2.52) it holds that

$$f\left(F^{-1}(\alpha)\right) \int_{F^{-1}(\alpha)}^{F^{-1}(\alpha)+\varepsilon_i} \left(F^{-1}(\alpha) + \varepsilon_i - x\right) \mathrm{d}x = f\left(F^{-1}(\alpha)\right) \frac{\varepsilon_i^2}{2}. \tag{2.53}$$

Using the mean value theorem we can approximate the second term of (2.52) as

$$\begin{aligned} &\left| \int_0^{\varepsilon_i} (\varepsilon_i - u)\left(f\left(F^{-1}(\alpha) + u\right) - f\left(F^{-1}(\alpha)\right)\right) \mathrm{d}u \right| \\ &\le \int_0^{\varepsilon_i} (u\varepsilon_i - u^2) \sup_{F^{-1}(\alpha)\le\theta\le F^{-1}(\alpha)+u} |f'(\theta)| \,\mathrm{d}u \\ &\le \int_0^{\varepsilon_i} (u\varepsilon_i - u^2) \sup_{\alpha\le\xi\le F(F^{-1}(\alpha)+\varepsilon_i)} \left|f'(F^{-1}(\xi))\right| \mathrm{d}u \\ &= \frac{|e_i|^3}{6} \sup_{\alpha\le\xi\le F(F^{-1}(\alpha)+\varepsilon_i)} \left|f'(F^{-1}(\xi))\right|. \end{aligned} \tag{2.54}$$

We shall prove that for some $C_\gamma < \infty$

$$\frac{(\alpha(1-\alpha))}{(f(F^{-1}(\alpha)))^2} \sup_{\alpha\le\xi\le F(F^{-1}(\alpha)+\varepsilon_i)} |f'(F^{-1}(\xi))| \le C_\gamma. \tag{2.55}$$

It holds that

$$
\begin{aligned}
&\frac{(\alpha(1-\alpha))}{(f(F^{-1}(\alpha)))^2}\sup_{\alpha\le\xi\le F(F^{-1}(\alpha)+\varepsilon_i)}|f'(F^{-1}(\xi))| \\
&\quad\le \sup_{\alpha\le\xi\le F(F^{-1}(\alpha)+\varepsilon_i)}\frac{\xi(1-\xi)|f'(F^{-1}(\xi))|}{(f(F^{-1}(\xi)))^2}\times \\
&\qquad\qquad\times\frac{1}{(f(F^{-1}(\alpha)))^2}\sup_{\alpha\le\xi\le F(F^{-1}(\alpha)+\varepsilon_i)}\frac{(f(F^{-1}(\xi)))^2}{\xi(1-\xi)} \\
&\quad= \frac{K_\gamma(\alpha(1-\alpha))}{(f(F^{-1}(\alpha)))^2}\sup_{\alpha\le\xi\le F(F^{-1}(\alpha)+\varepsilon_i)}\frac{(f(F^{-1}(\xi)))^2}{\xi(1-\xi)}. \qquad (2.56)
\end{aligned}
$$

Moreover

$$
\frac{\mathrm{d}}{\mathrm{d}x}\frac{(F(x)(1-F(x))^{1/2}}{f(x)} = (F(x)(1-F(x)))^{-1/2}\left(\frac{1}{2}-F(x)-\frac{F(x)(1-F(x))f'(x)}{f^2(x)}\right).
$$

Thus for $\gamma > -1/2$

$$
\lim_{x\to\infty}\left(\frac{1}{2}-F(x)-\frac{F(x)(1-F(x))f'(x)}{f^2(x)}\right)=\frac{1}{2}+\gamma>0
$$

and there exists some τ^* such that for all $\alpha \in [\tau^*, 1)$ the function $(f(F^{-1}(\cdot)))^2/(\cdot(1-\cdot))$ is decreasing and the supremum in (2.56) is attained at α. Hence on $[\tau^*, 1)$ from (2.56) follows that also (2.55) with $C_\gamma = K_\gamma$.

On the other hand for any $0 < \tau < \tau^*$ and $\alpha \in [\tau, \tau^*)$ we get by Lemma 2.3.2 that

$$
\begin{aligned}
&\sup_{\alpha\le\xi\le F(F^{-1}(\alpha)+\varepsilon_i)}\frac{\alpha(1-\alpha)(f(F^{-1}(\xi)))^2}{\xi(1-\xi)(f(F^{-1}(\alpha)))^2} \\
&\qquad\le \sup_{\alpha\le\xi\le\tau^*}\frac{\alpha(1-\alpha)\cdot(\xi(1-\alpha))^{2K_\gamma}}{\xi(1-\xi)\cdot(\alpha(1-\xi))^{2K_\gamma}} \\
&\qquad\le \frac{\sup_{\alpha\le\xi\le\tau^*}\xi^{2K_\gamma-1}(1-\alpha)^{2K_\gamma+1}}{\inf_{\alpha\le\xi\le\tau^*}\alpha^{2K_\gamma-1}(1-\xi)^{2K_\gamma+1}} \qquad (2.57) \\
&\qquad\le \frac{(\tau^*)^{2K_\gamma-1}(1-\tau)^{2K_\gamma+1}}{\tau^{2K_\gamma-1}(1-\tau^*)^{2K_\gamma+1}} =: K \qquad (2.58)
\end{aligned}
$$

for some $K < \infty$ and (2.55) follows with $C_\gamma = K \cdot K_\gamma$.
Therefore also the relation (2.55) holds for $\gamma > -1/2$ for all α on $[\tau, 1)$ for any $0 < \tau < 1/2$. Similarly we get that (2.55) also for the left tail under assumption that $\gamma_* > -1/2$. Combining (2.54) with (2.55) yields that for some $n \ge n_0$

$$
\left|\mathrm{E}Q_i - \sigma_\alpha^{-2}\frac{\varepsilon_i^2}{2}\right| \le \frac{C_\gamma}{6}n^{-3/2}(\alpha(1-\alpha))^{-1/2}|\mathbf{x}_i^\top\mathbf{t}|^3 \qquad (2.59)
$$

And by (F.4) combined with the definition of α_n^* in (2.44) we arrive to

$$\left|\mathrm{E}Q_i - \sigma_\alpha^{-2}\frac{\varepsilon_i^2}{2}\right| = \mathcal{O}_P(n^{-1+3\Delta-\delta/2}C_n^3). \tag{2.60}$$

Moreover from (2.60) follows

$$\left|\sum_{i=1}^n \left[\mathrm{E}Q_i - \frac{\varepsilon_i^2}{2}\sigma_\alpha^{-2}\right]\right| \leq \frac{1}{n}\sum_{i=1}^n |\mathbf{x}_i^\top \mathbf{t}|^3 \cdot \frac{C_\gamma n^{-1/2}}{6(\alpha(1-\alpha))^{1/2}} \tag{2.61}$$

and finally by (X.3) also

$$\begin{aligned}\left|\sum_{i=1}^n \left[\mathrm{E}Q_i - \frac{\varepsilon_i^2}{2}\sigma_\alpha^{-2}\right]\right| &= \mathcal{O}_P(C_n^3 n^{-1/2}(\alpha_n^*))^{-1/2}) \\ &= \mathcal{O}_P(C_n^3 n^{-\delta/2}) = o_P(1).\end{aligned} \tag{2.62}$$

In the following we shall first prove that

$$P\left\{\left|\sum_{i=1}^n (Q_i - \mathrm{E}Q_i)\right| \geq \zeta B_n\right\} \leq 2n^{-\zeta^2/4} \tag{2.63}$$

for any $\zeta > 0$ and $n \geq n_0$, where

$$B_n = n^{-\frac{1}{4}\delta+\frac{3}{2}\Delta}C_n^{\frac{3}{2}}(\log n)^{1/2} \tag{2.64}$$

Actually, by the third Bernstein inequality (see e.g. (2.13) in Hoeffding (1963)),

$$P\left\{\sum_{i=1}^n (Q_i - \mathrm{E}Q_i)) \geq nt\right\} \leq \exp\left\{-\tau\frac{\lambda}{2\left(1+\frac{1}{3}\lambda\right)}\right\}, \tag{2.65}$$

for $t < b$, provided $Q_i - \mathrm{E}Q_i \leq b$, $i = 1, \ldots, n$, where

$$\tau = \frac{nt}{b}, \qquad \lambda = \frac{bt}{\sigma^2}, \qquad \sigma^2 = \frac{1}{n}\sum_{i=1}^n \mathsf{var}\, Q_i. \tag{2.66}$$

By (2.49) and (2.51), as $n \to \infty$

$$Q_i \leq n^{-\frac{\delta}{2}+\Delta}C_n \tag{2.67}$$

and by (2.60)

$$\begin{aligned}\mathrm{E}Q_i &\leq n^{-1+2\Delta}C_n^2(1+n^{-\frac{\delta}{2}+\Delta}C_n) \\ &= n^{-1+2\Delta}C_n^2(1+o(1)),\end{aligned} \tag{2.68}$$

By (2.51) it is $Q_i - \mathrm{E}Q_i \leq Q_i$, thus one can set

$$b := b_n = n^{-\delta/2+\Delta}C_n \qquad \text{for}\, n \geq n_0 \tag{2.69}$$

By simple aritmetic

$$\mathsf{var}\, Q_i \leq \mathrm{E} Q_i^2 \leq (\alpha(1-\alpha))^{-1/2} \sigma_\alpha^{-1} |\varepsilon_i| \mathrm{E} Q_i$$

thus by (2.49) and (2.68)

$$\sigma^2 \leq n^{-1-\delta/2+3\Delta} C_n^3 (1+o(1)) \leq \frac{3}{2} n^{-1-\delta/2+3\Delta} C_n^3 \qquad \text{for } n \geq n_0 \tag{2.70}$$

Put $nt_n = \zeta B_n$, that is, $t = t_n = \zeta n^{-1-\frac{\delta}{4}+3\Delta/2} C_n^{3/2} (\log n)^{1/2}$. Then $t_n < b_n$ and using $\lambda = b_n t_n / \sigma^2$ one gets by (2.65)

$$\begin{aligned}
P\left\{\sum_{i=1}^n (Q_i - \mathrm{E}Q_i) \geq \zeta B_n\right\} &\leq \exp\left\{-\frac{nt_n^2}{2} \cdot \frac{1}{\sigma^2 + \frac{b_n t_n}{3}}\right\} \\
&\leq \exp\left\{-\frac{\zeta^2 n^{-1-\delta/2+3\Delta} C_n^3 (\log n)}{2(\frac{3}{2} n^{-1+\frac{\delta}{2}+3\Delta} C_n^3 + \frac{\zeta}{3} n^{-1-\frac{3}{4}\delta+\frac{5\Delta}{2}} C_n^{\frac{5}{2}} (\log n)^{\frac{1}{2}}}\right\} \\
&= \exp\left\{\frac{\zeta^2 (\log n)}{3 + \frac{2\zeta}{3} n^{-\frac{\delta}{4}-\frac{\Delta}{2}} C_n^{-\frac{1}{2}} (\log n)^{1/2}}\right\} \\
&\leq \exp\left\{\frac{-\zeta^2 \log n}{4}\right\} = n^{-\zeta^2/4},
\end{aligned} \tag{2.71}$$

for $n \geq n_0$. Because Q_i are non-negative random variables, we obtain analogous inequality for $P(\sum_{i=1}^n (Q_i - \mathrm{E}Q_i) \leq -\zeta B_n) \leq n^{-\zeta^2/4}$ and thus we arrive at (2.63). As

$$\left|\sum_{i=1}^n (Q_i - f(F^{-1}(\alpha)) \frac{\varepsilon_i^2}{2})\right| \leq \left|\sum_{i=1}^n (Q_i - \mathrm{E}Q_i)\right| + \left|\sum_{i=1}^n (\mathrm{E}Q_i - f(F^{-1}(\alpha)) \frac{\varepsilon_i^2}{2}\right| \tag{2.72}$$

and $B_n = n^{-\frac{1}{4}\delta + \frac{3}{2}\Delta} C_n^{\frac{3}{2}} (\log_n)^{1/2} \geq C_n^3 n^{-\delta/2}$, for some $n \geq n_0$, we finally get that regarding (2.49), (2.62), and (2.63) it holds

$$P\left\{|r_n(\mathbf{t}, \alpha)| \geq (\zeta+1) B_n\right\} \leq 2n^{-\zeta^2/4}, \tag{2.73}$$

for $n \geq n_0$, any $\zeta > 0$, and B_n given as (2.64).

Let us now choose intervals $[\alpha_\nu, \alpha_{\nu+1}]$ of length n^{-5} covering $[\alpha_n^*, 1-\alpha_n^*]$ and balls of radius n^{-5} covering $\{t : \|t\| \leq C_n\}$. Let $(\alpha_1, \alpha_2) \subset (\alpha_\nu, \alpha_{\nu+1})$ and let $\mathbf{t}_1$, $\mathbf{t}_2$ lie in the same ball. We shall prove that for some $0 < K < \infty$ and $n \geq n_0$

$$|\sigma_{\alpha_1}/\sigma_{\alpha_2} - 1| \leq K n^{-5} (\alpha_n^*)^{-1} \tag{2.74}$$

Let $\alpha := \alpha_1$. Then, there exists $\xi \in [-1, 1]$ such that $\alpha_2 = \alpha_1 + \xi n^{-5}$. By Lemma 2.3.2 we get

$$\begin{aligned}
\frac{\sigma_{\alpha_1}}{\sigma_{\alpha_2}} - 1 &= \frac{(\alpha_1(1-\alpha_1))^{1/2} f(F^{-1}(\alpha_2))}{(\alpha_2(1-\alpha_2))^{1/2} f(F^{-1}(\alpha_1))} - 1 \\
&\leq \frac{\alpha_1^{K_\gamma - 1/2}(1-\alpha_2)^{K_\gamma+1/2}}{\alpha_2^{K_\gamma-1/2}(1-\alpha_1)^{K_\gamma+1/2}} - 1 \\
&= \frac{\alpha_1^{K_\gamma - 1/2}(1-\alpha_2)^{K_\gamma+1/2} - \alpha_2^{K_\gamma-1/2}(1-\alpha_1)^{K_\gamma+1/2}}{\alpha_2^{K_\gamma-1/2}(1-\alpha_1)^{K_\gamma+1/2}} \qquad (2.75)
\end{aligned}$$

and comparing with n^{-5} by L'Hospital's rule yields

$$\begin{aligned}
&\lim_{n\to\infty} \frac{\alpha^{K_\gamma-1/2}(1-\alpha-\xi n^{-5})^{K_\gamma+1/2} - (\alpha+\xi n^{-5})^{K_\gamma-1/2}(1-\alpha)^{K_\gamma+1/2}}{n^{-5}(\alpha+\xi n^{-5})^{K_\gamma-1/2}(1-\alpha)^{K_\gamma+1/2}} \\
=\ &\lim_{n\to\infty}\Bigg[\frac{-\frac{5\xi}{n^6}(K_\gamma+\frac{1}{2})\alpha^{K_\gamma-\frac{1}{2}}(1-\alpha-\frac{\xi}{n^5})^{K_\gamma-\frac{1}{2}}}{\frac{5}{n^6}(\alpha+\frac{\xi}{n^5})^{K_\gamma-\frac{1}{2}}(1-\alpha)^{K_\gamma+\frac{1}{2}} + \frac{1}{n^{11}}5\xi(K_\gamma-\frac{1}{2})(\alpha+\frac{\xi}{n^5})^{K_\gamma-\frac{3}{2}}(1-\alpha)^{K_\gamma+\frac{1}{2}}} \\
&-\frac{\frac{5\xi}{n^6}(K_\gamma-\frac{1}{2})(\alpha+\frac{\xi}{n^5})^{K_\gamma-\frac{3}{2}}(1-\alpha)^{K_\gamma+\frac{1}{2}}}{\frac{5}{n^6}(\alpha+\frac{\xi}{n^5})^{K_\gamma-\frac{1}{2}}(1-\alpha)^{K_\gamma+\frac{1}{2}} + \frac{1}{n^{11}}5\xi(K_\gamma-\frac{1}{2})(\alpha+\frac{\xi}{n^5})^{K_\gamma-\frac{3}{2}}(1-\alpha)^{K_\gamma+\frac{1}{2}}}\Bigg] \\
=\ &\frac{\xi((K_\gamma-1/2)\alpha^{K_\gamma-3/2}(1-\alpha)^{K_\gamma+1/2} - (K_\gamma+1/2)(\alpha(1-\alpha))^{K_\gamma-1/2})}{\alpha^{K_\gamma-1/2}(1-\alpha)^{K_\gamma+1/2}} \\
=\ &\xi((K_\gamma-1/2)\alpha^{-1} + (K_\gamma+1/2)(1-\alpha)^{-1}). \qquad (2.76)
\end{aligned}$$

Hence it holds that

$$\begin{aligned}
\left|\frac{\sigma_{\alpha_1}}{\sigma_{\alpha_2}} - 1\right| &\leq n^{-5} \sup_{\alpha_n^* \leq \alpha \leq 1-\alpha_n*} \left|\xi((K_\gamma-1/2)\alpha^{-1} + (K_\gamma+1/2)(1-\alpha)^{-1})\right| \\
&= \mathcal{O}\left(n^{-5}(\alpha_n^*)^{-1}\right)
\end{aligned}$$

thus (2.74) holds.

Moreover, for $(\alpha_1, \alpha_2) \subset (\alpha_\nu, \alpha_{\nu+1})$ also holds

$$f(F^{-1}(\alpha_1))|F^{-1}(\alpha_1) - F^{-1}(\alpha_2)| \leq n^{-5} \qquad (2.77)$$

as by mean value theorem and again Lemma 4.5.2. in Csörgő and Révész (1981) there exist some $\eta > 0$ such that

$$\begin{aligned}
|F^{-1}(\alpha_1) - F^{-1}(\alpha_2)| &\leq |\alpha_1 - \alpha_2| \sup_{\xi\in(\alpha_1,\alpha_2)} \frac{f(F^{-1}(\alpha_1))}{f(F^{-1}(\xi))} \\
&\leq n^{-5} \sup_{\xi\in(\alpha_1,\alpha_2)} \left[\frac{(\xi\vee\alpha_1)(1-\xi\wedge\alpha_1)}{(\xi\wedge\alpha_1)(1-\xi\vee\alpha)}\right]^{K_\gamma} \\
&\leq n^{-5}\left[\frac{\sup_{\xi\in(\alpha_1,\alpha_2)}(\xi\vee\alpha_1)(1-\xi\wedge\alpha_1)}{\inf_{\xi\in(\alpha_1,\alpha_2)}(\xi\wedge\alpha_1)(1-\xi\vee\alpha)}\right]^{K_\gamma} \\
&= n^{-5}\left[\frac{\alpha_2(1-\alpha_1)}{\alpha_1(1-\alpha_2)}\right]^{K_\gamma} \\
&\leq (1+\eta)n^{-5}.
\end{aligned}$$

For fixed i, $1 \leq i \leq n$, write

$$|Q_i(\mathbf{t}_2,\alpha_2) - Q_i(\mathbf{t}_1,\alpha_1)| \leq |Q_i(\mathbf{t}_2,\alpha_2) - Q_i(\mathbf{t}_1,\alpha_2)| + |Q_i(\mathbf{t}_1,\alpha_2) - Q_i(\mathbf{t}_1,\alpha_1)| \quad (2.78)$$

and consider the terms on the right-hand side separetely. By (2.78), (2.49),

$$\begin{aligned} |Q_i(\mathbf{t}_2,\alpha_2) - Q_i(\mathbf{t}_1,\alpha_2)| &\leq (\alpha_2(1-\alpha_2))^{-1/2}\sigma_{\alpha_2}^{-1}|\varepsilon_{i\alpha_2\mathbf{t}_2} - \varepsilon_{i\alpha_2\mathbf{t}_1}| \\ &\leq n^{-1/2}(\alpha_2(1-\alpha_2))^{-1/2}|\mathbf{x}_i^\top(\mathbf{t}_2 - \mathbf{t}_1)| \\ &= \mathcal{O}(n^{-5-\delta/2+\Delta}) \qquad (2.79) \end{aligned}$$

For the corresponding centring term $\sigma_\alpha^{-2}\frac{\varepsilon_{i\alpha}^2}{2}$ we obtain the bound

$$\begin{aligned} \left|\frac{1}{2}\left(\varepsilon_{i\alpha_2\mathbf{t}_2}^2 - \varepsilon_{i\alpha_2\mathbf{t}_1}^2\right)\sigma_{\alpha_2}^{-2}\right| &= \left|\frac{1}{2}n^{-1}\left((\mathbf{x}_i^\top\mathbf{t}_2)^2 - (\mathbf{x}_i^\top\mathbf{t}_1)^2\right)\right| \\ &= \left|\frac{1}{2}n^{-1}(\mathbf{x}_i^\top(\mathbf{t}_1 - \mathbf{t}_2))\cdot(\mathbf{x}_i^\top(\mathbf{t}_1 + \mathbf{t}_2))\right| \\ &\leq \left|n^{-6+\Delta}n^\Delta C_n\right| = \mathcal{O}(n^{-6+2\Delta}C_n) \qquad (2.80) \end{aligned}$$

Consider the second term on the right-hand side of (2.78), which we denote Q^* for the sake of brevity. We should distinguish two case:

1. If $\varepsilon_{i\alpha_2\mathbf{t}_1} < E_{i\alpha_2} < 0$ and $\varepsilon_{i\alpha_1\mathbf{t}_1} < E_{i\alpha_1} < 0$ (or $0 < E_{i\alpha_2} < \varepsilon_{i\alpha_2\mathbf{t}_1}$ and $0 < E_{i\alpha_1} < \varepsilon_{i\alpha_1\mathbf{t}_1}$)

$$\begin{aligned} |Q^*| &= |(\alpha_2(1-\alpha_2))^{-1/2}\sigma_{\alpha_2}^{-1}E_{i\alpha_2} - (\alpha_1(1-\alpha_1))^{-1/2}\sigma_{\alpha_1}^{-1}E_{i\alpha_1})| \\ &\leq (\alpha_n^*)^{-1/2}(|(\sigma_{\alpha_2}^{-1} - \sigma_{\alpha_1}^{-1})E_{i\alpha_2}| + |\sigma_{\alpha_1}^{-1}(E_{i\alpha_1} - E_{i\alpha_2})|) \\ &\leq (\alpha_n^*)^{-1/2}(|(\sigma_{\alpha_2}^{-1} - \sigma_{\alpha_1}^{-1})|\cdot|\varepsilon_{i\alpha_2\mathbf{t}_1}| + |\sigma_{\alpha_1}^{-1}|\cdot|F^{-1}(\alpha_1) - F^{-1}(\alpha_2)|) \\ &\leq (\alpha_n^*)^{-1/2}(2n^{-1/2}|1 - (\sigma_{\alpha_2}/\sigma_{\alpha_1})|\cdot|\mathbf{x}_i^\top\mathbf{t}_1| + (\alpha_1(1-\alpha_1)^{-1/2}n^{-5}) \\ &= \mathcal{O}(C_n n^{-5.5+\Delta}(\alpha_n^*)^{-3/2}) = o(n^{-5.2}(\alpha_n^*)^{-3/2}) \qquad (2.81) \end{aligned}$$

   by (2.74), (2.77), and (X.3).

2. If $\varepsilon_{i\alpha_1\mathbf{t}_1} < E_{i\alpha_1} < 0$ and $0 < E_{i\alpha_2} < \varepsilon_{i\alpha_2\mathbf{t}_1}$ (or $\varepsilon_{i\alpha_2\mathbf{t}_1} < E_{i\alpha_2} < 0$ and $0 < E_{i\alpha_1\mathbf{t}_1} < \varepsilon_{i\alpha_1\mathbf{t}_1}$)

$$\begin{aligned} |Q^*| &\leq \left|\frac{\sigma_{\alpha_2}^{-1}(E_{i\alpha_2} - \varepsilon_{i\alpha_2\mathbf{t}_1})}{(\alpha_2(1-\alpha_2))^{1/2}}\right| + \left|\frac{\sigma_{\alpha_1}^{-1}(\varepsilon_{i\alpha_1\mathbf{t}_1} - E_{i\alpha_1})}{(\alpha_1(1-\alpha_1))^{1/2}}\right| \\ &\leq (\alpha_n^*)^{-1/2}\left(\sigma_{\alpha_2}^{-1}|\varepsilon_{i\alpha_1\mathbf{t}_1} - \varepsilon_{i\alpha_2\mathbf{t}_1}| + \sigma_{\alpha_1}^{-1}|\varepsilon_{i\alpha_1\mathbf{t}_1} - \varepsilon_{i\alpha_2\mathbf{t}_1}|\right) \\ &= (\alpha_n^*)^{-1/2}\left(n^{-1/2}|\mathbf{x}_i^\top\mathbf{t}_1|\cdot(|1 - \sigma_{\alpha_1}/\sigma_{\alpha_2}| + |1 - \sigma_{\alpha_2}/\sigma_{\alpha_1}|)\right) \\ &= \mathcal{O}(C_n n^{-4.5+\Delta}(\alpha_n^*)^{-1/2}) = o(n^{-4.2}(\alpha_n^*)^{-1/2}) \qquad (2.82) \end{aligned}$$

Finally notice that as the both centring terms are independent on α it holds in both cases

$$\frac{1}{2}|\sigma_{\alpha_2}^{-2}\varepsilon_{i\alpha_2\mathbf{t}_1}^2 - \sigma_{\alpha_1}^{-2}\varepsilon_{i\alpha_1\mathbf{t}_1}^2| = 0. \tag{2.83}$$

Let us fix one set S_ν in the decomposition of the set $[\alpha_n^*, 1-\alpha_n^*] \times \{\mathbf{t}, \|\mathbf{t}\| \le C_n\}$; the number of such sets is at most $(2C_n)^p n^{5(p+1)}$. As $\alpha_n^* \ge n^{-4}$, it follows from (2.78), (2.79), (2.80) (2.81), (2.82), and (2.83) that

$$\sup_{S_\nu} |r_n(\mathbf{t}_2, \alpha_2) - r_n(\mathbf{t}_1, \alpha_1)| \le K_1 n^{-1}, \tag{2.84}$$

where $0 < K_1 < \infty$. As $n^{-1} < B_n$ for some $n \ge n_0$ by (2.73) it holds

$$P\left\{\sup_{S_\nu} |r_n(\mathbf{t}, \alpha)| \ge (\zeta+2)B_n\right\} \le 2n^{-\zeta^2/4} \tag{2.85}$$

and finally

$$\begin{aligned} P\left\{\sup_{\|\mathbf{t}\|\le C_n, \alpha_n^*\le\alpha\le 1-\alpha_n^*} |r_n(\mathbf{t}, \alpha)| \ge (\zeta+2)\right\} &\le \sum_\nu P\left\{\sup_{S_\nu} |r_n(\mathbf{t}, \alpha)| \ge (\zeta+2)B_n\right\} \\ &\le 2^{p+1}C_n^p n^{5(p+1)} n^{-\zeta^2/4} = o_P(1) \end{aligned}$$

for $\zeta^2 > 20(p+1)$; and this entails

$$\begin{aligned} \sup\{|r_n(\mathbf{t}, \alpha) : \|\mathbf{t}\| \le C_n, \alpha_n^* \le \alpha \le 1-\alpha_n^*\} &= \mathcal{O}_P(B_n) \\ &= \mathcal{O}_P(n^{-\frac{1}{4}\delta+\frac{3}{2}\Delta} C_n^{\frac{3}{2}} (\log_n)^{\frac{1}{2}}) \end{aligned}$$

as $n \to \infty$.

□

Theorem 2.3.1. *Assume that the distribution function $F(x)$ of errors in model (2.1) fulfills conditions (F.1)–(F.4) and for the design matrix $\mathbf{X}$ it holds (X.1)–(X.4). Then for α_n^* as in (2.44) holds*

$$\begin{aligned} \sup_{\alpha_n^*\le\alpha\le 1-\alpha_n^*} \left\| n^{1/2} \frac{f(F^{-1}(\alpha))}{(\alpha(1-\alpha))^{1/2}} (\widehat{\boldsymbol\beta}_n\,(\alpha|\,\mathbf{Y}, \mathbf{x}) - \boldsymbol\beta(\alpha)) \right\| &= \\ \sup_{\alpha_n^*\le\alpha\le 1-\alpha_n^*} \left\| \frac{\mathbf{q}_n(\alpha)}{(\alpha(1-\alpha))^{1/2}} \right\| &= \mathcal{O}_P(\log\log n) \end{aligned} \tag{2.86}$$

and

$$\begin{aligned} &\frac{n^{1/2} f(F^{-1}(\alpha))}{(\alpha(1-\alpha))^{1/2}} \left(\widehat{\boldsymbol\beta}_n\,(\alpha|\,\mathbf{Y}, \mathbf{x}) - \boldsymbol\beta(\alpha)\right) \\ &\qquad = n^{-1/2}(\alpha(1-\alpha))^{-1/2} \mathbf{D}_n^{-1} \sum_{i=1}^n \mathbf{x}_{ni} \psi_\alpha(E_{i\alpha}) + o_P(1) \end{aligned} \tag{2.87}$$

uniformly in $\alpha_n^* \le \alpha \le 1-\alpha_n^*$, *where* $\boldsymbol{\beta}(\alpha) = (\beta_1 + F^{-1}(\alpha), \beta_2, \dots, \beta_p)^\top, E_{i\alpha} = E_i - F^{-1}(\alpha)$ *and* $C_n = C(\log\log n)^{1/2},\ 0 < C < \infty$.

Proof. If $\boldsymbol{\beta}_n(\alpha)$ minimizes (2.3) then

$$\mathbf{T}_{n\alpha} = n^{1/2}\sigma_\alpha^{-1}(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)) \tag{2.88}$$

minimizes the convex function

$$G_{n\alpha}(t) = (\alpha(1-\alpha))^{-1/2}\sigma_\alpha^{-1}\sum_{i=1}^{n}\left[\rho_\alpha(E_{i\alpha} - n^{-1/2}\sigma_\alpha\mathbf{x}_i^\top\mathbf{t}) - \rho_\alpha(E_{i\alpha})\right] \tag{2.89}$$

with respect to $\mathbf{t} \in \mathbb{R}^d$. By Lemma 2.3.3 for any fixed $C > 0$

$$\min_{\|\mathbf{t}\|<C} G_{n\alpha}(\mathbf{t}) = \min_{\|\mathbf{t}\|<C}\left\{-\mathbf{t}^\top\mathbf{Z}_{n\alpha} + \frac{1}{2}\mathbf{t}^\top\mathbf{D}_n\mathbf{t}\right\} + o_p(1) \tag{2.90}$$

uniformly in $\alpha_n^* \le \alpha \le 1-\alpha_n^*$, where

$$\mathbf{Z}_{n\alpha} = n^{-1/2}(\alpha(1-\alpha))^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\psi_\alpha(E_{i\alpha}). \tag{2.91}$$

We would like to examine the second minimization problem in (2.90). Denote its solution as

$$\mathbf{U}_{n\alpha} := \arg\min_{\mathbf{t}\in\mathbb{R}^p}\left\{-\mathbf{t}^\top\mathbf{Z}_{n\alpha} + \frac{1}{2}\mathbf{t}^\top\mathbf{D}_n\mathbf{t}\right\}. \tag{2.92}$$

Have $\mathbf{t} = \mathbf{u} + \mathbf{D}_n^{-1}\mathbf{Z}_{n\alpha}$ for any $\mathbf{u} \in \mathbb{R}^p$, then

$$\begin{aligned} -\mathbf{t}^\top\mathbf{Z}_{n\alpha} + \frac{1}{2}\mathbf{t}^\top\mathbf{D}_n\mathbf{t} &= -(\mathbf{u} + \mathbf{D}_n^{-1}\mathbf{Z}_{n\alpha})^\top\mathbf{Z}_{n\alpha} + \\ &\qquad \frac{1}{2}(\mathbf{u} + \mathbf{D}_n^{-1}\mathbf{Z}_{n\alpha})^\top\mathbf{D}_n(\mathbf{u} + \mathbf{D}_n^{-1}\mathbf{Z}_{n\alpha}) \\ &= -\frac{1}{2}\mathbf{Z}_{n\alpha}^\top\mathbf{D}_n^{-1}\mathbf{Z}_{n\alpha} + \frac{1}{2}\mathbf{u}^\top\mathbf{D}_n\mathbf{u}. \end{aligned} \tag{2.93}$$

$\mathbf{D}_n$ is positive definite matrix and hence $\frac{1}{2}\mathbf{u}^\top\mathbf{D}_n\mathbf{u} > 0$. It follows that $\mathbf{U}_{n\alpha} = \mathbf{D}_n^{-1}\mathbf{Z}_{n\alpha}$ and

$$\min_{t\in\mathbb{R}^p}\left\{-\mathbf{t}^\top\mathbf{Z}_{n\alpha} + \frac{1}{2}\mathbf{t}^\top\mathbf{D}_n\mathbf{t}\right\} = \frac{1}{2}\mathbf{Z}_{n\alpha}^\top\mathbf{D}_n^{-1}\mathbf{Z}_{n\alpha}. \tag{2.94}$$

It will be necessary to provide a probabilistic bound for $\sup\{\mathbf{U}_{n\alpha}, \alpha_n^* \le \alpha \le 1-\alpha_n^*\}$. As $\mathbf{D}_n \to \mathbf{D}$, this can be accomplished by a bound for $B := \sup\{\mathbf{Z}_{n\alpha}, \alpha_n^* \le \alpha \le 1-\alpha_n^*\}$. Rewriting (2.91) yields

$$\mathbf{Z}_{n\alpha} = \frac{n^{-1/2}}{(\alpha(1-\alpha))^{1/2}}\sum_{i=1}^{n}\mathbf{x}_i\left((1-\alpha)(I_{[F(E_i)\le\alpha]} - \alpha) + \alpha(I_{[F(E_i)\le 1-\alpha]} - (1-\alpha))\right),$$

and the invariance theorem of Shorack (1991) can be applied on process $\mathbf{Z}_{n\alpha}$. By conditions (X.3) and (X.4) for any $\alpha_n^* > n^{-1}$, equation (1.10) or (1.11) of Shorack (1991) implies that

$$B \leq \mathcal{O}_p(1) + c \sup_{\alpha_n^* \leq \alpha \leq 1-\alpha_n^*} (\alpha(1-\alpha))^{-1/2} W(s) \tag{2.95}$$

for some constant c where $W(s)$ is a Brownian Bridge. The supremum in (2.95) is bounded by $c(\log\log n)^{1/2} + \mathcal{O}_P(1)$ using the law of iterated logarithm (see, for example Shorack and Wellner (1986), pp. 599). Thus $\mathbf{Z}_{n\alpha} = \mathcal{O}_P((\log\log n)^{1/2})$ uniformly on $\alpha_n^* \leq \alpha \leq 1-\alpha_n^*$ and consequently $\mathbf{U}_{n\alpha} = \mathcal{O}_P\left((\log\log n)^{1/2}\right)$.

Using (2.94) we can rewrite the minimization (2.92) as

$$-\mathbf{t}^\top \mathbf{Z}_{n\alpha} + \frac{1}{2}\mathbf{t}^\top \mathbf{D}_n \mathbf{t} = \frac{1}{2}\left[(\mathbf{t}-\mathbf{U}_{n\alpha})^\top \mathbf{D}_n (\mathbf{t}-\mathbf{U}_{n\alpha}) - \mathbf{U}_{n\alpha}^\top \mathbf{D}_n \mathbf{U}_{n\alpha}\right] \tag{2.96}$$

and hence by rewriting (2.45) in the same fashion we come to

$$\begin{aligned}&\sup_{(\alpha,\mathbf{t})\in S} |r_n(\mathbf{t},\alpha)| = \\ &\qquad \sup_{(\alpha,\mathbf{t})\in S} \left| G_{n\alpha}(\mathbf{t}) - \frac{1}{2}\left[(\mathbf{t}-\mathbf{U}_{n\alpha})^\top \mathbf{D}_n(\mathbf{t}-\mathbf{U}_{n\alpha}) - \mathbf{U}_{n\alpha}^\top \mathbf{D}_n \mathbf{U}_{n\alpha}\right]\right| = o_P(1)\end{aligned} \tag{2.97}$$

for any set S of couples $(\alpha, \mathbf{t})$ fulfilling the assumptions of Lemma 2.3.3. Finally as we have $\mathbf{U}_{n\alpha} = \mathcal{O}_P\left((\log\log n)^{1/2}\right)$, inserting $\mathbf{t} = \mathbf{U}_{n\alpha}$ in (2.97) we further obtain

$$\sup_{\alpha_n^* \leq \alpha \leq 1-\alpha_n^*} \left\{ \left| G_{n\alpha}(\mathbf{U}_{n\alpha}) + \frac{1}{2}\mathbf{U}_{n\alpha}^\top \mathbf{D}_n \mathbf{U}_{n\alpha} \right| \right\} = o_P(1). \tag{2.98}$$

Using the previous, we shall show that

$$\sup_{\alpha_n^* \leq \alpha \leq 1-\alpha_n^*} \{|\mathbf{T}_{n\alpha} - \mathbf{U}_{n\alpha}|\} = o_P(1). \tag{2.99}$$

Have any $\varrho > 0$. Consider the ball $\mathbf{B}_{n\alpha}$ with center $\mathbf{U}_{n\alpha}$ and radius ϱ. Then, for $\mathbf{t} \in \mathbf{B}_{n\alpha}$,

$$\|\mathbf{t}\| \leq \|\mathbf{t}-\mathbf{U}_{n\alpha}\| + \|\mathbf{U}_{n\alpha}\| \leq \varrho + K_1(\log\log n)^{1/2}$$

for some K_1 with probability exceeding $1-\varepsilon$ for $n \geq n_0$. Hence, again by Lemma 2.3.3,

$$\Delta_{n\alpha} := \sup_{\alpha_n^* \leq \alpha \leq 1-\alpha_n^*} \sup_{t \in \mathbf{B}_{n\alpha}} |r_n(\mathbf{t},\alpha)| \xrightarrow[n\to\infty]{\mathrm{P}} 0 \tag{2.100}$$

Accordingly by (2.97) and (2.98) and by the fact that both terms on left side of (2.97) and (2.98) are positive (this is due to the fact that $\mathbf{D}_n$ is positive definite matrix).

$$\begin{aligned}\frac{1}{2}\left[(\mathbf{t}-\mathbf{U}_{n\alpha})^\top \mathbf{D}_n(\mathbf{t}-\mathbf{U}_{n\alpha}) - \mathbf{U}_{n\alpha}^\top \mathbf{D}_n \mathbf{U}_{n\alpha}\right] &\leq G_{n\alpha}(\mathbf{t}) + \Delta_{n\alpha}, &(2.101)\\ -\frac{1}{2}\mathbf{U}_{n\alpha}^\top \mathbf{D}_n \mathbf{U}_{n\alpha} &\geq G_{n\alpha}(\mathbf{U}_{n\alpha}) - \Delta_{n\alpha} &(2.102)\end{aligned}$$

Consider the behaviour of $G_{n\alpha}$ outside $\mathbf{B}_{n\alpha}$. Suppose $\mathbf{t}_\alpha = \mathbf{U}_{n\alpha} + k\xi$, $k > \varrho$ and $\|\xi\| = 1$. Let $\mathbf{t}^*_\alpha$ be the boundary point of $\mathbf{B}_{n\alpha}$ that lies on the line from $\mathbf{U}_{n\alpha}$ to $\mathbf{t}_\alpha$, i.e., $\mathbf{t}^*_\alpha = \mathbf{U}_{n\alpha} + \varrho\xi$. Then as it holds $\xi = \mathbf{t}_\alpha/k - \mathbf{U}_{n\alpha}/k$, it follows $\mathbf{t}^*_\alpha = (1-(\varrho/k))\mathbf{U}_{n\alpha} + (\varrho/k)\mathbf{t}_\alpha$ and hence, by the convexity of $G_{n\alpha}(\cdot)$

$$\varrho/kG_{n\alpha}(\mathbf{t}) + (1-\varrho/k)G_{n\alpha}(\mathbf{U}_{n\alpha}) \geq G_{n\alpha}(\mathbf{t}^*_\alpha). \tag{2.103}$$

Moreover using (2.101) and (2.102) we get that

$$\begin{aligned} G_{n\alpha}(\mathbf{t}^*_\alpha) &\geq \frac{1}{2}\left[(\mathbf{t}^*_\alpha - \mathbf{U}_{n\alpha})^\top \mathbf{D}_n(\mathbf{t}^*_\alpha - \mathbf{U}_{n\alpha}) - \mathbf{U}_{n\alpha}\mathbf{D}_n\mathbf{U}_{n\alpha}\right] - \Delta_{n\alpha} \\ &\geq \frac{1}{2}(\mathbf{t}^*_\alpha - \mathbf{U}_{n\alpha})^\top \mathbf{D}_n(\mathbf{t}^*_\alpha - \mathbf{U}_{n\alpha}) + G_{n\alpha}(\mathbf{U}_{n\alpha}) - 2\Delta_{n\alpha} \\ &= \frac{1}{2}\varrho^2\lambda_0 + G_{n\alpha}(\mathbf{U}_{n\alpha}) - 2\Delta_{n\alpha}. \end{aligned} \tag{2.104}$$

where $\lambda_0 > 0$ is the minimal eigenvalue of $\mathbf{D}$. Hence

$$\inf_{\|\mathbf{t}-\mathbf{U}_{n\alpha}\|>\varrho} G_{n\alpha}(\mathbf{t}) \geq G_{n\alpha}(\mathbf{U}_{n\alpha}) + (k/\varrho)(\frac{1}{2}\varrho^2\lambda_0 - 2\Delta_{n\alpha}). \tag{2.105}$$

As $1/2\varrho^2\lambda_0$ is positive and $\Delta_{n\alpha} \xrightarrow[n\to\infty]{\mathrm{P}} 0$, the last term in (2.105) is positive with the probability tending to one uniformly in $\alpha^*_n \leq \alpha \leq 1-\alpha^*_n$ for any fixed $\varrho > 0$. Therefore we have shown, that given any $\varrho > 0$ and $\varepsilon > 0$, there exists n_0 and $\eta > 0$ such that for $n \geq n_0$,

$$P\left(\inf_{\alpha^*_n \leq \alpha \leq 1-\alpha^*_n}\left[\inf_{\|\mathbf{t}-\mathbf{U}_{n\alpha}\|\geq\delta} G_{n\alpha}(\mathbf{t}) - G_{n\alpha}(\mathbf{U}_{n\alpha})\right] > \eta\right) > 1-\varepsilon \tag{2.106}$$

As $G_{n\alpha}(\mathbf{t})$ is positive function for $\mathbf{t} \in \mathbb{R}^p$ due to (2.106) $G_{n\alpha}$ must be minimized inside the ball of radius $\varrho > 0$ with a probability tending to one. Thus for any fixed $\varrho > 0$ $P\left(\sup_{\alpha^*_n \leq \alpha \leq 1-\alpha^*_n}\|\mathbf{T}_{n\alpha} - \mathbf{U}_{n\alpha}\| \leq \varrho\right) \to 1$ as $n \to \infty$, which is equivalent to (2.87). Finally, equation (2.86) follows from invariance theorem of Shorack (1991) as noted in lines after (2.91).

□

As we see in the proof of (2.3.3), the “censoring” parameter δ depends on the maximum of the components of the covariance matrix $\mathbf{X}$. If we assume somewhat stronger conditions instead of (X.4), we can obtain wider bounds. The previous statements holds even if the assumtions holds only almost surely. Suppose the following assumptions

(X.5) $n^{-1}\sum_{i=1}^n \|\mathbf{x}_{ni}\|^4 \overset{\text{a.s.}}{=} \mathcal{O}(1)$ as $n \to \infty$

(X.6) $\max_{1\leq i\leq n}\|\mathbf{x}_{ni}\| \overset{\text{a.s.}}{=} \mathcal{O}(C_n^\Delta)$ as $n \to \infty$ for some $\Delta > 0$.

and in accordance with (X.6) set

$$\alpha^*_n := \frac{(\log n)^{2+\delta}}{n}, \qquad \delta > 0 \tag{2.107}$$

Lemma 2.3.4. *Assume that the distribution function $F(x)$ of the errors in model (2.1) fulfills the conditions (F.1)–(F.4) and for the design matrix $\mathbf{X}$ it holds (X.1)–(X.2) and (X.5)–(X.6). Then for α_n^* as in (2.107), $t \in \mathbb{R}^p$ and $\alpha \in (0,1)$, and $r_n(\mathbf{t}, \alpha)$ as in (2.45) it holds*

$$\sup\left\{|r_n(\mathbf{t}, \alpha)| : \alpha_n^* \le \alpha \le 1 - \alpha_n^*, \|\mathbf{t}\| \le C_n\right\} = \mathcal{O}_P\left((\log n)^{1-\delta/4} C_n^{\frac{3}{2}}\right). \tag{2.108}$$

Proof. Have again ε_i and $Q_i(\mathbf{t}, \alpha)$ as in (2.49) and (2.50). As in the proof of Lemma 2.3.3 we get

$$\left|\mathrm{E}Q_i - \sigma_\alpha^{-2}\frac{\varepsilon_i^2}{2}\right| \le \frac{C_\gamma}{6} n^{-3/2} \alpha(1-\alpha))^{-1/2} |\mathbf{x}_i^\top \mathbf{t}|^3 \tag{2.109}$$

and thus also

$$\left|\sum_{i=1}^n \left[\mathrm{E}Q_i - \frac{\varepsilon_i^2}{2}\sigma_\alpha^{-2}\right]\right| \le \frac{1}{n}\sum_{i=1}^n |\mathbf{x}_i^\top \mathbf{t}|^3 \cdot \frac{C_\gamma n^{-1/2}}{6(\alpha(1-\alpha))^{1/2}}. \tag{2.110}$$

From (X.6) and (2.107) it follows

$$\left|\mathrm{E}Q_i - \sigma_\alpha^{-2}\frac{\varepsilon_i^2}{2}\right| = \mathcal{O}_P(n^{-1}(\log n)^{-1-\frac{\delta}{2}} C_n^{3+3\Delta}) \tag{2.111}$$

and from (X.5) and (2.107)

$$\left|\sum_{i=1}^n \left[\mathrm{E}Q_i - \frac{\varepsilon_i^2}{2}\sigma_\alpha^{-2}\right]\right| = \mathcal{O}_P(C_n^3(\log n)^{-1-\frac{\delta}{2}}) = o_P(1) \tag{2.112}$$

Similarly as in the proof of Lemma 2.3.3 we will use Bernstein inequality to prove

$$P\left\{\left|\sum_{i=1}^n (Q_i - \mathrm{E}Q_i)\right| \ge \zeta B_n\right\} \le 2n^{-\zeta^2/4} \tag{2.113}$$

for any $\zeta > 0$, and $n \ge n_0$, where

$$B_n = (\log n)^{-\delta/4} C_n^{3/2+3\Delta/2} = o(1). \tag{2.114}$$

By (2.49) and (2.50) and (X.6)

$$Q_i - \mathrm{E}Q_i \le Q_i \le C_n^{1+\Delta}(\log n)^{-1-\delta/2} =: b_n \tag{2.115}$$

and by (2.111) it holds almost surely

$$\begin{aligned}
\mathrm{E}Q_i &\leq n^{-1}(\log n)^{-1-\delta/2}C_n^{3+3\Delta} + n^{-1}C_n^{2+2\Delta} \\
&= n^{-1}C_n^{2+2\Delta}\left(1 + (\log n)^{-1-\frac{\delta}{2}}C_n^{1+\Delta}\right) \\
&= n^{-1}C_n^{2+2\Delta}\left(1 + o(1)\right). \qquad (2.116)
\end{aligned}$$

Hence as

$$\mathsf{var}\, Q_i \leq (\alpha(1-\alpha))^{-1/2}\sigma_\alpha^{-1}|\varepsilon_i|\mathrm{E}Q_i$$

we get by (2.116) that for σ^2 defined in (2.66) and for $n \geq n_0$ holds

$$\sigma^2 \leq n^{-1}(\log n)^{-1-\delta/2}C_n^{3+3\Delta}(1+o(1)) \leq \frac{3}{2}n^{-1}(\log n)^{-1-\delta/2}C_n^{3+3\Delta}. \qquad (2.117)$$

Put $nt_n = \zeta B_n$, that is $t = t_n = \zeta n^{-1}(\log n)^{-\delta/4}C_n^{3/2+3\Delta/2}$. Then $t_n < b_n$ and it yields for $n \geq n_0$ by (2.65)

$$\begin{aligned}
P\left\{\sum_{i=1}^n (Q_i - \mathrm{E}Q_i) \geq \zeta B_n\right\} &\leq \exp\left\{-\frac{nt_n^2}{2}\cdot\frac{1}{\sigma^2 + \frac{b_n t_n}{3}}\right\} \\
&\leq \exp\left\{-\frac{\zeta^2 n^{-1}(\log n)^{-1-\frac{\delta}{2}}C_n^{3+3\Delta}(\log n)}{2(\frac{3}{2}n^{-1}(\log n)^{-1-\frac{\delta}{2}}C_n^{3+3\Delta} + \frac{\zeta}{3}n^{-1}(\log n)^{-1-\frac{3}{4}\delta}C_n^{\frac{5}{2}+\frac{5}{2}\Delta}}\right\} \\
&= \exp\left\{\frac{\zeta^2(\log n)}{3 + \frac{2\zeta}{3}C_n^{-\frac{1}{2}-\frac{1}{2}\Delta}(\log n)^{-\frac{\delta}{4}}}\right\} \\
&\leq \exp\left\{\frac{-\zeta^2\log n}{4}\right\} = n^{-\zeta^2/4}. \qquad (2.118)
\end{aligned}$$

Again as Q_i are non-negative random variables, we obtain analogous inequality for $P(\sum_{i=1}^n (Q_i - \mathrm{E}Q_i) \leq -\zeta B_n) \leq n^{-\zeta^2/4}$ and we arrive at (2.113). By (2.72) and the fact that $B_n = (\log n)^{1-\delta/4}C_n^{3/2} \geq C_n^3(\log n)^{-\frac{1}{2}-\frac{\delta}{2}}$ we get from (2.49), (2.112), and (2.113)

$$P\left\{|r_n(\mathbf{t},\alpha)| \geq (\zeta+1)B_n\right\} \leq 2n^{-\zeta^2/4} \qquad (2.119)$$

for $n \geq n_0$, any $\zeta > 0$, and B_n given by (2.114).

As it holds that $\alpha_n^* \geq n^{-4}$ we finally get (2.108) by (2.78), (2.79), (2.80) (2.81), (2.82), and (2.83) using chaining arguments as in the proof of Lemma 2.3.3 after (2.73). □

Theorem 2.3.2. *Assume that the distribution function $F(x)$ fulfills conditions (F.1)-(F.4) and for the design matrix $\mathbf{X}$ it holds (X.1)–(X.2) and (X.5)–(X.6). Then for any $\delta > 4$ and $\alpha \in [\alpha_n^*, 1-\alpha_n^*]$ it holds*

$$\frac{n^{1/2}f(F^{-1}(\alpha))}{(\alpha(1-\alpha))^{1/2}}\left(\widehat{\boldsymbol{\beta}}_n\left(\alpha|\,\mathbf{Y},\mathbf{x}\right)-\boldsymbol{\beta}(\alpha)\right)$$
$$= n^{-1/2}(\alpha(1-\alpha))^{-1/2}\mathbf{D}_n^{-1}\sum_{i=1}^{n}\mathbf{x}_{ni}\psi_\alpha(E_{i\alpha})+o_P(1)$$

and

$$\sup_{\alpha_n^*\le\alpha\le 1-\alpha_n^*}\left\|n^{1/2}\frac{f(F^{-1}(\alpha))}{(\alpha(1-\alpha))^{1/2}}(\widehat{\boldsymbol{\beta}}_n\left(\alpha|\,\mathbf{Y},\mathbf{x}\right)-\boldsymbol{\beta}(\alpha))\right\| =$$
$$\sup_{\alpha_n^*\le\alpha\le 1-\alpha_n^*}\left\|\frac{\mathbf{q}_n(\alpha)}{(\alpha(1-\alpha))^{1/2}}\right\| = \mathcal{O}_P(\log\log n),$$

where $\alpha_n^*=\frac{(\log n)^{1+\delta}}{n}$.

Proof. It follows from the same convex arguments as in the proof of Theorem 2.3.1. □

A few notes should be made to the assumptions of Theorems 2.3.1 and 2.3.2 and the assumptions of the other approximations of the regression quantiles process established so far. Comparing (X.G.1)–(X.G.2) with (X.3)–(X.4) it yields that Theorem 3.1 of Gutenbrunner et al. (1993) is stronger in the sense that it admits more design matrices than our Theorem 2.3.1. On the contrary, our conditions (F.1–3) allow more general distribution functions than (F.G.1)–(F.G.4). Same conclusions hold also for the approximation theorem provided by Jurečková (1999) as the assumptions on design matrix are the same as in Gutenbrunner et al. (1993).

However, the situation in this case is not clear. Mainly, Lemma 2.2 of Jurečková (1999) is not an analogy of Lemma 3.1 of Gutenbrunner et al. (1993) (and our Lemma 2.3.3), as the sum of losses Q_i is not weighted by $(\alpha(1-\alpha))^{-1/2}$. A careful observation of the proof of Theorem 3.1. in Gutenbrunner et al. (1993) (similar as the convex arguments in the proof of our Theorem 2.3.1) reveals that Theorem 2.1. of Jurečková (1999) do not follow from the convex arguments and Lemma 2.2 stated therein. Precisely, an equivalent of $G_{n\alpha}(\mathbf{t}_\alpha^*)$ used by Jurečková (1999) would not be bounded away from zero in (2.105) and (2.105) due to the different weights. It is therefore questionable whether Theorem 2.1. of Jurečková (1999) really holds in the form given in the article. Moreover, the proof of Lemma 2.2. of Jurečková (1999) is not clear at some points, cf. the transition between (A.7) and (A.8) ibidem. Lemma 2.3.3 in this thesis can be thus seen as a correction of the assertions of Jurečková (1999). Different weights enable a relaxation of the assumptions on the distribution function and also a considerable extension of the interval where the relation (2.87) holds. This is at the cost of more strict conditions on the design matrix **X**.

Theorems 2.3.1 and 2.3.2 in fact enable an approximation to Brownian bridges as well. This is a consequence of Theorem 2.1. of Shorack (1979) as have been already mentioned in Gutenbrunner and Jurečková (1992).

Lemma 2.3.5. *Suppose that it holds*

(X.S.1) $n^{-1}\sum_{i=1}^{n}\|\mathbf{x}_i\|^2 = O(1)$ *as* $n \to \infty$.

(X.S.2) $\max_{1\le i\le n}\|\mathbf{x}_i\| = O(n^{1/2})$

(X.S.3) $\lim_{n\to\infty}\mathbf{D}_n = \mathbf{D}$, *where* $\mathbf{D}_n = n^{-1}\mathbf{X}_n^\top\mathbf{X}_n$ *and* $\mathbf{D}$ *is a positive definite* $(p\times p)$ *matrix.*

Then for the sequence

$$\mathbf{Z}_n(\alpha) := n^{-1/2}(\alpha(1-\alpha))^{-1/2}\sum_{i=1}^{n}\mathbf{x}_{ni}\psi_\alpha(E_{i\alpha})$$

holds

$$\mathbf{D}_n^{-1}\mathbf{Z}_n(\alpha) \to \mathbf{D}^{-1}\mathbf{B}_p(\alpha), \qquad \text{on } D[0,1]^p, \tag{2.120}$$

with $\mathbf{B}_p$ *being the vector of* p *independent Brownian bridges.*

Proof. See Shorack (1979), Theorem 2.1. □

Note that the conditions (X.S.1)–(X.S.3) follows from the already stronger assumptions (X.1)–(X.4). By Lemma 2.3.5 and Theorem 2.3.1 (or Theorem 2.3.2) we thus have that the regression quantile process is with $n\to\infty$ near to a p-dimensional Brownian bridge uniformly in $[\alpha_n^*, 1-\alpha_n^*]$.

While we are not able to establish the Bahadur representation of regression quantiles for $\alpha \in [1-\alpha_n^*, 1-1/n]$ (or $[\alpha_n^*, 1]$), we can at least establish a boundary for process $\overline{\mathbf{x}}^\top\mathbf{q}_n$, where

$$\overline{\mathbf{x}} := \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i. \tag{2.121}$$

Note that the first component $\overline{x}_1$ of $\overline{\mathbf{x}}$ is equal to one, as for the first component of $\mathbf{x}_i$ holds $x_{i,1} = 1$, $1\le i\le n$. The process $\overline{\mathbf{x}}^\top\mathbf{q}_n(\alpha)$ is monotone for $\alpha\in[0,1]$ which enables to provide a bound for $\overline{\mathbf{x}}^\top\mathbf{q}_n(\alpha)$ using relation $\overline{\mathbf{x}}^\top\mathbf{q}_n(\alpha) \le \overline{\mathbf{x}}^\top\mathbf{q}_n(1)$, where $\mathbf{q}_n(1) := n^{1/2}f\left(F^{-1}(1-1/n)\right)\left(\widehat{\boldsymbol\beta}_n(1)-\boldsymbol\beta(1)\right)$.

Theorem 2.3.3. *Suppose that the distribution of errors in model (2.1) fulfills the assumptions of Theorem 2.3.2, i.e. (F.1)–(F.4), (X.1)–(X.2), and (X.5)–(X.6). Moreover suppose that for* γ *in (F.4) it holds that* $\gamma > 0$ *and* $\max_{1\le i\le n}\|\mathbf{x}_{ni}\| = \mathcal{O}(1)$. *Then for* α_n^* *as in (2.107) follows*

$$\begin{aligned}\sup_{1-\alpha_n^*\le\alpha\le\frac{n-1}{n}}\left|\overline{\mathbf{x}}^\top\mathbf{q}_n(\alpha)\right| &= \sup_{1-\alpha_n^*\le\alpha\le\frac{n-1}{n}}\left|n^{1/2}f\left(F^{-1}(\alpha)\right)\overline{\mathbf{x}}^\top\left(\widehat{\boldsymbol\beta}_n(\alpha)-\boldsymbol\beta(\alpha)\right)\right| \\ &= \mathcal{O}_P(n^{-1/2}(\log n)^{(2+\delta)(1\vee\gamma^*)}) = o_P(1)\end{aligned} \tag{2.122}$$

and

$$\sup_{1/n\le\alpha\le\alpha_n^*}\left|\overline{\mathbf{x}}^\top\mathbf{q}_n(\alpha)\right| = \sup_{1/n\le\alpha\le\alpha_n^*}\left|n^{1/2}f\left(F^{-1}\left(\alpha\right)\right)\overline{\mathbf{x}}^\top\left(\widehat{\boldsymbol{\beta}}_n(\alpha)-\boldsymbol{\beta}(\alpha)\right)\right|$$
$$= \mathcal{O}_P(n^{-1/2}(\log n)^{(2+\delta)(1\vee\gamma_*)}) = o_P(1), \qquad (2.123)$$

where $\delta > 0$ is as in (2.107) and $\gamma^*, \gamma_* > 0$ as in (F.4).

Proof. It suffices to prove the relation for the upper tail, the case of the lower tail is analogous. Have for some $n_0 \in \mathbb{N}$ and $n \ge n_0$ sequence z_n defined such that

$$z_n := n\left(1 - \arg\max_{1-\alpha_n^*\le\alpha\le1-1/n} n^{1/2}f(F^{-1}(\alpha))\overline{\mathbf{x}}^\top\left(\widehat{\boldsymbol{\beta}}_n(\alpha)-\boldsymbol{\beta}(\alpha)\right)\right). \qquad (2.124)$$

As $(1-z_n/n)\in[1-\alpha_n^*, 1-1/n]$, by (2.107) follows that $1 \le z_n \le \log^{2+\delta}$. We shall prove that

$$n^{1/2}f(F^{-1}(1-z_n/n))\overline{\mathbf{x}}^\top\boldsymbol{\beta}(\alpha) = n^{1/2}f(F^{-1}(1-z_n/n))\overline{\mathbf{x}}^\top\boldsymbol{\beta}$$
$$+n^{1/2}f(F^{-1}(1-z_n/n))F^{-1}\left(1-z_n/n\right)\beta_1 = \mathcal{O}(n^{-1/2}z_n), \qquad (2.125)$$

As $\gamma^* > 0$ we get by von Mises condition (1.9) that it holds

$$n^{1/2}f\left(F^{-1}\left(1-\frac{z_n}{n}\right)\right)F^{-1}\left(1-\frac{z_n}{n}\right) = \mathcal{O}(z_n n^{-1/2}). \qquad (2.126)$$

Hence it follows also (2.125) as $\overline{\mathbf{x}}^\top\boldsymbol{\beta} \le \|\overline{\mathbf{x}}\|\cdot\|\boldsymbol{\beta}\| = \mathcal{O}(1)$. We shall see that also $n^{1/2}f(F^{-1}(1-z_n))\overline{\mathbf{x}}^\top\widehat{\boldsymbol{\beta}}_n(1-z_n/n)$ is tending to zero in infinity. First note that $\overline{\mathbf{x}}^\top\widehat{\boldsymbol{\beta}}_n(1-z_n/n) \le \overline{\mathbf{x}}^\top\widehat{\boldsymbol{\beta}}_n(1)$, which is due to the relation in Dodge and Jurečková (2000), pp. 127–128, saying that $\overline{\mathbf{x}}^\top\widehat{\boldsymbol{\beta}}_n(\alpha_1) \le \overline{\mathbf{x}}^\top\widehat{\boldsymbol{\beta}}_n(\alpha_2)$ whenever $0 \le \alpha_1 \le \alpha_2 \le 1$. Hence

$$\begin{aligned}
n^{1/2}f\left(F^{-1}\left(1-\frac{z_n}{n}\right)\right)\overline{\mathbf{x}}^\top\widehat{\boldsymbol{\beta}}_n\left(1-\frac{z_n}{n}\right) &\le n^{1/2}f\left(F^{-1}\left(1-\frac{z_n}{n}\right)\right)\overline{\mathbf{x}}^\top\widehat{\boldsymbol{\beta}}_n\left(1\right)\\
&\le \mathcal{O}\left(z_n^{\gamma^*}f\left(F^{-1}\left(1-\frac{1}{n}\right)\right)\left|\overline{\mathbf{x}}^\top\widehat{\boldsymbol{\beta}}_n(1)\right|\right)\\
&= \mathcal{O}\left(\frac{z_n^{\gamma^*}n^{-1/2}}{F^{-1}\left(1-1/n\right)}\left|\overline{\mathbf{x}}^\top\widehat{\boldsymbol{\beta}}_n(1)\right|\right),
\end{aligned}$$

where the second row follows from Lemma 2.3.2 (as we are applying the Lemma only on the right tail, where the density is continuous and monotone we can set $K_\gamma = \gamma^*$) and the third row from von Mises condition (1.9).

Now we shall prove that $\overline{\mathbf{x}}^\top\widehat{\boldsymbol{\beta}}_n(1) = \mathcal{O}_P(F^{-1}(1-1/n))$, which completes the proof. This can be done through the technique involving regression rank scores, which was used in Portnoy and Jurečková (2000).

Recall the definition of regression rank scores in (2.8). It follows that for the continuity points of any single regression quantile $\widehat{\boldsymbol{\beta}}_n(\alpha)$ it holds

$$\sum_{i=1}^{n} x_{ij} \frac{\mathrm{d}}{\mathrm{d}\alpha} \hat{a}_{ni}(\alpha) = -\sum_{i=1}^{n} x_{ij}, \quad 1 \leq i \leq n, 1 \leq j \leq p. \tag{2.127}$$

Note that there exists an $\varepsilon > 0$ such that $\hat{a}_{ni}(\alpha)$ are nonincreasing in $1-\varepsilon < \alpha < 1$, for $i = 1, \ldots, n$. Hence, by continuity of $\mathbf{a}(\alpha)$ it follows

$$\sum_{i=1}^{n} x_{ij} \hat{a}_i'(1) = -\sum_{i=1}^{n} x_{ij}, \quad j = 1, \ldots, p, \tag{2.128}$$

where $\hat{a}_i'(1)$ is the left-hand derivative of $\hat{a}_i(\alpha)$ at $\alpha = 1$). Hence, similarly as in the case of regression quantiles defined for $\alpha \in (0,1)$ we get the identity

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_n(1) = -\frac{1}{n} \sum_{i=1}^{n} Y_i \hat{a}_i'(1), \tag{2.129}$$

which by the definition of the model (2.1) implies

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^\top \left(\widehat{\boldsymbol{\beta}}_n(1) - \boldsymbol{\beta} \right) = -\frac{1}{n} \sum_{i=1}^{n} E_i \hat{a}_i'(1). \tag{2.130}$$

As the largest regression quantile is any solution of the linear program

$$\min_{\mathbf{b} \in \mathbb{R}^p} \sum_{i=1}^{n} \mathbf{x}_i^\top \mathbf{b}, \qquad \text{s.t. } Y_i \leq \mathbf{x}_i^\top \mathbf{b}, \quad i = 1, \ldots, n, \tag{2.131}$$

we get by duality theory that $(-\hat{a}_1'(1), \ldots, -a_n'(1))^\top$ is a solution of the linear programming problem

$$\begin{array}{rcl} \mathbf{Y}_n^\top \mathbf{A} & = & \max \\ \mathbf{X}_n^\top \mathbf{A} & = & \mathbf{X}_n^\top \mathbf{1}_n \\ A_i \geq 0, & & i = 1, \ldots, n. \end{array} \tag{2.132}$$

This implies (2.129) as the equality between the primal and the dual optima. It follows from the linear programming theory that there exists an optimal basis of problem (2.133) consisting of the p columns of $\mathbf{X}_n^\top$. Assume, without loss of generality, that the rows of $\mathbf{X}_n = \mathbf{X}$ are ordered such that

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix},$$

where $\mathbf{X}_1^\top$ is $p \times p$ matrix containing the optimal basis of the problem (2.133), $\mathbf{X}_2 = \mathbf{B}\mathbf{X}_1$ is the matrix of order $(n-p) \times p$, and $\mathbf{B}$ is a matrix of linear transformation of the basis contained in $\mathbf{X}_1$. Writing $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)$ we get the optimality criterion as $\mathbf{B}\mathbf{Y}_1 =$

$\mathbf{X}_2\mathbf{X}_1^{-1}\mathbf{Y}_1 \geq \mathbf{Y}_2$, where the inequality holds component wise. The optimal solution then would be

$$
\begin{aligned}
\hat{a}'_{p+1}(1) = \ldots = \hat{a}'_n(1) &= 0, \\
\left(\hat{a}'_1(1), \ldots, \hat{a}'_p(1)\right) &= \mathbf{1}_p^\top + \mathbf{1}_{n-p}^\top \mathbf{X}_2\mathbf{X}_1^{-1}.
\end{aligned} \tag{2.133}
$$

Rewriting (2.128) for $j = 1$ and the fact that the first column of $\mathbf{X}$ is intercept yields $\sum_{i=1}^n \hat{a}'_i(1) = -n$ and thus we have for $t \in \mathbb{R}$ that

$$
\begin{aligned}
P_{\boldsymbol{\beta}}\left(\sum_{i=1}^n \mathbf{x}_i^\top \left(\boldsymbol{\beta}(1) - \boldsymbol{\beta}\right) \geq nt\right) &= P_{\mathbf{0}}\left(\sum_{i=1}^n \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_n(1) \geq nt\right) \\
= P_{\mathbf{0}}\left(-\sum_{i=1}^n E_i\hat{a}'(1) \geq nt\right) &\leq P\left(E_{n:n} \geq t\right).
\end{aligned} \tag{2.134}
$$

Have any $\zeta \in \mathbb{R}$ and for $n \geq n_0$ set $t = t_n := \zeta F^{-1}(1 - 1/n)$. By (2.134) it follows

$$
P_{\boldsymbol{\beta}}\left(\overline{\mathbf{x}}^\top \left(\boldsymbol{\beta}(1) - \boldsymbol{\beta}\right) \geq \zeta F^{-1}\left(1 - \frac{1}{n}\right)\right) \leq P\left(\frac{E_{n:n}}{F^{-1}\left(1 - 1/n\right)} \geq \zeta\right)
$$

for any $\zeta \in \mathbb{R}$. But as $n \to \infty$

$$
P\left(\frac{E_{n:n}}{F^{-1}\left(1 - 1/n\right)} \geq \zeta\right) \xrightarrow[n\to\infty]{\mathcal{D}} 1 - \exp\left(-\zeta^{-\frac{1}{\gamma}}\right), \tag{2.135}
$$

see Beirlant et al. (2004), pp. 58. Letting $\zeta \to \infty$ gives that the right side of (2.135) converge to zero, hence $\overline{\mathbf{x}}^\top(\widehat{\boldsymbol{\beta}}_n(1) - \boldsymbol{\beta}) = \mathcal{O}_P(F^{-1}(1 - 1/n))$ which by $\|\overline{\mathbf{x}}\| = \mathcal{O}(1)$ and $F^{-1}(1 - 1/n) \to \infty$ finally gives

$$
\mathcal{O}\left(\frac{z_n^{\gamma^*} n^{-1/2}}{F^{-1}\left(1 - 1/n\right)} \left|\overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n(1)\right|\right) = \mathcal{O}_P\left(z_n^{\gamma^*} n^{-1/2}\right),
$$

which completes the proof. □

The bounds we have deduced for process $\mathbf{q}_n(\alpha)$ are not the optimal ones as they are linked to α_n^* of (2.107), which was defined in such order to enable the proof of Theorem 2.3.2 through the Bernstein inequality and the convex arguments.

We shall see that results have a profound theoretical impact on the following section, as they enable us to treat the tails of regression quantiles more precisely.

2.4 Tails of regression quantiles

The approximation theorems of the previous section open a possibility to approximate the tails of regression quantile and use this approximation to construct new class of estimates in the fashion described in section 1.5. Have again model (2.1) with i.i.d. errors, where $E_i \sim F, i = 1, \ldots, n$, $\mathbf{X}$ fulfill the assumptions (F.1)–(F.3) and $\mathbf{X}$ fulfills (X.1)–(X.4) (or

(X.1)–(X.2) and (X.5)–(X.6)). Throughout this section we shall moreover suppose that $\gamma^* = \gamma > 0$ (we are interested only in the right tail, so we may change without loss of generality the distribution near the left endpoint to reflect this fact), i.e. the distribution of errors in (2.1) belongs to the Fréchet domain.

While $\hat{\boldsymbol{\beta}}_n(\alpha) \in \mathbb{R}^p$, the key information about the distribution function is concentrated (at least under i.i.d. errors) in its first component, intercept. As Theorems 2.3.1 and 2.3.2 state, the theoretical counterpart of the first component of $\widehat{\boldsymbol{\beta}}_n(\alpha)$ is $F^{-1}(\alpha)+\beta_1$. Hence the tails of regression quantiles can be approximated with the probability tending to one uniformly on $[\alpha_n^*, 1-\alpha_n^*]$ similarily as in Theorems 1.5.1 and 1.5.2. Nevertheless, as we do not have the weighted Bahadur representation for regression quantiles which woul be analogous to Theorem 6.2.1 in Csörgő and Horváth (1993), see also de Haan and Ferreira (2006), Theorem 2.4.2, as well as we cannot use an easy approximation of regression quantiles by uniform quantile process, see Drees (1998*b*), Theorem 2.1., we cannot proceed directly. The cost is, that we obtain only an upper boundary for the desired regression tail quantile process.

Theorem 2.4.1. *Suppose that the distribution function F of errors in (2.1) satisfies (EVT.2) for some $\gamma \in \mathbb{R}$ and $\rho \le 0$. Suppose that the assumptions of Theorem 2.3.2 are fulfilled. Then we can define a Wiener processes $\{W_n(t)\}_{t\ge 0}$ such that for suitable chosen functions A, K, z_γ and $a(\cdot)$ as in (1.25) and on space $\mathcal{D}_{\gamma,h}$ equipped with metric seminorm*

$$\|z\|_{\gamma,h,\alpha_n^*} := \sup_{t\in[0,1]} t^\gamma h(t)|z(t)|. \tag{2.136}$$

defined for any $\varepsilon > 0$ through the space of weights as in (1.54) by function

$$h(t) = t^{1/2+\varepsilon}, \qquad t \in [0,1], \tag{2.137}$$

it holds

$$\begin{aligned}
&\Bigg\| k^{1/2}\left(\frac{\overline{\mathbf{x}}^\top\left(\widehat{\boldsymbol{\beta}}_n(\alpha)-\boldsymbol{\beta}\right)}{F^{-1}\left(1-\frac{k}{n}\right)} - t^{-\gamma}\right) - \gamma t^{-\gamma-1}W_n(t) \\
&\qquad\qquad\qquad\qquad -k^{1/2}A\left(\frac{k}{n}\right)t^{-\gamma}\frac{t^{-\rho}-1}{\rho}\Bigg\|_{\gamma,h} \\
&\le \left\|\gamma t^{-\gamma}(n/k)^{1/2}\overline{\mathbf{x}}^\top\mathbf{D}_n^{-1}\mathbf{Z}_n\left(1-\frac{tk}{n}\right)\right\|_{\gamma,h} + o_P(1),
\end{aligned} \tag{2.138}$$

$n\to\infty$, provided that $k = k(n)\to\infty$, $k/n\to 0$ and $\sqrt{k}A(k/n) = \mathcal{O}(1)$ and $k \ge \log^{\Delta(1\vee\gamma)}$ with $\Delta > 4+2\delta$.

Proof. If $\gamma > 0$ and (EVT.2) holds we can formulate Theorem 1.5.1 and write it in a slightly different form introduced by de Haan and Ferreira (2006), Theorem 2.4.8. Hence we have that for $E_{1:n}, \ldots, E_{n:n}$ we can define a sequence of Wiener processes $\{W_n(s)\}$ such that with A as in 1.25 we have for $\varepsilon > 0$ sufficiently small

$$\sup_{0\le t\le 1} t^{\gamma+1/2+\varepsilon} \Bigg| k^{1/2}\left(\frac{E_{n-[kt],n}}{F^{-1}\left(1-\frac{k}{n}\right)} - t^{-\gamma}\right) - \gamma t^{-\gamma-1} W_n(t) \\ -k^{1/2} A\left(\frac{k}{n}\right) t^{-\gamma}\frac{t^{-\rho}-1}{\rho}\Bigg| \xrightarrow[n\to\infty]{\mathrm{P}} 0. \tag{2.139}$$

By Theorem 6.1.5 of Csörgő and Horváth (1993) and by the law of the iterated logarithm for uniform quantile process, i.e. Theorem 5.2.4, ibidem, we get by Theorem 2.3.2

$$\overline{\mathbf{x}}^\top\left(\widehat{\boldsymbol{\beta}}_n\left(1-\frac{kt}{n}\right)-\boldsymbol{\beta}\right) - E_{n-[kt]:n} = \tag{2.140}$$

$$\frac{\overline{\mathbf{x}}^\top \mathbf{D}_n^{-1}\mathbf{Z}_n\left(1-\frac{tk}{n}\right)}{n^{1/2} f\left(F^{-1}\left(1-\frac{tk}{n}\right)\right)} + \frac{o_P(1)\left(\frac{kt}{n}\left(1-\frac{kt}{n}\right)\right)^{1/2}}{n^{1/2} f\left(F^{-1}\left(1-\frac{tk}{n}\right)\right)} \tag{2.141}$$

uniformly on $[\alpha_n^*/k, 1]$ with α_n^* as in (2.107) for any intermediate sequence $k \to \infty$, $k/n \to 0$ such that $k \le n\alpha_n^*$. Similarly Theorem 2.3.3 yields that uniformly on $[1/(n+1), n\alpha_n^*/k]$ holds

$$\overline{\mathbf{x}}^\top\left(\widehat{\boldsymbol{\beta}}_n\left(1-\frac{kt}{n}\right)-\boldsymbol{\beta}\right) - E_{n-[kt]:n} = \frac{\mathcal{O}_P(n^{-1}(\alpha_n^*)^{(1\vee\gamma)})}{f\left(F^{-1}\left(1-\frac{tk}{n}\right)\right)}. \tag{2.142}$$

For any k such that $k^{-1} = o((n\alpha_n^*)^{2(1\vee\gamma)})$, i.e. for $k \ge \log^{\Delta(1\vee\gamma)}$ with $\Delta > 4+2\delta$, where $\delta > 0$ is as in (2.107), it holds uniformly for $t \in [0,1]$ and any $\varepsilon > 0$ suitably small that

$$\begin{aligned} \frac{t^{\gamma+1/2+\varepsilon} k^{1/2}\mathcal{O}_P\left(n^{-1}(\alpha_n^*)^{(1\vee\gamma)}\right)}{F^{-1}\left(1-\frac{k}{n}\right) f\left(F^{-1}\left(1-\frac{tk}{n}\right)\right)} &\le \frac{t^{1/2+\varepsilon} k^{1/2}\mathcal{O}_P\left(n^{-1}(\alpha_n^*)^{(1\vee\gamma)}\right)}{F^{-1}\left(1-\frac{k}{n}\right) f\left(F^{-1}\left(1-\frac{k}{n}\right)\right)} \\ &\le t^{1/2+\varepsilon}\frac{\mathcal{O}_P\left((\alpha_n^*)^{(1\vee\gamma)}\right)}{k^{1/2}} = o_P(1), \end{aligned} \tag{2.143}$$

where the first inequality is by Lemma 2.3.2 and the second by von Mises condition (1.9). Similarly

$$\begin{aligned} \frac{o_P(1)\, t^{\gamma+1/2+\varepsilon} k^{1/2}\left(\frac{kt}{n}\left(1-\frac{kt}{n}\right)\right)^{1/2}}{n^{1/2}F^{-1}\left(1-\frac{k}{n}\right) f\left(F^{-1}\left(1-\frac{tk}{n}\right)\right)} &\le \frac{o_P(1)\, t^{1/2+\varepsilon} k^{1/2}\left(\frac{kt}{n}\left(1-\frac{kt}{n}\right)\right)^{1/2}}{n^{1/2}F^{-1}\left(1-\frac{k}{n}\right) f\left(F^{-1}\left(1-\frac{k}{n}\right)\right)} \\ &\le \frac{o_P(1)\, t^{1+\varepsilon}\left(\frac{k}{n}\right)}{F^{-1}\left(1-\frac{k}{n}\right) f\left(F^{-1}\left(1-\frac{k}{n}\right)\right)} \\ &= t^{1+\varepsilon} o_P(1) = o_P(1). \end{aligned} \tag{2.144}$$

uniformly $t \in [0,1]$ again by Lemma 2.3.2 and by von Mises condition (1.9). Hence it follows

$$\sup_{t\in[0,1]} t^{\gamma+1/2+\varepsilon} \max\{M_n(t), N_n(t)\} = o_P(1), \tag{2.145}$$

where

$$M_n(t) := \frac{k^{1/2}\mathcal{O}_P\left(n^{-1}(\alpha_n^*)^{(1\vee\gamma)}\right)}{F^{-1}\left(1-\frac{k}{n}\right) f\left(F^{-1}\left(1-\frac{tk}{n}\right)\right)}$$

and

$$N_n(t) := \frac{o_P(1)\, t^{\gamma+1/2+\varepsilon} k^{1/2}\left(\frac{kt}{n}\left(1-\frac{kt}{n}\right)\right)^{1/2}}{n^{1/2}F^{-1}\left(1-\frac{k}{n}\right) f\left(F^{-1}\left(1-\frac{tk}{n}\right)\right)}.$$

Moreover by Lemma 2.3.2 and von Mises condition (1.9) we have

$$\begin{aligned} L_n(t) &:= t^{1/2+\varepsilon+\gamma}\left(\frac{(k)^{1/2}}{n^{1/2}F^{-1}\left(1-\frac{k}{n}\right) f\left(F^{-1}\left(1-\frac{tk}{n}\right)\right)} - \gamma\left(\frac{n}{k}\right)^{1/2} t^{-\gamma}\right) \\ &\qquad\qquad\qquad\qquad\qquad \overline{\mathbf{x}}^\top \mathbf{D}_n^{-1}\mathbf{Z}_n\left(1-\frac{tk}{n}\right) \\ &= o(1)\, t^{1/2\varepsilon}\left(\frac{n}{k}\right)^{1/2} \overline{\mathbf{x}}^\top \mathbf{D}_n^{-1}\mathbf{Z}_n\left(1-\frac{tk}{n}\right) \end{aligned} \tag{2.146}$$

By Lemma 2.3.5 we have that $\mathbf{D}_n^{-1}\mathbf{Z}_n \to \mathbf{D}_n^{-1}\mathbf{B}_p$ on $D[0,1]^p$ with $\mathbf{B}_p$ being a vector of p independent Brownian bridges on $[0,1]$. Similary as in 1-diemensional case it holds by the defintion for p-dimensional Brownian bridges that

$$\mathbf{B}\left(1-\frac{kt}{n}\right) \overset{\mathcal{D}}{=} \mathbf{B}\left(\frac{kt}{n}\right) \overset{\mathcal{D}}{=} \mathbf{W}\left(\frac{kt}{n}\right) - \frac{kt}{n}\mathbf{W}(1), \tag{2.147}$$

where $\mathbf{W}$ is a vector of p independent Wiener processes. Note that as k is an intermediate seequence it is

$$\sup_{t\in[0,1]}\left\|\left(\frac{kt}{n}\right)^{1/2} t^{\varepsilon}\mathbf{W}(1)\right\| = o_P(1). \tag{2.148}$$

Moreover as $\sqrt{n/k}\mathbf{W}(kt/n) \xrightarrow[n\to\infty]{\mathcal{W}} \mathbf{W}(t)$ and $\|\overline{\mathbf{x}}^\top\mathbf{D}_n^{-1}\| = \mathcal{O}(1)$, we have by Lemma 2.3.5, (2.146), (2.147) and (2.148) that for the process $L_n(t)$ holds uniformly in $t\in[0,1]$

$$L_n(t) \xrightarrow[n\to\infty]{\mathcal{W}} 0.$$

which yields

$$\sup_{t\in[0,1]} |L_n(t)| \xrightarrow[n\to\infty]{\mathrm{P}} 0. \tag{2.149}$$

By (2.139), (2.141), (2.142), (2.145), and (2.149) we finally get that

$$\sup_{k^{-1}\le t\le 1} t^{\gamma+1/2+\varepsilon}\left|k^{1/2}\left(\frac{\overline{\mathbf{x}}^\top\left(\widehat{\boldsymbol{\beta}}_n(\alpha)-\boldsymbol{\beta}\right)}{F^{-1}\left(1-\frac{k}{n}\right)}-t^{-\gamma}\right)-\gamma t^{-\gamma-1}W_n(t)\right.$$
$$\left.-k^{1/2}A\left(\frac{k}{n}\right)t^{-\gamma}\frac{t^{-\rho}-1}{\rho}\right|$$
$$\le \sup_{0\le t\le 1} t^{\gamma+1/2+\varepsilon}\left|\gamma t^{-\gamma}(n/k)^{1/2}\overline{\mathbf{x}}^\top\mathbf{D}_n^{-1}\mathbf{Z}_n\left(1-\frac{tk}{n}\right)\right|+o_P(1). \tag{2.150}$$

To extend relation (2.150) over the whole unit interval it suffices to show that

$$\sup_{1-\frac{1}{n+1}\le\alpha\le 1-\frac{1}{n}}\left|n^{1/2}f\left(F^{-1}(\alpha)\right)\left(\overline{\mathbf{x}}^\top\left(\widehat{\boldsymbol{\beta}}_n(\alpha)-\boldsymbol{\beta}\right)-E_{n:n}\right)\right|=\mathcal{O}_P\left(n^{-1/2}\right),$$

which can be done using the arguments in the proof of Theorem 2.3.3 (note that if $(1-z_n/n)\in[1-1/(n+1),1-1/n]$ then $z_n=O(1)$). □

Theorem 2.4.2. *Suppose that the assumptions of Theorem 2.3.2 are fulfilled (thus F fullfils (EVT.1)). Assume that $\gamma>0$. Then we can define a Wiener processes $\{W_n(t)\}_{t\ge 0}$ such that on $\mathcal{D}_{\gamma,h}$ equipped with metric seminorm $\|\cdot\|_{\gamma,h,\varepsilon}$ defined in (2.136) it holds for any $\varepsilon>0$*

$$\left\|k^{1/2}\left(\frac{\overline{\mathbf{x}}^\top\left(\widehat{\boldsymbol{\beta}}_n(\alpha)-\boldsymbol{\beta}\right)}{F^{-1}\left(1-\frac{k}{n}\right)}-t^{-\gamma}\right)-\gamma t^{-\gamma-1}W_n(t)\right\|_{\gamma,h}$$
$$\le\left\|\gamma t^{-\gamma}(n/k)^{1/2}\overline{\mathbf{x}}^\top\mathbf{D}_n^{-1}\mathbf{Z}_n\left(1-\frac{tk}{n}\right)\right\|_{\gamma,h}+o_P(1)$$
$$+\mathcal{O}\left(k^{1/2}\sup_{x\in(0,1+\varepsilon)}x^{\gamma+\frac12}\left|R\left(\frac{k}{n},x\right)\right|\right) \tag{2.151}$$

$n\to\infty$, provided that $k=k(n)\to\infty$, $k/n\to 0$ and $k\ge\log^{\Delta(1\vee\gamma)}$ with $\Delta>4+2\delta$.

Proof. Follows analogously as Theorem 2.4.1 from the different paramaterization of (2.138) introduced in Theorem 1.5.2. □

Although Theorem 2.4.1 is not a direct analogue of Theorem 1.5.1 as it provides only an upper bound for the probability, we can use it to deduce some important facts about smooth functionals of the tail regression quantiles applied on *tail regression quantile process* defined as

$$\hat{Q}_{\overline{\mathbf{x}},n,k}(t,\mathbf{X},\mathbf{Y})=\hat{Q}_{\overline{\mathbf{x}},n,k}(t):=\left\{\overline{\mathbf{x}}^\top\left(\widehat{\boldsymbol{\beta}}_n\left(1-\frac{kt}{n}\right)\right)\right\}_{t\in[0,1]}. \tag{2.152}$$

Define following Gaussian processes:

$$
\begin{aligned}
V_\gamma(t) &:= t^{-(\gamma+1)}W(t), \qquad t \in [0,1] \\
U_{\mathbf{D}}(t) &:= t^{-\gamma}\overline{\mathbf{x}}^\top \mathbf{D}^{-1}W(t), \qquad t \in [0,1].
\end{aligned} \tag{2.153}
$$

It holds By Lemma 2.3.5 that $(n/k)^{1/2}\mathbf{D}_n^{-1}\mathbf{Z}_n \to \mathbf{D}^{-1}\mathbf{W}$ weakly in $D[0,1]^p$. Hence, $(n/k)^{1/2}\overline{\mathbf{x}}^\top \mathbf{D}_n^{-1}\mathbf{Z}_n \to \overline{\mathbf{x}}^\top \mathbf{D}^{-1}\mathbf{W}$ weakly in $D[0,1]$, which yields $t^{-\gamma}(n/k)^{1/2}\overline{\mathbf{x}}^\top \mathbf{D}_n^{-1}\mathbf{Z}_n \to t^{-\gamma}\overline{\mathbf{x}}^\top \mathbf{D}^{-1}\mathbf{W}$ in $\mathcal{D}_{\gamma,h}$ and thus

$$
\gamma t^{-\gamma}(n/k)^{1/2}\overline{\mathbf{x}}^\top \mathbf{D}_n^{-1}\mathbf{Z}_n\left(1-\frac{tk}{n}\right) \to t^{-\gamma}\overline{\mathbf{x}}^\top \mathbf{D}_n^{-1}\mathbf{W}(t) \tag{2.154}
$$

weakly on $D[0,1]$ as well as on $\mathcal{D}_{\gamma,h}$. Using this result we can establish the consistency of any suitable smooth scale and location invariant functional applied on $\hat{Q}_{\overline{\mathbf{x}},n,k}$.

Theorem 2.4.3. *Suppose that the assumptions of Theorem 2.4.1 or Theorem 2.4.2 are fulfilled, $k = k_n$ is an intermediate sequence and T satisfies conditions (T.1)–(T.3). Moreover assume that $T_{|\mathcal{D}_{\gamma,h}}$ is continuous in z_γ, then*

$$
T\left(\hat{Q}_{\overline{\mathbf{x}},n,k}\right) \xrightarrow[n\to\infty]{\mathrm{P}} \gamma. \tag{2.155}
$$

Proof. By Theorem 2.4.1 or Theorem 2.4.2 we get due to (2.154) that

$$
\frac{\hat{Q}_{\overline{\mathbf{x}},n,k} - \overline{\mathbf{x}}^\top \boldsymbol{\beta}}{F^{-1}\left(1-\frac{k}{n}\right)} \xrightarrow[n\to\infty]{} z_\gamma + 1/\gamma \tag{2.156}
$$

weakly in $\mathcal{D}_{\gamma,h}$ and thus the result hold due to the continuity of T in z_γ and its location and scale invariance. $\square$

Under Hadamard differentiability, we can also deduce an asymptotic normality of any functional fulfilling (T.1)–(T.3).

Theorem 2.4.4. *Assume that $T : span(\mathcal{D}_{\gamma,h}, 1) \to \mathbb{R}$ satisfies conditions (T.1)–(T.4). Moreover, suppose that the assumptions of Theorem 2.4.1 or Theorem 2.4.2 are fulfilled and let $\sigma_{T,\gamma}$ be as in (1.62), $\mu_{T,\gamma,\rho}$ as in (1.63),*

$$
\varsigma_{T,\mathbf{X},\gamma} := \int_{[0,1]\times[0,1]} \overline{\mathbf{x}}^\top \mathbf{D}^{-1}(st)^{-\gamma}\min(s,t)\nu_{T,\gamma}^2\left(\mathrm{d}\, s, \mathrm{d}\, t\right). \tag{2.157}
$$

Then under the condition (EVT.2) on distribution function F of errors in model (2.1) and $\lim_{n\to\infty} k^{1/2}A\left(k/n\right) = \lambda \in [0,\infty]$ follows

(i) $\lambda \in (0,\infty)$

$$
k^{1/2}\left(T\left(\hat{Q}_{\overline{\mathbf{x}},n,k}\right) - \gamma\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(\lambda\mu_{T,\gamma,\rho}, \tilde{\sigma}_{T,\gamma}^2). \tag{2.158}
$$

for some $\tilde{\sigma}_{T,\gamma}^2 \le \sigma_{T,\gamma}^2 + \varsigma_{T,\mathbf{X},\gamma}^2$.

(ii) $\lambda = \infty$

$$k^{1/2}\left(T\left(\hat{Q}_{\overline{\mathbf{x}},n,k}\right)-\gamma\right)\xrightarrow[n\to\infty]{\mathrm{P}}\mu_{T,\gamma,\rho} \tag{2.159}$$

If (EVT.2) does not hold and it holds (1.66), i.e. $\sup_{x\in(0,1+\varepsilon]} x^{\gamma+1/2}\left|R\left(k_n/n,x\right)\right| = o\left(k_n^{-1/2}\right)$ *for some* $\varepsilon > 0$*, then*

$$k^{1/2}\left(T\left(\hat{Q}_{\overline{\mathbf{x}},n,k}\right)-\gamma\right)\xrightarrow[n\to\infty]{\mathcal{D}}\mathcal{N}(0,\tilde{\sigma}^2_{T,\gamma}). \tag{2.160}$$

with some $\tilde{\sigma}^2_{T,\gamma} \le \sigma^2_{T,\gamma} + \varsigma^2_{T,\mathbf{X},\gamma}$.

Proof. By the proof of Theorem 2.4.1 we get that there are some versions of $\overline{U}$ of $U_{\mathbf{D}}$ and $\overline{V}$ of $V_{\gamma,t}$ and $\hat{Q}^*_{\overline{\mathbf{x}},n,k}$ such that

$$k_n^{1/2}\left(\frac{\hat{Q}^*_{\overline{\mathbf{x}},n,k}-\overline{\mathbf{x}}^\top\boldsymbol{\beta}}{F^{-1}\left(1-\frac{k}{n}\right)}-\overline{z}_\gamma\right)=\overline{U}+\overline{V}+\lambda K+o_P(1).$$

Note that $T(z_\gamma) = T(\overline{z}_\gamma)$ for z_γ and $\overline{z}_\gamma$ defined in (1.24) and (1.68) and that the limiting process is continuous. Therefore Hadamard differentiability (1.61) implies that with $\varepsilon = k_n^{-1/2}\downarrow 0$ and $y_n = k_n^{1/2}\left(\left(\hat{Q}^*_{\overline{\mathbf{x}},n,k}-\overline{\mathbf{x}}^\top\boldsymbol{\beta}\right)/F^{-1}(1-k/n)-\overline{z}_\gamma\right)$ implies

$$k_n^{1/2}\left(T\left(\frac{\hat{Q}^*_{\overline{\mathbf{x}},n,k}-\overline{\mathbf{x}}^\top\boldsymbol{\beta}}{F^{-1}(1-k/n)}\right)-T\left(z_\gamma\right)\right)\xrightarrow[n\to\infty]{\mathrm{P}}T'(\overline{V}+\overline{U}+\lambda K(t)).$$

and thus by location and scale invariancy of T and $T(z_\gamma)=\gamma$ it follows

$$k_n^{1/2}\left(T\left(\hat{Q}_{\overline{\mathbf{x}},n,k}\right)-\gamma\right)\xrightarrow[n\to\infty]{\mathcal{W}}\int_0^1 V_\gamma+U_{\mathbf{D}}+\lambda K(t)\mathrm{d}\nu_{T,\gamma}. \tag{2.161}$$

Note that $\lambda K(t)$ is a deterministic process. Hence, if we assume that the processes V_γ and $U_{\mathbf{D}}$ are not dependent, we get from the fact that the both processes are Gaussian that the right side of (2.161) is a normal distributed random variable with mean $\mu_{T,\gamma}$ and variance $\sigma^2_{T,\gamma}+\varsigma^2_{T,\mathbf{X},\gamma}$, see Proposition 2.2.1 in Shorack and Wellner (1986). The assumption that the processes V_γ and U_γ are independent cannot be deduced in the proof of Theorem 2.4.1, thus we got for the variance only the upper bound.

The other assertions of the theorem follows analogously from Theorem 2.4.2. □

Theorems 2.4.3 and 2.4.4 provide a basic framework to establish the asymptotic properties of various estimators based on smooth functionals of $\hat{Q}_{\overline{\mathbf{x}},n,k}$. It follows that for functionals such as (1.52) or (1.76) that $T(\hat{Q}_{\overline{\mathbf{x}},n,k})$ is a consistent and asymptotically normal estimate of γ, with a specified mean and variance limited by the upper boundary. An important conclusion of Theorem 2.4.4 is that the regression quantile estimation (i.e. using $T(\hat{Q}_{\overline{\mathbf{x}},n,k})$ instead of $T(\hat{Q}_{n,k})$, where $\hat{Q}_{n,k}$ is the empirical tail quantile function of errors in (2.1)) can affect only the variance of the resulting γ-estimator and not the bias. Hence the quantile regression case in model (2.1) can be reduced with ease to the univariate *i.i.d.* case. As an example, consider

$$\hat{\gamma}^{\mathrm{RQ,Pick}}_{k,n,\overline{\mathbf{x}}} := \frac{1}{\log 2} \log \left(\frac{\overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n \left(\tau_{m-[k/4]}\right) - \overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n \left(\tau_{m-[k/2]}\right)}{\overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n \left(\tau_{m-[k/2]}\right) - \overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n \left(\tau_{m-k}\right)} \right), \tag{2.162}$$

where $0 < \tau_1 < \ldots < \tau_m < 1$ corresponds to $m(n, \mathbf{Y}, \mathbf{X})$ unique solutions $\widehat{\boldsymbol{\beta}}_n(\tau_i), i = 1, \ldots, m$ of minimization problem (2.3). Hence $\hat{Q}_{\overline{\mathbf{x}},n,k}$ is a step function on $[0,1]$ as k is an intermediate sequence and $m(n) \to \infty$. Hence estimator (2.162) is an analogue to Pickands estimator, as $\hat{\gamma}^{\mathrm{RQ,Pick}}_{k,n,\overline{\mathbf{x}}} = T_{\mathrm{Pick}}(\hat{Q}_{\overline{\mathbf{x}},n,k})$, with T_{Pick} introduced in (1.52). It is consistent and has an asymptotic normal distribution according to Theorems 2.4.3 and 2.4.4 provided that $[n(1-\tau_{m-k})]$ is a suitable intermediate sequence. This estimator is identical for some $\tilde{k}$ with the estimator

$$\hat{\gamma}^{\mathrm{RQ,Pick}}_{\tilde{k},n,\overline{\mathbf{x}}} := \frac{1}{\log 2} \log \left(\frac{\overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n \left(1 - \frac{\tilde{k}}{4n}\right) - \overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n \left(1 - \frac{\tilde{k}}{2n}\right)}{\overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n \left(1 - \frac{\tilde{k}}{2n}\right) - \overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n \left(1 - \frac{\tilde{k}}{n}\right)} \right). \tag{2.163}$$

Two estimators (2.162) and (2.163) differ only in their parametrization of the intermediate sequence. We get the consistency and asymptotical normality of (2.163) if $\tilde{k}$ is a suitable intermediate sequence. The estimator (2.163) is similar to the one proposed by Chernozhukov (2005), c.f. statistic (6.1) therein and his Theorem 6.1. He uses an arbitrary vector $\mathbf{x}$ to have a general theory of $\hat{\gamma}^{\mathrm{RQ,Pick}}_{k,n,\mathbf{x}}$. The theory for his estimator is more general and he obtained its consistency under various cases of heteroscedasticity.

On the other hand our improvement is that we obtained consistency and asymptotic normality for a much larger class of similar estimators as have been suggested by the literature so far. Our method is convenient particularly for the estimators based on functionals which can be written as $T\left(\int_0^1 h_1 d\nu_1\right)$ or $T\left(\int_0^1 h_1(t)\mathrm{d}\nu_1, \int_0^1 h_2 \mathrm{d}\nu_2\right)$, where ν_i are suitable finite signed Borel measures on $[0,1]$ and h_i are suitable measurable functions. Hill's estimator (with $h_1 = \log^+(\cdot)$ and ν_1 being Lebesgue measure on $[0,1]$) as well as PWM-estimator (see (1.76)) can be written in this fashion. Unlike of Pickands estimator, the asymptotic properties of these functionals cannot be covered by the older results of Chernozhukov (2005). For example let

$$P^{\mathrm{RQ}}_{n,j} := \frac{1}{j} \sum_{i=1}^{j} \left(\overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n(\tau_{m-i+1}) - \overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n(\tau_{m-k}) \right) \tag{2.164}$$

and

$$Q^{\mathrm{RQ}}_{n,j} := \frac{1}{j} \sum_{i=1}^{j} \frac{i}{j} \left(\overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n(\tau_{m-i+1}) - \overline{\mathbf{x}}^\top \widehat{\boldsymbol{\beta}}_n(\tau_{m-k}) \right), \tag{2.165}$$

where $0 < \tau_1 < \ldots < \tau_m < 1$ corresponds to $m(n, \mathbf{Y}, \mathbf{X})$ unique solutions $\widehat{\boldsymbol{\beta}}_n(\tau_i), i = 1, \ldots, m$ of minimization problem (2.3). Note that $P^{\mathrm{RQ}}_{n,k} = \int_0^1 \hat{Q}_{\overline{\mathbf{x}},n,k}(t)\mathrm{d}\mu(t)$ and $P^{\mathrm{RQ}}_{n,k} =$

$\int_0^1 t\hat{Q}_{\overline{\mathbf{x}},n,k}(t)\mathrm{d}\mu(t)$ where $\mu(t)$ is Lebesgue measure on $[0,1]$. Then we can define for a suitable intermediate sequence k estimator

$$\hat{\gamma}_{n,k}^{\mathrm{RQ,PWM}} := \frac{P_{n,k}^{\mathrm{RQ}} - 4Q_{n,k}^{\mathrm{RQ}}}{P_{n,k}^{\mathrm{RQ}} - 2Q_{n,k}^{\mathrm{RQ}}}. \tag{2.166}$$

It holds $\hat{\gamma}_{n,k}^{\mathrm{RQ,PWM}} = T_{PWM}\left(\hat{Q}_{\overline{\mathbf{x}},n,k}\right)$ and thus by Theorems 2.4.3 and 2.4.4 this estimator is consistent and asymptotic normal. Note also that for $k \in \mathbb{N}$ and $k/n \to 0$ the process $\hat{Q}_{\overline{\mathbf{x}},n,k}$ is a step function with $j(k,\mathbf{Y},\mathbf{X}) \geq k$ steps. Comparing estimator $\hat{\gamma}_{n,k}^{\mathrm{RQ,PWM}}$ with $\hat{\gamma}_{n,k}^{\mathrm{PWM}}$, i.e. classical estimate applied on errors of the model, we get that $\hat{\gamma}_{n,k}^{\mathrm{RQ,PWM}}$ is calculated from different number of values $j(k,\mathbf{Y},\mathbf{X})$, than estimator $\hat{\gamma}_{n,k}^{\mathrm{PWM}}$. The beauty of our functional representation lies in the fact, that it is universal for the both cases. Indeed, asymptotic properties of $\hat{\gamma}_{n,k}^{\mathrm{RQ,PWM}}$ and $\hat{\gamma}_{n,k}^{\mathrm{PWM}}$ are similar except of the asymptotic variance. In a similar way we can obtain a consistency and an asymptotic normality of the functionals which are defined only in an implicit way, which is the case of ML-estimator generated by functionals (1.77) and (1.79).

Note also that the information about the regular variation of the tails can be concentrated just to the intercept, as was already indicated by the asymptotic properties of the largest regression quantile. Consider following reparametrization of the model (2.1):

$$\begin{aligned} \beta_1^* &= \beta_1 + \overline{x}_2\beta_2 + \ldots + \overline{x}_p\beta_p \\ \beta_i^* &= \beta_2, \qquad i = 2,\ldots,p, \\ &\text{and} \\ x_{1i}^\star &= 1 = x_{1i}, \qquad i = 1,\ldots,n \\ x_{ji}^\star &= x_{ji} - \frac{1}{n}\sum_{i=1}^n x_{ji}, \qquad i = 1,\ldots,n. \end{aligned} \tag{2.167}$$

Then the model (2.1) can be rewritten as

$$\mathbf{Y} = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{E}. \tag{2.168}$$

It is easy to see that if $\frac{1}{n}\mathbf{X}^\top\mathbf{X} \to \mathbf{D}$, where $\mathbf{D}$ is a positive definite matrix, then also $\frac{1}{n}\mathbf{X}^{*\top}\mathbf{X}^* \to \mathbf{D}^*$, where $\mathbf{D}^*$ differs from $\mathbf{D}$ only in the first row and column which equals to $(1,0,\ldots,0)^\top \in \mathbb{R}^p$, see Dodge and Jurečková (2000), pp. 128–129. It follows that under this reparametrization $\overline{\mathbf{x}}^{*\top}\left(\widehat{\boldsymbol{\beta}}_n^*(\alpha) - \boldsymbol{\beta}^*\right) = \beta_1^* = \beta_1 + \overline{x}_2\beta_2 + \ldots + \overline{x}_p\beta_p$ and it holds that for any two quantile regression estimates of the intercept

$$\hat{\beta}_{n,1}^*(\alpha_1) \leq \hat{\beta}_{n,1}^*(\alpha_2), \qquad 0 \leq \alpha_1 \leq \alpha_2 \leq 1. \tag{2.169}$$

As $\overline{\mathbf{x}}^* = (1,0,\ldots,0)^\top \in \mathbb{R}^p$ we obtain the consistency and asymptotic normality of estimators based on regression quantile estimates of intercept in model (2.168). The estimator (2.162) can be thus simplified to

$$\hat{\gamma}_{k,n}^{*\mathrm{RQ,Pick}} := \frac{1}{\log 2} \log \left(\frac{\hat{\boldsymbol{\beta}}^*_{n,1}\left(1 - \frac{k}{4n}\right) - \hat{\boldsymbol{\beta}}^*_{n,1}\left(1 - \frac{k}{2n}\right)}{\hat{\boldsymbol{\beta}}^*_{n,1}\left(1 - \frac{k}{2n}\right) - \hat{\boldsymbol{\beta}}^*_{n,1}\left(1 - \frac{k}{n}\right)} \right). \tag{2.170}$$

We can write in a similar fashion the other estimators such as (2.166) as well.

.

Chapter 3

Residuals and two-step regression quantiles

It is fairly complicate to establish a theory of "extreme" intercepts of regression quantiles. Moreover, one can ask, whether it is desirable at all. The regression quantiles for α near to the largest regression quantile, for quantiles $\alpha < \alpha_{\max}$, where

$$\alpha_{\max} := \min_{\alpha\in[0,1]} \left\{ \alpha \left| I[\widehat{\boldsymbol{\beta}}_n(\alpha) - \widehat{\boldsymbol{\beta}}_n(1) = 0] \right. \right\},$$

with $\widehat{\boldsymbol{\beta}}_n(1)$ being the largest regression quantile (2.16), are sensitive to few extremal data even for moderate sample sizes. As we mentione Smith (1994) alongside with Portnoy and Jurečková (2000) proved the consistency of the largest regression quantile, but due to the fact that for any $\alpha \in (0,1)$ at least $d+1$ data points lie on the line $\hat{\boldsymbol{\beta}}_n(\alpha)$, the largest regression quantiles usually remain far away from their theoretical counterparts. The question is whether this behaviour reflects more the properties of the design matrix $\mathbf{X}$ or the extremal behaviour of the underlying error distribution.

To reveal more about the distribution of the very high regression quantiles we shall discuss the properties of exceedances over given quantile regression threshold. The method has been already discussed in the literature, cf. Beirlant et al. (2004), pp. 226–229, but their approach has not been supported by theoretical asymptotical results. The simplest settings can be described as follows. Have a linear model as in (2.1) with the intercept β_1 (i.e. the first column of $\mathbf{X}_n$ is $\mathbf{1}$) and denote by $\underline{\mathbf{X}}$ the matrix $\mathbf{X}$ without its first column (i.e. $\underline{\mathbf{X}}$ is $n \times (p-1)$ matrix and we shall call it's rows as $\underline{\mathbf{x}}_i$). Consider the set of positive residuals calculated with respect to one specific (theoretical) regression quantile $\boldsymbol{\beta}(\alpha) = (\beta_1 + F^{-1}(\alpha), \beta_2, \cdots, \beta_p)$

$$\mathbf{R}(\alpha) := \{Y_i - \mathbf{x}_i^\top \boldsymbol{\beta}(\alpha); Y_i - \mathbf{x}_i^\top \boldsymbol{\beta}(\alpha) > 0\}$$

and their order statistics $R_{i:l} = R_{i:l}(\alpha) \in \mathbf{R}(\alpha)$, where $l = card(\mathbf{R}(\alpha))$ and $i = 1, \ldots, l$.

From the EVT point of view such residuals contain for suitable α whole information about the extremes of the errors E_i as the errors itself. A natural approach is to describe how much of the information about the tails is preserved in

$$\hat{\mathbf{R}}(\alpha) := \{Y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}(\alpha); Y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}(\alpha) > 0\}, \tag{3.1}$$

where $\hat{\boldsymbol{\beta}}(\alpha)$ is some empirical quantile regression estimator. Apart from the regression quantiles of Koenker and Basset (1978) there are also other possibilities how to estimate theoretical regression quantile $\boldsymbol{\beta}(\alpha)$. The approach sketched by (3.1) itself have very much in common with two-step regression quantiles introduced by Jurečková and Picek (2005). These quantiles are defined as R-estimates and their residuals and they are asymptotically equivalent to the regression quantiles of Koenker and Basset (1978). The two-step regression approach is worth of further examination. The following lines are reworked version of the joint work of the author with Jan Picek which appeared in Picek and Dienstbier (2010).

3.1 Extremes of two-step regression quantiles

The two-step regression quantiles have been introduced by Jurečková and Picek (2005) as an alternative of the α-regression quantiles of Koenker and Basset (1978). They are defined as follows:

Let $\hat{\boldsymbol{\beta}}_{nR}(\alpha)$ be an appropriate R-estimate of the slope parameter $\boldsymbol{\beta}$ and let $\tilde{\beta}_{n,1}$ denote $[n\alpha]$-order statistic of the residuals $Y_i - \underline{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_{nR}(\alpha)$, then the vector

$$\tilde{\boldsymbol{\beta}}_n(\alpha) := \left(\tilde{\beta}_{n,1}, \widehat{\boldsymbol{\beta}}_{nR}(\alpha)\right)^\top \tag{3.2}$$

is called *two-step α-regression quantile.*

The initial R-estimator of the slope parameters is constructed as an inverse of the rank test statistic calculated in the Hodges-Lehmann manner, see Hodges and Lehmann (1963): Denote Denote $R_{ni}(\mathbf{Y} - \underline{\mathbf{X}}\mathbf{b})$ the rank of $Y_i - \underline{\mathbf{x}}_i^\top \mathbf{b}$ among $(Y_1 - \underline{\mathbf{x}}_1^\top \mathbf{b}, \ldots, Y_n - \underline{\mathbf{x}}_n^\top \mathbf{b}), \mathbf{b} \in \mathbb{R}^{p-1}$, $i = 1, \ldots, n$. Note that $R_{ni}(\mathbf{Y} - \underline{\mathbf{X}}\mathbf{b})$ is also the rank of $Y_i - b_0 - \underline{\mathbf{x}}_i^\top \mathbf{b}$ among $(Y_1 - b_1(\alpha) - \underline{\mathbf{x}}_1^\top \mathbf{b}, \ldots, Y_n - b_1(\alpha) - \underline{\mathbf{x}}_n^\top \mathbf{b})$ for any $\alpha \in (0,1)$ because the ranks are translation invariant. Hence the initial estimation of the slope parameter is invariant to any shift of the data. Consider the vector $\mathbf{S}_n(\mathbf{b}) = (S_{n,1}(\mathbf{b}), \ldots, S_{n,p-1}(\mathbf{b}))^\top$ of the linear rank statistics, where

$$S_{n,j}(\mathbf{b}) = \sum_{i=1}^n x_{i,j+1} \psi_\alpha \left(\frac{R_{ni}(\mathbf{Y} - \underline{\mathbf{X}}\mathbf{b})}{n+1} \right), \quad \mathbf{b} \in \mathbb{R}^{p-1}, \quad j = 1, \ldots, p-1. \tag{3.3}$$

and $\psi_\alpha(x) = \alpha - I[x < 0], x \in \mathbb{R}$ as in (2.46). Then the estimator $\hat{\boldsymbol{\beta}}_{nR}$ is defined as

$$\widehat{\boldsymbol{\beta}}_{nR} = \operatorname{argmin}_{\mathbf{b} \in \mathbb{R}^{p-1}} \|\mathbf{S}_n(\mathbf{b})\|_1, \tag{3.4}$$

where $\|\mathbf{S}\|_1 = \sum_{j=1}^{p-1} |S_j|$ is the L_1 norm of $\mathbf{S}$, see Jurečková (1971); or

$$\widehat{\boldsymbol{\beta}}_{nR} = \text{argmin}_{\mathbf{b}\in\mathbb{R}^{p-1}} \mathcal{D}_n(\mathbf{b}), \tag{3.5}$$

where

$$\mathcal{D}_n(\mathbf{b}) = \sum_{i=1}^{n} (Y_i - \underline{\mathbf{x}}_i^\top \mathbf{b}) \psi_\alpha \left(\frac{R_{ni}(\mathbf{Y} - \underline{\mathbf{X}}\mathbf{b})}{n+1} \right) \tag{3.6}$$

is the Jaeckel's measure of rank dispersion, see Jaeckel (1972).

The estimate $\widehat{\boldsymbol{\beta}}_{nR}$ estimates only the slope parameters and the computation is invariant to the size of the intercept. Solutions (3.4) and (3.5) are generally not unique, nevertheless the asymptotic representations apply to any of such solution; e.g. one can take the center of gravity of the set of all solutions $\mathcal{B}_n$ or the expecatation of a random vector uniformly distributed over $\mathcal{B}_n$ to obtain suitable single value.

The asymptotic of the two-step regression quantiles coincides with the asymptotics of the R-estimators. Assume that the design matrix $\mathbf{X}$ fulfills the conditions (X.1–3) and the distribution function F of errors in model (2.1) fulfills the condition

(F.F1) F has a continuous density f that is positive on the support of F and has finite Fisher's information, i.e. $0 < \int \left(\frac{f'(x)}{f(x)} \right)^2 dF(x) < \infty$.

Under these conditions the R-estimator (3.4) and (3.5) admits the following asymptotic representation,

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{nR} - \underline{\boldsymbol{\beta}}) = \\ n^{-\frac{1}{2}} (f(F^{-1}(\alpha))^{-1} \underline{\mathbf{D}}^{-1} \sum_{i=1}^{n} \underline{\mathbf{x}}_i \left(\alpha - I[E_i < F^{-1}(\alpha)] \right) + o_p(n^{-1/4}), \tag{3.7}$$

where again $\underline{\mathbf{D}} = \lim_{n\to\infty} \underline{\mathbf{D}}_{\mathbf{n}}$, $\underline{\mathbf{D}}_n = \frac{1}{n}\sum_{i=1}^n \underline{\mathbf{x}}_i \underline{\mathbf{x}}_i^\top$, and $\underline{\boldsymbol{\beta}} = (\beta_2, \ldots, \beta_p)$ is the slope component of $\boldsymbol{\beta}$, for detailed proof see Jurečková and Sen (1996). Using this relation Jurečková and Picek (2005) showed that also the two-step regression quantiles defined in (3.2) are asymptotically equivalent to the regression quantiles of Koenker and Basset (1978), i.e.

$$n^{1/2} \left\| \hat{\boldsymbol{\beta}}_n(\alpha) - \tilde{\boldsymbol{\beta}}_n(\alpha) \right\| \xrightarrow[n\to\infty]{\text{P}} 0, \qquad \text{as } n \to \infty, \tag{3.8}$$

see Jurečková and Picek (2005), Corollary 2.1. Not only asymptotically, two-step regression quantiles lies close to regression quantiles of Koenker and Basset (1978) also numerically as indicate simulation studies. While they are more difficult to calculate and their definition is somehow tricky, they are also more simple to deal with. The main advantage is that their number is identical with that of order statistics of the errors. This fact enables to establish a connection between the multivariate distribution of two-step regression quantiles and the univariate distribution of the errors in model (2.1). The fact plays an important role in establishing the extreme value theory on two-step regression quantiles.

We have previously seen that the population counterpart of α-regression quantile is vector $\boldsymbol{\beta}(\alpha) = (\beta_1 + F^{-1}(\alpha), \beta_2, \ldots, \beta_p)^\top$. The difference between empirical regression quantile and its theoretical population counterpart is $\mathcal{O}_P(n^{-1/2}(\log\log n)^{1/2})$ under general conditions on $\mathbf{X}$ and F, cf. Theorem 2.3.1 and the previous discussion. The coincidence of two-step regression quantiles is stressed by the fact, that the order statistics of residuals can be written in terms of regression quantiles as

$$E_{[n\alpha]} = \arg\min_b \left\{ \sum_{i=1}^n \rho_\alpha \left(Y_i - b - \underline{\mathbf{x}}_i^\top \underline{\boldsymbol{\beta}} \right), \, b \in \mathbb{R} \right\}.$$

While we usually do not know the value of $\underline{\boldsymbol{\beta}}$, we can replace it by an R-estimate, which (as we have seen) is ivariant to the exact value of the intercept containing the information about quantiles of the error distribution. Under the condition that $\beta_1 = 0$ in (2.1) the intercept of α-two-step quantile can be thus seen as an estimate of the $[n\alpha]$-order staitistic of errors.

$$\hat{E}_{[n\alpha]} = \hat{E}_{[n\alpha]}(\widehat{\boldsymbol{\beta}}_{nR}(\alpha)) = \arg\min_b \left\{ \sum_{i=1}^n \rho_\alpha \left(Y_i - b - \underline{\mathbf{x}}_i^\top \widehat{\boldsymbol{\beta}}_{nR}(\alpha) \right), \, b \in \mathbb{R} \right\}. \tag{3.9}$$

Nevertheless, there is a considerable question about the tail behaviour of such estimates as we do not know much about the asymptotic properties of extreme R-estimates. For our applications it is enough to develop a less ambitious theory dealing with extreme of two-step regression quantiles. In fact we shall not diverge from the previous developments of the theory.

In Jurečková and Picek (2005) the authors considered *the extreme two-step quantile* $\hat{E}_{n:n}$, which they define as the maximum of the residuals

$$\hat{E}_{n:n} = \max\{Y_1 - \underline{\mathbf{x}}_1^\top \widehat{\boldsymbol{\beta}}_{nR}, \ldots, Y_n - \underline{\mathbf{x}}_n^\top \widehat{\boldsymbol{\beta}}_{nR}\} \tag{3.10}$$

calculated with respect to an appropriate R-estimate $\widehat{\boldsymbol{\beta}}_{nR}$ of $\underline{\boldsymbol{\beta}}$. Similarly as in the case of regression quantiles if one admits the assumption controlling maximum over $\|\mathbf{x}_i\|$, $i = 1, \ldots, n$ as in (X.4), $\hat{E}_{n:n}$ is a consistent estimate of $E_{n:n} + \beta_1$ and

$$|\hat{E}_{n:n} - E_{n:n} - \beta_1| = \mathcal{O}_p(n^{-\delta}) \quad \text{as } \; n \to \infty, \; 0 < \delta < \tfrac{1}{2} \tag{3.11}$$

Let $\widehat{\boldsymbol{\beta}}_{nR}^+$ be the initial R-estimate generated by the score function $\varphi_{1-\frac{1}{n}}(u) = I[u \geq 1 - \frac{1}{n}] - \frac{1}{n}$, $0 < u < 1$. In this case the Jaeckel measure of the rank dispersion (3.6) takes the form

$$\max_{1 \leq i \leq n} \{Y_i - \underline{\mathbf{x}}_i^\top \mathbf{b}\} - \bar{Y}_n, \tag{3.12}$$

where $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Hence,

$$\widehat{\boldsymbol{\beta}}_{nR}^{+} = \min_{\mathbf{b}\in\mathbb{R}^{p-1}} \sum_{i=1}^{n} \left(Y_i - \underline{\mathbf{x}}_i^\top \mathbf{b}\right)^+ . \tag{3.13}$$

Then we can define *the largest two-step regression quantile* as $\left(\hat{E}_{n:n}, \widehat{\boldsymbol{\beta}}_{nR}^{+}\right)$. For this estimate it holds that $\hat{E}_{n:n} + \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_{nR}^{+} \geq Y_i, i = 1, \ldots, n$, while for some i_0 the inequality reduces to equality. Hence the largest two-step regression quantile coincides with the largest regression quantile considered by Portnoy and Jurečková (2000) defined as

$$\widehat{\boldsymbol{\beta}}(1) := \arg\min_{\mathbf{b}\in\mathbb{R}^p} \left\{ \sum_{i=1}^{n} \mathbf{x}_i^\top \mathbf{b} \,\middle|\, Y_i \leq \mathbf{x}_i^\top \mathbf{b}, \quad i = 1, \ldots, n \right\}$$

This estimator can be obtained from the usual definition of regression quantile by letting $\alpha \to 1$. The properties of the largest regression quantile have been already studied from the extremal point of view in Jurečková (2007), where the author derives a simple test on domain of attraction. The two-step regression quantile coincides exactly with the extreme regression quantile considered in Portnoy and Jurečková (2000). However, as the methods of EVT are often based not only on the maxima but also on other higher empirical quantiles, we shall proceed beyond largest two-step regression quantile. In the following we consider the properties of residuals appertaining to any single R-estimate discussed above. Denote $\left\{\hat{E}_1, \ldots, \hat{E}_n\right\}$ the set of residuals $\left\{Y_1 - \underline{\mathbf{x}}_1^\top(\widehat{\boldsymbol{\beta}}_{nR} - \underline{\boldsymbol{\beta}}), \ldots, Y_n - \underline{\mathbf{x}}_n^\top(\widehat{\boldsymbol{\beta}}_{nR} - \underline{\boldsymbol{\beta}})\right\}$. The following lemma shows that k-th ordered residual $\hat{E}_{k:n}$ is an appropriate estimate of $E_{k:n}$.

Lemma 3.1.1. *Let $\widehat{\boldsymbol{\beta}}_{nR}$ be an R-estimate of $\underline{\boldsymbol{\beta}}$, generated by a fixed nondecreasing and integrable score function $\varphi : (0,1) \mapsto \mathbb{R}$, independent of n, as in (3.3) and (3.4). Assume the conditions (A1) – (A3) and*

$$\max_{1\leq i\leq n} \|\underline{\mathbf{x}}_i\| = \mathcal{O}\left(n^{\frac{1}{2}-\delta}\right) \quad as \quad n \to \infty, \ 0 < \delta < \tfrac{1}{2}. \tag{3.14}$$

then

$$\sup_{1\leq k\leq n} \left|\hat{E}_{k:n} - E_{k:n} - \beta_1\right| = \mathcal{O}_P(n^{-\delta}), \qquad as\ n \to \infty \tag{3.15}$$

Proof. Let $D_1, \ldots, D_n$ denote the antiranks of $E_1, \ldots, E_n$, i.e. the indices satisfying $E_{i:n} = E_{D_i}$, $i = 1, \ldots, n$. Moreover for an R-estimate $\widehat{\boldsymbol{\beta}}_{nR}$ of the slope components of $\boldsymbol{\beta}$ and $n \in \mathbb{N}$

$$u_n := u_n(\widehat{\boldsymbol{\beta}}_{nR}) := \max_{i=1,\ldots,n} |\underline{\mathbf{x}}_i^\top(\widehat{\boldsymbol{\beta}}_{nR} - \underline{\boldsymbol{\beta}})|.$$

From the asymptotic representation of $\widehat{\boldsymbol{\beta}}_{nR}$ (3.7) and (3.14) we get $u_n = \mathcal{O}_P(n^{-\delta})$ as $n \to \infty$.

Notice that $\hat{E}_{1:n} \leq E_{1:n} + \beta_1 + u_n$, because the opposite case $\hat{E}_{1:n} > E_{1:n} + \beta_1 + u_n$ implies

$$\hat{E}_{\underline{D}_1} = E_{1:n} + \beta_1 + \mathbf{x}_{\underline{D}_1}(\underline{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{nR}) \leq E_{1:n} + \beta_1 + u_n < \hat{E}_{1:n}.$$

Hence, $\hat{E}_{1:n}$ is the smallest observation among $\left\{\hat{E}_i, i = 1, \dots, n\right\}$, therefore it cannot be greater than $\hat{E}_{\underline{D}_1}$.

Similarly, $\hat{E}_{2:n} \leq E_{2:n} + \beta_1 + u_n$ because $\hat{E}_{2:n} > E_{2:n} + \beta_1 + u_n$ leads to

$$\hat{E}_{\underline{D}_2} = E_{2:n} + \beta_1 + \mathbf{x}_{\underline{D}_2}(\underline{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{nR}) \leq E_{2:n} + \beta_1 + u_n < \hat{E}_{2:n}$$

and

$$\hat{E}_{\underline{D}_1} = E_{1:n} + \beta_1 + \mathbf{x}_{\underline{D}_1}(\underline{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{nR}) \leq E_{2:n} + \beta_1 + u_n < \hat{E}_{2:n}.$$

If we proceed analogously, we get

$$\hat{E}_{i,n} \leq E_{i,n} + \beta_1 + u_n, \qquad i = 1, \dots, n. \tag{3.16}$$

On the other hand, it holds for the highest two-step ordered residual $\hat{E}_{n:n} \geq E_{n:n} + \beta_1 - u_n$, because $\hat{E}_{n:n} < E_{n:n} + \beta_1 - u_n$ implies

$$\hat{E}_{\underline{D}_n} = E_{n:n} + \beta_1 + \mathbf{x}_{\underline{D}_n}(\underline{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{nR}) \geq E_{n:n} + \beta_1 - u_n > \hat{E}_{n:n}.$$

We get by the similar arguments as in (3.16)

$$\hat{E}_{i,n} \geq E_{i,n} + \beta_1 - u_n, \qquad i = 1, \dots, n. \tag{3.17}$$

Finally, $u_n = \mathcal{O}_p(n^{-\delta})$ together with (3.17) and (3.16) imply (3.15). $\square$

In the introductory chapter we have seen that many EVT estimators can be written as smooth functionals $T(Q_n)$ of the *empirical tail quantile function*. If we would have been able to observe the errors of (2.1) directly, the appropriate tail quantile function would take the form

$$Q_n(t) := F_n^{-1}\left(1 - \frac{k_n}{n} t\right) = E_{n-[k_n t]:n}, \quad t \in [0,1],$$

In the next step, we replace the unobservable errors with the residuals $\hat{E}_1, \dots, \hat{E}_n$ and establish EVT for the tail quantile function of the residuals. Let any $k \in \mathbb{N}$ be such that $\hat{E}_{k:n} > 0$. We define the *tail quantile function of the residuals* as

$$\hat{Q}_{n,k}(t) := \hat{E}_{n-[kt]:n}, \qquad t \in [0,1]. \tag{3.18}$$

By Lemma (3.1.1) $\hat{Q}_{n,k}$ is the consistent estimate of the empirical tail function of the errors $Q_{n,k}(t) = E_{n-[kt]:n}$ uniformly in $t \in [0,1]$. Hence following Drees (1998*b*), we can provide an approximation of $\hat{Q}_{n,k}$ for the intermediate sequences of $k(n)$. Suppose that F in (2.1) fulfills (EVT.2) with a, A, and K as in (1.25), i.e. there is some is some $\rho \le 0$ that K takes the form (1.26), (1.27), or (1.28) respectively to γ and ρ. Define again seminorm $\|\cdot\|_{\gamma,h}$ on metric space $\mathcal{D}_{\gamma,h}$ for any $h \in \mathcal{H}$ defined as in (1.55), (1.56), and (1.54). Then Theorem 1.5.1 can be reformulated to establish an approximation of $\hat{Q}_{n,k}$ defined in (3.18).

Theorem 3.1.1. *Suppose that the distribution function F of errors in (2.1) satisfies (EVT.2) for some $\gamma \in \mathbb{R}$ and $\rho \le 0$. Suppose that the assumptions of Lemma 3.1.1 are fulfilled. Then we can define a sequence of Wiener processes $\{W_n(t)\}_{t\ge 0}$ such that for suitable chosen functions A, K, and a as in (1.25), metric seminorm $\|\cdot\|_{\gamma,h}$ as in (1.55) and each $\varepsilon > 0$,*

$$\left\| \frac{\hat{Q}_{n,k}(t) - F^{-1}\left(1-\frac{k}{n}\right) - \beta_1}{a\left(\frac{k}{n}\right)} - \left(z_\gamma(t) - k^{-1/2} t^{-(\gamma+1)} W_n(t) + A\left(\frac{k}{n}\right) K(t) \right) \right\|_{\gamma,h} = o_P\left(k^{-1/2} + |A(k/n)|\right) \quad (3.19)$$

$n \to \infty$, provided that $k = k(n) \to \infty$, $k/n \to 0$ and $\sqrt{k}A(k/n) = \mathcal{O}(1)$

Proof. Immediately follows from (3.15) and the approximation of the tail quantile function derived in Theorem 2.1 of Drees (1998*b*). □

If it holds only the domain of attraction condition (EVT.1) an analogue of Theorem 1.5.2 can be formulated.

Theorem 3.1.2. *Suppose that it holds (EVT.1). Then we can define Wiener processes $\{W_n(t)\}_{t\ge 0}$ such that almost surely for all $h \in \mathcal{H}$ as in (1.54) and $\varepsilon > 0$*

$$\left\| \frac{\hat{Q}_{n,k}(t) - F^{-1}\left(1-\frac{k}{n}\right) - \beta_1}{a\left(\frac{k}{n}\right)} - \left(z_\gamma(t) - k^{-1/2} t^{-(\gamma+1)} W_n(t) \right) \right\|_{\gamma,h} = o_P\left(k^{-1/2}\right) + \mathcal{O}\left(\sup_{x \in (0,1+\varepsilon)} x^{\gamma+\frac{1}{2}} \left| R\left(\frac{k}{n}, x\right) \right| \right) \quad (3.20)$$

where $k = k(n)$ is an intermediate sequence $k \to \infty$, $k/n \to 0$, $n \to \infty$.

Proof. See Theorem 2.1 in Drees (1998*b*). □

We have already shown in the first chapter that various estimators of γ can be in the i.i.d. case written as functionals of the empirical tail quantile function, i.e. $\hat{\gamma}_{n,k} = T(\hat{Q}_{n,k})$. The asymptotic properties of these estimators are under suitable conditions on T given by the Theorem 1.5.1, cf. Theorem 1.5.3 and Drees (1998*b*). Theorems 3.1.1 and 3.1.2 enable to define the estimates of γ as functionals of $\hat{Q}_{n,k}$. While $Q_{n,k}$ is not observable in the regression case, it can be estimated using $\hat{Q}_{n,k}$. Again we shall establish the convergence using weighted supremum seminorm $|\cdot|_{\gamma,h}$ on $\mathcal{D}_{\gamma,h}$ with $h \in \mathcal{H}$ for any fixed $\gamma \in \mathbb{R}$. Moreover have $C_{\gamma,h} := \left\{z \in \mathcal{D}_{\gamma,h} | z_{|(0,1]} \in C(0,1]\right\}$ be a subset of continuous functions on $(0,1]$ of $\mathcal{D}_{\gamma,h}$. The following theorem shows consistence and asymptotic normality of a broad class of functionals $\hat{Q}_{n,k}$.

Theorem 3.1.3. *Suppose that the assumptions of Lemma 3.1.1 are fulfilled. If for the distribution function F of errors in model (2.1 holds $F \in MDA(G_\gamma)$, $k = k_n$ is an intermediate sequence, T satisfies conditions (T.1)–(T.3), and in addition $T_{|\mathcal{D}_{\gamma,h}}$ is continuous in z_γ, then*

$$T\left(\hat{Q}_{k,n}\right) \xrightarrow[n\to\infty]{\mathrm{P}} \gamma. \tag{3.21}$$

Proof. Immediately follows from Lemma 3.1.1 and Theorem 1.5.3 by Theorem 3.1.1 and Theorem 3.1.2. □

Similarly as in the location i.i.d. model an asymptotical normality of any estimator $T(\hat{Q}_{k,n})$ follows from the second order condition under Hadamard differentiability of T, which was introduced in the condition (T.4) on page 28.

Theorem 3.1.4. *Assume that $T : span(\mathcal{D}_{\gamma,h}, 1) \to \mathbb{R}$ satisfies conditions (T.1)–(T.4). Let $\sigma_{T,\gamma}$ be as in (1.62) and $\mu_{T,\gamma,\rho}$ as in (1.63). Then under the condition (EVT.2) on distribution function F of errors in model (2.1) and $\lim_{n\to\infty} kA(k/n) = \lambda \in [0,\infty]$ follows*

(i) $\lambda \in (0,\infty)$
$$k^{1/2}\left(T(\hat{Q}_{k,n}) - \gamma\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(\lambda\mu_{T,\gamma,\rho}, \sigma^2_{T,\gamma}). \tag{3.22}$$

(ii) $\lambda = \infty$
$$k^{1/2}\left(T(\hat{Q}_{k,n}) - \gamma\right) \xrightarrow[n\to\infty]{\mathrm{P}} \mu_{T,\gamma,\rho} \tag{3.23}$$

Moreover, if only (EVT.1) holds and it holds (1.66), i.e. $\sup_{x\in(0,1+\varepsilon]} x^{\gamma+1/2} |R(k_n/n, x)| = o\left(k_n^{-1/2}\right)$ *for some* $\varepsilon > 0$*, then*

$$k^{1/2}\left(T(\hat{Q}_{k,n}) - \gamma\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2_{T,\gamma}). \tag{3.24}$$

Proof. Immediately follows from Lemma 3.1.1 and Theorem 1.5.4 by Theorem 3.1.1 and Theorem 3.1.2. □

Theorems 3.1.3 and 3.1.4 have profound practical implications – virtually any estimator, i.e. anyone which can be written in the form of a smooth tail quantile function functional, gives consistent and asymptotically normal results when applied on extreme two-step regression quantiles. The class of estimators covers practically all frequently used estimators as discussed in section 1.5. While we consider only the location and scale invariant estimators of γ, the nuisance parameter of real intercept β_1 does not play any role in the estimation. Nevertheless a suitable standardization, e.g. through subtracting the regression median, allows to generalize to the case of estimators which are only scale invariant similarly as in Drees (1998a). However, any such extension cannot be recommended from the practical point of view. As we shall see in the next section, simulation results show that such estimators have a large bias. While Hill's estimator is consistent for a suitable $k = k(n)$ and large $n \to \infty$ any shift in location can severely affect the estimation and any error in the estimation of β_1 can multiply the bias of the estimation of γ.

3.2 Residuals of regression quantiles

We defined extreme two-step regression quantiles as the residuals over a given R-estimate. Nevertheless, the assertions in the previous section hold for any consistent estimator of the slope. The proof of Lemma 3.1.1 is quite universal and can be immediately rewritten in terms of residuals of regression quantiles if $\widehat{\boldsymbol{\beta}}_{nR}$ is replaced by $\hat{\boldsymbol{\beta}}_n(\alpha)$, $\alpha \in [0,1]$. Consider the set of nonnegative residuals over this regression quantile threshold

$$\tilde{\mathbf{E}}(\alpha) = \left\{ \left(Y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_n(\alpha) \right)^+, i = 1, \ldots, n \right\}, \tag{3.25}$$

where $x^+ := \max(0, x)$. We shall denote the order of lowest positive residual as $k = k(\alpha) \in \{1, \ldots, n\}$ for any $n \in \mathbb{N}$, i.e. it holds

$$\tilde{E}_{n-k-1,n} = 0 \qquad \text{and} \qquad \tilde{E}_{n-k,n} > 0.$$

In such case we can estimate the tail quantile function as

$$\tilde{Q}_{n,k}(t) := \tilde{E}_{n-[kt]:n} = \tilde{E}_{n-[k(\alpha)t]:n}, \qquad t \in [0,1]. \tag{3.26}$$

For suitably chosen $\alpha = \alpha(n) \in [0,1]$ we get $k(\alpha)$ as an intermediate sequence. Accordingly we can base the estimation of γ appertaining to distribution function F of the errors in model (2.1) on $k(\alpha)$ positive residuals $E_{n-k,n}, \ldots, E_{n,n}$ or, using functional notation, any such estimator can be written as $T(\tilde{Q}_{k(\alpha),n})$, for some functional functional fulfilling (T.1)–(T.3).

Another possibility is to choose a fixed regression quantile threshold $\hat{\boldsymbol{\beta}}_n(\tau)$ for some $\tau \in [0,1]$. In this case one gets $R_\tau \leq n$ positive residuals $\breve{E}_i = Y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_n(\tau)$, $i \in 1, \ldots, R_\tau$, where R_τ tends to infinity with $n \to \infty$. The appropriate estimate of the tail quantile function shall be the k largest residuals, i.e.

$$\breve{Q}_{n,k}(t) := \breve{E}_{n-[k_n t]:n}, \qquad t \in [0,1]. \tag{3.27}$$

for a suitable intermediate sequence $k = k_n$. Again the consistency of the estimate $\breve{Q}_{n,k}$ immediately follows from an analogue of Lemmma 3.1.1.

In the situation of model 2.1 with i.i.d. errors it does not matter if the estimate $\hat{Q}_{n,k}$ or $\breve{Q}_{n,k}$ is used. However in real data analysis a suitable selection of the regression quantile threshold can filter out irregularities among the distributions of error $E_i,\ i = 1, \dots, n$ occurring bellow high quantiles.

The approach has been already discussed in the literature, c.f. Beirlant et al. (2004), pp. 226–229 and also Kyselý et al. (2010) or Northrop and Jonathan (2011), but the asymptotic properties of such estimators have not been considered. The methodological approach based on functionals of the estimated tail quantile function allows to develop an asymptotic theory of such estimators in a simple case of linear model (2.1) with i.i.d. errors. The consistency and asymptotical normality of the estimators based on residuals of the regression quantiles immediately follows from the theory explained in the previous pages. An extension of the uniform approximation of the regression quantile process as it was introduced by Theorems 2.3.1 and 2.3.2 on the other hand allows to admit intermediate sequences of $\alpha_n = k_n/n$ in (3.25).

While these methods have proven to provide reasonable solutions in the real data modelling, considerable theoretical problems remain unresolved. The most important is the behaviour of the tails of regression quantile process under various forms of a dependency in the errors of model (2.1). The property of the regression quantile process has been studied in the case of long range dependence by Koul and Mukherjee (1994), which represent a generalization of Gutenbrunner and Jurečková (1992).

.

Chapter 4

Real data and simulations

In this chapter we shall discuss the methods of an estimation of γ developed in the previous chapters from the computational point of view. We shall again work in the linear model setting (2.1) and our interest shall be to estimate the extremal properties of errors $E_1, \dots, E_n$ of the model. As it is difficult to test the efficiency of the inference procedures of EVT on the real data – typically the datasets consists of only limited number of observations and/or it is difficult to obtain more "rare" observations of the given random sample – the simulation results are very important decisive tool to evaluate the performance of the inferential methods.

4.1 Simulations

4.1.1 Intercepts of regression quantiles

We tested the estimators of γ on simple linear models with a different distribution functions of error. There have been generated 1000 observations from the selected models.

(a) $Y = 20 - \frac{1}{10}X_{i,j} + E_{i,j}, \qquad X_{i,j} = i, \quad i = 1, \dots, 20, \quad j = 1, \dots, 50$

(b) $Y = 1 + \frac{1}{2}X_i + E_i, \qquad X_i \sim U(0,5), \quad i = 1, \dots, 1000,$

The first model is with a fixed covariate matrix and the second is with a random covariate matrix. Both covariate matrix fulfills conditions (X.1)–(X.2) and (X.5)–(X.6) introduced in section 2.3.

Each of the models was generated with a two different distributions of errors:

(i) Burrleigh distribution Burr$(1, 2, 1)$

(ii) Fréchet distribution with a shape parameter $\alpha = 3$

$$F(x) = \exp\left(-x^{-3}\right), \quad x > 0.$$

We reparametrized the models using (2.167) to get an alternate expression of (a) and (b) in the form

(a) $Y_{ij} = 20 - \frac{1}{10}\left(X_{i,j} - \frac{21}{2}\right) + E_{ij}, \qquad i = 1,\dots,20, \quad j = 1,\dots,50$

(b) $Y_i = 1 + \frac{1}{2}\left(X_i - \frac{1}{1000}\sum_{i=1}^{1000} X_i\right) + E_i, \qquad i = 1,\dots,1000.$

It follows that $\overline{\mathbf{x}}^\top = (1,0) \in \mathbb{R}^2$ in the both cases and hence due to the theorems of the previous chapter we can base our estimation on the intercepts of regression quantiles, which form a non-increasing process $q^*_{n,1,k}(t) := \hat{\beta}_{n,1}(1 - \frac{tk}{n}) \in \mathbb{R}^1$ for $t \in [0,1]$ for any k being an intermediate sequence (note that we omit the asterisk notation $\hat{\beta}_{n,1}$ we introduced in the section 2.4 for the purpose of this chapter as we consider her only the reparametrized model).

As for each n, $\mathbf{Y}_n = Y_1,\dots,Y_n$ and $\mathbf{X}_{n\times p}$ there exists $m \in \mathbb{N}$ unique solutions of the minimization problem (2.3) and $\hat{\beta}_{n,1}(\tau)$ forms for $\tau \in [0,1]$ a step function with m steps. As we have already noted previously m depends on the exact numerical form of the matrix $\mathbf{X}$ and vector $\mathbf{Y}$, see Koenker (2005), pp. 34–38. For the sake of the notation we assign each step the lowest probability level $\{\tau_j\}_{j\in\{1,\dots,m\}}$ such that $\hat{\beta}_{n,1}(\tau_i)$ equals to the i-th step. More precisely this means $0 = \tau_1 < \tau_2 < \dots < \tau_m \le 1$ and for each $i \in \{1,\dots,m\}$ it holds $\hat{\beta}_{n,1}(\alpha) = \hat{\beta}_{n,1}(\tau_i)$, $i = 1,\dots,n$ for each α in some right neighborhood of τ_i and $\hat{\beta}_{n,1}(\alpha) = \hat{\beta}_{n,1}(\tau_{i-1}) < \hat{\beta}_{n,1}(\tau_i)$, $i = 2,\cdots,n$ for α in an arbitrary left neighborhood of τ_i.

We calculated our estimators for different $k = 1,\dots,l \in \mathbb{N}$ such that $k = \tau_{m-k}$, hence they are based on the k largest unique solutions of the process of regression quantiles intercept. This can be done due to our construction of the estimators, it is $T(q^*_{n,1,k_1}) = T(q^*_{n,1,k_2})$, for any $\tau_i \le k_1/n \le k_2/n\tau_{i+1}$, $i = 1,\dots,n$.

We used the following estimators, which can be deduced by applying functionals (1.77) and (1.79), (1.76), (1.52) on the realizations of $q^*_{n,1,k}$:

(1) Maximum likelihood estimator (ML) $\hat{\gamma}^{\mathrm{RQ,ML}}_{m,k}$ (*ML-estimator*) of the extreme value index based on the k largest unique estimates of $\hat{\beta}_{n,1}(\tau)$, $\quad \tau \in (0,1)$, i.e. the estimator fits generalized Pareto distribution (GPD) on the exceedances of $\{\hat{\beta}_{n,1}(\tau_j)\}_{j=m-k,\dots,m}$ over $\hat{\beta}_{n,1}(\tau_{m-k-1})$. To calculate ML-estimate we used `evir` package from `R`.

(2) Probability weighted moments estimator (PWM)

$$\hat{\gamma}^{\mathrm{RQ,PWM}}_{m,k} = \frac{\frac{1}{k}\sum_{j=1}^{k}\left(4\frac{j}{k+1} - 3\right)\hat{\beta}_{n,1}(\tau_{m-i+1})}{\frac{1}{k}\sum_{j=1}^{k}\left(2\frac{j}{k+1} - 1\right)\hat{\beta}_{n,1}(\tau_{m-i+1})}$$

(3) Pickands estimator

$$\hat{\gamma}^{\mathrm{RQ,P}}_{m,k} = \frac{1}{\log 2}\log\left(\frac{\hat{\beta}_{n,1}(\tau_{m-[k/4]}) - \hat{\beta}_{n,1}(\tau_{m-[k/2]})}{\hat{\beta}_{n,1}(\tau_{m-[k/2]}) - \hat{\beta}_{n,1}(\tau_{m-k})}\right)$$

For the sake of comparison we also observed the properties of estimators, which are not scale invariant. We used functional generating Hill's estimator (1.51) to obtain:

(4) Hill's estimator

$$\hat{\gamma}_{n,k}^{H_Q} = \frac{1}{k}\sum_{i=0}^{k}\Big(\log\hat{\beta}_{n,1}(\tau_{m-i}) - \log\hat{\beta}_{n,1}(\tau_{m-k})\Big)$$

(4) An attempt to fix the problem with intercept part of the estimation by subtracting the median of intercepts and plug the corrected data to Hill's estimator

$$\hat{\gamma}_{n,k}^{H_Q} = \frac{1}{k}\sum_{i=0}^{k}\Big(\log\hat{\hat{\beta}}_{n,1}(\tau_{m-i}) - \log\hat{\hat{\beta}}_{n,1}(\tau_{m-k})\Big) \tag{4.1}$$

where we used

$$\hat{\hat{\beta}}_{n,1}(\,\cdot\,) := \hat{\beta}_{n,1}(\,\cdot\,) - \hat{\beta}_{n,1}(1/2) \tag{4.2}$$

We randomly generated (a) and (b) 1000 times for both distributions of error and calculated estimates for $k = 1, \ldots, 600$. Afterwards we calculated the medians and the 5th, 10th, 90th, and 95th percentiles of the estimates for each $k = 1, \ldots, 600$. The results are shown on the accompanying figures (in each of them these medians and percentiles are plotted by the solid line, correct number of gamma is indicated by the straight line). For the sake of comparison we also estimated γ from the errors E_{ij}, $i = 1, \ldots, 20$, $j = 1, \ldots, 50$ (respectively E_i, $i = 1, \ldots, 1000$) we used to construct our models. We applied respective classical version of the estimators: ML-estimate defined in (1.44), PWM-estimate (1.45), Pickands estimate (1.39), and Hill's estimate (1.32). Again we calculated these estimates for each replication of the model and $k = 1, \ldots, 600$. We calculated the medians and the 5th, 10th, 90th, and 95th percentiles of the estimates for each $k = 1, \ldots, 600$ and included them in each estimator plot for comparison (dashed lines).

The figures demonstrate that our estimators are adequate for the regression situation (2.1). Their performance is (on average) only slightly worse than the performance of their univariate analogies applied on the errors of the model. This holds with the exception of Hill's estimate (Figure 4.3) as well as with the version of Hill's estimate which is applied on the data "corrected" by subtracting the median (Figure 4.4). This is clearly due to the fact that Hill's estimator is not location invariant. As our estimates are based on the intercept, whose theoretical counterpart is $\beta_1 + F^{-1}(\alpha)$ and not just $F^{-1}(\alpha)$, β_1 represent a nuisance factor in our estimations. While this does not matter in the case of location and scale invariant estimators, any such nuisance can enlarge the bias of the estimation in the case of estimators, which are only scale and not location invariant. This bias can be eliminated only in the case when n is large and k is sufficiently small with a rate indicated by our theoretical results, compare our Theorem 1.5.3 with Theorem 2.1 in Drees (1998*a*). However, this is at the cost of an increased variance of the estimation. Note that this behaviour is nicely demonstrated on Figure 4.3: the median of the estimations is close to the correct value of γ if k is small, but the variance is very large in that area. Note that this problem cannot be solved by any simple correction as is the subtraction of the medians (or means) – we do not subtract only the nuisance parameter β_1 but medians of E_i as well. This stands behind the bias on Figure 4.4. The Figure indicate that this is the reason of the bias of our estimate – the bias is same in the case estimation performed on i.i.d. errors, where we also subtracted the medians.

On the contrary to the case of Hill's estimator, the other figures demonstrate that the properties of the location and scale invariant estimators are very similar if the generating functional is applied on empirical tail quantile function as well as on extreme regression quantile process. It seems that the estimation in a linear model only slightly increases the variance of the estimation, if the model is correctly reparametrized in the sense of (2.167), while the bias of the estimation remains unchanged. This fact is supported by our theoretical results. Hence, the location and scale invariant estimators are clearly preferable to the estimators, which are only scale invariant.

4.1.2 Residuals of R-estimates

For the sake of comparison, we also calculated estimates introduced in Chapter 3. These estimators were already examined in Picek and Dienstbier (2010), where is included a small simulation study. The results indicate that the estimators have a nice performance, which is similar to the performance of the respective estimators on a simple i.i.d. sample.

Therefore we restricted ourselves to a direct comparison of these estimators with the estimators based on intercepts of regression quantiles. We used three estimators, equivalents to ML-estimator, PWM-estimator and Pickands estimator. They were calculated by applying functionals (1.77) and (1.79), (1.76), (1.52) on $\hat{Q}_{n,k}(t) = \hat{Q}_{n,k,\hat{\boldsymbol{\beta}}_{nR}}(t) := \hat{E}_{n-[kt]:n}$, with $\hat{E}_{n-k:n}$ being the k-th largest ordered residual $Y_i - \underline{x}_i^\top \hat{\boldsymbol{\beta}}_{nR}$, for R-estimate of the slope $\hat{\boldsymbol{\beta}}_{nR}$ defined in (3.5). For our estimation we used $\hat{\boldsymbol{\beta}}_{nR}(1/2)$ and compared it also with $\hat{\boldsymbol{\beta}}_{nR}(4/5)$. Note that these estimators differs in their form of Jaeckel's measure of dispersion $\mathcal{D}_n(\alpha)$ defined in (3.6). From the computational point of view, there was no considerable difference in both methods, except that $\hat{\boldsymbol{\beta}}_{nR}(1/2)$ is naturaly more stable for small data samples. The results presented are calculated with $\hat{\boldsymbol{\beta}}_{nR}(1/2)$.

Hence we obtained following estimators:

(1) Maximum likelihood estimator (ML) $\hat{\gamma}_{n,k}^{\mathrm{2RQ,ML}}$ (*ML-estimator*) of the extreme value index based on the $k+1$ largest residuals $\hat{E}_{n:n}, \ldots, \hat{E}_{n-k:n}$ from $\left\{Y_i - \underline{\mathbf{x}}_i^\top \hat{\boldsymbol{\beta}}_{nR}\right\}_{i=1,\ldots,n}$, i.e. the estimator fits generalized Pareto distribution (GPD) to these residuals. To calculate ML-estimate we used `evir` package from `R`.

(2) Probability weighted moments estimator (PWM)

$$\hat{\gamma}_{n,k}^{\mathrm{2RQ,PWM}} = \frac{\frac{1}{k}\sum_{j=1}^{k}\left(4\frac{j}{k+1}-3\right)\hat{E}_{n-i+1:n}}{\frac{1}{k}\sum_{j=1}^{k}\left(2\frac{j}{k+1}-1\right)\hat{E}_{n-i+1:n}}$$

(3) Pickands estimator

$$\hat{\gamma}_{n,k}^{\mathrm{2RQ,P}} = \frac{1}{\log 2}\log\left(\frac{\hat{E}_{n-[k/4]:n}-\hat{E}_{n-[k/2]:n}}{\hat{E}_{n-[k/2]:n}-\hat{E}_{n-k:n}}\right)$$

We applied these estimators on the data simulated from models (a) and (b) introduced in the previous subsection. Both models were generated with Fréchet and Burrleigh distribution of errors. The results in both cases were almost identical (similary as it is in the case of estimators based on regression quantile intercepts), therefore we present only the results for model (b) with a random covariate matrix. The results are on Figures 4.15 – 4.20. We did not applied an analogy of Hill's estimate, as the result is similar as can bee seen on Figure 4.3. Scale invariant estimators are severely biased by the approach similarly as by the estimation through intercepts of the regression quantiles. This was already noted in Picek and Dienstbier (2010).

We see that the results are indeed almost identical with the results calculated by the respective ML, PWM or Pickands estimator applied on the errors of our model (grey dashed lines). It seems that this is due to a fast convergence $\hat{\boldsymbol{\beta}}_{nR} \to \boldsymbol{\beta}$. From the comparison with the estimators based on intercepts of regression quantiles it appears that it is clearly preferable to use residuals of R-estimates rather than the intercepts of regression quantiles. This observation is supported by our theoretical results which states that the estimator $T(\hat{Q}_{n,k}(t))$ have the same asymptotic bias and variance as the estimator $T(E_{n-[kt]:n})$ for any functional T fulfilling (T.1)–(T.4). We were not able to prove this relation in case of regression quantiles and their intercepts, as we calculated only the upper bound for the variance of the functional. Nevertheless the approach based on intercepts of regression quantiles is more general and offers more possibilities to generalize the results if some of the assumptions on the model are not met (e.g. autoregressive dependence of the errors). The approach based on the residuals of R-estimate seems to be a less robust to these violations.

4.2 Condroz data

The Condroz data consists of the pH-values and the Calcium (Ca) contents in soil samples, collected in different districts of the Condroz region in Belgium. The data have been introduced in Goegebeur et al. (2005) and subsequently studied by various authors, see e.g. Vandewalle et al. (2004), Beirlant et al. (2004), pp. 34–39 and 226–240, and Hubert and Vandervieren (2006). Ca-content is expressed in mg/kg (values are 10 times multiplied in the data set) of dry soil. The dataset contains 1505 observations and it is a part of a larger database collected by the Belgian non-profit organization REQUASUD (Réseau Qualité Sud).[1]

As in Goegebeur et al. (2005) our main interest shall be to describe the extremal properties of Ca levels with pH-values as an explanatory variable. The data set is plotted on Figures 4.21 and 4.22. The pictures indicate that there are very few observations for the lowest as well as the highest pH levels (particularly only 5 observations for pH = 5.1).

We do not provide an exhaustive study concerning this dataset. Instead of that, we just apply the methods, we have developed in the previous chapters in some sort of a "naive" uninformed approach about the dataset. Consecutively, we shall try to analyze the results in an attempt to either justify them or reject.

We suppose, that there is some additive model between the levels of pH and Ca and we also suppose that the extremal properties of Ca remains the same for different pH levels, i.e. γ is same throughout all levels of pH. We will later see, whether our results indicate the assumption or not. As the Figure 4.21 does not indicate that the situation can be described by a linear trend, we fitted a linear model with quadratic trend:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + E_i, \qquad i = 1, \dots, n. \tag{4.3}$$

where $X_i = x_i - \frac{1}{n}\sum_{i=1}^n x_i$ and $X_i^2 = x_i^2 - \frac{1}{n}\sum_{i=1}^n x_i^2$ with x_i being pH levels in the dataset and Y_i observations of Ca $i = 1, \dots, n$. Initially there was $n = 1505$ values of Y_i and X_i in the dataset. We calculated the process of regression quantile intercepts obtaining $m = 1872$ unique intercepts, i.e. $\hat{\beta}_1(\tau_m) > \hat{\beta}_1(\tau_{m-1}) > \dots > \hat{\beta}_1(\tau_1)$ for some suitable sequence $1 > \tau_m > \tau_{m-1} > \dots > \tau_1 > 0$. We used three different estimators on the data set: ML-estimator, PWM-estimator, Pickands-estimator, i.e.

(1) Maximum likelihood estimator (ML) $\hat{\gamma}_{m,k}^{\text{RQ,ML}}$ (*ML-estimator*) of the extreme value index based on the k largest unique estimates of $\hat{\beta}_{n,1}(\tau), \quad \tau \in (0,1)$, i.e. the estimator fits generalized Pareto distribution (GPD) on the exceedances of $\{\hat{\beta}_{n,1}(\tau_j)\}_{j=m-k,\dots,m}$ over $\hat{\beta}_{n,1}(\tau_{m-k-1})$. To calculate ML-estimate we used `evir` package from `R`.

(2) Probability weighted moments estimator (PWM)

$$\hat{\gamma}_{m,k}^{\text{RQ,PWM}} = \frac{\frac{1}{k}\sum_{j=1}^k \left(4\frac{j}{k+1} - 3\right)\hat{\beta}_{n,1}(\tau_{m-i+1})}{\frac{1}{k}\sum_{j=1}^k \left(2\frac{j}{k+1} - 1\right)\hat{\beta}_{n,1}(\tau_{m-i+1})}$$

(3) Pickands estimator

$$\hat{\gamma}_{m,k}^{\text{RQ,P}} = \frac{1}{\log 2}\log\left(\frac{\hat{\beta}_{n,1}(\tau_{m-[k/4]}) - \hat{\beta}_{n,1}(\tau_{m-[k/2]})}{\hat{\beta}_{n,1}(\tau_{m-[k/2]}) - \hat{\beta}_{n,1}(\tau_{m-k})}\right)$$

[1]The data set is available as a supplement of Beirlant et al. (2004) on `http://lstat.kuleuven.be/Wiley/`.

We obtained different estimates with respect to k, which are plotted on Figure 4.23 and 4.24. Except of the well known fact about the large variance of Pickands estimators, there seems to be a stability region indicating that $\gamma \doteq 0.5$ is a reasonable estimate. However, we should ask, if this result is not affected by the flat tails (or is it by the low number of observations?) in the regions where pH is small.

To answer on this question we calculated linear model (4.3) for pH levels > 6.5. We got $n = 956$ observations resulting in $m = 1154$ unique intercept values. Again, we plugged intercepts in ML, PWM and Pickands estimator and plotted the estimates versus k, the results are on Figure 4.25. We see that similarly to Figure 4.23 there is a region of stability interrupted only briefly by a narrow peak for $k \doteq 200$. Otherwise the estimate of γ remains principally the same. We can again accept that $\gamma \doteq 0.5$.

Finally, we restricted pH levels one more time such that $6.8 < pH < 7.4$ and obtained $n = 544$ observations resulting in $m = 616$ unique intercepts of regression quantiles. The results on Figure 4.26 again indicate that γ lies in the vicinity of 0.5. The fact that Pickands estimator fall off to the region bellow zero can be attributed to its larger variance in comparison to PWM and ML estimators.

We said in the beginning, that the Condroz dataset was analysed many times. Hence, there are different estimates of γ based on different assumptions, which authors made. In Beirlant et al. (2004), pp. 226–231, the authors fitted the data assuming that $\gamma(pH) = \exp(\beta_0 + \beta_1 pH)$, i.e. that gamma depends on pH. This was supported by the non-parametric approach estimating γ for each level of pH, see Beirlant et al. (2004), pp. 238–241. As we have seen this extremal dependence of γ on pH can be attributed to the different number of observations in each pH level. In Vandewalle et al. (2004) the authors proposed, that the largest observations are in fact outliers, claiming that they appertain to the border regions of Condroz and have therefore a different distribution. Using robust methods removing (or down-weighting) these "outliers" they got an estimate of $\gamma \doteq 0.2$. While, we can criticize this approach on the ground that it is unreliable that Ca contents and pH are linked in a different way across Belgium, we should realize that also our approach is based on the assumption that there is a quadratic (or linear) dependence between pH and Ca. Moreover, the given task seems to be odd from the beginning – pH is rather a function of Ca levels (and other elements in the soil) than the physical level of Ca is derived from pH. These considerations one more time stresses the importance of assumptions, which are in EVT crucial and in practical data analysis much more important than any theoretical results. However, this could be stated about any method of mathematical statistics.

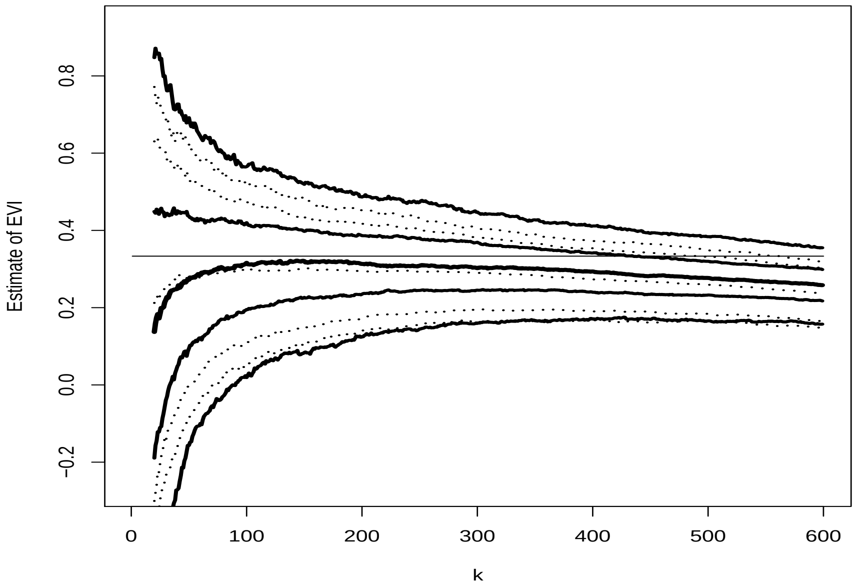

Figure 4.1: Quantile regression ML-estimator applied in model (a) with Fréchet distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown vs k (solid lines). The estimator is compared with ML-estimator applied on the sample of errors, again the 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations are indicated for each k (dashed lines).

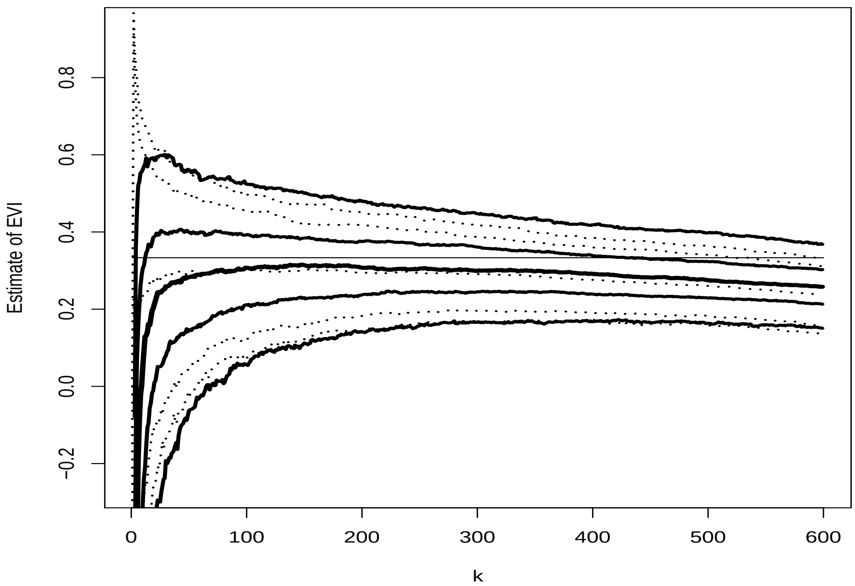

Figure 4.2: Quantile regression PWM-estimator applied in model (a) with Fréchet distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown vs k (solid lines). The estimator is compared with PWM-estimator applied on the sample of errors, again the 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (dashed lines).

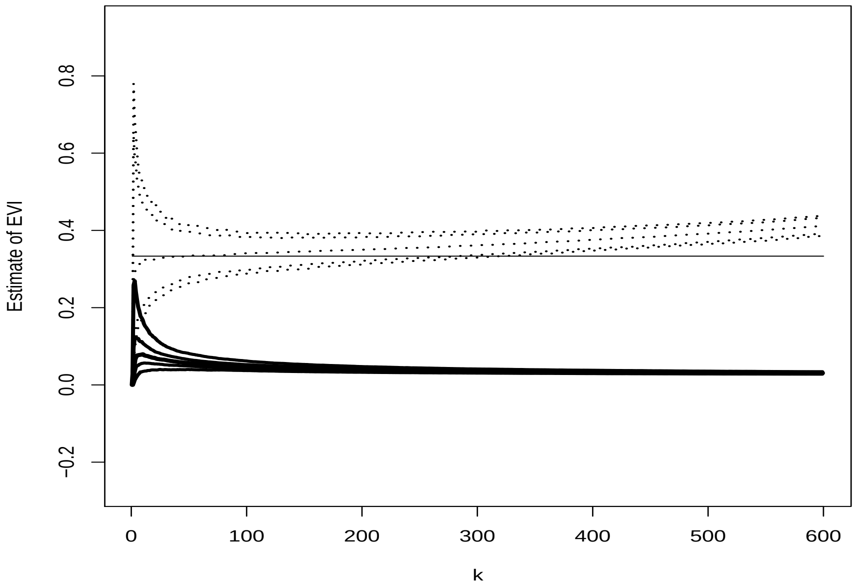

Figure 4.3: Quantile regression Hill's estimator applied in model (a) with Fréchet distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown versus growing k (solid lines). The estimator is compared with Hill's estimator applied on the sample of the errors, again the 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (dashed lines).

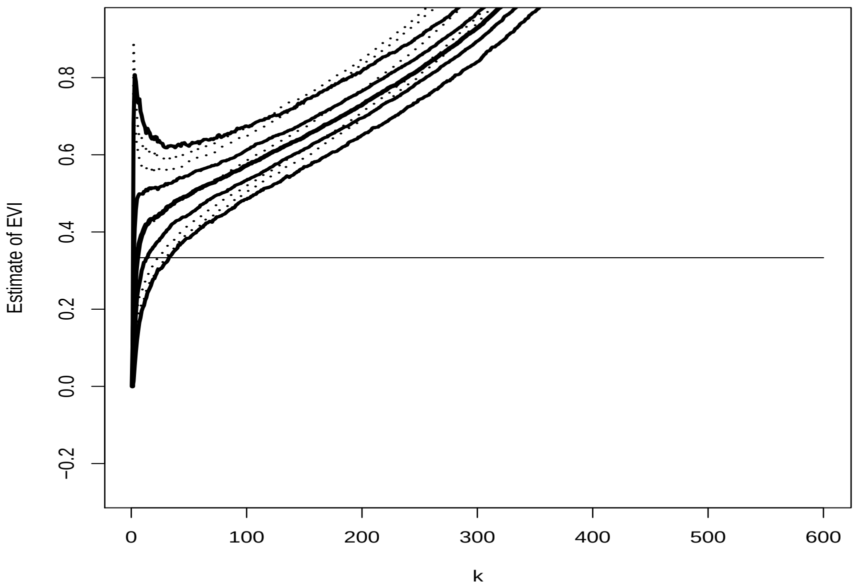

Figure 4.4: Quantile regression Hill's estimator applied in model (a) with Fréchet distribution of errors. The Medians and the 5th, 10th, 90th, and 95th percentiles of 1000 replicates are shown versus growing k (solid lines). The estimated intercepts were normalized thus obtaining the estimator (4.1). The estimator is compared with Hill's estimator applied on normalized errors, again the medians and the 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (dashed lines). For the results of Hill's estimator applied on the same unaltered errors see Figure 4.3 (dashed).

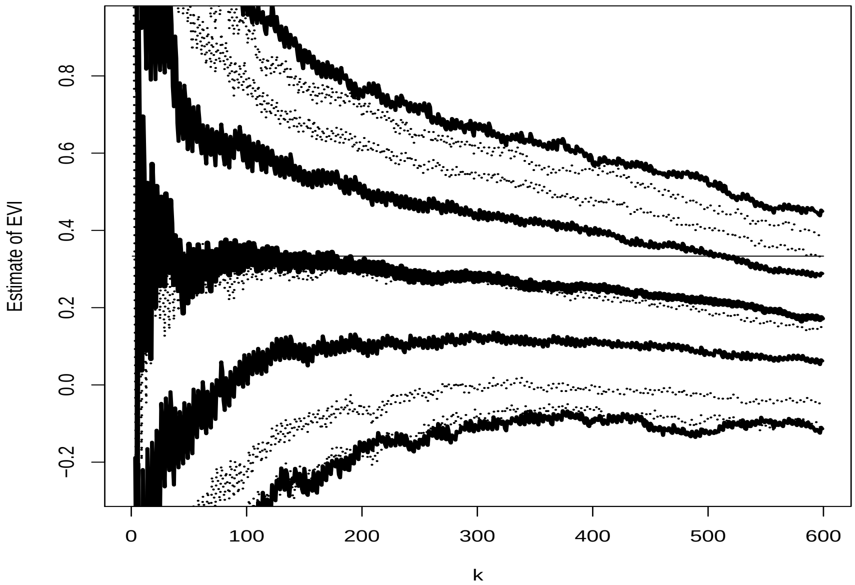

Figure 4.5: Quantile regression Pickands estimator applied in model (a) with Fréchet distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown versus growing k (solid lines). The estimator is compared with Pickands estimator applied on the sample of errors, again 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (dashed lines).

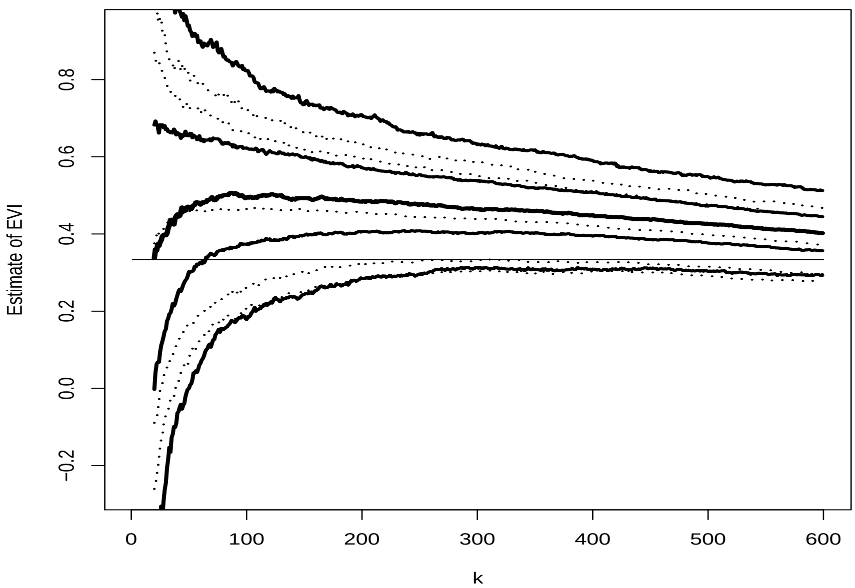

Figure 4.6: Quantile regression ML estimator applied in model (a) with Burrleigh distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown versus growing k (solid lines). The estimator is compared with ML estimator applied on the sample of errors, again 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (dashed lines).

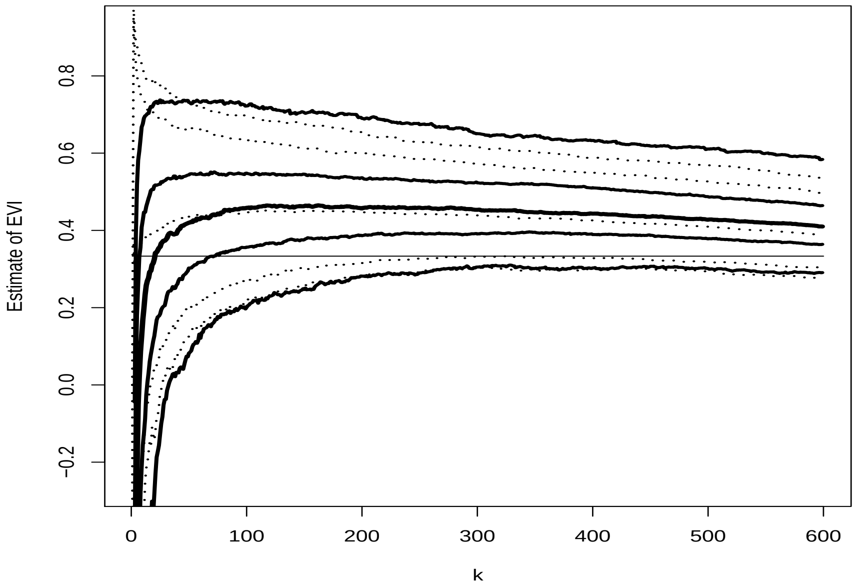

Figure 4.7: Quantile regression PWM estimator applied in model (a) with Burrleigh distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown versus growing k (solid lines). The estimator is compared with PWM estimator applied on the sample of errors, again 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (dashed lines).

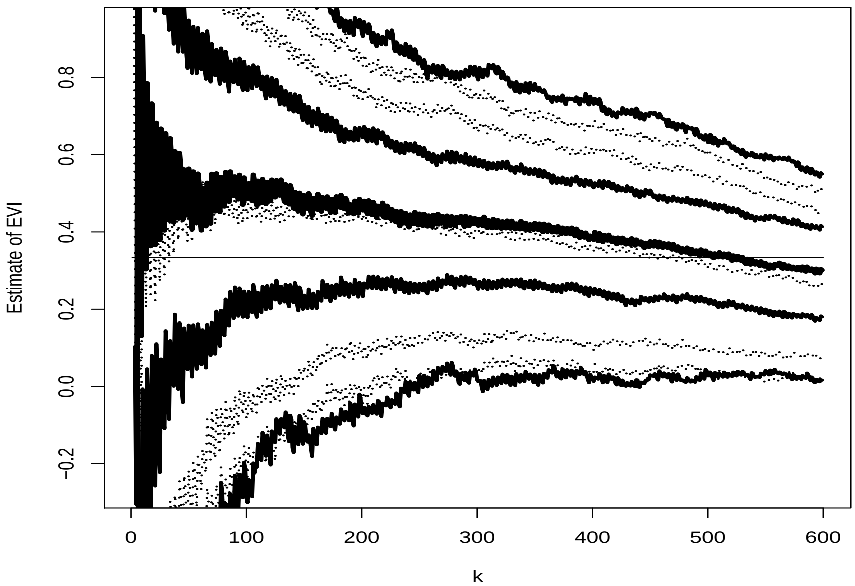

Figure 4.8: Quantile regression Pickands estimator applied in model (a) with Burrleigh distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown versus growing k (solid lines). The estimator is compared with Pickands estimator applied on the sample of errors, again 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (dashed lines).

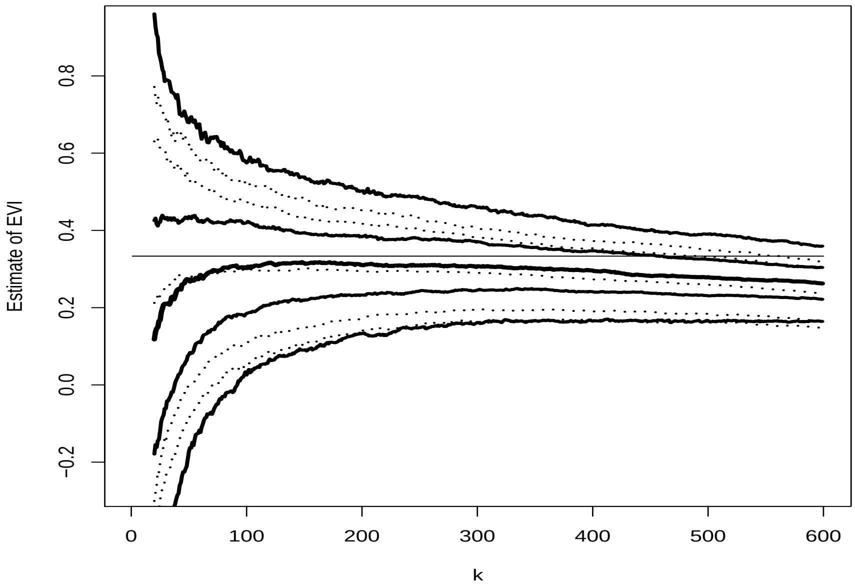

Figure 4.9: Quantile regression ML estimator applied in model (b) with Fréchet distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown versus growing k (solid lines). The estimator is compared with ML estimator applied on the sample of errors, again 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (dashed lines).

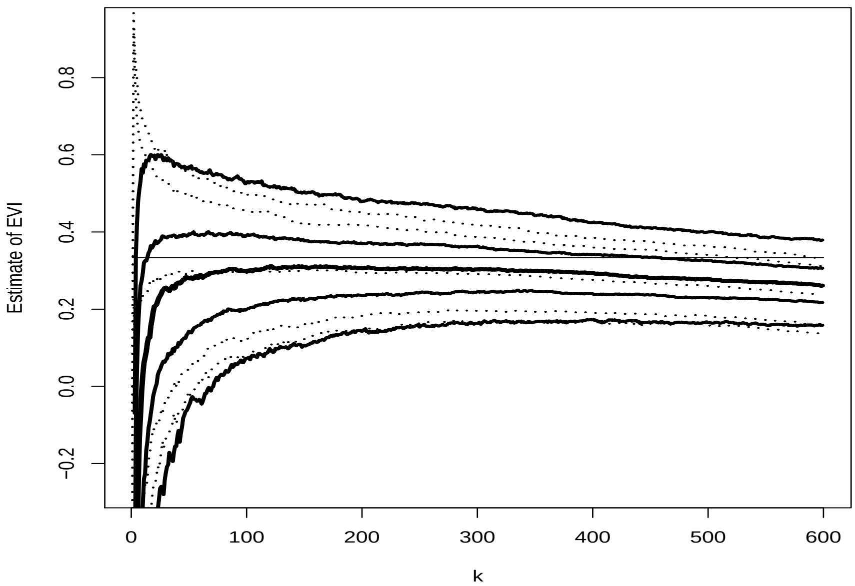

Figure 4.10: Quantile regression PWM estimator applied in model (b) with Fréchet distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown versus growing k (solid lines). The estimator is compared with PWM estimator applied on the sample of errors, again 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (dashed lines).

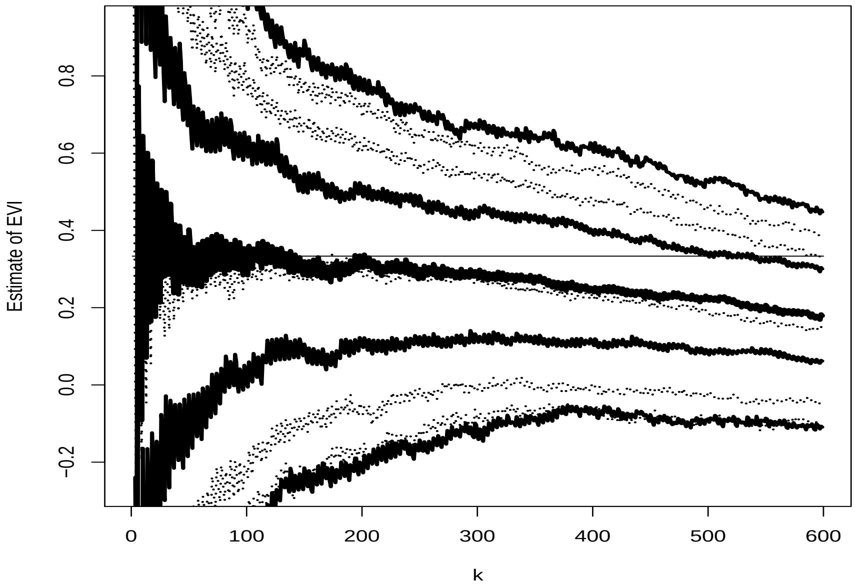

Figure 4.11: Quantile regression Pickands estimator applied in model (b) with Fréchet distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown versus growing k (solid lines). The estimator is compared with ML estimator applied on the sample of errors, again 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (dashed lines).

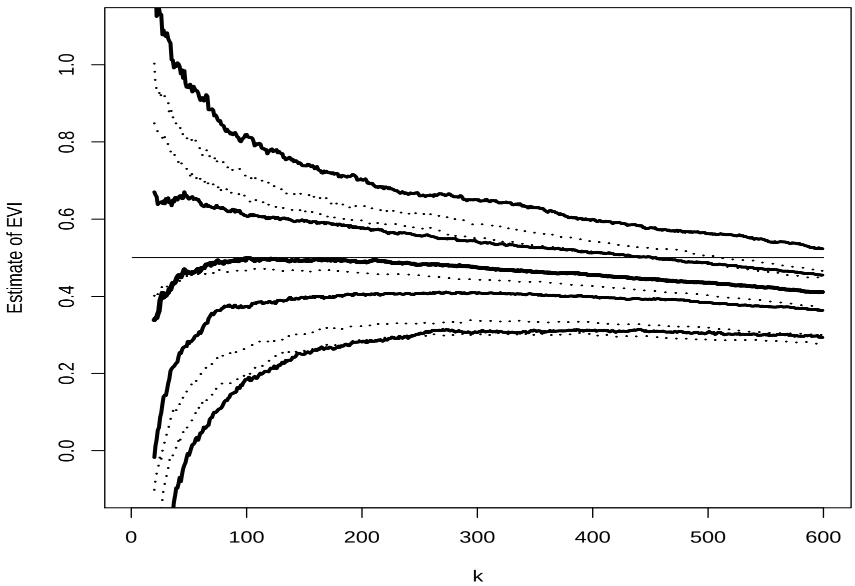

Figure 4.12: Quantile regression PWM estimator applied in model (b) with Burrleigh distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown versus growing k (solid lines). The estimator is compared with PWM estimator applied on the sample of errors, again 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (dashed lines).

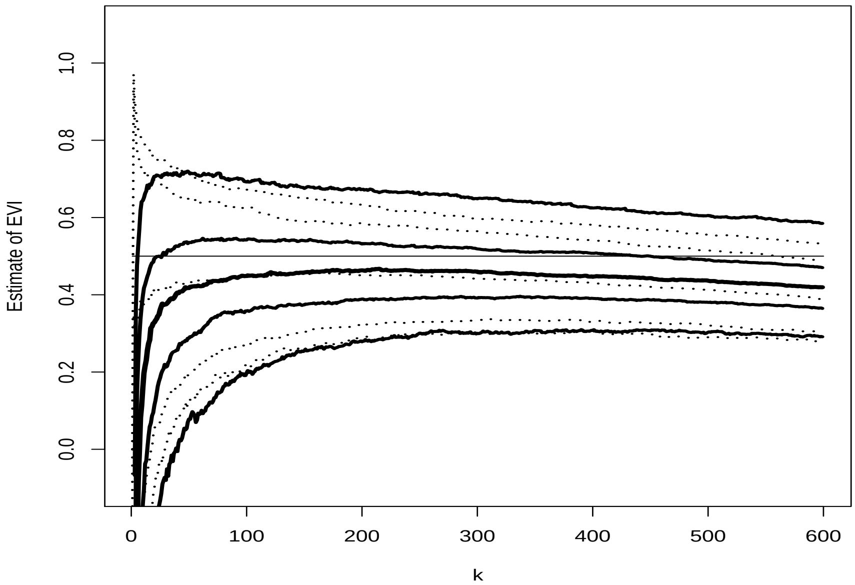

Figure 4.13: Quantile regression PWM estimator applied in model (b) with Burrleigh distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown versus growing k (solid lines). The estimator is compared with ML estimator applied on the sample of errors, again 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (dashed lines).

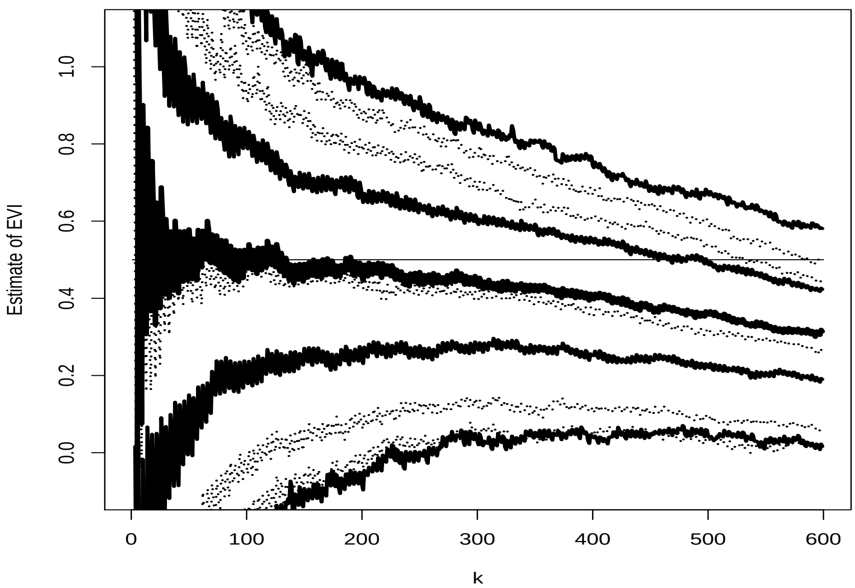

Figure 4.14: Quantile regression PWM estimator applied in model (b) with Burrleigh distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown versus growing k (solid lines). The estimator is compared with PWM estimator applied on the sample of errors, again 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (dashed lines).

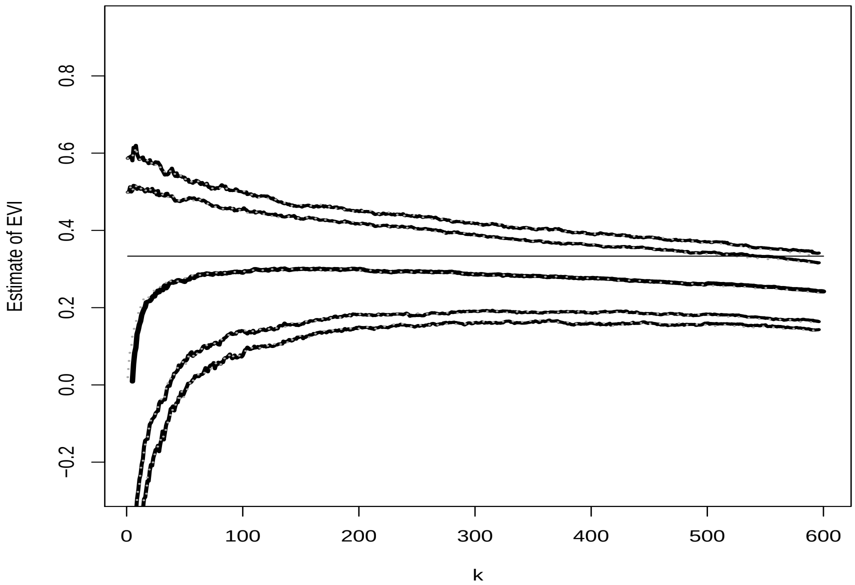

Figure 4.15: Residuals of $\hat{\boldsymbol{\beta}}_{nR}$ ML-estimator applied in model (b) with Fréchet distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown vs k (solid lines). The estimator is compared with ML-estimator applied on the sample of errors, again the 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations are indicated for each k (grey dashed lines).

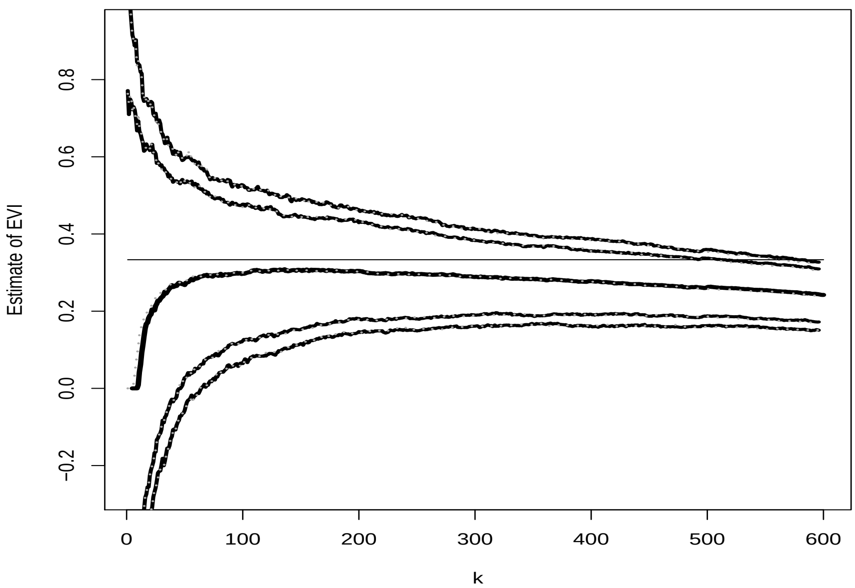

Figure 4.16: Residuals of $\hat{\boldsymbol{\beta}}_{nR}$ PWM-estimator applied in model (b) with Fréchet distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown vs k (solid lines). The estimator is compared with PWM-estimator applied on the sample of errors, again the 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (grey dashed lines).

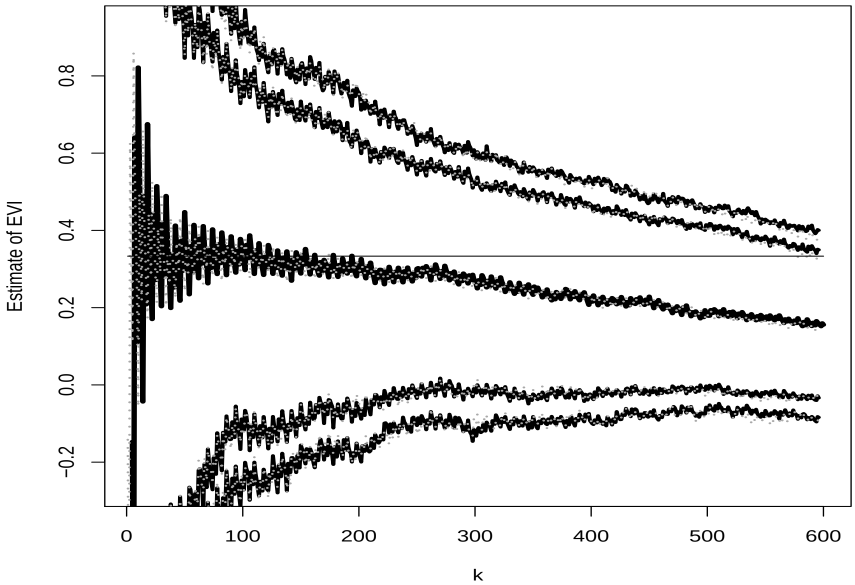

Figure 4.17: Residuals of $\hat{\boldsymbol{\beta}}_{nR}$ Pickands estimator applied in model (b) with Fréchet distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown vs k (solid lines). The estimator is compared with Pickands estimator applied on the sample of errors, again the 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations are indicated for each k (grey dashed lines).

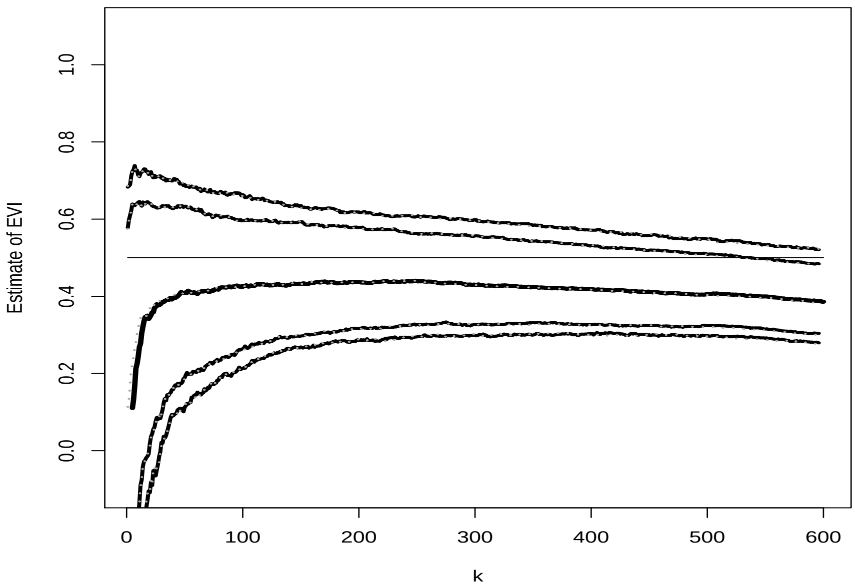

Figure 4.18: Residuals of $\hat{\boldsymbol{\beta}}_{nR}$ ML-estimator applied in model (b) with Burrleigh distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown vs k (solid lines). The estimator is compared with ML-estimator applied on the sample of errors, again the 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (grey dashed lines).

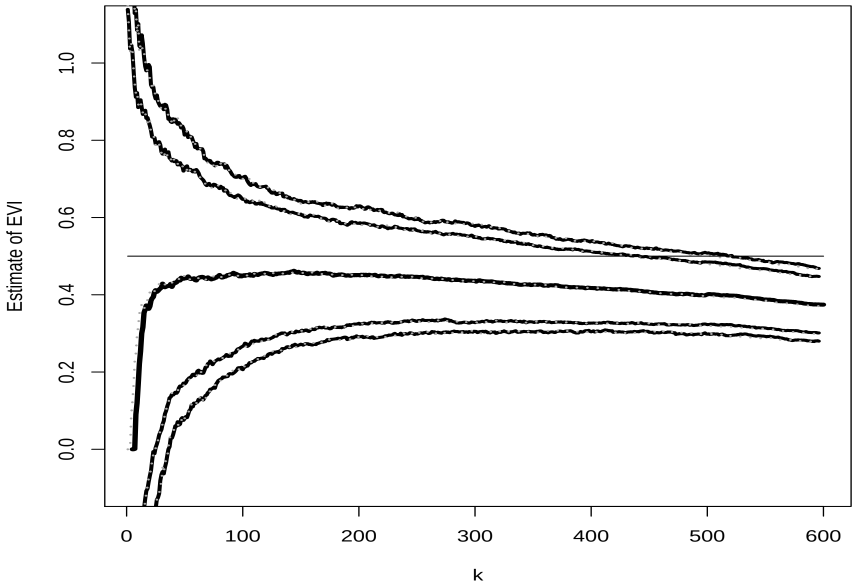

Figure 4.19: Residuals of $\hat{\boldsymbol{\beta}}_{nR}$ PWM-estimator applied in model (b) with Burrleigh distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown vs k (solid lines). The estimator is compared with PWM-estimator applied on the sample of errors, again the 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations are indicated for each k (grey dashed lines).

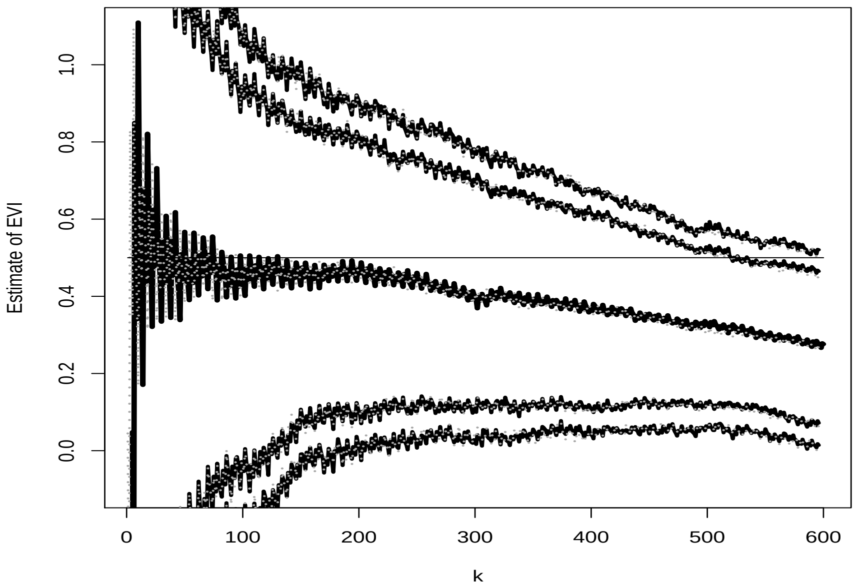

Figure 4.20: Residuals of $\hat{\boldsymbol{\beta}}_{nR}$ Pickands estimator applied in model (b) with Burrleigh distribution of errors. The 5th, 10th, 50th, 90th, and 95th percentiles of 1000 replicates are shown vs k (solid lines). The estimator is compared with Pickands estimator applied on the sample of errors, again the 5th, 10th, 50th, 90th, and 95th percentiles of 1000 estimations for each k are indicated (grey dashed lines).

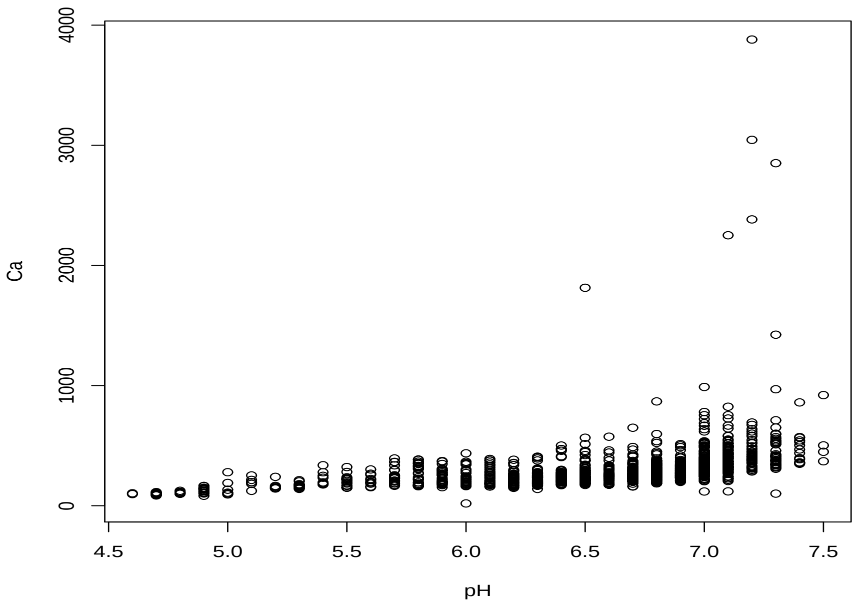

Figure 4.21: Condroz data: 1505 observations of Ca-levels in various district of Condroz region plotted against their respective pH.

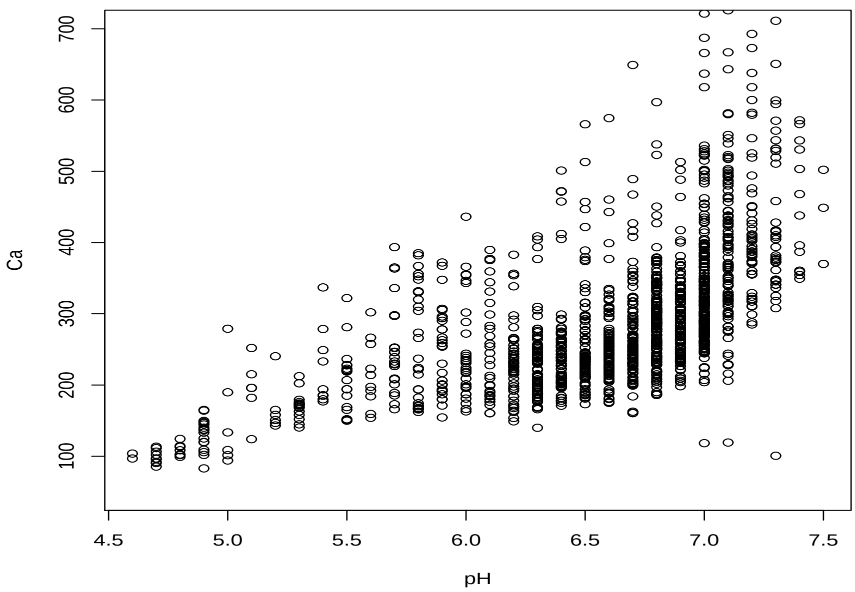

Figure 4.22: Condroz data: Only observations of Ca levels bellow 700 are shown on this plot. It is clearly visible that there is very few observations of some pH levels particularly in the lower part of the plot.

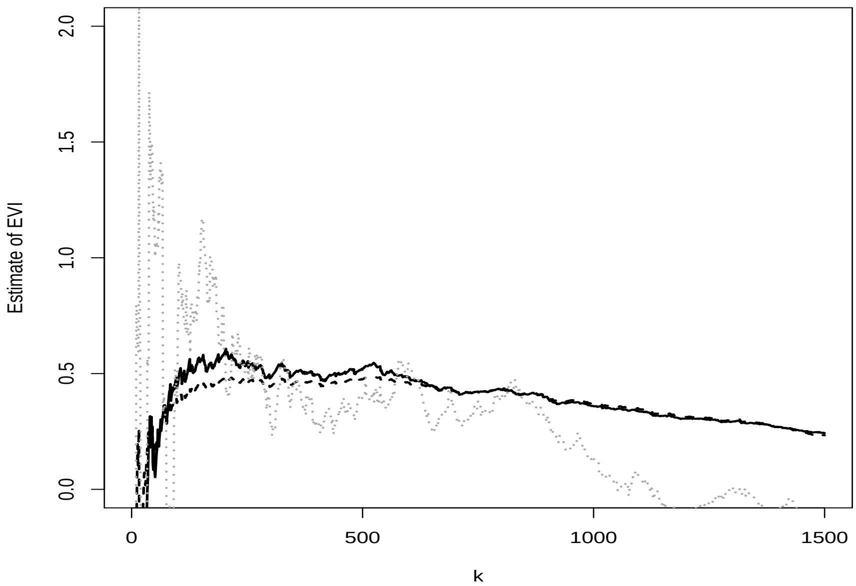

Figure 4.23: Estimator plots of ML-estimator (solid), PWM-estimator (dashed) and Pickands estimator (grey, dotted) performed on intercepts of regression quantiles on normalized in reparametrized quadratic model.

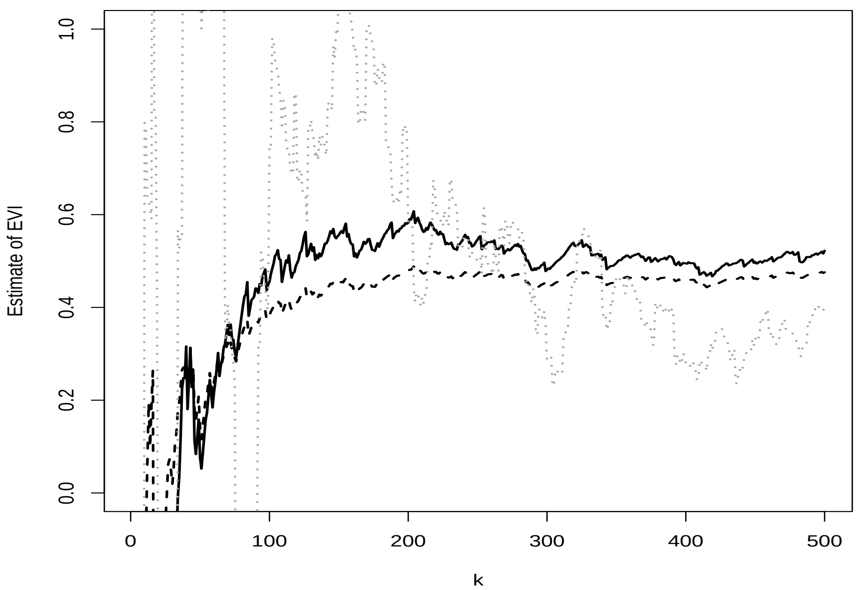

Figure 4.24: Estimator plots of ML-estimator (solid), PWM-estimator (dashed) and Pickands estimator (grey, dotted) performed on intercepts of regression quantiles in reparametrized quadratic model, $k \leq 500$.

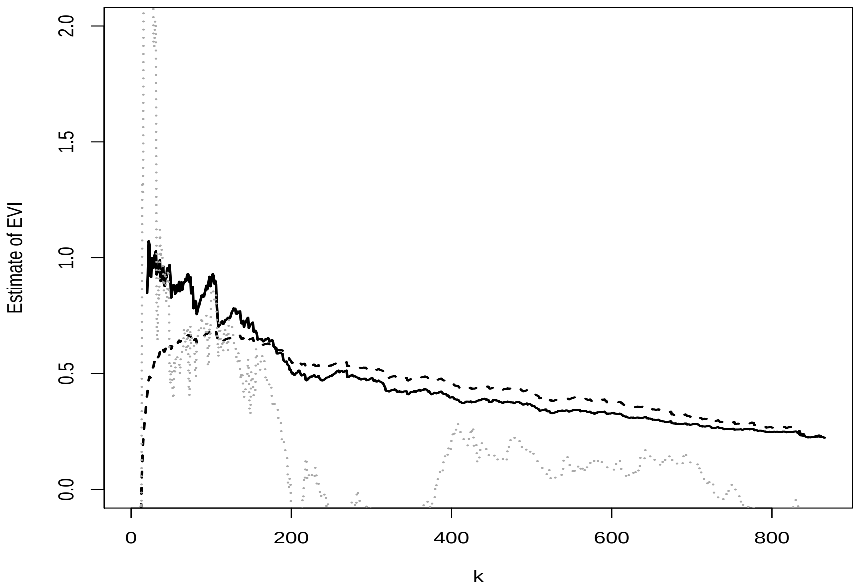

Figure 4.25: Estimator plots of ML-estimator (solid), PWM-estimator (dashed) and Pickands estimator (grey, dotted) performed on intercepts of regression quantiles on normalized in reparametrized quadratic model.

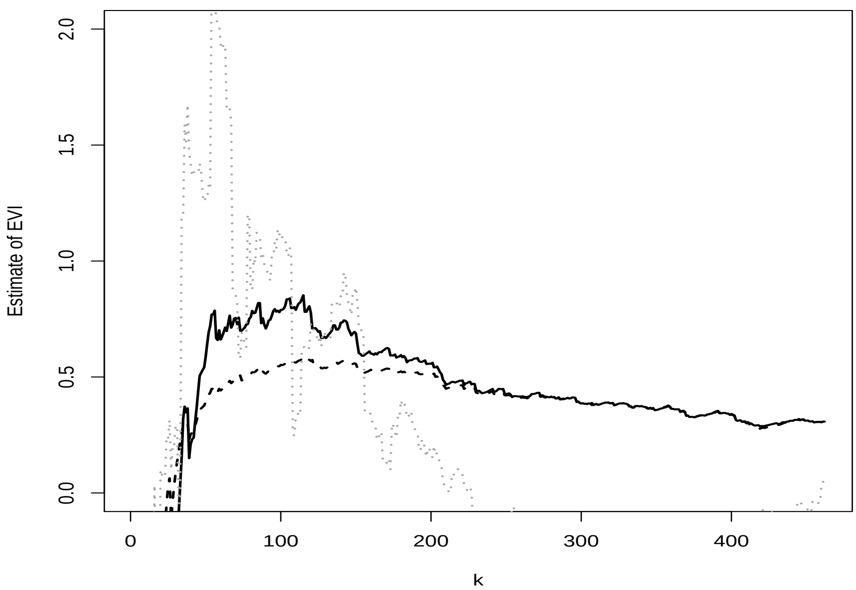

Figure 4.26: Estimator plots of ML-estimator (solid), PWM-estimator (dashed) and Pickands estimator (grey, dotted) performed on intercepts of regression quantiles in reparametrized quadratic model.

Appendix - Regular variation

Here we provide only a brief introduction into the concept of regular variation. For the more detailed explanation and proofs we refer to de Haan and Ferreira (2006), pp. 361–407. A reader can find a brief summary also in Beirlant et al. (2004), pp. 76–80.

Let f be an ultimately positive function on $\mathbb{R}^+$. We say that the function f is *regularly varying* at ∞ if and only if there exists a real constant α for which

$$\lim_{x\to\infty} \frac{f(xt)}{f(x)} = t^\alpha, \qquad \text{for all } t > 0.$$

We call α the index of regular variation and write $f \in \mathcal{R}_\alpha$. In the case $\alpha = 0$, i.e.

$$\lim_{x\to\infty} \frac{f(xt)}{f(x)} = 1, \qquad \text{for all } t > 0,$$

we call the f to be *slowly varying*. We will reserve symbol ℓ for such functions. The class of slowly varying functions $\mathcal{R}_0$ is an important one in the context of EVT and have a lot of properties of the class appear in the proofs. We note just a few of them.

(i) $\mathcal{R}_0$ is closed under addition, multiplication and division.

(ii) If ℓ is slowly varying, then ℓ^a is slowly varying for all $a \in \mathbb{R}$

(iii) If $\ell \in \mathcal{R}_0$, then

$$\lim_{x\to\infty} \frac{\ell(xt)}{\ell(x)} = 1$$

uniformly for $t \in S$ for any compact $S \subset \mathbb{R}^+$.

(iv) If ℓ is slowly varying, then

$$\lim_{x\to\infty} \frac{\ell(x)}{\log x} = 0.$$

(v) If $f \in \mathcal{R}_\alpha$ then for any $\ell \in \mathcal{R}_0$

$$\lim_{x\to\infty} \frac{f(x)}{x^\alpha \ell(x)} = 1$$

hence as the regular variation is a property holding only if $x \to \infty$, we can write that $f(x) = x^\alpha \ell_f(x)$ for some $\ell_f(x) \in \mathcal{R}_0$.

To link two regularly varying functions we introduce the notation of *de Bruyn conjugate*:

If $\ell(x)$ is a slowly varying function, then there exists a slowly varying $\ell^*(x)$, the de Bruyn conjugate of ℓ such that

$$\lim_{x\to\infty} \ell(x)\ell^*(x\ell(x)) = 1$$

The de Bruyn conjugate is asymptotically unique in the sense that if also $\tilde{\ell}$ is a slowly varying function and $\ell(x)\tilde{\ell}(x\ell(x)) \to 1$ then

$$\lim_{x\to\infty} \frac{\ell^*(x)}{\tilde{\ell}(x)} = 1.$$

Furthermore

$$\lim_{x\to\infty} \frac{(\ell^*)^*(x)}{\ell(x)} = 1.$$

The concept of regular variation can be easily modified for any $x \to \zeta \in \mathbb{R}$ such that f is ultimately positive on some left or right neighbourhood of ζ. We say that f is regularly varying at ζ^+ with some index α, i.e. $f \in \mathcal{R}_\alpha(\zeta^+)$, if and only if

$$\lim_{x\downarrow\zeta} \frac{f(tx)}{f(x)} = t^\alpha, \qquad \text{for all } t > 0.$$

Similarly we can define the concept of regular variation on the left neighbourhood of ζ, we say that f is regularly varying in ζ^- for some index α if and only if

$$\lim_{x\uparrow\zeta} \frac{f(tx)}{f(x)} = t^\alpha, \qquad \text{for all } t > 0.$$

Bibliography

Bahadur, R. R. (1966), 'A note on quantiles in large samples', *Annals of Mathematical Statistics* **37**, 577–580.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004), *Statistics of Extremes, Theory and Application*, John Wiley & Sons, Chichester.

Buch-Kromann, T. (2006), 'Claim cost estimation of large insurance losses', *Journal of Actuarial Practice* **13**, 199–219.

Chernozhukov, V. (2005), 'Extremal quantile regression', *Annals of Statistics* **33**, 806–839.

Csörgő, M. and Horváth, L. (1993), *Weighted Approximations in Probability and Statistics*, John Wiley & Sons, Chichester.

Csörgő, M. and Révész, P. (1981), *Strong Approximations in Probability and Statistics*, Akedémiai Kiadó, Budapest.

de Haan, L. and Ferreira, A. (2006), *Extreme Value Theory, An Introduction*, Springer, New York.

de Haan, L. and Resnick, S. I. (1996), 'Second-order regular variation and rates of convergence in extreme-value theory', *Annals of Probability* **24, Number 1**, 97–124.

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

Dienstbier, J. (2009), Problems with extremes, *in* 'Proceedings of 15th summer school of JČ(S)MF ROBUST 2008'.

Dietrich, D., de Haan, L. and Hüsler, J. (2002), 'Testing extreme value conditions', *Extremes* **5:1**, 71–85.

Dodge, Y. and Jurečková, J. (2000), *Adaptive Regression*, Springer, New York.

Drees, H. (1998*a*), 'A general class of estimators of the extreme value index', *Journal of Statistical Planning and Inference* **66**, 95–112.

Drees, H. (1998*b*), 'On smooth statistical tail functionals', *Scandinavian Journal of Statistics* **25**, 187–210.

Drees, H. (1998*c*), 'Optimal rates of convergence for estimates of the extreme value index', *Annals of Statistics* **26, No. 1**, 434–448.

Drees, H. (2003), 'Extreme quantile estimation for dependent data, with applications to finance', *Bernoulli* **9**, 617–657.

Drees, H., de Haan, L. and Li, D. (2006), 'Approximations to the tail empirical distribution function with application to testing extreme value conditions', *Journal of Statistical Planning and Inference* **136**, 3498–3538.

Goegebeur, Y., Planchon, V., Beirlant, J. and Oger, R. (2005), 'Quality assesment of pedochemical data using extreme value methodology', *Journal of Applied Sciences* **5 (6)**, 1092–1102.

Gutenbrunner, C. and Jurečková, J. (1992), 'Regression rank scores and regression quantiles', *Annals of Statistics* **20**, 305–330.

Gutenbrunner, C., Jurečková, J., Koenker, R. and Portnoy, S. (1993), 'Tests of linear hypotheses based on regression rank scores', *Journal of Nonparametric Statistic* **2**, 307–331.

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

Hájek, J., Šidák, Z. and Sen, P. K. (1999), *Theory of Rank Tests*, Academic Press, San Diego.

Hodges, J. L. and Lehmann, E. L. (1963), 'Estimation of location based on rank tests', *Annals of Mathematical Statistics* **34**, 598–611.

Hoeffding, W. (1963), 'Probability inequalities for sums of bounded random variables', *Journal of the American Statistical Association* **56**, 13–30.

Hüsler, J. and Li, D. (2006), 'On testing extreme value conditions', *Extremes* **9**, 69–86.

Huber, P. J. (1981), *Robust Statistics*, Wiley, New York.

Hubert, M. and Vandervieren, E. (2006), An adjusted boxplot for skewed distribution, Technical report, Katholieke Universiteit Leuven, Department of Mathematics, Section of Statistics.

Jaeckel, L. A. (1972), 'Estimating regression coefficients by minimizing the dispersion of the residuals.', *Annals of Mathematical Statistics* **43**, 1149–1459.

Jurečková, J. (1971), 'Nonparametric estimate of regression coefficients', *Annals of Mathematical Statistics* **42**, 1328–1338.

Jurečková, J. (1999), 'Regression rank scores testes against heavy-tailed alternatives', *Bernoulli* **5**, 659–676.

Jurečková, J. (2007), 'Remark on extreme regression quantile', *Sankhyā: The Indian Journal of Statistics* **69**, 87–100.

Jurečková, J. and Picek, J. (2005), 'Two-step regression quantiles', *Sankhyā: The Indian Journal of Statistics* **67**, 227–252.

Jurečková, J. and Sen, P. K. (1996), *Robust Statistical Procedures, Asymptotics and Interrelations*, Wiley, New York.

Koenker, R. (2005), *Quantile Regression*, Cambridge University Press, Cambridge.

Koenker, R. (2006), 'Quantile regression in r: A vignette'.
URL: *http://www.econ.uiuc.edu/ roger/research/rq/vig.pdf*

Koenker, R. and Basset, G. (1978), 'Regression quantiles', *Econometrica* **46**, 35–50.

Koenker, R. and Machado, J. A. F. (1999), 'Goodness of fit and related inference process for quantile regression', *Journal of the American Statistical Association* **94**, 1296–1310.

Koenker, R. and Xiao, Z. (2002), 'Inference on the quantile regression process', *Econometrica* **70**, 1583–1612.

Koul, H. L. and Mukherjee, K. (1994), 'Regression quantiles and related processes under long range dependent errors', *Journal of Multivariate Analysis* **51**, 318–337.

Kyselý, J., Picek, J. and Beranová, R. (2010), 'Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold', *Global and Planetary Change* **72**, 55–68.

Mason, D. M. (1982), 'Laws of large numbers for sums of extreme values', *Annals of Probability* **10**, 754–764.

Neves, C. and Fraga-Alves, M. I. (2008), 'Testing extreme value conditions - an overview and recent approaches', *REVSTAT* **6**, 83–100.

Northrop, P. J. and Jonathan, P. (2011), 'Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights', *Environmetrics* **(accepted for publication)**.

Parzen, E. (1979), 'Nonparametric statistical data modeling', *Journal of the American Statistical Association* **74, No. 365**, 105–121.

Picek, J. and Dienstbier, J. (2010), *Nonparametrics and Robustenss in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková*, Institute of Mathematical Statistics, Beachwood, Ohio, USA, chapter Extremes of two-step regression quantiles, pp. 204–214.

Portnoy, S. (1991), 'Asymptotic behavior of the number of regression quantile breakpoints', *SIAM Journal on Scientific Computing* **12**, 867–883.

Portnoy, S. and Jurečková, J. (2000), 'On extreme regression quantiles', *Extremes* **2:3**, 227–243.

Resnick, S. I. (2007), *Heavy-Tail Phenomena, Probabilistic and Statistical Modeling*, Springer, New York.

Rieder, H. (1994), *Robust asymptotic statistics*, Springer, New York.

Shorack, G. R. (1979), 'The weighted empirical process of row independent random variables with arbitrary distribution functions', *Statistica Neerlandica* **33**, 169–189.

Shorack, G. R. (1991), 'Embedding the finite sampling process at a rate', *Annals of Probability* **19**, 826–842.

Shorack, G. R. (2000), *Probability for Statisticians*, Springer, New York.

Shorack, G. R. and Wellner, J. A. (1986), *Empirical Processes with Applications to Statistics*, Wiley, New York.

Smith, R. L. (1994), 'Nonregular regression', *Biometrika* **81**, 173–183.

Vandewalle, B., Beirlant, J. and Hubert, M. (2004), *Theory and Applications of Recent Robust Methods*, Birkhäuser Verlag, Basel, chapter A robust estimator of the tail index based on an exponential regression model, pp. 367–376.

Zhou, C. (2010), 'The extent of the maximum likelihood estimator for the extreme value index', *Journal of Multivariate Analysis* **101**, 971–983.

Address:

Department of probability and mathematical statistics,
Faculty of Mathematics and Physics,
Charles University in Prague
Sokolovská 83,
186 75, Prague 8
E-mail: dienstbier.jan@gmail.com

This thesis was typeset with LaTeX 2_ε[2] by author.

[2] LaTeX 2_ε is an extension of LaTeXwhich is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society.