People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Mohamed Khider University - Biskra

Order Number: ....
Series: ....

Faculty of Exact Sciences and Sciences of Nature and Life
Department of Mathematics

THESIS

In Candidacy for the Degree of

DOCTOR 3^{rd} CYCLE IN MATHEMATICS

In the Field of **Statistics**

By

SAIDA MANCER

TITLE

New approach for estimating the distribution tails for incomplete data

In front of the jury composed of:

Prof.	Djamel Meraghni	University of Biskra	President
Prof.	Abdelhakim NECIR	University of Biskra	Supervisor
Prof.	Djabrane Yahia	University of Biskra	Examiner
Prof.	Zaher Mohdeb	University of Constantine 3	Examiner

February 23, 2023

Abstract

Our work is situated in the field of extreme values' statistics for incomplete data namely the truncation and the censoring. In this context, several approaches for estimating distribution tails under random truncation have recently been developed: Gardes & Stupfler (2015) [18], Benchaira et al. (2015) [5], Benchaira et al. (2016a) [6], Benchaira et al. (2016b) [7], et Haouas et al. (2018) [21].

The first objective of this thesis is to define a new method " the semi-parametric method" to estimate the tail index of the distribution, while the majority of the existing method depend on the non-parametric estimator of the tail distribution index such as LyndeBell and Woodroofe, the ours is based on the semi-parametric estimator defined in Wang 1989 [48] that allows us introducing new estimators with high efficiency.

For the second objective, at this point, we are interested in correcting the error of kernel estimators, such as Benchaira et al. (2016b)'s estimator, so we have introduced a new kernel estimator with reduced bias at the same time.

Without forgetting the complete data, in the third objective of this thesis we add a new estimator of the extreme value's index beside the well-known estimators such as Hill, Peng, ... etc. The new one is characterized by its robustness and stability and was developed by using the idea which was presented in Basu 1998 [2] based on the density power divergence function.

RÉSUME

Il s'agit de la statistique des valeurs extrêmes pour les données incomplètes a savoir la troncature et le censure. Dans ce contexte, plusieurs approches d'estimation des queues de distribution sous troncature aléatoires sont récemment développées: Gardes & Stupfler (2015) [18], Benchaira et al. (2015) [5], Benchaira et al. (2016a) [6], Benchaira et al. (2016b) [7], et Haouas et al. (2018) [21].

Le premier objectif de cette thèse est de définir une nouvelle méthode "la méthode semi-paramétrique" pour estimer l'indice de la queue de distribution, alors que la majorité des méthodes existantes dépendent de l'estimateur non paramétrique de la fonction de distribution tel que LyndeBell et Woodroofe, le nôtre est basé sur l'estimateur semi-paramétrique défini par Wang 1989 [48]. qui nous permet d'introduire de nouveaux estimateurs avec une forte efficacité.

Pour le deuxième objectif, dans ce point, nous nous intéressons à la correction du biais des estimateurs a noyaux, comme l'estimateur de Benchaira et al. (2016b), nous avons donc introduit en même temps un nouvel estimateur a noyau et a biais réduit.

Sans oublier les données complètes, dans le troisième volet de cette thèse nous ajoutons un nouvel estimateur de l'indice des valeurs extrêmes, caractérisant par sa robustesse et sa stabilité et a été développé en utilisant l'idée qui a été présentée dans Basu 1998 [2] basant sur la fonction de divergence de puissance de la densité.

DEDICATION

I would like to dedicate this humble work to:
My dear parents
To my brothers and sisters
To my dear brother Nour el-Islam
To my beloved son Yahia

And I do not forget those who left us last year
I dedicate this thesis to the spirit of my grandmother, my aunt, and my uncle
"may God have mercy on them".

Saida...

Acknowledgements

Firstly, and foremost, I would like to thank God for giving me the strength, knowledge, ability, and opportunity to undertake this research study and to persevere and complete it satisfactorily.

I would like to take this opportunity to express my gratitude to my supervisor **Prof. Abdelhakim Necir** for the continuous support of my PHD study, and for his patience and motivation. Thank you Sir for your encouragement, advice, and support through ups and downs all those years.

I also thank those who agreed to be the referees of this thesis and allocated their valuable time in order to evaluate the quality of this work: **Porf. Djamel Meraghni**, **Porf. Djabrane Yahia**, **Porf. Fateh Benatia** and **Prof. Zaher Mohdeb** for their examination of the report and their valuable comments.

I wish to express my special appreciation to my sister Dr. Soumia Mancer and my brother Dr. M'hamed Mancer for their consistent support during my study years, and also my dearest brother Haroune who always gave me a hand when I was stacked.

My sincere thanks go to my family ever, my parents, my brother, my sisters, and my son for their deep love and sincere support for me. Thank you so much.

Abbreviations and Notations

The various abbreviations and notations used throughout this thesis are explained below.

$\bar{F}$	tail of the F distribution F
x_F	the end point of F, equals to $\sup\{x : F(x) < 1\}$
$(X_{1,n}, \ldots, X_{n,n})$	order statistics associated to $(X_1, \ldots, X_n)$
$a_n = o(b_n)$	$a_n/b_n \to 0$ as $n \to \infty$
$a_n = O(b_n)$	a_n/b_n is bounded
$\xrightarrow{P}$	convergence in probability
$\xrightarrow{\mathfrak{D}}$	convergence in distribution
$a_n = o_p(b_n)$	$a_n/b_n \xrightarrow{P} 0$
A. bias	Absolute bias
RMSE	Root Mean Squared Error
MSE	Mean Squared Error

TABLE OF CONTENTS

Page

List of Tables

List of Figures

Figure **Page**

Introduction

Since time immemorial, human has sought to develop his knowledge, techniques and sciences in order to facilitate his daily life. One of these sciences is statistics. People use statistics for a variety of purposes, including as predicting the weather and preparing for natural disasters like earthquakes, floods...etc. Also, it makes it possible to help humanity to prevent certain diseases by determining the probability for a person to develop the disease or to study the evolution of its waves over time in terms of its rate of spread and the severity of its effects, as was recently the case with Corona virus.

Global Overview and Motivation

Generally, several areas of research are developed in statistics namely data analysis, inferential statistics, order statistics, and the extreme values theory (EVT). The EVT theory has received much attention from many researchers, such as Laurens de Hann who is known as the father of this theory. Various subjects were developed such as extreme value index estimation (Hill, 1975) and the Peaks-Over-Threshold (POT) method (Balkema and de Haan, 1975). Remember that the principle of EVT is to study events with a low probability of occurrence having serious consequences for human beings, property, and the environment. From this point, the applications of this theory become to cover many fields, for example, in hydrology to predict floods, in insurance to predict major disasters, in oceanography to study rogue waves, in meteorology, demography, etc..

The major aim of EVT is to know the asymptotic behaviour of extremes (maximum and minimum). In other words, it allows the study of the behaviour of the distribution tails from the largest observed data. Researchers Fisher and

Tippett prove that the distributions tail of the extremes is one of the three cases (according to the value of index γ called extreme value index): Weibull ($\gamma < 0$), Gumbull ($\gamma = 0$), or Frechet ($\gamma > 0$). Thereby, estimating the γ index becomes quite important.

In fact, literature is full of estimators of such an index. We can mention here the works of Hill 1975, Pickand, Peng,...etc. Such estimators deal in particular with complete data. Recently, and even for the case of incomplete data, researchers have proposed a very specific estimators of the tail index.

It should be noted here that incomplete data represent the case in which we lose some observations about the phenomenon under study. There are basicly two forms of this data:

- ***Censoring data***: some observations will be censored, meaning that we only know that they are below (or above) some bound. This can for instance occur if we measure the concentration of a chemical in a water sample. If the concentration is too low, the laboratory equipment cannot detect the presence of the chemical. It may still be present though, so we only know that the concentration is below the laboratory's detection limit.

- ***Truncation data***: the process generating the data is such that it only is possible to observe outcomes above (or below) the truncation limit. This can for instance occur if measurements are taken using a detector that only is activated if the signals it detects are above a certain limit. There may be lots of weak incoming signals, but we can never tell using this detector.

Main Objectives

The goal of this thesis is to introduce a new approach method to estimate the tail index under randomly right truncated data.

- The first objective of the thesis is to derive an estimator for the tail index of Pareto -type distribution that is randomly right truncated based on the semi-parametric estimation method and establish its consistency and asymptotic normality.

- The second objective is to introduce a bias reduction for the developed kernel estimators of the tail index of under right truncated Pareto-type distributions.

- The third objective is the robustness of smoothed estimators for the Pareto-tail index.

Thesis Structure

This thesis is organized as follows:

Chapter 1 is dedicated to the preliminary concepts concerning the EVT, regular variation, order statistics, censoring, and truncated data, which will be used through the thesis. It ends by surveying the estimation for the tail index under the right truncated data.

Chapter 2 provides a full description of the semi-parametric estimation for the extreme value index of the **Pareto-type distribution** of randomly right truncated data, a simulation study, and an application to a real dataset of induction times of AIDS diseases is done.

Chapter 3 introduces a bias reduction to a kernel estimation of the tail index of randomly right truncated **Pareto-type distribution**, and its asymptotic normality is made. The finite sample behaviour of the proposed estimator is checked by a simulation study. We end by giving an application to a real dataset of lifetimes of auto-mobile brake pads.

Chapter 4 presents the robustness of the smooth estimators by beginning to introduce the crucial function called **weighted minimum density power divergence** leading to this class of estimators. We also establish the consistency and asymptotic normality of this estimator, and its finite sample behaviour is carried out by simulation study.

Publications

As a result of this thesis the following publications were produced:

- Mancer, S., Necir, A., & Benchaira, S. (2022). Semiparametric tail-index estimation for randomly right-truncated heavy-tailed data. Arab Journal of Mathematical Sciences, (ahead-of-print).
- Mancer, S., Necir, A., & Benchaira, S. (2022). Bias Reduction in Kernel Tail Index Estimation for Randomly Truncated Pareto-Type Data. Sankhya A, 1-38.
- A weighted minimum density power divergence estimator for the Pareto tail index, with Prof. **Abdelhakim Necir** and Prof. **Djamel Meraghni**. *(in preparation)*.

CHAPTER 1

PRELIMINARIES

It seemed necessary to us before approaching the subject to recall some basic notions so that the reader understands the sequence of this thesis. For this, the extreme values theory and their properties are presented in section 1. Then, the incomplete data are reviewed in sections 2 and 3 by interesting in their definitions with explanatory examples. We end the chapter with results already obtained by other authors concerning the estimation of the index of extreme values under randomly right-truncated data.

1.1 Extreme value theory

Extreme value theory is concerned with the behaviour of tails of distributions. In other words, it studies phenomena whose probability of occurrence is negligible. It is particularly interested in the asymptotic law of the maximum and the minimum of the observations. As $\min(X_1,...,X_n) = -\max(-X_1,...,-X_n)$, then it suffices to study the behaviour of the maximum (max), then reformulate the results for the minimum (min).

Throughout this section, we denote by X a random variable (rv) defined over some probability space $(\Omega, \mathscr{A}, \mathbf{P})$ with continuous cumulative distribution function $\mathbf{F}$ (cdf) and its survival function or the well-known tail distribution: $\overline{\mathbf{F}} = \mathbf{1} - \mathbf{F}$. We also introduce $x_F := \sup\{x \in \mathbb{R},\ F(x) < 1\}$ the end point of F, and $u \leq x_F$ a real number called threshold.

1.1.1 Limit law distribution of Maxima

Let $X_1,...,X_n$ be a sample of independent identically distributed (iid) rv's from the distribution $\mathbf{F}$ and $X_{1,n},...,X_{n,n}$ the order statistics associated, where

$$X_{1,n} := \min(X_1,...,X_n) \text{ and } X_{n,n} := \max(X_1,...,X_n)$$

We know that the law of the maximum is defined as follows:

$$\begin{aligned} F_{X_{n,n}}(t) &= P(X_{n,n} \leq t) = P(X_1 \leq t,...,X_n \leq t) \\ &= [\mathbf{F}(t)]^n \end{aligned}$$

then,

$$\lim_{n\to\infty} F_{X_{n,n}}(x) = \lim_{n\to\infty} [\mathbf{F}(x)]^n = \begin{cases} 1 & \text{if } x \geq x_F \\ 0 & \text{if } x < x_F \end{cases}$$

In practice, the law $\mathbf{F}$ is unknown, then the behavior of $F_{X_{n,n}}$ will be even more difficult to study because this distribution is a degenerate law and this result provides very little information on the asymptotic behaviour of the max $X_{n,n}$.

Our goal is to obtain a non-degenerate law, but how?

Well, the idea is to carry out a transformation, the best known in statistics is the normalization illustrated through the Central Limit Theorem TCL which gives a non-degenerate asymptotic distribution of the mean $\overline{X}_n$ of n rv's.

Theorem 1.1 (Central Limit Theorem). *Let $X_1,...,X_n$ be a sequence of iid rv's with mean μ and finite variance σ^2, then*

$$\frac{\sqrt{n}\left(\overline{X}_n-\mu\right)}{\sigma}\xrightarrow[n\to\infty]{D}N(0,1)$$

equivalent to

$$\frac{\overline{X}_n-b_n}{a_n}\xrightarrow[n\to\infty]{D}N(0,1)$$

where: $a_n=\frac{\sigma}{\sqrt{n}}>0$ and $b_n\in\mathbb{R}$ are called normalization constant.

The proof of this theorem could be found in any standard book of statistics, see for instance, Saporta, G.(1990) page 66.
Similarly, for the maximum $X_{n,n}$ there exists $a_n>0$ and $b_n\in\mathbb{R}$ such that:

$$\lim_{n\to\infty}P\left(\frac{X_{n,n}-b_n}{a_n}\leq x\right)=\lim_{n\to\infty}\mathbf{F}^n(a_nx+b_n)=H(x),\ \forall x\in\mathbb{R} \tag{1.1}$$

where: H is a non-degenerate distribution function called the extreme value distribution.

It is worthy to know that Fisher was the first who observed the behaviour of the maximum and its variations. The crucial theorem in EVT is the following.

Theorem 1.2 (Fisher and Tippett). *If cdf $\mathbf{F}$ satisfies assumption (1.1), then cdf H is the same, up to location and scale, as one of the following distribution:*

Gumbel: $\Lambda(x)=\exp(-\exp(-x)),\ x\in\mathbb{R}.$

Frechet: $\Phi_\xi(x)=\begin{cases}0 & ,x\leq 0\\ \exp(-x^{-\xi}) & ,x>0\end{cases},\xi>0$

Weibull: $\Psi_\xi(x)=\begin{cases}\exp(-\left(-x^{\xi}\right)) & ,x\leq 0\\ 1 & ,x>0\end{cases},\xi>0$

In the following paragraph, we consider rv $(X_1,...,X_n)$ iid as well as their maximum $X_{n,n}$. We are looking for sequences $a_n > 0$ and $b_n \in \mathbb{R}$, such that the sequence $\left(a_n^{-1}\left(X_{n,n} - b_n\right)\right)$ converges in law to a non-degenerate limit. We consider random variables with uniform, exponential and Cauchy distributions.

1.1.2 Generalized extreme value distribution

Definition 1.1 (Generalized extreme value distribution)**.** *The generalized extreme value distribution GEVD is a df H_γ defined for all $x \in \mathbb{R}$ such that $1 + \gamma x > 0$, as follows*

$$H_\gamma(x) = \begin{cases} \exp\left(-\left(1+\gamma x\right)^{-1/\gamma}\right) & , \text{if } \gamma \neq 0 \\ \exp(-e^{-x}) & , \text{if } \gamma = 0 \end{cases}$$

the real parameter γ is called extreme value index, tail index or also shape parameter.

In the following proposition, we express the three extreme value distribution Λ, Φ_ξ and Ψ_ξ in terms of the GEVD H_γ.

Proposition 1.1. *We note that:*

$$\begin{cases} \Phi_\xi(x) = H_{1/\xi}(\xi(x-1)) & , x > 0 \\ \Psi_\xi(x) = H_{-1/\xi}(\xi(x+1)) & , x < 0 \\ \Lambda(x) = H_0(x) & , x \in \mathbb{R} \end{cases}$$

Remark 1.1. *Given the above proposition, we can write:*

$$H_\gamma = \begin{cases} \Psi_{-1/\gamma}(x) & , \gamma < 0 \\ \Lambda & , \gamma = 0 \\ \Phi_{1/\gamma}(x) & , \gamma > 0 \end{cases}$$

we recall that Gumbel type corresponds to $\gamma = 0$ *, Frechet type to $\gamma > 0$ and Weibull type to $\gamma < 0$.*

1.1.2.1 Properties of each distribution

$\gamma > 0$: the endpoint associated to the distribution $\Phi_{\frac{1}{\gamma}}$ is $x_F = +\infty$, i.e., this distribution has a heavy tail and the moments of order greater than or equal to $\frac{1}{\gamma}$ do not exist.

$\gamma = 0$: the endpoint associated to the distribution Λ is $x_F \leq +\infty$, but, this distribution has a light tail "Light tail" and all the moments exist.

$\gamma < 0$: the endpoint associated to the distribution $\Psi_{-\frac{1}{\gamma}}$ is $x_F = -\frac{1}{\gamma}$, so this distribution has a "Short tail".

1.1.3 Regular variation and Domains of attraction

The EVT could not have been developed and come into being without the use and application of the theory of regularly varying functions (RV) and its properties. The notion of regularly varying functions was introduced by J. Karamata in 1930 in the case of continuous functions, while the case of measurable functions was treated by Korevaar in 1949; these functions are continuous, differentiable, ... etc. For more details you can see [9].

The set of regularly varying functions has generalizations and extensions they are called first and second-order condition. These last are the most substantial notions for which the EVT should be well understood to study.

1.1.3.1 Regular variation

The regularly varying functions are exactly the functions that model the tails $\overline{\mathbf{F}}$ of the distributions $\mathbf{F}$ and the first-order condition is exactly the necessary and sufficient condition which ensures that the distribution $\mathbf{F}$ is in the domains of attraction of the max. The role of this condition appeared in the construction of the estimators of the tail index, and the second-order condition gives the consistency and the asymptotic normality of this estimators.

Definition 1.2. *f is said to be a function with regular variation at infinity of* *index ρ (noted $f \in RV_\rho$) if it satisfies:*

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} = x^\rho \quad \forall x > 0$$

Definition 1.3. *f is said to be a function with regular variation to the right of 0 with index $(-\rho)$, if for all $x > 0$:*

$$\lim_{s\to 0} \frac{f(sx)}{f(s)} = x^{-\rho},$$

and we note $f \in RV_{-\rho}(+0)$.

Remark 1.2.

- *If $\ell \in VR_0$, then we say that ℓ is a slowly varying function at infinity.*
- *Any measurable function, positive, and its limit is positive (in particular, the positive constants), is a slowly varying function at infinity.*
- *If $f \in RV_\rho$ and $\ell \in RV_0$, then $f(x) = x^\rho \ell(x)$.*

Theorem 1.3 (Karamata's representation)**.** *We say that ℓ is a slowly varying function at infinity iff:*

$$\ell(t) = c(t)\exp\left\{\int_a^t \varepsilon(u)du/u\right\} \quad ,t \geq a$$

where $c(t) \to c > 0$ and $\varepsilon(t) \to 0$ as $t \to \infty$.

Example 1.1.

- *The functions: $x^\rho \log(1+x)$, $x^\rho \log\log(e+x)$, $x^\rho \log(x)$, and $x^\rho \exp\{(\log x)^\alpha\}$, $0 < \alpha < 1$ are regularly varying functions at infinity.*
- *Whereas, $2+\sin x$, $2+\sin\log x$, and $x\exp\sin\log x$ are not regularly varying functions.*

Theorem 1.4 (Uniform convergence theorem)**.** *If f is a regularly varying functions at infinity of index ρ, then:*

$$\sup_{x\in A} \left| \frac{f(\lambda x)}{f(x)} - x^\rho \right| \to 0 \;\; as \;\; x \to \infty,$$

where:

$$A=\begin{cases}[a,b] & \text{if } \rho=0.\\ ]0,b] & \text{if } \rho>0, \text{ suppose that } f \text{ is bounded on }]0,b].\\ [a,\infty[& \text{if } \rho<0.\end{cases}$$

Proposition 1.2.

1. *If $f \in RV_\rho$, then $f^\alpha \in RV_{\alpha\rho}$.*

2. *If f_1, $f_2 \in RV_\rho$ and $f_2(x) \to \infty$ as $x \to \infty$, then $f_1 \circ f_2 \in RV_{\rho_1\rho_2}$.*

3. *If f_1, $f_2 \in RV_\rho$, then $f_1 + f_2 \in RV_\rho$ where $\rho = \max(\rho_1, \rho_2)$.*

4. *(Potter's inequality, 1942) If $f \in RV_\rho$, then there exists t_0 such that, for all $x \geq 1$, $t \geq t_0$:*

$$(1-\varepsilon)x^{\rho-\varepsilon} < \tfrac{f(tx)}{f(t)} < (1+\varepsilon)x^{\rho+\varepsilon}, \quad \forall \varepsilon > 0.$$

Proposition 1.3 (Drees(1998))**.** *If $f \in RV_\rho$, then there exists $t_0 = t_0(\varepsilon,\delta)$ such that, for $t, tx \geq t_0$:*

$$\left|\tfrac{f(tx)}{f(t)} - x^\rho\right| \leq \varepsilon \max\left(x^{\rho+\delta}, x^{\rho-\delta}\right) \quad \forall \varepsilon, \delta > 0.$$

Theorem 1.5 (Karamata's theorem)**.** *Let $f \in RV_\sigma$, and locally bounded on $x_0 \leq x < \infty$.*

If $\sigma \geq -1$, then:

$$\lim_{x\to\infty} \frac{xf(x)}{\int_{x_0}^{x} f(t)dt} = \sigma + 1 \tag{1.2}$$

If $\sigma < -1$ (If $\sigma = -1$ and $\int_x^\infty f(t)dt < +\infty$), then:

$$\lim_{x\to\infty} \frac{xf(x)}{\int_{x}^{\infty} f(t)dt} = -(\sigma + 1) \tag{1.3}$$

Conversely, if f verifies (1.2) for $-1 < \sigma < \infty$, then $f \in RV_\sigma$. While if f verifies (1.3) with $-\infty < \sigma < -1$, then $f \in RV_\sigma$.

1.1.3.2 Domains of attraction

In order to characterize the domains of attraction, we need to propose necessary and sufficient conditions on F so that it can be in $D_M\ (H_\gamma)$. these conditions are known by the ***conditions of von Mises (1936)***. The interested reader is referred to [24] for more details about such conditions. Since our work focuses on the case of heavy-tailed distributions, we are only interested in the domains of attraction of the latter.

Definition 1.4. *The set of laws F satisfying (1.1) is called the domain of attraction of H_γ, and we write $F \in D_M\ (H_\gamma)$.*

Let U be a quantile function such that $U(t) = F^{-1}(1-\frac{1}{t}), \forall t > 1$. For all $\gamma \in \mathbb{R}$, we have:

$$F \in\ D_M\ (H_\gamma) \Leftrightarrow U \in RV_\gamma.$$

i-e, for all $x > 0$:

$$\lim_{t\to\infty} \frac{U(tx)-U(t)}{a(t)} = \begin{cases} \frac{x^\gamma-1}{\gamma} & ,\gamma \neq 0 \\ \ln x & ,\gamma = 0. \end{cases} \tag{1.4}$$

In this case, (1.4) is the first order condition for the regularly varying function; we have a special case for $\gamma > 0$ where $a(t) = \gamma U(t)$.
By replacing $a(t) = \gamma U(t)$ in (1.4) we get:

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma \quad \forall x > 0. \tag{1.5}$$

Thus, the first order condition for regularly varying functions is defined, *i.e*:

$$F \in D_M(\Phi_{\frac{1}{\gamma}}) \Leftrightarrow U \in RV_\gamma. \tag{1.6}$$

Depending on $\bar{F}$:

$$F \in D_M(\Phi_{\frac{1}{\gamma}}) \Leftrightarrow \bar{F} \in RV_{-\frac{1}{\gamma}}. \tag{1.7}$$

i-e, for all $x > 0$:

$$\lim_{t\to\infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\frac{1}{\gamma}} \Longleftrightarrow \bar{F}(x) = x^{-\frac{1}{\gamma}}\ell(x)$$

where: $\ell \in RV_0$.

1.1.4 Estimation of the extreme value index

There are several methods and techniques to estimate the extreme value index. In this part of chapter, we are limited to three principle methods.

Let $X_1,...,X_n$ be independent and identically distributed (iid) of non-negative random variables (rv's) as n copies of a rv X, defined over some probability space $(\Omega, \mathscr{A}, \mathbf{P})$, with cumulative distribution function (cdf) F. We assume that the tail distribution $\overline{F} := 1 - F$ is regularly varying at infinity with negative index $(-1/\gamma)$, *i.e.*, for every $x > 0$,

$$\frac{\overline{F}(xu)}{\overline{F}(u)} \rightarrow x^{-1/\gamma}, \text{ as } u \rightarrow \infty. \tag{1.8}$$

The parameter $\gamma > 0$ is called the shape parameter or the tail index or the extreme value index (EVI). It plays a very crucial role in the analysis of extremes as it governs the thickness of the distribution right-tail. The problem of estimating the EVI has received a lot of attention in the last four decades.

1.1.4.1 Hill estimator

The most popular estimator of γ is Hill's estimator [25], defined by

$$\widehat{\gamma}_k^{(H)} := \frac{1}{k} \sum_{i=1}^{k} \log \frac{X_{n-i+1:n}}{X_{n-k:n}} = \sum_{i=1}^{k} \frac{i}{k} \log \frac{X_{n-i+1:n}}{X_{n-i:n}},$$

where $X_{1:n} \leq ... \leq X_{n:n}$ denote the order statistics pertaining to the sample $X_1,...,X_n$ and $k = k_n$ is an integer sequence satisfying $1 < k < n$, $k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$. The discrete character and non-stability of Hill's estimator present major drawbacks. Indeed, adding a single large-order statistic in the calculation of the estimator, that is, increasing k by 1, may deviate from the true value of the estimate substantially. Thus, the plotting of this estimator as a function of the upper order statistics often gives a zig-zag figure. To overcome this issue, the authors introduce the following estimator

1.1.4.2 Csörgő et *al.*(1985) estimator

Csörgő et *al.*(1985) (CDM) introduced more general weighs instead of the natural one i/k that appears in the second formula of $\widehat{\gamma}_k^{(H)}$, to define the following kernel estimator

$$\widehat{\gamma}_{k,K}^{(CDM)} := \sum_{i=1}^{k} \frac{i}{k} K\left(\frac{i}{k}\right) \log \frac{X_{n-i+1:n}}{X_{n-i:n}}, \tag{1.9}$$

where K is a continuous nonnegative nonincreasing function on $(0,1)$ such that $\int_0^1 K(s)ds = 1$.

1.1.4.3 Hüsler *et al.*(2006) estimator

The work presented in [28] used the weighted least squares estimator (WLSE) given by

$$\widehat{\gamma}_{k,J} := \frac{\frac{1}{k}\sum_{i=1}^{k} J\left(\frac{i}{k}\right) \log \frac{X_{n-i+1:n}}{X_{n-k:n}}}{\int_0^1 J(s) \log s^{-1} ds}, \tag{1.10}$$

where J is a suitable continuous nonnegative nonincreasing function defined on $(0,1)$ such that $\int_0^1 J(s)ds = 1$. Similar estimators to $\widehat{\gamma}_{k,J}$ are also considered in [**?**], [**?**] and recently [11]. The authors pointed out that the least squares estimator $\widehat{\gamma}_{k,J}$ may be rewritten into CDM's one $\widehat{\gamma}_{k,K}^{(CDM)}$ for the kernel function

$$K(s) = \frac{1}{s} \int_0^s J(t)dt \left(\int_0^1 J(t) \log t^{-1} dt \right)^{-1}, \ s \in (0,1). \tag{1.11}$$

For example, let $J_{\log}(s) = -\mathbf{1}_{(0,1)}(s) \log s$, we get $K(s) = \mathbf{1}_{(0,1)}(s)(1 - \log s)$. The converse is not necessarily true, *i.e.*, it may be K is a kernel function without the function J being. For example, taking $K(s) := -\mathbf{1}_{(0,1)}(s) \log s$, we get $J(s) = 0$. The notation $\mathbf{1}_A$ stands for the indicator function of a set A. The commonly used kernel functions J are: the indicator $J_0 := \mathbf{1}_{(0,1)}$, the linear-weight, biweight, triweight and quadweight functions defined on $0 < s < 1$ by

$$\begin{gathered} J_1(s) := 2(1-s), \qquad J_2(s) := \frac{15}{8}\left(1-s^2\right)^2, \\ J_3(s) := \frac{35}{16}\left(1-s^2\right)^3, \quad J_4(s) := \frac{315}{128}\left(1-s^2\right)^4, \end{gathered} \tag{1.12}$$

and zero elsewhere respectively, where $\mathbf{1}_A$ stands for the indicator function of a set A. For the use of this type of weight functions one refers to [19] and [28].

The nice properties of this type of estimators are the smoothness and the stability, contrary to Hill's one which rather exhibits fluctuations along the range of upper extreme values. However, these estimators do not take into account possible deviations from assumed extreme value models. These may arise as a result of possible outliers in the data that may (or may not) have been recorded in error. In such a dataset, the estimators mentioned above are known to be sensitive to these outlying observations, affecting their quality. In addition, small errors in the estimation of model parameters, such as the tail index, can cause significant errors in the estimation of extreme events such as high quantiles and exceedance probabilities (see e.g., Brazauskas and Serfling 2000). And this what we will see in the last chapter of this thesis.

1.2 Incomplete data

In this section, our interest is to present briefly the incomplete data which includes censoring and truncating data. To truncate data is to completely delete a value from the dataset, whereas to censor data is to simply capture a portion of information about a value. Both of them induce information loss in a dataset, however truncating causes more information loss since it requires completely eliminating some data values.

1.2.1 Cencoring data

Censoring data values means collecting only partial information that is higher or smaller than a certain value. We may clarify this definition using the example below.

Example 1.2 (A vitamin D test). *A vitamin **D** test, also known as hydroxyvitamin D(25OH), is one of the best ways to monitor our body's vitamin **D** levels. The test determines whether these levels are too high, too low, or normal. The analysis is performed by taking a simple blood sample from the vein of the patient who needs to know his vitamin **D** level. It is measured according to the Office of Dietary Supplements (ODS) by measuring the level of calcifediol in units (nmol/L) or (ng/mL).*

In medical laboratories, we can record the following values:

- *Patient 1:* 10.23 *ng*/*mL*.
- *Patient 2:* < 8 *ng*/*mL*.

In this case, for Patient 2, we only know that his vitamin level is less than 8*, and we cannot know the exact value as we knew it for Patient 1.*

1.2.1.1 Types of censoring

In literature we have three types of consoring.

1. *Right censoring:*

   The variable of interest is said to be right-censored if the individual concerned has no information about its last sighting. Thus, in the presence of the right censorship, the variables of interest are not all observed. A typical example is where the event considered is the death of a sick patient and the duration of observation is the total duration of hospitalization. We also find this type of phenomenon in reliability studies when the failure of a device or an electronic component does not allow continued observation of another device or component. We can also find these kinds of phenomena in hydrology, rainfall, etc. The experimenter can set an end date of the experiment and the observations for the individuals for whom the event of interest was not observed before that date will be censored at right.

2. *Left censoring:*

   There is left censoring when the individual has already undergone the event before he is observed. We only know that the variable of interest is less than or equal to a known variable. For example, if we want to study the reliability of a certain electronic component that is connected in parallel with one or more other components: the system can continue to operate, although aberrant way until this failure is detected (for example during control or in the event of a system shutdown). Thus, the duration observed for this component is left censored. In everyday life, there are several phenomena that present both right and left censored data.

3. *Interval censoring:*

   In this case, as its name suggests, we observe both a lower bound and an upper bound of the variable of interest. We find this model usually in medical follow-up studies where patients are checked periodically if a patient does not show up for one or more check-ups and then presents itself after the event of interest has occurred. We also have this kind of data which is right censored or, more rarely, to the left. An advantage of this type is that it allows data to be presented right or left censored by intervals of the type $[c, +\infty[$ and $[0, c]$ respectively.

1.2.2 Truncated data

Truncation of data values means removing valuesfrom a set of data that are less than or greater than a certain value. For concrete examples of truncated data in medical treatments one refers, among others, to [31] and [48]. Truncated data schemes may also occur in many other fields, namely actuarial sciences, astronomy, demography and epidemiology, see for instance the textbook of [32].

Also, we have three types of the truncated data which are:

1. *Left truncated:*

   When people below a threshold are absent from the sample, our data are left-truncated. For instance, fish smaller than the net grid won't be included in our sample if we want to determine the size of a certain fish using specimens caught using a net.

2. *Right truncated:*

   Think about the AIDS study that is discussed in chapter 2's section 5. Here, samples of people with AIDS caused by transfusions were taken. The amount of time between an infection at the moment of transfusion and the onset of clinical AIDS was calculated retrospectively using the transfusion times. Only people who had experienced AIDS before their waiting period from transfusion to June 30, 1986 the date the registry was sampled were accessible for observation. Patients who received blood transfusions before June 30, 1986 but contracted AIDS after that date were not tracked down and are right-truncated.

3. *Interval truncated:*

   Or if an individual is potentially seen and only if its failure-time falls inside a specific range, specific to that individual, doubly shortened failure-time develops. The statistical analysis of astronomical observations and survival analysis both heavily rely on doubly truncated data.

In this thesis, we are interested in the randomly right-truncated, for that we drive, in the next section, the related works with Tail index estimation under right truncating data.

1.3 Tail index estimation under right truncating data

This section discusses the estimators for the tail index that the authors recently proposed. Before that, let us present the most important elements used from this part to the end of the thesis.

Let $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, ..., N \geq 1$ be a sample from a couple $(\mathbf{X}, \mathbf{Y})$ of independent positive random variables (rv's) defined over a probability space $(\Omega, \mathscr{A}, \mathbf{P})$, with continuous distribution functions (df's) $\mathbf{F}$ and $\mathbf{G}$ respectively. Suppose that $\mathbf{X}$ is right-truncated by $\mathbf{Y}$, in the sense that $\mathbf{X}_i$ is only observed when $\mathbf{X}_i \leq \mathbf{Y}_i$. Thus, let us denote (X_i, Y_i), $i = 1, ..., n$ to be the observed data, as copies of a couple of dependent rv's (X, Y) corresponding to the truncated sample $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, ..., N$, where $n = n_N$ is a random sequence of discrete rv's. By the weak law of large numbers, we have

$$n/N \overset{\mathbf{P}}{\to} p := \mathbf{P}(\mathbf{X} \leq \mathbf{Y}) = \int_0^\infty \mathbf{F}(w) d\mathbf{G}(w), \text{ as } N \to \infty, \tag{1.13}$$

where the notation $\overset{\mathbf{P}}{\to}$ stands for the convergence in probability. The constant p corresponds to the probability of observed sample which is supposed to be

non-null, otherwise nothing is observed. The truncation phenomena frequently occurs in medical studies, when one wants to study the length of survival after the start of the disease: if **Y** denotes the elapsed time between the onset of the disease and death, and if the follow-up period starts **X** units of time after the onset of the disease then, clearly, **X** is right-truncated by **Y**.
From [18] the marginal df's F^* and G^* corresponding to the joint df of (X,Y) are given by

$$F^*(x) := p^{-1}\int_0^x \overline{\mathbf{G}}(w)d\mathbf{F}(w) \text{ and } G^*(x) := p^{-1}\int_0^x \mathbf{F}(w)d\mathbf{G}(w).$$

By the previous first equation we derive a representation of the underlying df **F** as follows:

$$\mathbf{F}(x) = p\int_0^x \frac{dF^*(w)}{\overline{\mathbf{G}}(w)}, \tag{1.14}$$

which will be for a great interest thereafter. In the sequel, we are dealing with the concept of regular variation. A function φ is said to be regularly varying at infinity with negative index $-1/\eta$, notation $\varphi \in RV\left(-1/\eta\right)$, if

$$\varphi(st)/\varphi(t) \to s^{-1/\eta}, \text{ as } t \to \infty, \tag{1.15}$$

for $s > 0$. This relation is known as the first-order condition of regular variation and the corresponding uniform convergence is formulated in terms of "Potter's inequalities" as follows: for any small $\epsilon > 0$, there exists $t_0 > 0$ such that for any $t \geq t_0$ and $s \geq 1$, we have

$$(1-\epsilon)s^{-1/\eta-\epsilon} < \varphi(st)/\varphi(t) < (1+\epsilon)s^{-1/\eta+\epsilon}. \tag{1.16}$$

See for instance Proposition B.1.9 (assertion 5, page 367) in [24]. The second-order condition ,see [20], expresses the rate of the convergence (1.15) above. For any $x > 0$, we have

$$\frac{\varphi(tx)/\varphi(t) - x^{-1/\eta}}{A(t)} \to x^{-1/\eta}\frac{x^{\tau/\eta}-1}{\tau\eta}, \text{ as } t \to \infty, \tag{1.17}$$

where $\tau < 0$ denotes the second-order parameter and A is a function tending to zero and not changing signs near infinity with regularly varying absolute value with positive index τ/η. A function φ that satisfies assumption (4.9) is denoted $\varphi \in RV_2\left(-1/\eta;\tau,A\right)$. We now have enough material to tackle the main goal of the paper. To begin, let us assume that the tails of both df's **F** and **G** are regularly varying. That is

$$\overline{\mathbf{F}} \in RV\left(-1/\gamma_1\right) \text{ and } \overline{\mathbf{G}} \in RV\left(-1/\gamma_2\right), \text{ with } \gamma_1, \gamma_2 > 0. \tag{1.18}$$

Under this assumption, [18] showed that

$$\overline{F}^{*} \in RV\left(-1/\gamma\right) \text{ and } \overline{G}^{*} \in RV\left(-1/\gamma_2\right), \tag{1.19}$$

where

$$\gamma := \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2}. \tag{1.20}$$

For details on the proof of this statement, on refers to [6] (Lemma A1).

1.3.1 Gardes and Stupfler estimator

Recently Gardes and Stupfler(2015) introduced an estimator of γ_1 defined by:

$$\hat{\gamma}_1^{(\mathbf{GS})}(k_1, k_2) := \frac{\hat{\gamma}_2(k_2)\hat{\gamma}(k_1)}{\hat{\gamma}_2(k_2) - \hat{\gamma}(k_1)},$$

where k_1 and k_2 are two distinct sample fraction used respectively in Hill's estimator of tail indices γ and γ_2

$$\hat{\gamma}(k_1) := \frac{1}{k_1} \sum_{i=1}^{k_1} \log X_{n-i+1:n} - \log X_{n-k_1:n}$$

and

$$\hat{\gamma}_2(k_2) := \frac{1}{k_2} \sum_{i=1}^{k_2} \log Y_{n-i+1:n} - \log Y_{n-k_2:n}$$

These estimators are based on the top order statistics $X_{n-k:n} \leq \ldots \leq X_{n:n}$ and $Y_{n-k:n} \leq \ldots \leq Y_{n:n}$ pertaining to the samples $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ respectively.

1.3.2 Benchaira *et al* estimator

Benchaira *et al*.(2015) in [5] considered a sample fraction $k = k_1 = k_2$ satisfying $k \to \infty$ and $k/n \to 0$ as $n \to \infty$ and they defined the following estimators of γ and γ_2

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

and

$$\hat{\gamma}_2 := \frac{1}{k}\sum_{j=1}^{k} \log Y_{n-j+1:n} - \log Y_{n-k:n}$$

and

$$\hat{\gamma}_1^{\mathbf{(BMN1)}} := \frac{\frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k} \log \frac{X_{n-i+1:n}}{X_{n-k:n}} - \log \frac{Y_{n-j+1:n}}{Y_{n-k:n}}}{\sum_{i=1}^{k} \log\left(\frac{Y_{n-i+1:n}X_{n-k:n}}{Y_{n-k:n}X_{n-i+1:n}}\right)}.$$

they also provided a Gaussian representation in terms of two-parameter Wiener process which leads to consistency and asymptotic normality of $\hat{\gamma}_1^{(BMN1)}$.

1.3.3 Worms and Worms estimator

By using a Lynden-bell integral, Worms and Worms(2016) in [50] proposed the following estimator for the tail index γ_1 :

$$\widehat{\gamma}_1^{\mathbf{(W)}}(u) := \frac{1}{\overline{\mathbf{F}}_n^{\mathbf{(1)}}(u)} \sum_{i=1}^{n} \mathbf{1}(X_i > u) \frac{\mathbf{F}_n^{\mathbf{(1)}}(X_i)}{C_n(X_i)} \log \frac{X_i}{u},$$

for a given deterministic threshold $u > 0$, where

$$\mathbf{F}_n^{\mathbf{(1)}}(x) := \prod_{X_i > x} \left[1 - \frac{1}{nC_n(X_i)}\right],$$

is the popular nonparametric maximum likelihood estimator of cdf $\mathbf{F}$ introduced in the well-known work Lynden-Bell(1971) [36], with

$$C_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le x \le Y_i).$$

Independently, Benchaira *et al*.(2016a) in [6] used a Woodroofe-integral with a random threshold, to derive the following estimator

$$\widehat{\gamma}_1^{\mathbf{(BMN2)}} := \frac{1}{\overline{\mathbf{F}}_n^{\mathbf{(2)}}(X_{n-k:n})} \sum_{i=1}^{k} \frac{\mathbf{F}_n^{\mathbf{(2)}}(X_{n-i+1:n})}{C_n(X_{n-i+1:n})} \log \frac{X_{n-i+1:n}}{X_{n-k:n}}, \tag{1.21}$$

where

$$\mathbf{F}_n^{\mathbf{(2)}}(x) := \prod_{X_i > x} \exp\left\{-\frac{1}{nC_n(X_i)}\right\},$$

is the so-called Woodroofe's nonparametric estimator of df $\mathbf{F}$.

1.3.4 Kernel estimator

Benchaira *et al.*(2016b) in [7] proposed a Kernel (smoothed) version to $\widehat{\gamma}_1$ given by

$$\widehat{\gamma}_{1,K} := \frac{1}{n\overline{\mathbf{F}}_n\left(X_{n-k:n}\right)} \sum_{i=1}^{k} \frac{\mathbf{F}_n\left(X_{n-i+1:n}\right)}{C_n\left(X_{n-i+1:n}\right)} g_K\left(\frac{\overline{\mathbf{F}}_n\left(X_{n-i+1:n}\right)}{\overline{\mathbf{F}}_n\left(X_{n-k:n}\right)}\right) \log \frac{X_{n-i+1:n}}{X_{n-k:n}}, \qquad (1.22)$$

where g_K is the Lebesgue derivative of function $s \to sK(s)$ and K is a nonnegative kernel function satisfying the following assumptions:

- $[A1]$ $K(s) \geq 0$ for $s \in (0,1]$, otherwise $K(s) = 0$.
- $[A2]$ g_K is positive nonincreasing over some interval in $(0,1]$.
- $[A3]$ $\int_{\mathbb{R}} K(s)\,ds = 1$.
- $[A4]$ K and its first and second Lebesgue derivatives K' and K'' are bounded.
- $[A5]$ K is nonincreasing.

The commonly used kernel functions are: the indicator, the biweight, triweight and quadweight kernels respectively defined, for $0 < s \leq 1$, by

$$K_1 := \mathbf{1}\{[0,1)\},\ K_2(s) := \frac{15}{8}\left(1-s^2\right)^2,$$
$$K_3(s) := \frac{35}{16}\left(1-s^2\right)^3,\ K_4(s) := \frac{315}{128}\left(1-s^2\right)^4,$$

and zero elsewhere, where $\mathbf{1}\{A\}$ stands for the indicator function of a set A. It is worth mentioning that assumptions $[A1]-[A4]$ are usually made to construct and investigate the asymptotic behavior of the tail index kernel estimators, see for instance [13], [19] and [12]. Assumption $[A5]$ will be used later on to define a weighted estimator for the asymptotic bias of $\widehat{\gamma}_{1,K}$. Notice that the second assumption $[A2]$ allows to assign weights which preserve the decreasing nature of the tail $\overline{\mathbf{F}}$. Assuming the second-order conditions

$$\overline{\mathbf{F}} \in RV_2\left(-1/\gamma_1; \tau_1, A_{\mathbf{F}}\right) \text{ and } \overline{\mathbf{G}} \in RV_2\left(-1/\gamma_2; \tau_2, A_{\mathbf{G}}\right), \qquad (1.23)$$

Benchaira *et al.*(2016b) in [7] also showed that, whenever $\gamma_1 < \gamma_2$, one has

$$\sqrt{k}\left(\widehat{\gamma}_{1,K} - \gamma_1\right) \overset{\mathcal{D}}{=} \mathcal{N}\left(0, \left(\gamma^2/\gamma_1\right)^2 \int_0^1 \varphi_K^2(s)\,ds\right) + \eta_{1,K}\sqrt{k}A_{\mathbf{F}}(n/k) + o_{\mathbf{P}}(1), \qquad (1.24)$$

as $n \to \infty$, provided that $\sqrt{k} A_{\mathbf{F}}(n/k) = O(1)$, where $\eta_{1,K} := \eta_{1,K}(\tau_1) = \int_0^1 s^{-\tau_1} K(s) ds$, and

$$\varphi_K(s) := s^{-1} \int_0^s t^{-\gamma/\gamma_2} \left\{ K\left(t^{\gamma/\gamma_1}\right) - \frac{\gamma_1}{\gamma_2} t^{-\gamma_2/\gamma_1} K\left(t^{\gamma/\gamma_1}\right) + t^{\gamma/\gamma_1} K'\left(t^{\gamma/\gamma_1}\right) \right\} dt. \quad (1.25)$$

The condition $\gamma_1 < \gamma_2$ ensures that we have at our disposal enough observations pertaining to the right-tail of $\mathbf{X}$.

1.3.5 Haouas *et al.* estimator

Haouas et al.(2019) in [22] respectively proposed a Kernel-smoothed and a reduced-biais versions of this estimator and establish their consistency and asymptotic normality. It is worth mentioning that Lynden-Bell integral estimator $\widehat{\gamma}_1^{(\mathbf{W})}(u)$ with a random threshold $u = X_{n-k:n}$ becomes

$$\widehat{\gamma}_1^{(\mathbf{W})} := \frac{1}{\overline{\mathbf{F}}_n^{(\mathbf{1})}(X_{n-k:n})} \sum_{i=1}^{k} \frac{\mathbf{F}_n^{(\mathbf{1})}(X_{n-i+1:n})}{C_n(X_{n-i+1:n})} \log \frac{X_{n-i+1:n}}{X_{n-k:n}}. \quad (1.26)$$

In a simulation study, [21] compared this estimator with $\widehat{\gamma}_1^{(\mathbf{BMN})}$. They pointed out that both estimators have similar behaviors in terms of biases and mean squared errors.

CHAPTER 2

SEMIPARAMETRIC TAIL-INDEX ESTIMATION FOR RANDOMLY RIGHT-TRUNCATED HEAVY-TAILED DATA

The present chapter deals with semi-parametric estimation methods that are used for models which are partly parametric and partly non parametric. The choise of this method seems the best when parametric information of the truncation distribution is available. Indeed, the section 1 derive the semiparametric estimator for the distribution function **F** *based on the Conditional Maximum Likelihood method proposed by Wang. Section 2 present the methodology estimation of the new estimator for the tail index under randomly right-truncated heavy-tailed data and we establish in section 3 their consistency and asymptotic normality. The performance of the proposed estimator is checked by simulation in section 4. An application to a real dataset composed of induction times of AIDS diseases is given in section 5.*

2.1 Semi-parametric estimator of the truncation distribution function F

Recall that the nonparametric Lynden-Bell estimator $\mathbf{F}_n^{(1)}$ was constructed on the basis of the fact that **F** and **G** are both unknown. In this paper, we are dealing with the situation when **F** is unknown but **G** is parametrized by a known model $\mathbf{G}_\theta$, $\theta \in \Theta \subset \mathbb{R}^d$, $d \geq 1$ having a density $\mathbf{g}_\theta$ with respect to Lebesgue measure. [48] considered this assumption and introduced a semiparametric estimator for df **F** defined by

$$\mathbf{F}_n\left(x;\widehat{\theta}_n\right) := P_n\left(\widehat{\theta}_n\right)\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}\left(X_i \leq x\right)}{\overline{\mathbf{G}}_{\widehat{\theta}_n}\left(X_i\right)}, \tag{2.1}$$

where

$$1/P_n\left(\widehat{\theta}_n\right) := n^{-1}\sum_{i=1}^{n} 1/\overline{\mathbf{G}}_{\widehat{\theta}_n}\left(X_i\right)$$

and

$$\widehat{\theta}_n := \arg\max_{\theta\in\Theta}\prod_{i=1}^{n} g_\theta\left(Y_i\right)/\overline{\mathbf{G}}_\theta\left(X_i\right), \tag{2.2}$$

denoting the conditional maximum likelihood estimator (CMLE) of θ, which is consistent and asymptotically normal, see for instance [1]. On the other hand, [48] showed that $\mathbf{F}_n\left(x;\widehat{\theta}_n\right)$ is an uniformly consistent estimator over the x-axis and established, under suitable regularity assumptions, its asymptotic normality. [48] and [40] pointed out that the semiparametric estimate has greater efficiency uniformly over the x-axis. In the light of a simulation study, the authors suggest that the semiparametric estimate is a better choice when parametric information of the truncation distribution is available. Since the apparition of this estimation method many papers are devoted to the statistical inference with truncation data, see for instance: [8], [34], [42], [44], [41], and [45].

2.2 Construction of the new estimator

Motivated by the features of the semiparametric estimation, we next propose a new estimator for γ_1 by means of a suitable functional of $\mathbf{F}_n\left(x;\widehat{\theta}_n\right)$. We start our construction by noting that from Theorem 1.2.2 in de [24], the first-order condition (1.18) (for $\mathbf{F}$) implies that

$$\lim_{t\to\infty} \frac{1}{\overline{\mathbf{F}}(t)} \int_t^{\infty} \log(x/t)\, d\mathbf{F}(x) = \gamma_1. \tag{2.3}$$

In other words, γ_1 may viewed as a functional $\psi_t(\mathbf{F})$, for a large t, where

$$\psi_t(\mathbf{F}) := \frac{1}{\overline{\mathbf{F}}(t)} \int_t^{\infty} \log(x/t)\, d\mathbf{F}(x).$$

Replacing $\mathbf{F}$ by $\mathbf{F}_n\left(\cdot;\widehat{\theta}_n\right)$ and letting $t = X_{n-k:n}$ yield

$$\begin{aligned} \widehat{\gamma}_1 &= \psi_{X_{n-k:n}}\left(\mathbf{F}_n\left(\cdot;\widehat{\theta}_n\right)\right) \\ &= \frac{1}{\overline{\mathbf{F}}_n\left(X_{n-k:n};\widehat{\theta}_n\right)} \int_{X_{n-k:n}}^{\infty} \log(x/X_{n-k:n})\, d\mathbf{F}_n\left(x;\widehat{\theta}_n\right), \end{aligned} \tag{2.4}$$

as new estimator for γ_1. Observe that

$$\begin{aligned} &\int_t^{\infty} \log(x/t)\, d\mathbf{F}_n\left(x;\widehat{\theta}_n\right) \\ &= P_n\left(\widehat{\theta}\right) \int_{X_{n-k:n}}^{\infty} \log(x/X_{n-k:n})\, \mathbf{1}(x \geq X_{n-k})\, d\mathbf{F}_n\left(x;\widehat{\theta}_n\right), \end{aligned}$$

which may be rewritten into

$$\begin{aligned} &\frac{P_n\left(\widehat{\theta}_n\right) 1}{n} \sum_{i=1}^{n} \int_{X_{n-k:n}}^{\infty} \frac{\log(x/X_{n-k:n})\, \mathbf{1}(x \geq X_{n-k})}{\overline{\mathbf{G}}_{\widehat{\theta}_n}(X_i)} d\mathbf{1}(X_i \leq x) \\ &= P_n\left(\widehat{\theta}_n\right) \frac{1}{n} \sum_{i=1}^{k} \frac{\log(X_{n-i+1}/X_{n-k:n})}{\overline{\mathbf{G}}_{\widehat{\theta}_n}(X_{n-i+1:n})}. \end{aligned}$$

On the other hand, $\mathbf{F}\left(X_{n-k:n};\widehat{\theta}_n\right)$ equals

$$P_n\left(\widehat{\theta}_n\right) \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{1}(X_{i:n} \leq X_{n-k:n})}{\overline{\mathbf{G}}_{\widehat{\theta}_n}(X_{i:n})} = P_n\left(\widehat{\theta}_n\right) \frac{1}{n} \sum_{i=1}^{n-k} 1/\overline{\mathbf{G}}_{\widehat{\theta}_n}(X_{i:n}).$$

Hence

$$\begin{aligned} \overline{\mathbf{F}}\left(X_{n-k:n};\widehat{\theta}_n\right) &= \frac{\frac{1}{n}\sum_{i=1}^{n} 1/\overline{\mathbf{G}}_{\widehat{\theta}_n}(X_{i:n}) - \frac{1}{n}\sum_{i=1}^{n-k} 1/\overline{\mathbf{G}}_{\widehat{\theta}_n}(X_{i:n})}{\frac{1}{n}\sum_{i=1}^{n} 1/\overline{\mathbf{G}}_{\widehat{\theta}_n}(X_{i:n})} \\ &= P_n\left(\widehat{\theta}_n\right) \frac{1}{n} \sum_{i=1}^{k} 1/\overline{\mathbf{G}}_{\widehat{\theta}_n}(X_{n-i+1:n}). \end{aligned}$$

Thereby, the form of our new estimator is

$$\widehat{\gamma}_1 = \frac{\sum_{i=1}^{k} \left(\overline{\mathbf{G}}_{\widehat{\theta}_n}(X_{n-i+1:n})\right)^{-1} \log(X_{n-i+1}/X_{n-k:n})}{\sum_{i=1}^{k} \left(\overline{\mathbf{G}}_{\widehat{\theta}_n}(X_{n-i+1:n})\right)^{-1}}. \tag{2.5}$$

2.3 Main results and Proofs

In this section, we present Theorems that are dealing with the consistency and asymptotic normality of our new estimator by giving their proofs. The asymptotic behavior of $\widehat{\gamma}_1$ will be established by means of the following tail empirical process

$$\mathbf{D}_n\left(x;\widehat{\theta}_n;\gamma_1\right) := \sqrt{k}\left(\frac{\overline{\mathbf{F}}_n\left(xX_{n-k:n};\widehat{\theta}_n\right)}{\overline{\mathbf{F}}_n\left(X_{n-k:n};\widehat{\theta}_n\right)} - x^{-1/\gamma_1}\right), \text{ for } x > 1.$$

This method was already used to establish the asymptotic behavior of Hill's estimator for complete data ([24], page 162) that we will adapt to the truncation case. Indeed, by using an integration by parts and a change of variables of the integral (2.4), one gets

$$\widehat{\gamma}_1 = \int_1^\infty x^{-1} \frac{\overline{\mathbf{F}}_n\left(xX_{n-k:n};\widehat{\theta}_n\right)}{\overline{\mathbf{F}}_n\left(X_{n-k:n};\widehat{\theta}_n\right)} dx,$$

and therefore

$$\sqrt{k}\left(\widehat{\gamma}_1 - \gamma_1\right) = \int_1^\infty x^{-1} \mathbf{D}_n\left(x;\widehat{\theta}_n;\gamma_1\right) dx. \tag{2.6}$$

Thus for a suitable weighted weak approximation to $\mathbf{D}_n\left(\cdot;\widehat{\theta}_n;\gamma_1\right)$, we may easily deduce the consistency and asymptotic normality of $\widehat{\gamma}_1$. This process may also contribute to the goodness-of-fit test to fitting heavy-tailed distributions via, among others, the Kolmogorov-Smirnov and Cramer-Von Mises type statistics

$$\sup_{x>1} \left|\mathbf{D}_n\left(x;\widehat{\theta}_n,\widehat{\gamma}_1\right)\right| \text{ and } \int_1^\infty \mathbf{D}_n^2\left(x;\widehat{\theta}_n,\widehat{\gamma}_1\right) dx^{-1/\widehat{\gamma}_1}.$$

We note that, the regularity assumptions, denoted $[A0]$, concerning the existence, consistency and asymptotic normality of the CLME estimator $\widehat{\theta}_n$, given in (2.2), are discussed in [1]. Here we only state additional conditions on df $\mathbf{G}_\theta$ corresponding to Pareto-type models which are required to establish the asymptotic behavior of our newly estimator $\widehat{\gamma}_1$.

- $[A1]$ For each fixed y, the function $\theta \to \mathbf{G}_\theta(y)$ is continuously differentiable of partial derivatives $G_\theta^{(j)} =: \partial G_\theta / \partial \theta_j$, $j = 1, ..., d$.
- $[A2]$ $\overline{\mathbf{G}}_\theta^{(j)} \in RV\left(-1/\gamma_2\right)$.
- $[A3]$ $y^{-\epsilon}\overline{\mathbf{G}}_\theta^{(j)}(y)/\overline{\mathbf{G}}_\theta(y) \to 0$, as $y \to \infty$, for any $\epsilon > 0$.

For common Pareto-type models, one may easily check that there exist some constants $a_j \geq 0$, c_j and d_j, such that $\overline{\mathbf{G}}_\theta^{(j)}(y) \sim c_j\left(y^{-1/\gamma_2} + d_j\right)\log y$, for all large x. Then one may consider that the assumptions $[A1]-[A3]$ are not very restrictive and they may be acceptable in the extreme value theory.

2.3.1 Important Lemma

Lemma 2.1. *For any small $\epsilon > 0$, we have*

$$\frac{\overline{F}_n^*(X_{n-k:n}w)}{\overline{F}^*(X_{n-k:n})} = O_{\mathbf{P}}\left(w^{-1/\gamma+\epsilon/2}\right), \text{ uniformly on } w \geq 1.$$

Proof. Let $V_n(t) := n^{-1}\sum_{i=1}^n \mathbf{1}(\xi_i \leq t)$ be the uniform empirical df pertaining to the sample $\xi_i := \overline{F}^*(X_i)$, $i = 1, ..., n$, of iid uniform$(0,1)$ rv's. It is clear that, for an arbitrary x, we have $V_n\left(\overline{F}^*(x)\right) = \overline{F}_n^*(x)$ almost surely. From Assertion 7 in [46] (page 415), $V_n(t)/t = O_{\mathbf{P}}(1)$ uniformly on $1/n \leq t \leq 1$, this implies that

$$\frac{\overline{F}_n^*(X_{n-k:n}w)}{\overline{F}^*(X_{n-k:n}w)} = O_{\mathbf{P}}(1), \text{ uniformly on } w \geq 1. \tag{2.7}$$

On the other hand, by applying Potter's inequalities (1.16) to $\overline{F}^*$, we get

$$\frac{\overline{F}^*(X_{n-k:n}w)}{\overline{F}^*(X_{n-k:n})} = O_{\mathbf{P}}\left(w^{-1/\gamma+\epsilon/2}\right), \text{ uniformly on } w \geq 1. \tag{2.8}$$

Combining the two statements (2.7) and (2.8) gives the desired result. ∎

2.3.2 Theorems and Proofs

Theorem 2.1. *Assume that $\overline{\mathbf{F}} \in RV_2\left(-1/\gamma_1; \rho_1, \mathbf{A}\right)$ and $\mathbf{G}_\theta \in RV\left(-1/\gamma_2\right)$ satisfying the assumptions $[A0]-[A3]$, and suppose that $\gamma_1 < \gamma_2$. Then on the probability* *space* $(\Omega, \mathscr{A}, \mathbf{P})$, *there exists a standard Wiener process* $\{W(s), 0 \leq s \leq 1\}$ *such that, for any small* $0 < \epsilon < 1/2$, *we have*

$$\sup_{x>1} x^{\epsilon} \left| \mathbf{D}_n\left(x; \widehat{\theta}_n, \gamma_1\right) - \Gamma(x; W) - x^{-1/\gamma_1} \frac{x^{\rho_1/\gamma_1} - 1}{\rho_1 \gamma_1} \sqrt{k} \mathbf{A}(a_k) \right| \xrightarrow{\mathbf{P}} 0,$$

provided that $\sqrt{k} \mathbf{A}(a_k) = O(1)$, *where*

$$\begin{aligned} \Gamma(x; W) &:= \frac{\gamma}{\gamma_1} x^{-1/\gamma_1} \left\{ x^{1/\gamma} W\left(x^{-1/\gamma}\right) - W(1) \right\} \\ &\quad + \frac{\gamma}{\gamma_1 + \gamma_2} x^{-1/\gamma_1} \int_0^1 s^{-\gamma/\gamma_2 - 1} \left\{ x^{1/\gamma} W\left(x^{-1/\gamma} s\right) - W(s) \right\} ds, \end{aligned}$$

is a centred Gaussian process and $a_k := F^{*\leftarrow}(1 - k/n)$[1].

Proof. Let us first notice that the semiparametric estimator of df $\mathbf{F}$ given in (2.2) may be rewritten into

$$\mathbf{F}_n\left(x; \widehat{\theta}_n\right) = P_n\left(\widehat{\theta}_n\right) \int_0^x \frac{dF_n^*(w)}{\overline{\mathbf{G}}_{\widehat{\theta}_n}(w)}, \tag{2.9}$$

and $1/P_n\left(\widehat{\theta}\right) = \int_0^\infty dF_n^*(w)/\overline{\mathbf{G}}_{\widehat{\theta}_n}(w)$, where $F_n^*(w) := n^{-1} \sum_{i=1}^n \mathbf{1}(X_i \leq w)$ denotes the usual empirical df pertaining to the observed sample $X_1, ..., X_n$. It is worth mentioning that by using the strong law of large numbers $P_n\left(\widehat{\theta}_n\right) \to P(\theta)$ (almost surely) as $n \to \infty$, where $P(\theta) = 1/\int_0^\infty dF^*(w)/\overline{\mathbf{G}}_\theta(w)$ (see e.g. Lemma 3.2 in [48]). On the other hand from equation (1.14), we deduce that $p = 1/\int_0^\infty dF^*(w)/\overline{\mathbf{G}}(w)$, it follows that $p \equiv P(\theta)$ because we already assumed that $\mathbf{G} \equiv \mathbf{G}_\theta$. Next we use the distribution tail

$$\overline{\mathbf{F}}(x) = P(\theta) \int_x^\infty \frac{dF^*(w)}{\overline{\mathbf{G}}_\theta(w)}, \tag{2.10}$$

and its empirical counterpart

$$\overline{\mathbf{F}}_n\left(x; \widehat{\theta}_n\right) = P_n\left(\widehat{\theta}_n\right) \int_x^\infty \frac{dF_n^*(w)}{\overline{\mathbf{G}}_{\widehat{\theta}_n}(w)}.$$

We begin by decomposing $k^{-1/2} \mathbf{D}_n\left(x; \widehat{\theta}_n\right)$, for $x > 1$, into the sum of

$$\mathbf{M}_{n1}(x) := x^{-1/\gamma_1} \frac{\overline{\mathbf{F}}_n\left(x X_{n-k:n}; \widehat{\theta}_n\right) - \overline{\mathbf{F}}_n\left(x X_{n-k:n}; \theta\right)}{\overline{\mathbf{F}}(x X_{n-k:n})},$$

[1] $F^{*\leftarrow}(s) := \inf\{x : F^*(x) \geq s\}$, $0 < s < 1$, denotes the quantile (or the generalized inverse) function pertaining to df F^*.

$$\mathbf{M}_{n2}(x) := x^{-1/\gamma_1} \frac{\overline{\mathbf{F}}_n(xX_{n-k:n};\theta) - \overline{\mathbf{F}}(xX_{n-k:n})}{\overline{\mathbf{F}}(xX_{n-k:n})},$$

$$\mathbf{M}_{n3}(x) := -\frac{\overline{\mathbf{F}}(xX_{n-k:n})}{\overline{\mathbf{F}}_n(X_{n-k:n};\theta)} \frac{\overline{\mathbf{F}}_n(X_{n-k:n};\theta) - \overline{\mathbf{F}}(X_{n-k:n})}{\overline{\mathbf{F}}(X_{n-k:n})},$$

$$\mathbf{M}_{n4}(x) := \left(\frac{\overline{\mathbf{F}}(xX_{n-k:n})}{\overline{\mathbf{F}}_n(X_{n-k:n};\theta)} - x^{-1/\gamma_1}\right) \frac{\overline{\mathbf{F}}_n(xX_{n-k:n};\theta) - \overline{\mathbf{F}}(xX_{n-k:n})}{\overline{\mathbf{F}}(xX_{n-k:n})}$$

and

$$\mathbf{M}_{n5}(x) := \frac{\overline{\mathbf{F}}(xX_{n-k:n})}{\overline{\mathbf{F}}(X_{n-k:n})} - x^{-1/\gamma_1}.$$

Our goal is to provide a weighted weak approximation to the tail empirical process $\mathbf{D}_n\left(x;\widehat{\theta}_n;\gamma_1\right)$. Let $\xi_i := \overline{F}^*(X_i)$, $i = 1,...,n$ be a sequence of independent and identically distributed rv's. Recall that both df's $\mathbf{F}$ and $\mathbf{G}_\theta$ are assumed to be continuous, this implies that F^* is continuous as well, therefore $\mathbf{P}(\xi_i \le u) = u$, this means that $(\xi_i)_{i=1,n}$ are uniformly distributed on $(0,1)$. Let us now define the corresponding uniform tail empirical process

$$\alpha_n(s) := \sqrt{k}\,(\mathbf{U}_n(s) - s), \text{ for } 0 \le s \le 1, \tag{2.11}$$

where

$$\mathbf{U}_n(s) := k^{-1} \sum_{i=1}^{n} \mathbf{1}(\xi_i < ks/n), \tag{2.12}$$

denotes the tail empirical df pertaining to the sample $(\xi_i)_{i=1,n}$. In view of Proposition 3.1 of [17], there exists a Wiener process W such that for every $0 \le \epsilon < 1/2$,

$$\sup_{0\le s<1} s^{-\epsilon} |\alpha_n(s) - W(s)| \overset{\mathbf{P}}{\to} 0, \text{ as } n \to \infty. \tag{2.13}$$

Let us fix a sufficiently small $0 < \epsilon < 1/2$. We will successively show that, under the first-order conditions of regular variation (1.18), we have, uniformly on $x \ge 1$, for all large n :

$$\sqrt{k}\mathbf{M}_{n2}(x) = \frac{\gamma}{\gamma_1} x^{1/\gamma_2} W\left(t^{-1/\gamma}\right) + \frac{\gamma}{\gamma_1} \int_{x^{1/\gamma_2}}^{\infty} W\left(t^{-\gamma_2/\gamma}\right) dt + o_{\mathbf{P}}\left(x^{\frac{1}{2}\left(\frac{1}{\gamma_2} - \frac{1}{\gamma_1}\right)+\epsilon}\right) \tag{2.14}$$

and

$$\sqrt{k}\mathbf{M}_{n3}(x) = -x^{-1/\gamma_1}\left(\frac{\gamma}{\gamma_1} W(1) + \frac{\gamma}{\gamma_1} \int_1^{\infty} W\left(t^{-\gamma_2/\gamma}\right) dt\right) + o_{\mathbf{P}}\left(x^{-1/\gamma_1+\epsilon}\right), \tag{2.15}$$

while

$$\sqrt{k}\mathbf{M}_{n1}(x) = o_{\mathbf{P}}\left(x^{-1/\gamma_1+\epsilon}\right),\ \sqrt{k}\mathbf{M}_{n4}(x) = o_{\mathbf{P}}\left(x^{\frac{1}{2}\left(\frac{1}{\gamma_2}-\frac{1}{\gamma_1}\right)+\epsilon}\right), \tag{2.16}$$

and

$$\sqrt{k}\mathbf{M}_{n5}(x) = x^{-1/\gamma_1}\frac{x^{\rho_1/\gamma_1}-1}{\rho_1\gamma_1}\sqrt{k}\mathbf{A}(a_k) + o_{\mathbf{P}}\left(x^{-1/\gamma_1}\right). \tag{2.17}$$

Throughout the proof, without loss of generality, we assume that $a\epsilon \equiv \epsilon$, for any constant $a > 0$. We point out that all the rest terms of the previous approximations are negligible in probability, uniformly on $x > 1$. Let us begin by the term $\mathbf{M}_{n1}(x)$ which may be made into

$$\frac{x^{-1/\gamma_1}}{\overline{\mathbf{F}}(xX_{n-k:n})}P_n\left(\widehat{\theta}_n\right)\left(\int_x^\infty \frac{dF_n^*(X_{n-k:n}w)}{\overline{\mathbf{G}}_{\widehat{\theta}}(X_{n-k:n}w)} - \int_x^\infty \frac{dF_n^*(X_{n-k:n}w)}{\overline{\mathbf{G}}_\theta(X_{n-k:n}w)}\right)$$
$$= \frac{x^{-1/\gamma_1}}{\overline{\mathbf{F}}(xX_{n-k:n})}P_n\left(\widehat{\theta}_n\right)\int_x^\infty\left(\frac{1}{\overline{\mathbf{G}}_{\widehat{\theta}}(X_{n-k:n}w)} - \frac{1}{\overline{\mathbf{G}}_\theta(X_{n-k:n}w)}\right)dF_n^*(X_{n-k:n}w).$$

Applying the mean value theorem (for several variables) to function $\theta \to 1/\overline{\mathbf{G}}_\theta(\cdot)$, yields

$$\frac{1}{\overline{\mathbf{G}}_{\widehat{\theta}}(z)} - \frac{1}{\overline{\mathbf{G}}_\theta(z)} = \sum_{i=1}^d\left(\widehat{\theta}_{i,n}-\theta_i\right)\frac{\overline{\mathbf{G}}_{\widetilde{\theta}}^{(i)}(z)}{\overline{\mathbf{G}}_{\widetilde{\theta}}^2(z)},\text{ for any } z>1,$$

where $\widetilde{\theta}_n$ is such that $\widetilde{\theta}_{i,n}$ is between θ_i and $\widehat{\theta}_{i,n}$, for $i = 1,...,d$, therefore

$$\mathbf{M}_{n1}(x) = \frac{x^{-1/\gamma_1}}{\overline{\mathbf{F}}(xX_{n-k:n})}P_n\left(\widehat{\theta}_n\right)\sum_{i=1}^d\left(\widehat{\theta}_i-\theta_i\right)\int_x^\infty\frac{\overline{\mathbf{G}}_{\widetilde{\theta}}^{(i)}(X_{n-k:n}w)}{\overline{\mathbf{G}}_{\widetilde{\theta}}^2(X_{n-k:n}w)}dF_n^*(X_{n-k:n}w).$$

Recall that by assumptions (1.18) and [$A2$] both $\overline{\mathbf{G}}_\theta$ and $\overline{\mathbf{G}}_\theta^{(i)}$ are regularly varying with the same index $\left(-1/\gamma_2\right)$ and on the other hand, $X_{n-k:n} \xrightarrow{\mathbf{P}} \infty$ and $w > 1$, imply that $X_{n-k:n}w \xrightarrow{\mathbf{P}} \infty$. Applying Pooter's inequalities (1.16), we get

$$\frac{\overline{\mathbf{G}}_{\widetilde{\theta}}(X_{n-k:n}w)}{\overline{\mathbf{G}}_{\widetilde{\theta}}(X_{n-k:n})} = (1+o_{\mathbf{P}}(1))w^{-1/\gamma_2+\epsilon} = \frac{\overline{\mathbf{G}}_{\widetilde{\theta}}^{(i)}(X_{n-k:n}w)}{\overline{\mathbf{G}}_{\widetilde{\theta}}^{(i)}(X_{n-k:n})},$$

it follows that

$$\mathbf{M}_{n1}(x) = (1+o_{\mathbf{P}}(1))P_n\left(\widehat{\theta}_n\right)\frac{x^{-1/\gamma_1}}{\overline{\mathbf{G}}_{\widetilde{\theta}}(X_{n-k:n})\overline{\mathbf{F}}(xX_{n-k:n})}$$
$$\times\sum_{i=1}^d\frac{\overline{\mathbf{G}}_{\widetilde{\theta}}^{(i)}(X_{n-k:n})}{\overline{\mathbf{G}}_{\widetilde{\theta}}(X_{n-k:n})}\left|\widehat{\theta}_{i,n}-\theta_i\right|\int_x^\infty w^{1/\gamma_2-\epsilon}dF_n^*(X_{n-k:n}w).$$

Under some regularity assumptions, [1] stated that $\sqrt{n}\left(\widehat{\theta}_n-\theta\right)$ is asymptotically a centred multivariate normal rv, which implies that $\widehat{\theta}_{i,n}-\theta_i=O_{\mathbf{P}}\left(n^{-1/2}\right)$ and thus $\widehat{\theta}_n\overset{\mathbf{P}}{\to}\theta$. On the other hand, by the law of large numbers $P_n(\theta)\overset{\mathbf{P}}{\to}P(\theta)$ as $n\to\infty$, then we may readily show that $P_n\left(\widehat{\theta}_n\right)\overset{\mathbf{P}}{\to}P(\theta)$ as $n\to\infty$ as well. Note that since $\widehat{\theta}_n$ is a consistent estimator of θ then $\widetilde{\theta}_n$ is too. Then by using the fact that $X_{n-k:n}\overset{\mathbf{P}}{\to}\infty$ and both conditions $[A1]$ and $[A3]$, we show readily that

$$(X_{n-k:n})^{-\epsilon}\frac{\overline{\mathbf{G}}^{(i)}_{\widetilde{\theta}_n}(X_{n-k:n})}{\overline{\mathbf{G}}_{\widetilde{\theta}_n}(X_{n-k:n})}\overset{\mathbf{P}}{\to}0,\text{ as }n\to\infty,$$

and $\overline{\mathbf{G}}_{\theta}(X_{n-k:n})/\overline{\mathbf{G}}_{\widetilde{\theta}_n}(X_{n-k:n})\overset{\mathbf{P}}{\to}1$. In view of Lemma A1 in [6], we infer that $X_{n-k:n}=(1+o_{\mathbf{P}}(1))(k/n)^{-\gamma}$, thus

$$\mathbf{M}_{n1}(x)=(k/n)^{-\epsilon\gamma}o_{\mathbf{P}}\left(n^{-1/2}\right)\widetilde{\mathbf{M}}_{n1}(x),$$

where

$$\widetilde{\mathbf{M}}_{n1}(x):=\frac{x^{-1/\gamma_1}P(\theta)}{\overline{\mathbf{G}}_{\theta}(X_{n-k:n})\overline{\mathbf{F}}(xX_{n-k:n})}\int_x^{\infty}w^{1/\gamma_2-\epsilon}dF_n^*(X_{n-k:n}w).$$

Making use of representation (2.10), we write

$$\widetilde{\mathbf{M}}_{n1}(x)=x^{-1/\gamma_1}\left(\int_x^{\infty}\frac{\overline{\mathbf{G}}_{\theta}(X_{n-k:n})}{\overline{\mathbf{G}}_{\theta}(X_{n-k:n}w)}d\frac{F^*(X_{n-k:n}w)}{\overline{F}^*(X_{n-k:n})}\right)^{-1}\times\left(\int_x^{\infty}w^{1/\gamma_2-\epsilon}d\frac{F_n^*(X_{n-k:n}w)}{\overline{F}^*(X_{n-k:n})}\right).\tag{2.18}$$

Once again by using the routine manipulations of Potter's inequalities, we show that the first integral in (2.18) is equal to

$$(1+o_{\mathbf{P}}(1))\int_x^{\infty}w^{1/\gamma_2+\epsilon/2}d\frac{F^*(X_{n-k:n}w)}{\overline{F}^*(X_{n-k:n})}.$$

An integration by parts to the previous integral yields

$$x^{1/\gamma_2+\epsilon/2}\frac{\overline{F}^*(X_{n-k:n}x)}{\overline{F}^*(X_{n-k:n})}+\left(1/\gamma_2+\epsilon/2\right)\int_x^{\infty}w^{1/\gamma_2+\epsilon/2-1}\frac{\overline{F}^*(X_{n-k:n}w)}{\overline{F}^*(X_{n-k:n})}dw.$$

Recall that from (1.19) we have $\overline{F}^*\in RV_{(-1/\gamma)}$, then

$$\frac{\overline{F}^*(X_{n-k:n}w)}{\overline{F}^*(X_{n-k:n})}=(1+o_{\mathbf{P}}(1))w^{-1/\gamma+\epsilon/2},$$

uniformly on $w > 1$. Therefore the previous quantity reduces into

$$(1 + o_{\mathbf{P}}(1)) \left(1 + \frac{1/\gamma_2 + \epsilon/2}{-1/\gamma_1 + \epsilon}\right) x^{-1/\gamma_1 + \epsilon}.$$

Thereby the first expression between two brackets in (2.18) equals $O_{\mathbf{P}}\left(x^{1/\gamma_1 - \epsilon}\right)$. Let us consider the second factor in (2.18). By similar arguments as used for the first factor, we show that

$$x^{1/\gamma_2 + \epsilon/2} \frac{\overline{F}_n^*(X_{n-k:n}x)}{\overline{F}^*(X_{n-k:n})} + \left(1/\gamma_2 + \epsilon/2\right) \int_x^\infty w^{1/\gamma_2 + \epsilon/2} \frac{\overline{F}_n^*(X_{n-k:n}w)}{\overline{F}^*(X_{n-k:n})} dw,$$

multiplied by $(1 + o_{\mathbf{P}}(1))$, uniformly on $x > 1$. From Lemma 2.1**,** we have

$$\frac{\overline{F}_n^*(X_{n-k:n}w)}{\overline{F}^*(X_{n-k:n})} = O_{\mathbf{P}}\left(w^{-1/\gamma + \epsilon/2}\right),$$

which implies that the previous expression equals $O_{\mathbf{P}}\left(x^{-1/\gamma_1 + \epsilon}\right)$, thus $\widetilde{\mathbf{M}}_{n1}(x) = O_{\mathbf{P}}\left(x^{-1/\gamma + \epsilon}\right)$ and therefore

$$\sqrt{k}\mathbf{M}_{n1}(x) = (k/n)^{1/2 - \epsilon\gamma} O_{\mathbf{P}}\left(x^{-1/\gamma_1 + \epsilon}\right).$$

By assumption $k/n \to 0$, it follows that $\sqrt{k}\mathbf{M}_{n1}(x) = o_{\mathbf{P}}\left(x^{-1/\gamma_1 + \epsilon}\right)$ which meets the result of (2.18). Let now consider the second term $\mathbf{M}_{n2}(x)$ which may be rewritten into

$$-x^{-1/\gamma_1} \frac{k/n}{\overline{F}^*(X_{n-k:n})} \frac{\overline{\mathbf{F}}(X_{n-k:n})}{\overline{\mathbf{F}}(xX_{n-k:n})} \frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})/\overline{F}^*(X_{n-k:n})}{\overline{\mathbf{F}}(X_{n-k:n})}$$
$$\times \int_x^\infty \frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}w)} d\frac{\overline{F}_n^*(X_{n-k:n}w) - \overline{F}^*(X_{n-k:n}w)}{k/n}.$$

In view of Potter's inequalities, it is clear that

$$\frac{\overline{\mathbf{F}}(X_{n-k:n})}{\overline{F}^*(X_{n-k:n})/\overline{\mathbf{G}}_\theta(X_{n-k:n})} \xrightarrow{\mathbf{P}} \frac{\gamma_1}{\gamma} P(\theta)$$

and

$$\frac{\overline{\mathbf{F}}(X_{n-k:n})}{\overline{\mathbf{F}}(xX_{n-k:n})} \xrightarrow{\mathbf{P}} x^{1/\gamma_1}.$$

Smirnov's lemma (see, e.g., Lemma 2.2.3 in de Haan and Feriera, 2006) with the fact that $\overline{F}^*(X_{n-k:n}) \overset{d}{=} \xi_{k+1:n}$ imply that $\frac{n}{k}\xi_{k+1:n} \overset{\mathbf{P}}{\to} 1$, hence $\frac{n}{k}\overline{F}^*(X_{n-k:n}) = 1 + o_{\mathbf{P}}(1)$. Therefore

$$\mathbf{M}_{n2}(x) = -(1 + o_{\mathbf{P}}(1))\frac{\gamma}{\gamma_1}\int_x^\infty \frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}w)} d\frac{\overline{F}_n^*(X_{n-k:n}w) - \overline{F}^*(X_{n-k:n}w)}{k/n}.$$

On the other hand, using an integration by parts yields

$$\mathbf{M}_{n2}(x) = (1 + o_{\mathbf{P}}(1))\frac{\gamma_1}{\gamma}\left(\mathbf{M}_{n2}^{(1)}(x) + \mathbf{M}_{n2}^{(2)}(x)\right),$$

where

$$\mathbf{M}_{n2}^{(1)}(x) := \int_x^\infty \frac{\overline{F}_n^*(X_{n-k:n}w) - \overline{F}^*(X_{n-k:n}w)}{k/n} d\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}w)}$$

and

$$\mathbf{M}_{n2}^{(2)}(x) := \frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)} \frac{\overline{F}_n^*(X_{n-k:n}x) - \overline{F}^*(xX_{n-k:n})}{k/n}.$$

By using the change of variables $t = \overline{\mathbf{G}}_\theta(X_{n-k:n})/\overline{\mathbf{G}}_\theta(X_{n-k:n}w)$, it is easy to verify that

$$\mathbf{M}_{n2}^{(1)}(x) = \int_{\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}}^\infty \frac{n}{k}\left\{\overline{F}_n^*\left(\mathbf{G}_\theta^{\leftarrow}\left(1 - \overline{\mathbf{G}}_\theta(X_{n-k:n})t^{-1}\right)\right) - \overline{F}^*\left(\mathbf{G}_\theta^{\leftarrow}\left(1 - \overline{\mathbf{G}}_\theta(X_{n-k:n})t^{-1}\right)\right)\right\}dt.$$

Observe that

$$\mathbf{M}_{n2}^{(1)}(x) = \int_{\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}}^\infty (\mathbf{U}_n(\vartheta_n(t;\theta)) - \vartheta_n(t;\theta))dt,$$

where $\vartheta_n(t;\theta) := \frac{n}{k}\overline{F}^*\left(\mathbf{G}_\theta^{\longleftarrow}\left(1 - \overline{\mathbf{G}}_\theta(X_{n-k:n})t^{-1}\right)\right)$ and $\mathbf{U}_n$ are the tail empirical df given in (2.12). Thereby

$$\sqrt{k}\mathbf{M}_{n2}^{(1)}(x) = \int_{\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}}^\infty \alpha_n(\vartheta_n(t;\theta))dt,$$

with α_n being the tail empirical process defined in (2.11). Let us decompose the previous integral into

$$\int_{\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}}^\infty (\alpha_n(\vartheta_n(t;\theta)) - W(\vartheta_n(t;\theta)))dt + \int_{\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}}^\infty W(\vartheta_n(t;\theta))dt$$
$$= S_n + R_n.$$

By applying weak approximation (2.13) we get

$$S_n = o_{\mathbf{P}}(1) \int_{\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}}^{\infty} (\vartheta_n(t;\theta))^{1/2-\epsilon} dt.$$

Observe that $\overline{F}^*\left(\mathbf{G}_\theta^{\leftarrow}\left(1-\overline{\mathbf{G}}_\theta(X_{n-k:n})\right)\right) = \overline{F}^*(X_{n-k:n})$, thereby

$$\vartheta_n(t;\theta) = \frac{n}{k}\overline{F}^*(X_{n-k:n}) \frac{\overline{F}^*\left(\mathbf{G}_\theta^{\leftarrow}\left(1-\overline{\mathbf{G}}_\theta(X_{n-k:n})t^{-1}\right)\right)}{\overline{F}^*\left(\mathbf{G}_\theta^{\leftarrow}\left(1-\overline{\mathbf{G}}_\theta(X_{n-k:n})\right)\right)}.$$

It is easy to check that $\overline{F}^*\left(\mathbf{G}_\theta^{\leftarrow}(1-\cdot)\right) \in RV\left(\gamma_2/\gamma\right)$, then once again by means of Pooter's inequality, we show that $\vartheta_n(t;\theta) = (1+o_{\mathbf{P}}(1))\, t^{-\gamma_2/\gamma+\epsilon}$, therefore

$$S_n = o_{\mathbf{P}}(1) \int_{\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}}^{\infty} \left(t^{-\gamma_2/\gamma+\epsilon}\right)^{1/2-\epsilon} dt.$$

By using an elementary integration we get

$$S_n = o_{\mathbf{P}}(1) \left(\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}\right)^{(-\gamma_2/\gamma+\epsilon)(1/2-\epsilon)+1} = o_{\mathbf{P}}\left(x^{\frac{1}{\gamma_2}-\frac{1}{2\gamma}+\epsilon}\right).$$

By replacing γ by its by its expression given in (1.20), we end up with

$$S_n = o_{\mathbf{P}}\left(x^{\frac{1}{2}\left(\frac{1}{\gamma_2}-\frac{1}{\gamma_1}\right)+\epsilon}\right).$$

The term R_n may be decomposed into

$$\int_{\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}}^{x^{1/\gamma_2}} W(\vartheta_n(t;\theta))\, dt + \int_{x^{1/\gamma_2}}^{\infty} W(\vartheta_n(t;\theta))\, dt = R_{n1} + R_{n2}.$$

It is clear that

$$|R_{n1}| < \left\{ \sup_{t > \frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}} \frac{|W(\vartheta_n(t;\theta))|}{(\vartheta_n(t;\theta))^\epsilon} \right\} \int_{\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}}^{x^{1/\gamma_2}} (\vartheta_n(t;\theta))^\epsilon dt.$$

It is ready to check, by using the change of variables $\vartheta_n(t;\theta) = s$, that the previous first factor between the curly brackets equals

$$\sup_{0 < s < \frac{n}{k}\overline{F}^*(X_{n-k:n}x;\theta)} \frac{|W(s)|}{s^\epsilon} < \sup_{0 < s < \frac{n}{k}\overline{F}^*(X_{n-k:n};\theta)} \frac{|W(s)|}{s^\epsilon}.$$

From Lemma 3.2 in [17] $\sup_{0<s\leq 1} s^{-\delta}|W(s)| = O_{\mathbf{P}}(1)$, for any $0<\delta<1/2$, then since $n\overline{F}^{*}(X_{n-k:n};\theta)/k \xrightarrow{\mathbf{P}} 1$, as $n\to\infty$, we infer that

$$\sup_{0<s<\frac{n}{k}\overline{F}^{*}(X_{n-k:n};\theta)} s^{-\epsilon}|W(s)| = O_{\mathbf{P}}(1).$$

for all large n. On the other hand, we already pointed out above that

$$\vartheta_n(t;\theta) = (1+o_{\mathbf{P}}(1))\, t^{-\gamma_2/\gamma+\epsilon},$$

which implies that the second factor is equal to

$$O_{\mathbf{P}}(1)\int_{\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}}^{x^{1/\gamma_2}} \left(t^{-\gamma_2/\gamma+\epsilon}\right)^{\epsilon} dt = O_{\mathbf{P}}(1)\int_{\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}}^{x^{1/\gamma_2}} t^{-\epsilon\gamma_2/\gamma+\epsilon} dt,$$

which after integration yields

$$O_{\mathbf{P}}(1)\left\{\left(\frac{\overline{\mathbf{G}}_\theta(X_{n-k:n})}{\overline{\mathbf{G}}_\theta(X_{n-k:n}x)}\right)^{-\epsilon\gamma_2/\gamma+\epsilon+1} - \left(x^{-1/\gamma}\right)^{-\epsilon\gamma_2/\gamma+\epsilon+1}\right\}.$$

Recall that from formula (1.20) we have $\gamma_2/\gamma>1$, then by using the mean value theorem and Pooter's inequalities, we get $R_{n1} = o_{\mathbf{P}}(x^{-\epsilon})$. The second term R_{n2} may be decomposed into

$$R_{n2} = \int_{x^{1/\gamma_2}}^{\infty}\left(W(\vartheta_n(t;\theta)) - W\left(t^{-\gamma_2/\gamma}\right)\right)dt + \int_{x^{1/\gamma_2}}^{\infty} W\left(t^{-\gamma_2/\gamma}\right)dt.$$

From Proposition B.1.10 in [24], we have with high probability,

$$c_n(t;\theta) := \left|\vartheta_n(t;\theta) - t^{-\gamma_2/\gamma}\right| \leq \epsilon t^{-\gamma_2/\gamma-\epsilon}, \text{ as } n\to\infty, \tag{2.19}$$

this means that $\sup_{x>1}\sup_{t>x^{1/\gamma_2}} c_n(t;\theta) \xrightarrow{\mathbf{P}} 0$, as $n\to\infty$. This implies by using Levy's modulus of continuity of the Wiener process see, e.g., Theorem 1.1.1 in [14], that

$$\left|W(\vartheta_n(t;\theta)) - W\left(t^{-\gamma_2/\gamma}\right)\right| \leq 2\sqrt{c_n(t;\theta)\log(1/c_n(t;\theta))},$$

with high probability. By using the fact that $\log s < \epsilon s^{-\epsilon}$, for $s\downarrow 0$ together with inequality (2.19), we show that

$$\left|W(\vartheta_n(t;\theta)) - W\left(t^{-\gamma_2/\gamma}\right)\right| < 2\epsilon t^{-(\gamma_2/\gamma-\epsilon)/2},$$

uniformly on $t > x^{1/\gamma_2}$, it follows that

$$\left|\int_{x^{1/\gamma_2}}^{\infty}\left(W(\vartheta_n(t;\theta)) - W\left(t^{-\gamma_2/\gamma}\right)\right)dt\right| = o_{\mathbf{P}}(1)\left|\int_{x^{1/\gamma_2}}^{\infty} t^{-(\gamma_2/\gamma-\epsilon)/2}dt\right|.$$

Recall that the assumption $\gamma_1 < \gamma_2$ together with the equation $1/\gamma = 1/\gamma_1 + 1/\gamma_2$, imply that $\gamma_2/(2\gamma) > 1$, thus $-(\gamma_2/\gamma - \epsilon)/2 + 1 < 0$, therefore $\left|\int_{x^{1/\gamma_2}}^{\infty} t^{-(\gamma_2/\gamma-\epsilon)/2}dt\right| = o_{\mathbf{P}}\left(x^{-1/\gamma_1-\epsilon}\right)$. Then we showed that

$$R_{n1} = o_{\mathbf{P}}\left(x^{-\epsilon}\right) \text{ and } R_{n2} = \int_{x^{1/\gamma_2}}^{\infty} W\left(t^{-\gamma_2/\gamma}\right)dt + o_{\mathbf{P}}\left(x^{-1/\gamma_1-\epsilon}\right),$$

hence

$$\sqrt{k}\,\mathbf{M}_{n2}^{(1)}(x) = R_n + S_n = \int_{x^{1/\gamma_2}}^{\infty} W\left(t^{-\gamma_2/\gamma}\right)dt + o_{\mathbf{P}}\left(x^{-1/\gamma_1-\epsilon}\right) + o_{\mathbf{P}}\left(x^{\frac{1}{2}\left(\frac{1}{\gamma_2}-\frac{1}{\gamma_1}\right)+\epsilon}\right).$$

It is clear that

$$\left(-\frac{1}{\gamma_1} - \epsilon\right) - \left(\frac{1}{2}\left(\frac{1}{\gamma_2} - \frac{1}{\gamma_1}\right) + \epsilon\right) = -\frac{\gamma_1 + \gamma_2 + 4\epsilon\gamma_1\gamma_2}{2\gamma_1\gamma_2} < 0.$$

then

$$\sqrt{k}\,\mathbf{M}_{n2}^{(1)}(x) = \int_{x^{1/\gamma_2}}^{\infty} W\left(t^{-\gamma_2/\gamma}\right)dt + o_{\mathbf{P}}\left(x^{\frac{1}{2}\left(\frac{1}{\gamma_2}-\frac{1}{\gamma_1}\right)+\epsilon}\right).$$

By using similar arguments we end up with

$$\sqrt{k}\,\mathbf{M}_{n2}^{(2)}(x) = x^{1/\gamma_2}W\left(t^{-1/\gamma}\right) + o_{\mathbf{P}}\left(x^{-\frac{1}{\gamma_1}+\epsilon}\right),$$

therefore we omit further details. Finally we have

$$\sqrt{k}\,\mathbf{M}_{n2}(x) = \frac{\gamma}{\gamma_1}x^{1/\gamma_2}W\left(t^{-1/\gamma}\right) + \frac{\gamma}{\gamma_1}\int_{x^{1/\gamma_2}}^{\infty} W\left(t^{-\gamma_2/\gamma}\right)dt + o_{\mathbf{P}}\left(x^{\frac{1}{2}\left(\frac{1}{\gamma_2}-\frac{1}{\gamma_1}\right)+\epsilon}\right).$$

Let us now focus on the term $\mathbf{M}_{n3}(x)$. From the latter approximation, we infer that

$$\begin{aligned}\sqrt{k}\,\mathbf{M}_{n2}(1) &= \sqrt{k}\,\frac{\overline{\mathbf{F}}_n(X_{n-k:n};\theta) - \overline{\mathbf{F}}(X_{n-k:n})}{\overline{\mathbf{F}}(X_{n-k:n})} \\ &= \frac{\gamma}{\gamma_1}W(1) + \frac{\gamma}{\gamma_1}\int_1^{\infty} W\left(t^{-\gamma_2/\gamma}\right)dt + o_{\mathbf{P}}(1), \qquad (2.20)\end{aligned}$$

which implies that

$$\sqrt{k}\,\frac{\overline{\mathbf{F}}_n(X_{n-k:n};\theta) - \overline{\mathbf{F}}(X_{n-k:n})}{\overline{\mathbf{F}}(X_{n-k:n})} = O_{\mathbf{P}}(1).$$

In other words, we have

$$\frac{\overline{\mathbf{F}}_n\left(X_{n-k:n};\theta\right)}{\overline{\mathbf{F}}\left(X_{n-k:n}\right)}=1+O_{\mathbf{P}}\left(k^{-1/2}\right). \tag{2.21}$$

The regular variation of $\overline{\mathbf{F}}(\cdot)$ and (2.21) together imply that

$$\frac{\overline{\mathbf{F}}\left(xX_{n-k:n}\right)}{\overline{\mathbf{F}}_n\left(X_{n-k:n};\theta\right)}=x^{-1/\gamma_1}+o_{\mathbf{P}}\left(x^{-1/\gamma_1+\epsilon}\right). \tag{2.22}$$

By combining the results (4.27) and (4.18) we get

$$\sqrt{k}\,\mathbf{M}_{n3}(x)=-x^{-1/\gamma_2}\left(\frac{\gamma}{\gamma_1}W(1)+\frac{\gamma}{\gamma_1}\int_1^\infty W\left(t^{-\gamma_2/\gamma}\right)dt\right)+o_{\mathbf{P}}\left(x^{-1/\gamma_1+\epsilon}\right).$$

For the fourth term $\mathbf{M}_{n4}(x)$ we write

$$\sqrt{k}\,\mathbf{M}_{n4}(x)=\left(\frac{\overline{\mathbf{F}}\left(xX_{n-k:n}\right)}{\overline{\mathbf{F}}_n\left(X_{n-k:n};\theta\right)}-x^{-1/\gamma_1}\right)\left(\sqrt{k}\,\frac{\overline{\mathbf{F}}_n\left(xX_{n-k:n};\theta\right)-\overline{\mathbf{F}}\left(xX_{n-k:n}\right)}{\overline{\mathbf{F}}\left(xX_{n-k:n}\right)}\right).$$

From (4.18) the first factor of the previous equation equals $o_{\mathbf{P}}\left(x^{-1/\gamma_1+\epsilon}\right)$. On the other hand, the change of variables $s=t^{-\gamma_2/\gamma}$, yields

$$\int_{x^{1/\gamma_2}}^\infty W\left(t^{-\gamma_2/\gamma}\right)dt=\frac{\gamma}{\gamma_2}\int_0^{x^{-1/\gamma}}s^{-\gamma/\gamma_2-1}W(s)\,ds.$$

Since $\sup_{0<s<1}s^{-1/2+\epsilon}\left|W(s)\right|=O_{\mathbf{P}}(1)$, then we easily show that

$$\int_{x^{1/\gamma_2}}^\infty W\left(t^{-\gamma_2/\gamma}\right)dt=O_{\mathbf{P}}\left(x^{\frac{1}{2}\left(\frac{1}{\gamma_2}-\frac{1}{\gamma_1}\right)+\epsilon}\right),$$

it follows that $\sqrt{k}\,M_{n2}(x)=O_{\mathbf{P}}\left(x^{\frac{1}{2}\left(\frac{1}{\gamma_2}-\frac{1}{\gamma_1}\right)+\epsilon}\right)$ as well. Therefore

$$\sqrt{k}\,\frac{\overline{\mathbf{F}}_n\left(xX_{n-k:n};\theta\right)-\overline{\mathbf{F}}\left(xX_{n-k:n}\right)}{\overline{\mathbf{F}}\left(xX_{n-k:n}\right)}=x^{1/\gamma_1}O_{\mathbf{P}}\left(x^{\frac{1}{2}\left(\frac{1}{\gamma_2}-\frac{1}{\gamma_1}\right)+\epsilon}\right)=O_{\mathbf{P}}\left(x^{\frac{1}{2\gamma}+\epsilon}\right).$$

Hence we have

$$\sqrt{k}\,\mathbf{M}_{n4}(x)=o_{\mathbf{P}}\left(x^{-1/\gamma_1+\epsilon}\right)O_{\mathbf{P}}\left(x^{\frac{1}{2\gamma}+\epsilon}\right)=o_{\mathbf{P}}\left(x^{\frac{1}{2}\left(\frac{1}{\gamma_2}-\frac{1}{\gamma_1}\right)+\epsilon}\right).$$

By assumption $\overline{\mathbf{F}}$ satisfies the second-order condition of regular variation (4.9), this means that for

$$\lim_{t\to\infty}\frac{\overline{\mathbf{F}}(tx)/\overline{\mathbf{F}}(t)-x^{-1/\gamma_1}}{\mathbf{A}(t)}=x^{-1/\gamma_1}\frac{x^{\rho_1/\gamma_1}-1}{\rho_1\gamma_1}, \tag{2.23}$$

for any $x > 0$, where $\rho_1 < 0$ is the second-order parameter and $\mathbf{A}$ is $RV(\rho_1/\gamma_1)$. The uniform inequality corresponding to (2.23) says: there exist $t_0 > 0$, such that for any $t > t_0$, we have

$$\left| \frac{\overline{\mathbf{F}}(tx)/\overline{\mathbf{F}}(t) - x^{-1/\gamma_1}}{\mathbf{A}(t)} - x^{-1/\gamma_1} \frac{x^{\rho_1/\gamma_1} - 1}{\rho_1\gamma_1} \right| < \epsilon x^{-1/\gamma_1 + \rho_1/\gamma_1 + \epsilon},$$

see for instance assertion (2.3.23) of Theorem 2.3.9 in [24]. It is easy to check that the latter inequality implies that

$$\begin{aligned}\sqrt{k}\mathbf{M}_{n5}(x) &= \sqrt{k}\left(\frac{\overline{\mathbf{F}}(xX_{n-k:n})}{\overline{\mathbf{F}}(X_{n-k:n})} - x^{-1/\gamma_1}\right) \\ &= x^{-1/\gamma_1}\frac{x^{\rho_1/\gamma_1} - 1}{\rho_1\gamma_1}\sqrt{k}\mathbf{A}(X_{n-k:n}) + o_{\mathbf{P}}\left(x^{-1/\gamma_1}\frac{x^{\rho_1/\gamma_1} - 1}{\rho_1\gamma_1}\sqrt{k}\mathbf{A}(X_{n-k:n})\right).\end{aligned}$$

Recall that $a_k = F^{*\leftarrow}(1 - k/n)$ and notice that $X_{n-k:n}/a_k \overset{\mathbf{P}}{\to} 1$ as $n \to \infty$, then in view of the regular variation of $\mathbf{A}$ we infer that $\mathbf{A}(X_{n-k:n}) = (1 + o_{\mathbf{P}}(1))\mathbf{A}(a_k)$. On the other hand, by assumption $\sqrt{k}\mathbf{A}(a_k)$ is asymptotically bounded, therefore

$$\sqrt{k}\mathbf{M}_{n5}(x) = x^{-1/\gamma_1}\frac{x^{\rho_1/\gamma_1} - 1}{\rho_1\gamma_1}\sqrt{k}\mathbf{A}(a_k) + o_{\mathbf{P}}\left(x^{-1/\gamma_1}\right).$$

To summarize, at this stage we showed that

$$\begin{aligned}\mathbf{D}_n\left(x;\widehat{\theta}\right) &= \frac{\gamma}{\gamma_1}x^{1/\gamma_2}W\left(t^{-1/\gamma}\right) + \frac{\gamma}{\gamma_1}\int_{x^{1/\gamma_2}}^{\infty} W\left(t^{-\gamma_2/\gamma}\right)dt \\ &\quad - x^{-1/\gamma_2}\left(\frac{\gamma}{\gamma_1}W(1) + \frac{\gamma}{\gamma_1}\int_1^{\infty} W\left(t^{-\gamma_2/\gamma}\right)dt\right) \\ &\quad + x^{-1/\gamma_1}\frac{x^{\rho_1/\gamma_1} - 1}{\rho_1\gamma_1}\sqrt{k}\mathbf{A}(a_k) + \varsigma(x),\end{aligned}$$

where $\varsigma(x) := o_{\mathbf{P}}\left(x^{-1/\gamma_1+\epsilon}\right) + o_{\mathbf{P}}\left(x^{-1/\gamma_1}\right) + o_{\mathbf{P}}\left(x^{\frac{1}{2}\left(\frac{1}{\gamma_2} - \frac{1}{\gamma_1}\right)+\epsilon}\right)$. By using a change of variables, we show that sum of the first three terms equals the Gaussian process $\Gamma(x;W)$ stated in Theorem 3.1. Recall that $\gamma_1 < \gamma_2$ and

$$\frac{1}{2}\left(\frac{1}{\gamma_2} - \frac{1}{\gamma_1}\right) + \epsilon < 0,$$

then it is easy to verify that $\varsigma(x) = o_{\mathbf{P}}\left(x^{\frac{1}{2}\left(\frac{1}{\gamma_2}-\frac{1}{\gamma_1}\right)+\epsilon}\right)$. It follows that

$$x^{\epsilon}\left\{\mathbf{D}_n\left(x;\widehat{\theta}\right) - \Gamma(x;W) - x^{-1/\gamma_1}\frac{x^{\rho_1/\gamma_1}-1}{\rho_1\gamma_1}\sqrt{k}\mathbf{A}(a_k)\right\}$$
$$= o_{\mathbf{P}}\left(x^{\frac{1}{2}\left(\frac{1}{\gamma_2}-\frac{1}{\gamma_1}\right)+2\epsilon}\right) = o_{\mathbf{P}}(1),$$

uniformly on $x > 1$, therefore

$$\sup_{x>1} x^{\epsilon}\left|\mathbf{D}_n\left(x;\widehat{\theta}\right) - \Gamma(x;W) - x^{-1/\gamma_1}\frac{x^{\rho_1/\gamma_1}-1}{\rho_1\gamma_1}\sqrt{k}\mathbf{A}(a_k)\right| = o_{\mathbf{P}}(1),$$

for any samll $0 < \epsilon < 1/2$, which completes the proof of Theorem 3.1. ■

Applying the weak approximation presented in Theorem 3.1, we establish both consistency and asymptotic normality of our new estimator $\widehat{\gamma}_1$, that we state in the second Theorem.

Theorem 2.2. *Under the assumptions of Theorem* 3.1, *we have*

$$\widehat{\gamma}_1 - \gamma_1$$
$$= k^{-1/2}\int_1^{\infty} x^{-1}\Gamma(x;W)\,dx + \mathbf{A}(a_k)\int_1^{\infty} x^{-1/\gamma_1 - 1}\frac{x^{\rho_1/\gamma_1}-1}{\rho_1\gamma_1}dx + o_{\mathbf{P}}\left(k^{-1/2}\right),$$

this implies that $\widehat{\gamma}_1 \overset{\mathbf{P}}{\to} \gamma_1$. *Whenever* $\sqrt{k}\mathbf{A}(a_k) \to \lambda < \infty$, *we get*

$$\sqrt{k}\left(\widehat{\gamma}_1 - \gamma_1\right) \overset{\mathcal{D}}{\to} \mathcal{N}\left(\frac{\lambda}{1-\rho_1}, \sigma^2\right),$$

where $\sigma^2 := \gamma^2\left(1+\gamma_1/\gamma_2\right)\left(1+\left(\gamma_1/\gamma_2\right)^2\right)\left(1-\gamma_1/\gamma_2\right)^3$, *and* $\mathbf{1}(\mathscr{A})$ *stands for the indicator function pertaining to a set* $\mathscr{A}$.

Proof. From the representation (2.6) we write

$$\widehat{\gamma}_1 - \gamma_1 = T_{n1} + T_{n2} + T_{n3},$$

where

$$T_{n1} := k^{-1/2}\int_1^{\infty} x^{-1}\left\{\mathbf{D}_n\left(x;\widehat{\theta};\gamma_1\right) - \Gamma(x;W) - x^{-1/\gamma_1}\frac{x^{\rho_1/\gamma_1}-1}{\rho_1\gamma_1}\sqrt{k}\mathbf{A}(a_k)\right\}dx$$

$$T_{n2} := k^{-1/2} \int_1^\infty x^{-1} \Gamma(x;W)\, dx$$

and

$$T_{n3} := -\mathbf{A}(a_k) \int_1^\infty x^{-1/\gamma_1 - 1} \frac{x^{\rho_1/\gamma_1} - 1}{\rho_1 \gamma_1} dx.$$

Using Theorem 3.1 yields $T_{n1} = o_{\mathbf{P}}\left(k^{-1/2}\right) \int_1^\infty x^{-1+\epsilon} dx = o_{\mathbf{P}}\left(k^{-1/2}\right) = o_{\mathbf{P}}(1)$. Since $\mathbf{E}\,|W(s)| \leq s^{1/2}$, then it is easy to show that $\int_1^\infty x^{-1}\Gamma(x;W)\,dx = O_{\mathbf{P}}(1)$, it follows that $T_{n2} = O_{\mathbf{P}}\left(k^{-1/2}\right) = o_{\mathbf{P}}(1)$. Using an elementary integration, we get $T_{n3} = \mathbf{A}(a_k)/\left(1-\rho_1\right)$ which tends to zero as $n \to \infty$, because $a_k \to \infty$ and $|\mathbf{A}|$ is regularly varying with negative index. Therefore $\widehat{\gamma}_1 \overset{\mathbf{P}}{\to} \gamma_1$, as $n \to \infty$ which gives the first result of Theorem. To establish the asymptotic normality, we write

$$\sqrt{k}\left(\widehat{\gamma}_1 - \gamma_1\right) = \sqrt{k}\, T_{n1} + \sqrt{k}\, T_{n2} + \sqrt{k}\, T_{n3},$$

where

$$\sqrt{k}\, T_{n1} = o_{\mathbf{P}}(1), \sqrt{k}\, T_{n2} = \int_1^\infty x^{-1}\Gamma(x;W)\, dx$$

and

$$\sqrt{k}\, T_{n3} = \frac{\sqrt{k}\,\mathbf{A}(a_k)}{1-\rho_1}.$$

Note that $\Gamma(x;W)$ is a centred Gaussian process and by using the assumption $\sqrt{k}\,\mathbf{A}(a_k) \to \lambda < \infty$, we end up with

$$\sqrt{k}\left(\widehat{\gamma}_1 - \gamma_1\right) \overset{\mathscr{D}}{\to} \mathscr{N}\left(\frac{\lambda}{1-\rho_1}, \mathbf{E}\left[\int_1^\infty x^{-1}\Gamma(x;W)\, dx\right]^2\right).$$

By elementary calculations (we omit the details) we show that

$$\mathbf{E}\left[\int_1^\infty x^{-1}\Gamma(x;W)\, dx\right]^2 = \sigma^2.$$

■

2.4 Simulation study

In this section we will perform a simulation study in order to compare the finite sample behavior of our new semiparametric estimator $\widehat{\gamma}_1$, given in (2.5), with Woodrofee and Lynden-Bell integral estimators $\widehat{\gamma}_1^{(\mathbf{BMN})}$ and $\widehat{\gamma}_1^{(\mathbf{W})}$, given respectively in (1.21) and (1.26). The truncation and truncated distributions functions $\mathbf{F}$ and $\mathbf{G}$, will be chosen among the following two models:

- Burr (γ, δ) distribution with right-tail function:

$$\overline{H}(x) = \left(1 + x^{1/\delta}\right)^{-\delta/\gamma}, \ x \geq 0, \ \delta > 0, \ \gamma > 0;$$

- Fréchet (γ) distribution with right-tail function:

$$\overline{H}(x) = 1 - \exp\left(-x^{-1/\gamma}\right), x > 0, \gamma > 0.$$

The simulation study be made in fours scenarios following to the choice of the underlying df's $\mathbf{F}$ and $\mathbf{G}_\theta$:

- [$S1$] Burr (γ_1, δ) truncated by Burr (γ_2, δ); with $\theta = (\gamma_2, \delta)$
- [$S2$] Fréchet (γ_1) truncated by Fréchet (γ_2); with $\theta = \gamma_2$
- [$S3$] Fréchet (γ_1) truncated by Burr (γ_2, δ); with $\theta = (\gamma_2, \delta)$
- [$S4$] Burr (γ_1, δ) truncated by Fréchet (γ_2); with $\theta = \gamma_2$

To this end, we fix $\delta = 1/4$ and choose the values 0.6 and 0.8 for γ_1 and 55% and 90% for the portions of observed truncated data given in (1.13) so that the assumption $\gamma_1 < \gamma_2$ stated in Theorem 3.1 holds. In other words the values of p have to be greater than 50%. For each couple (γ_1, p), we solve the equation (1.13) to get the pertaining γ_2-value, which we summarize as follows:

$$(p, \gamma_1, \gamma_2) = (55\%, 0.6, 1.4), (90\%, 0.6, 5.4), (55\%, 0.8, 1.9), (90\%, 0.8, 7.2). \quad (2.24)$$

For each scenario, we simulate 1000 random samples of size $N = 300$ and compute the root mean squared error (RMSE) and the absolute bias (ABIAS) corresponding to each estimator $\widehat{\gamma}_1$, $\widehat{\gamma}_1^{(\mathbf{BMN})}$ and $\widehat{\gamma}_1^{(\mathbf{W})}$. The comparison is done by plotting the ABIAS and RMSE as functions of the sample fraction k which varies from 2 to 120. This range is chosen so that it contains the optimal number of upper extremes k^* used in the computation of the tail index estimate. There are many heuristic methods to select k^*, see for instance [10], here we use the algorithm proposed by [43] in page 137, which is incorporated in the R software "Xtremes"package. Note that the computation the CMLE of θ is made by means of the syntax "maxLik" of the MaxLik R software package. The optimal sample fraction k^* is defined, in this procedure, by

$$k^* := \arg\min_{1<k<n} \frac{1}{k} \sum_{i=1}^{k} i^{\omega} \left| \widehat{\gamma}(i) - \text{median}\left\{\widehat{\gamma}(1), ..., \widehat{\gamma}(k)\right\} \right|,$$

for suitable constant $0 \leq \omega \leq 1/2$, where $\widehat{\gamma}(i)$ corresponds to an estimator of tail index γ, based on the i upper order statistics, of a Pareto-type model. We observed, in our simulation study, that $\omega = 0.3$ allows better results both in terms of bias and rmse. It is worth mentioning that making N vary did not provide notable findings, therefore we kept the size N fixed. The finite sample behaviors of the above mentioned estimators are illustrated in Figures 2.1-2.8.

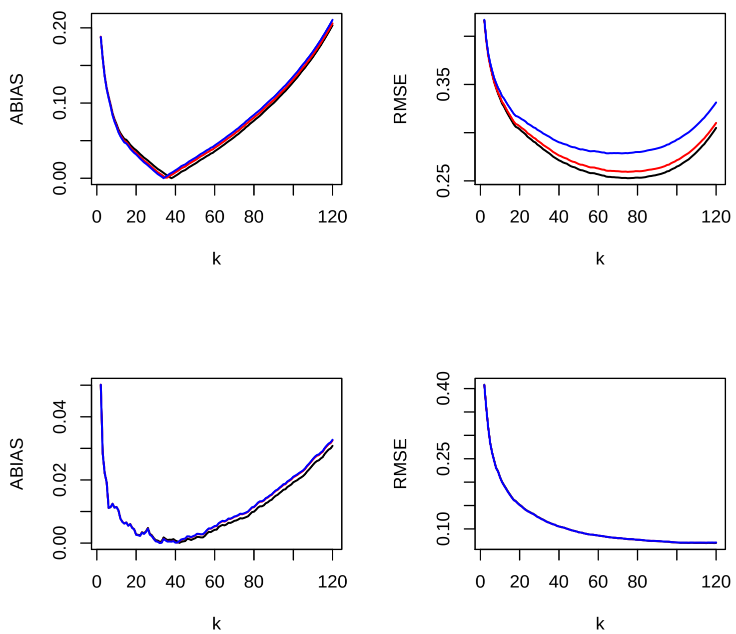

Figure 2.1: Absolute bias (left two panels) and RMSE (right two panels) of $\widehat{\gamma}_1$ (black) and $\widehat{\gamma}_1^{(\mathbf{BMN})}$ (red) and $\widehat{\gamma}_1^{(\mathbf{W})}$(blue), corresponding to two situations of scenario $S_1 : \left(\gamma_1 = 0.6, p = 55\%\right)$ (top two panels) and $\left(\gamma_1 = 0.6, p = 90\%\right)$ (bottom two panels) based on 1000 samples of size 300.

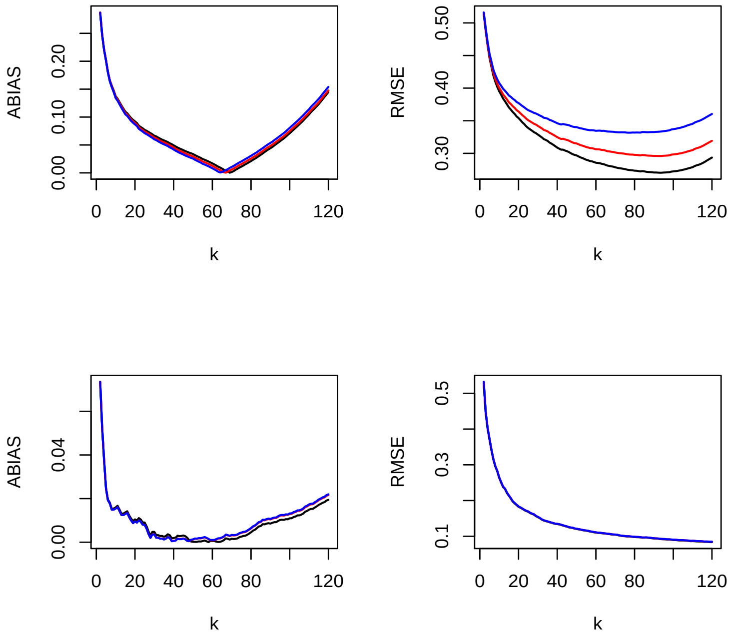

Figure 2.2: Absolute bias (left two panels) and RMSE (right two panels) of $\widehat{\gamma}_1$ (black) and $\widehat{\gamma}_1^{\mathbf{(BMN)}}$ (red) and $\widehat{\gamma}_1^{\mathbf{(W)}}$(blue), corresponding to two situations of scenario $S_1 : (\gamma_1 = 0.8, p = 55\%)$ (top two panels) and $(\gamma_1 = 0.8, p = 90\%)$ (bottom two panels) based on 1000 samples of size 300.

The overall conclusion is that the biases of three estimators are almost equal, however in the case of medium truncation ($p \approx 50\%$) the RMSE of our new semiparametric $\widehat{\gamma}_1$ is clearly the smallest compared that of $\widehat{\gamma}_1^{\mathbf{(BMN)}}$ and $\widehat{\gamma}_1^{\mathbf{(W)}}$.

Actually, the medium truncation situation is the most frequently encountered in real data, while the strong truncation ($p >> 50\%$) remains, up to our knowledge, theoretical. In this sense, we may consider that the semiparametric estimator is more efficient than the two other ones.

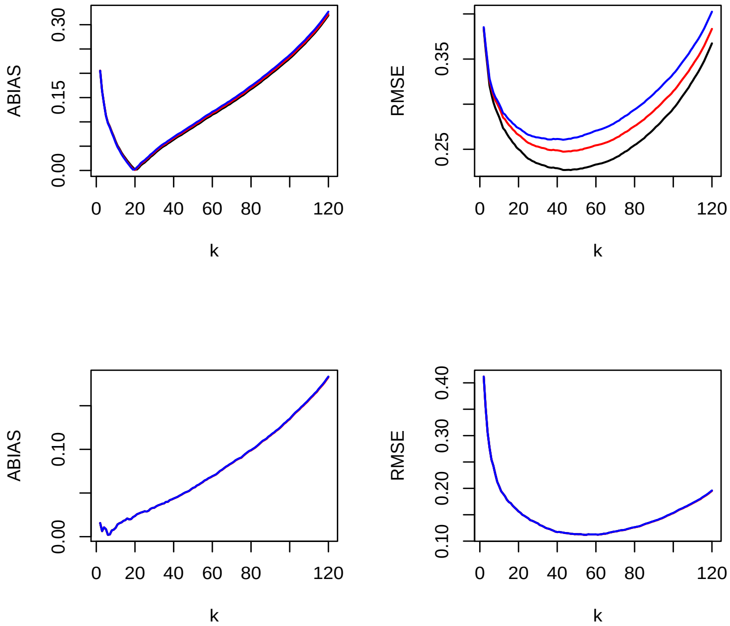

Figure 2.3: Absolute bias (left two panels) and RMSE (right two panels) of $\widehat{\gamma}_1$ (black) and $\widehat{\gamma}_1^{\mathbf{(BMN)}}$ (red) and $\widehat{\gamma}_1^{\mathbf{(W)}}$(blue), corresponding to two situations of scenario $S_2 : (\gamma_1 = 0.6, p = 55\%)$ (top two panels) and $(\gamma_1 = 0.6, p = 90\%)$ (bottom two panels) based on 1000 samples of size 300.

We point out that the two estimators $\widehat{\gamma}_1^{\mathbf{(BMN)}}$ and $\widehat{\gamma}_1^{\mathbf{(W)}}$ have almost the same behavior which actually was noticed before by [21]. The optimal sample fractions and estimate values of the tail index obtained through the three estimators are given in Tables 2.1-2.4.

	k^*	$\widehat{\gamma}_1$	k^*	$\widehat{\gamma}_1^{\mathbf{(BMN)}}$	k^*	$\widehat{\gamma}_1^{\mathbf{(W)}}$
$S1$	44	0.600	41	0.599	40	0.600
$S2$	18	0.601	17	0.600	16	0.597
$S3$	21	0.601	20	0.601	19	0.599
$S4$	30	0.603	27	0.600	25	0.598

Table 2.1: Optimal sample fractions and estimate values of the tail index $\gamma_1 = 0.6$ based on 1000 samples of size 300 for the four scenarios with $p = 0.55$.

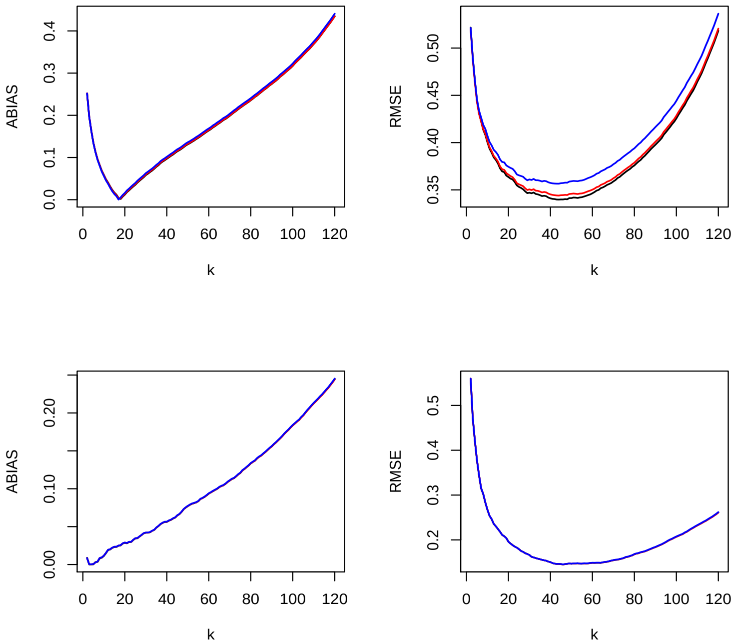

Figure 2.4: Absolute bias (left two panels) and RMSE (right two panels) of $\widehat{\gamma}_1$ (black) and $\widehat{\gamma}_1^{(\mathbf{BMN})}$ (red) and $\widehat{\gamma}_1^{(\mathbf{W})}$(blue), corresponding to two situations of scenario $S_2 : \left(\gamma_1 = 0.8, p = 55\%\right)$ (top two panels) and $\left(\gamma_1 = 0.8, p = 90\%\right)$ (bottom two panels) based on 1000 samples of size 300.

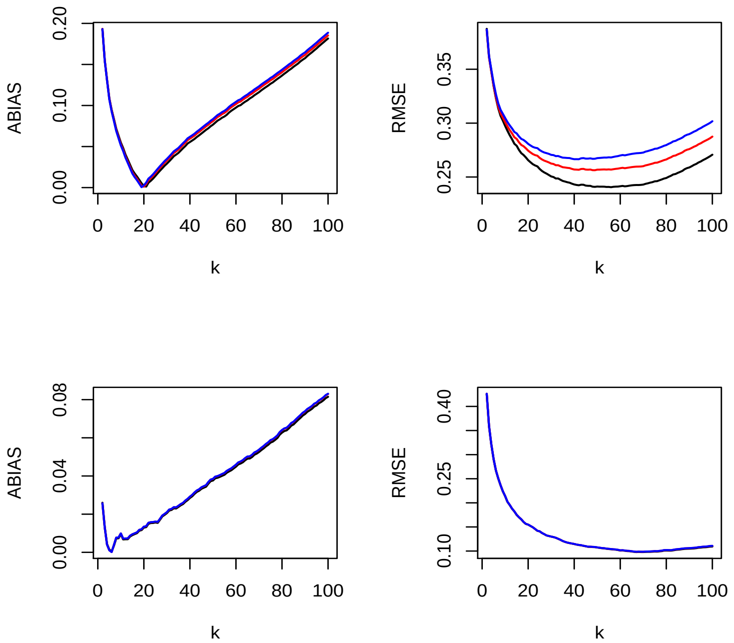

Figure 2.5: Absolute bias (left two panels) and RMSE (right two panels) of $\widehat{\gamma}_1$ (black) and $\widehat{\gamma}_1^{(\mathbf{MBN})}$ (red) and $\widehat{\gamma}_1^{(\mathbf{W})}$(blue), corresponding to two situations of scenario $S_3 : \left(\gamma_1 = 0.6, p = 55\%\right)$ (top two panels) and $\left(\gamma_1 = 0.6, p = 90\%\right)$ (bottom two panels) based on 1000 samples of size 300.

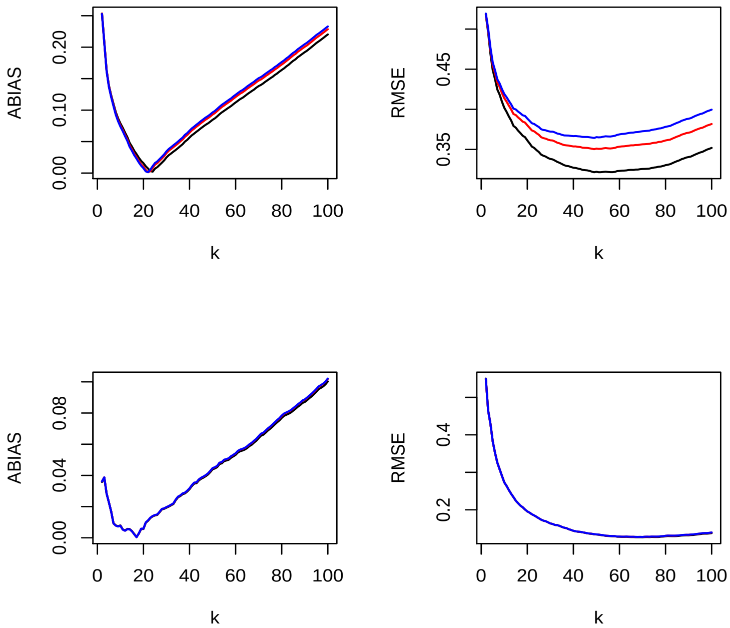

Figure 2.6: Absolute bias (left two panels) and RMSE (right two panels) of $\widehat{\gamma}_1$ (black) and $\widehat{\gamma}_1^{(\mathbf{BMN})}$ (red) and $\widehat{\gamma}_1^{(\mathbf{W})}$(blue), corresponding to two situations of scenario $S_3 : \left(\gamma_1 = 0.8, p = 55\%\right)$ (top two panels) and $\left(\gamma_1 = 0.8, p = 90\%\right)$ (bottom two panels) based on 1000 samples of size 300.

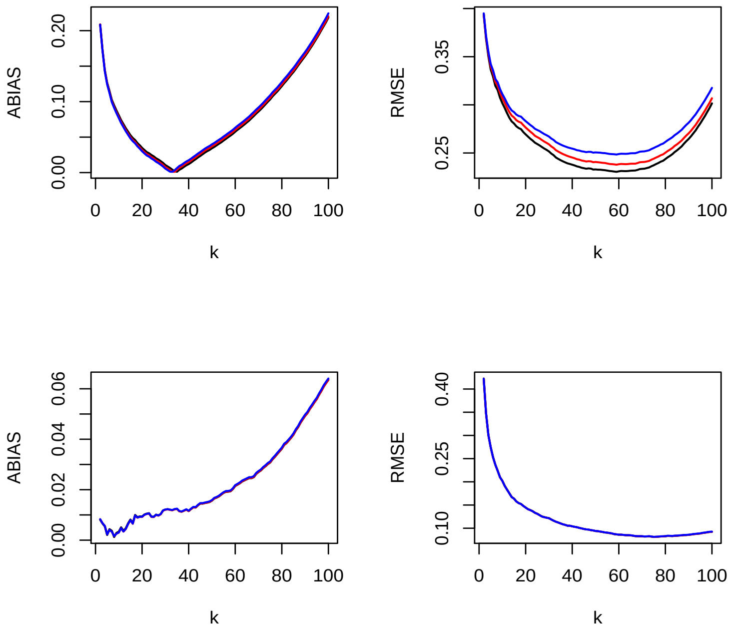

Figure 2.7: Absolute bias (left two panels) and RMSE (right two panels) of $\widehat{\gamma}_1$ (black) and $\widehat{\gamma}_1^{\mathbf{(BMN)}}$ (red) and $\widehat{\gamma}_1^{\mathbf{(W)}}$(blue), corresponding to two situations of scenario $S_4 : \left(\gamma_1 = 0.6, p = 55\%\right)$ (top two panels) and $\left(\gamma_1 = 0.6, p = 90\%\right)$ (bottom two panels) based on 1000 samples of size 300.

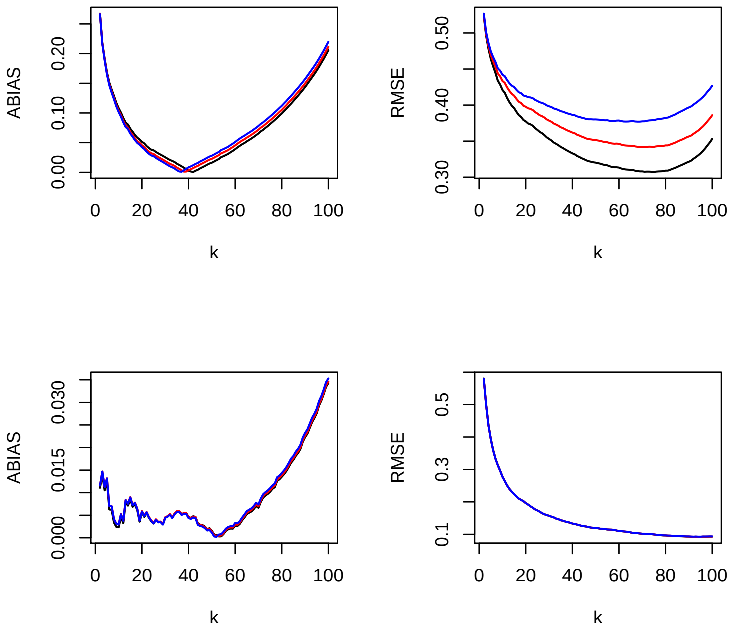

Figure 2.8: Absolute bias (left two panels) and RMSE (right two panels) of $\widehat{\gamma}_1$ (black) and $\widehat{\gamma}_1^{(\mathbf{BMN})}$ (red) and $\widehat{\gamma}_1^{(\mathbf{W})}$(blue), corresponding to two situations of scenario $S_4 : \left(\gamma_1 = 0.8, p = 55\%\right)$ (top two panels) and $\left(\gamma_1 = 0.8, p = 90\%\right)$ (bottom two panels) based on 1000 samples of size 300.

	k^*	$\widehat{\gamma}_1$	k^*	$\widehat{\gamma}_1^{(\mathbf{BMN})}$	k^*	$\widehat{\gamma}_1^{(\mathbf{W})}$
$S1$	82	0.610	82	0.611	82	0.611
$S2$	37	0.640	37	0.640	37	0.640
$S3$	46	0.633	37	0.625	37	0.625
$S4$	52	0.610	52	0.610	52	0.610

Table 2.2: Optimal sample fractions and estimate values of the tail index $\gamma_1 = 0.6$ based on 1000 samples of size 300 for the four scenarios with $p = 0.9$.

	k^*	$\widehat{\gamma}_1$	k^*	$\widehat{\gamma}_1^{(\mathbf{BMN})}$	k^*	$\widehat{\gamma}_1^{(\mathbf{W})}$
$S1$	59	0.799	57	0.800	54	0.799
$S2$	21	0.803	21	0.803	20	0.799
$S3$	24	0.802	22	0.798	22	0.801
$S4$	51	0.799	52	0.800	50	0.801

Table 2.3: Optimal sample fractions and estimate values of the tail index $\gamma_1 = 0.8$ based on 1000 samples of size 300 for the four scenarios with $p = 0.55$.

	k^*	$\widehat{\gamma}_1$	k^*	$\widehat{\gamma}_1^{(\mathbf{BMN})}$	k^*	$\widehat{\gamma}_1^{(\mathbf{W})}$
$S1$	90	0.804	90	0.806	90	0.807
$S2$	34	0.845	34	0.846	34	0.846
$S3$	40	0.831	40	0.831	40	0.831
$S4$	71	0.814	71	0.814	71	0.815

Table 2.4: Optimal sample fractions and estimate values of the tail index $\gamma_1 = 0.8$ based on 1000 samples of size 300 for the four scenarios with $p = 0.9$.

2.5 Real data example

In this section, we give an application to the AIDS data set, available in the "DTDA" R package and the textbook of [29] (page 19) and already used by [31]. The data present the infection and induction times for $n = 258$ adults who were infected with HIV virus and developed AIDS by June 30, 1986. The variable of interest here is the time of induction $\mathbf{T}$ of the disease duration which elapses between the date of infection M and the date $M + T$ of the declaration of the disease. The sample $(T_1, M_1), ..., (T_n, M_n)$ are taken between two fixed dates: "0" and "8", i.e. between April 1, 1978, and June 30, 1986. The initial date "0" denotes an infection occurring in the three months: from April 1, 1978, to June 30, 1978. Let us assume that M and T are the observed rv's, corresponding to the underlying rv's $\mathbf{M}$ and $\mathbf{T}$, given by the truncation scheme $0 \leq \mathbf{M} + \mathbf{T} \leq 8$, which in turn may be rewritten into

$$0 \leq M \leq S, \tag{2.25}$$

where $S := 8 - T$. To work within the framework of the present paper, let us make the following transformations:

$$X := \frac{1}{S + \epsilon} \text{ and } Y := \frac{1}{M + \epsilon}, \tag{2.26}$$

where $\epsilon = 0.05$ so that the two denominators be non-null. Thus, in view of (2.25), we have $X \leq Y$, which means that X is randomly right-truncated by Y. Thereby, for the given sample $(T_1, M_1), ..., (T_n, M_n)$, from (T, M), the previous transformations produce a new one $(X_1, Y_1), ..., (X_n, Y_n)$ from (X, Y).

Let us now denote by $\mathbf{F}$ and $\mathbf{G}$ the df's of the underling rv's $\mathbf{X}$ and $\mathbf{Y}$ corresponding to the truncated rv's X and Y, respectively. By using parametric likelihood methods, [35] fit both df's of $\mathbf{M}$ and $\mathbf{S}$ by the two-parameter Weibull model, this implies that the df's of $\mathbf{F}$ and $\mathbf{G}$ by may be fitted by two-parameter Fréchet model, namely $\mathbf{H}_{(a.r)}(x) = \exp(-a^r x^{-r})$, $x > 0$, $a > 0$, $r > 0$, hence both $\mathbf{F}$ and $\mathbf{G}$ are heavy-tailed. The estimated parameters corresponding to the fitting of df $\mathbf{G}$ are $a_0 = 0.004$ and $r_0 = 2.1$, see also [31] page 520. Thus on may consider that df $\mathbf{G}$ is known and equals $\mathbf{G}_\theta = \mathbf{H}_{(a_0, r_0)}$, where $\theta = (a_0, r_0)$. By using the Thomas and Reiss algorithm, given above, we compute the optimal sample fraction k^* corresponds to the tail index estimator $\widehat{\gamma}_1$ of df $\mathbf{F}$ is γ_1. We find

$$k^* = 19,\ X_{n-k:n} = 0.356 \text{ and } \widehat{\gamma}_1 = 0.917. \tag{2.27}$$

The well-known Weissman estimator [49] of the high quantile, $q_v := \mathbf{F}^{-1}(1 - v_n)$, corresponding to the underling df $\mathbf{F}$ is given by

$$\widehat{q}_v := X_{n-k:n} \left(\frac{v}{\overline{\mathbf{F}}_n (X_{n-k:n})} \right)^{-\widehat{\gamma}_1},$$

where $v = 1/(2n)$, and $\mathbf{F}_n$ is the semiparametric estimator of df $\mathbf{F}$ of $\mathbf{X}$ given in (2.1). From the values (2.27), we get $\widehat{q}_v = 0.061$. Let us now compute the high quantile of $\mathbf{T}$ based on the original data, $T_1,...,T_n$. Recall that $\mathbf{P}(\mathbf{X} \geq q_v) = v$ and $\mathbf{X} = \mathbf{1}/(8 - \mathbf{T} + \epsilon)$, this implies that $\mathbf{P}(\mathbf{T} \geq 1/q_v - 8 + \epsilon) = v$, this means that $1/q_v - 8 + \epsilon$ is the high quantile of $\mathbf{T}$, which corresponds to the end-time t_{end} that we want to estimate. Thereby $\widehat{t}_{end} = 1/\widehat{q}_v - 8 + 10^{-2} = 1/0.061 - 8 + 10^{-2} = 8.40$, the value the end time of induction of AIDS is: 8 years, 4 months and 24 days.

CHAPTER 3

BIAS REDUCTION IN KERNEL TAIL INDEX ESTIMATION FOR RANDOMLY TRUNCATED PARETO-TYPE DATA

The current chapter, allows to introduce A bias reduction to a kernel estimator of the tail index of randomly right-truncated Pareto-type distributions is made. The asymptotic normality of the derived estimator is established by assuming the second-order condition of regular variation. A simulation study is carried out to evaluate the finite sample behavior of the proposed estimator and compare it to those with non-reduced bias. An application to a real dataset of lifetimes of automobile brake pads is done.

3.1 Bias reduction of $\widehat{\gamma}_{1,K}$

In this section we propose a bias reduction to $\widehat{\gamma}_{1,K}$ by means of a weighted estimator to the rate of convergence $A_{\mathbf{F}}(n/k)$ given in asymptotic approximation (1.24). More precisely, given estimators for $\eta_{1,K}$ and $A_{\mathbf{F}}(n/k)$, denoted $\widehat{\eta}_{1,K}$ and $\widehat{A}_{\mathbf{F},K}(n/k)$, we propose an asymptotically centred normal estimator of γ_1 defined by

$$\widehat{\gamma}_{1,K}^{*} := \widehat{\gamma}_{1,K} - \widehat{\eta}_{1,K}\widehat{A}_{\mathbf{F},K}(n/k). \tag{3.1}$$

For the construction of $\widehat{A}_{\mathbf{F},K}(n/k)$, we opt for a similar approach as the one used in [4] to the bias-reduction in tail index estimation for censored data by introducing a weight function. To this end, let us define

$$L_{t,K} := \int_1^{\infty} \frac{\overline{\mathbf{F}}(tx)}{\overline{\mathbf{F}}(t)} K\left(\frac{\overline{\mathbf{F}}(tx)}{\overline{\mathbf{F}}(t)}\right) \frac{dx}{x} \text{ and } E_{t,K}(\beta) := 1 - \beta \int_1^{\infty} x^{-\beta-1} K^{*}\left(\frac{\overline{\mathbf{F}}(tx)}{\overline{\mathbf{F}}(t)}\right) dx, \tag{3.2}$$

for $\beta > 0$, where K^* is a measurable positive weight function that depends on K. The weight function K^* has to be chosen so as to improve the estimation of $A_{\mathbf{F}}(n/k)$ leading to the accuracy of the bias-reduced estimator $\widehat{\gamma}_{1,K}^{*}$ as function of the tuning parameter β. The first condition which K^* has to fulfill is that its derivative be positive nonincreasing in order to assign less weight to the distribution tail corresponding to the estimator $\widehat{E}_{k,K}(\beta)$ of $E_{t,K}(\beta)$, given in (3.6) below. It is convenient that both $\widehat{\gamma}_{1,K}$ and $\widehat{A}_{\mathbf{F},K}(n/k)$ be related to two weight functions depending on the same kernel function K in order to get an easy-to-use formula to $\widehat{\gamma}_{1,K}^{*}$. At first glance, we tried $K^*(s) = sK(s)$, as in $L_{t,K}$, but this did not give satisfaction in the finite sample behavior (in terms of bias and mean squared error as well). Our second choice fell on the form $K^*(s) = \int_0^s K(t)dt$, where, in addition to assumptions $[A1]-[A4]$, the kernel K needs here to satisfy the fifth one $[A5]$. This provides interesting results and improves those of the non-weighted case (see Section of simulation study). Of course, this form is not unique and one can suggest other forms to K^*. It is clear that when using the indicator kernel $K = K_1$ in (3.2), both the tail functionals $L_{t,K}$ and $E_{t,K}(\beta)$ meet those introduced, without weight functions, by [4]. One may check that the three kernel functions K_i, $i = 2, 3, 4$ fulfill the conditions $[A1]-[A5]$.

In proposition 3.1, we state that $L_{t,K} \to \gamma_1$ and $E_{t,K}(\beta) \to \eta_{2,K}$ as $t \to \infty$, for a fixed $\beta > 0$, where

$$\eta_{2,K} = \eta_{2,K}(\beta;\gamma_1) := 1 - \beta\gamma_1 \int_0^1 s^{\gamma_1\beta-1} K^*(s)\,ds.$$

Moreover, we show that

$$A_{\mathbf{F}}(t) = \frac{E_{t,K}(\beta) - f_*(L_{t,K})}{\eta_{3,K} - f'_*(\gamma_1)\,\eta_{1,K}} (1 + o(1)), \text{ as } t \to \infty, \tag{3.3}$$

where

$$f_*(x) := 1 - \beta x \int_0^1 s^{\beta x-1} K^*(s)\,ds, \text{ for } \beta > 0, \tag{3.4}$$

and

$$\eta_{3,K} = \eta_{3,K}(\beta;\tau_1,\gamma_1) := \frac{\beta}{\tau_1} \int_0^1 s^{\gamma_1\beta} \left(1 - s^{-\tau_1}\right) K(s)\,ds.$$

Let us set

$$\eta_{4,K} = \eta_{4,K}(\beta;\gamma_1) := f'_*(\gamma_1) = -\beta \int_0^1 s^{\beta\gamma_1-1} (\beta\gamma_1 \log s + 1) K^*(s)\,ds.$$

To obtain an estimator to $A_{\mathbf{F}}(n/k)$ it suffices to substitute, in (3.3), both γ_1 and $L_{t,K}$ by $\widehat{\gamma}_{1,K}$ and replace $\eta_{i,K}$, $i = 1,...,4$ by their respective estimators $\widehat{\eta}_{1,K} := \eta_{1,K}(\widehat{\tau}_1)$, $\widehat{\eta}_{2,K} := \eta_{2,K}(\beta;\widehat{\gamma}_1)$, $\widehat{\eta}_{3,K} := \eta_{3,K}(\beta;\widehat{\tau}_1,\widehat{\gamma}_1)$ and $\widehat{\eta}_{4,K} := \eta_{4,K}(\beta;\widehat{\gamma}_1)$, where $\widehat{\tau}_1$ is the consistent estimator for second-order parameter τ_1 first proposed by [22]. It is worth mentioning that $\widehat{\eta}_{2,K}$ coincides with $f_*(\widehat{\gamma}_{1,K})$, this means that $\widehat{\eta}_{2,K}$ is an estimator for $f_*(L_{t,K})$. For the estimation of $E_{t,K}(\beta)$, we substitute $\mathbf{F}$ by $\mathbf{F}_n$ and t by $X_{n-k:n}$, to get

$$\widehat{E}_{k,K}(\beta) = 1 - \beta \int_1^\infty K^*\left(\frac{\overline{\mathbf{F}}_n(xX_{n-k:n})}{\overline{\mathbf{F}}_n(X_{n-k:n})}\right) x^{-\beta-1}\,dx, \tag{3.5}$$

which by an integration by parts becomes

$$\widehat{E}_{k,K}(\beta) = \int_{X_{n-k:n}}^\infty K\left(\frac{\overline{\mathbf{F}}_n(x)}{\overline{\mathbf{F}}_n(X_{n-k:n})}\right) (x/X_{n-k:n})^{-\beta}\, d\frac{\mathbf{F}_n(x)}{\overline{\mathbf{F}}_n(X_{n-k:n})}.$$

The relation between df's $\mathbf{F}$ and F^* is $\int_x^\infty d\mathbf{F}(y)/\mathbf{F}(y) = \int_x^\infty dF^*(y)/C(y)$, where $C(y) := \mathbf{P}(X \le y \le Y)$ denotes the theoretical counterpart of C_n defined above in Woodroofe's nonparametric estimator $\mathbf{F}_n$, see for instance equation (1) in [47].

Differentiating leads to the following crucial equation $C(x)d\mathbf{F}(x) = \mathbf{F}(x)dF^*(x)$, which implies that $C_n(x)d\mathbf{F}_n(x) = \mathbf{F}_n(x)dF_n^*(x)$, where $F_n^*(x) := n^{-1}\sum_{i=1}^n \mathbf{1}(X_i \leq x)$ being the empirical counterpart of df F^*. This allows to rewrite (3.5) into

$$\widehat{E}_{k,K}(\beta) = \int_{X_{n-k:n}}^{\infty} \frac{\mathbf{F}_n(x)}{C_n(x)} K\left(\frac{\overline{\mathbf{F}}_n(x)}{\overline{\mathbf{F}}_n(X_{n-k:n})}\right)(x/X_{n-k:n})^{-\beta} d\frac{F_n^*(x)}{\overline{\mathbf{F}}_n(X_{n-k:n})},$$

which may be made into

$$\widehat{E}_{k,K}(\beta) = \frac{1}{n\overline{\mathbf{F}}_n(X_{n-k:n})}\sum_{i=1}^{k}\frac{\mathbf{F}_n(X_{n-i+1:n})}{C_n(X_{n-i+1:n})}K\left(\frac{\overline{\mathbf{F}}_n(X_{n-i+1:n})}{\overline{\mathbf{F}}_n(X_{n-k:n})}\right)\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{-\beta}. \quad (3.6)$$

A smoothed estimator

$$\widehat{A}_{\mathbf{F},K}(n/k) := \frac{\widehat{E}_{k,K}(\beta) - \widehat{\eta}_{2,K}}{\widehat{\eta}_{3,K} - \widehat{\eta}_{4,K}\widehat{\eta}_{1,K}},$$

is now constructed of $A_{\mathbf{F}}(n/k)$. Substituting this expression in (3.1) we end up with the new bias-reduced estimator of γ_1:

$$\widehat{\gamma}_{1,K}^*(\beta) := \widehat{\gamma}_{1,K} - \widehat{\eta}_{1,K}\frac{\widehat{E}_{k,K}(\beta) - \widehat{\eta}_{2,K}}{\widehat{\eta}_{3,K} - \widehat{\eta}_{4,K}\widehat{\eta}_{1,K}}, \text{ for } \beta > 0. \quad (3.7)$$

In particular, considering the indicator kernel function K_1, we get

$$\eta_1 = \frac{1}{1-\tau_1},\ \eta_2 = \frac{1}{\beta\gamma_1 + 1},$$

$$\eta_3 = \frac{-\beta}{(\beta\gamma_1+1)(\beta\gamma_1 - \tau_1 + 1)} \text{ and } \eta_4 = -\frac{\beta}{(\beta\gamma_1+1)^2}.$$

Once again, substituting γ_1 by $\widehat{\gamma}_1$ and τ_1 by $\widehat{\tau}_1$, we derive estimators to $\widehat{\eta}_i$ for η_i, $i = 1, ..., 4$, leading to the corresponding estimator of $A_{\mathbf{F}}(n/k)$ given by

$$\widehat{A}_{\mathbf{F}}(n/k) := \frac{(1-\widehat{\tau}_1)(\beta\widehat{\gamma}_1+1)^2(\beta\widehat{\gamma}_1 - \widehat{\tau}_1 + 1)}{\beta^2\widehat{\tau}_1\widehat{\gamma}_1}\left(\widehat{E}_k(\beta) - \frac{1}{1+\beta\widehat{\gamma}_1}\right),$$

where $\widehat{E}_k(\beta) := \widehat{E}_{k,K_1}(\beta)$. Thereby, we define two reduced estimators to γ_1, which are less performing then $\widehat{\gamma}_{1,K}^*(\beta)$, given by

$$\overline{\gamma}_{1,K}^*(\beta) := \widehat{\gamma}_{1,K} - \widehat{\eta}_{1,K}\widehat{A}_{\mathbf{F}}(n/k) \text{ and } \widetilde{\gamma}_1^*(\beta) := \widehat{\gamma}_1 - \frac{1}{1-\widehat{\tau}_1}\widehat{A}_{\mathbf{F}}(n/k). \quad (3.8)$$

	non-reduced estimator of γ_1	estimator of $A_{\mathbf{F}}(n/k)$
$\widehat{\gamma}^*_{1,K}(\beta)$	K	K^*
$\overline{\gamma}^*_{1,K}(\beta)$	K	no weight
$\widetilde{\gamma}^*_1(\beta)$	no weight	no weight
$\widehat{\gamma}_{1,K}$	K	/
$\widehat{\gamma}_1$	no weight	/

Table 3.1: Tail index estimators of γ_1 according to the assigned weights.

The estimator $\overline{\gamma}^*_{1,K}(\beta)$ is half-weighted in the sense that $\widehat{\gamma}_{1,K}$ is assigned by the kernel function $K \neq K_1$ and $\widehat{A}_{\mathbf{F}}(n/k)$ is without kernel, while in $\widetilde{\gamma}^*_1(\beta)$ both $\widehat{\gamma}_1$ and $\widehat{A}_{\mathbf{F}}(n/k)$ are without kernel. Both estimators are introduced to illustrate the performance of the bias reduction following the weights that are assigned to each one of $\widehat{\gamma}_1$ and $\widehat{A}_{\mathbf{F}}(n/k)$. In the simulation study, we will show that indeed $\overline{\gamma}^*_{1,K}(\beta)$ performs better than $\widetilde{\gamma}^*_1(\beta)$ and that $\widehat{\gamma}^*_{1,K}(\beta)$ in turn improves $\overline{\gamma}^*_{1,K}(\beta)$. Notice that the second formula in (3.8) was first introduced by [4] to reduce the bias in tail index estimation for censored data that we adapt to the truncation case and generalize to the kernel estimation framework. The aforementioned five estimators of γ_1 are summarized in Table 4.3 according to the assigned weights to the non-reduced estimator of γ_1 and the estimator of $A_{\mathbf{F}}(n/k)$.

3.2 Main results and proof

Theorem 3.1. *Assume that both second-order conditions* (1.23) *hold. Let $k = k_n$ be a sequence of integer such that $k \to \infty$, $k/n \to 0$ and $\sqrt{k}A_{\mathbf{F}}(n/k)$ is asymptotically bounded. Then, for a given nonincreasing kernel function K satisfying assumptions* $[A1]-[A5]$ *and for a fixed $\beta > 0$, we have*

$$\sqrt{k}\left(\widehat{\gamma}^*_{1,K}(\beta) - \gamma_1\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \left(\gamma^2/\gamma_1\right)^2 \int_0^1 q^2_{\beta,K}(s)ds\right),$$

as $n \to \infty$, provided that $\gamma_1 < \gamma_2$, where

$$q_{\beta,K}(s) := \frac{\gamma}{\gamma_1} K_\beta\left(s^{\gamma/\gamma_1}\right) - \frac{\gamma}{\gamma_2} s^{-\gamma/\gamma_1} \int_0^{s^{\gamma/\gamma_1}} u^{-\gamma_2/\gamma} K_\beta(u)du, \qquad (3.9)$$

with $K_\beta(u) := u^{-1}\int_0^u \left(1 + \rho_K \eta_{4,K} + \rho_K s^{\beta\gamma_1}\right) K(s)ds$ and $\rho_K := \left(\eta_{3,K}/\eta_{1,K} - \eta_{4,K}\right)^{-1}$.

Remark 3.1. *Note that* $\widehat{\gamma}_{1,K}^{*}(0)$ *reduces to* $\widehat{\gamma}_{1,K}$ *before the bias reduction. For this particular case,* $\beta = 0$, *we have* $\eta_{4,K} = 0$ *and* $K_0 = K$, *therefore* $q_{0,K} = \varphi_K$, *given in* (1.25). *Thus Theorem* 3.1 *meets that of [7] stated in* (1.24).

3.2.1 Instrumental result

Proposition 3.1. *Assume that* $\overline{\mathbf{F}} \in RV_2\left(-1/\gamma_1; \tau_1, A_{\mathbf{F}}\right)$ *and* K *satisfies assumptions* $[A1]-[A4]$, *then: (i)* $E_t\left(\beta\right) = \eta_2 + \eta_3 A_{\mathbf{F}}(t)(1+o(1))$, *for* $\beta > 0$, *and for a given twice-differentiable function* f : *(ii)* $f\left(L_{t,K}\right) = f\left(\gamma_1\right) + f'\left(\gamma_1\right)\eta_1 A_{\mathbf{F}}(t)(1+o(1))$, *where* $\eta_i = \eta_{i,K}$, $i = 1,2,3$ *are stated in (1.24) and subsection* 3.3.1 *respectively. Moreover*

$$(iii)\ A_{\mathbf{F}}(t) = \frac{E_t\left(\beta\right) - f_*\left(L_t\right)}{\eta_3 - f_*'\left(\gamma_1\right)\eta_1}(1+o(1)),\ as\ t \to \infty,$$

where f_* *is as in* (3.4), $E_t\left(\beta\right) := E_{t,K}\left(\beta\right)$ *and* $L_t := L_{t,K}$.

Proof. Recall that (3.2), and let us decompose $E_t\left(\beta\right)$ into the sum of

$$E_t^{(1)}\left(\beta\right) := 1 - \beta \int_1^\infty x^{-\beta-1} K^*\left(x^{-1/\gamma_1}\right) dx$$

and

$$E_t^{(2)}\left(\beta\right) := -\beta \int_1^\infty x^{-\beta-1}\left(K^*\left(\frac{\overline{\mathbf{F}}(tx)}{\overline{\mathbf{F}}(t)}\right) - K^*\left(x^{-1/\gamma_1}\right)\right) dx.$$

Using the change of variables $s = x^{-1/\gamma_1}$ we readily show that $E_t^{(1)}\left(\beta\right) = \eta_2$, for $\beta > 0$. Applying Taylor's expansion to K^*, we may rewrite $E_t^{(2)}\left(\beta\right)$ into the sum of

$$E_t^{(2,1)}\left(\beta\right) := -\beta \int_1^\infty x^{-\beta-1}\left(\frac{\overline{\mathbf{F}}(tx)}{\overline{\mathbf{F}}(t)} - x^{-1/\gamma_1}\right) K\left(x^{-1/\gamma_1}\right) dx$$

and

$$E_t^{(2,2)}\left(\beta\right) := -\frac{\beta}{2} \int_1^\infty x^{-\beta-1}\left(\frac{\overline{\mathbf{F}}(tx)}{\overline{\mathbf{F}}(t)} - x^{-1/\gamma_1}\right)^2 K'\left(\xi_t(x)\right) dx,$$

where $\xi_t(x)$ is between $\overline{\mathbf{F}}(tx)/\overline{\mathbf{F}}(t)$ and x^{-1/γ_1}. Since $\overline{\mathbf{F}} \in \mathscr{RV}_2\left(-1/\gamma_1; \tau_1, A_{\mathbf{F}}\right)$, then making use of Potters inequalities corresponding to the second-order condition of df F, we write

$$\left| \frac{\frac{\overline{\mathbf{F}}(tx)}{\overline{\mathbf{F}}(t)} - x^{-1/\gamma_1}}{A_{\mathbf{F}}(t)} - x^{-1/\gamma_1}\frac{x^{\tau_1/\gamma_1} - 1}{\tau_1\gamma_1} \right| \le \epsilon x^{-1/\gamma_1 + \tau_1/\gamma_1 + \epsilon}, \tag{3.10}$$

for any $\epsilon > 0$, for all $x > 1$ and for all large t, see for instance

Theorem 2.3.9 in [24]. Using this inequality, we end up with

$$E_t^{(2,1)}(\beta) = -(1+o(1))A_{\mathbf{F}}(t)\beta\int_1^\infty x^{-\beta-1/\gamma_1-1}\frac{x^{\tau_1/\gamma_1}-1}{\tau_1\gamma_1}K\left(x^{-1/\gamma_1}\right)dx.$$

Once again, using change of variables $s = x^{-1/\gamma_1}$, we show easily that the previous integral equals to η_3. In view of assumption $[A4]$, the function K' is bounded, it follows that

$$E_t^{(2,2)}(\beta) = O(1)\int_1^\infty x^{-\beta-1}\left|\frac{\overline{\mathbf{F}}(tx)}{\overline{\mathbf{F}}(t)} - x^{-1/\gamma_1}\right|^2 dx.$$

Let $\epsilon > 0$ so small such that $-1/\gamma_1 + \epsilon < 0$. The Potters inequalities corresponding to the first order condition of regularly varying functions says:

$$\left|\frac{\overline{\mathbf{F}}(tx)}{\overline{\mathbf{F}}(t)} - x^{-1/\gamma_1}\right| \le \epsilon x^{-1/\gamma_1+\epsilon} < \epsilon,$$

for all $x \ge 1$ and for all large t, see for instance

Proposition B.1.9, assertion 5 in [24]. It follows that

$$E_t^{(2,2)}(\beta) = o(1)\int_1^\infty x^{-\beta-1}\left|\frac{\overline{\mathbf{F}}(tx)}{\overline{\mathbf{F}}(t)} - x^{-1/\gamma_1}\right| dx.$$

Let us write

$$E_t^{(2,2)}(\beta) = o(A_{\mathbf{F}}(t))\int_1^\infty x^{-\beta-1}\left|\frac{\frac{\overline{\mathbf{F}}(tx)}{\overline{\mathbf{F}}(t)} - x^{-1/\gamma_1}}{A_{\mathbf{F}}(t)}\right| dx.$$

Subtracting $x^{-1/\gamma_1}\frac{x^{\tau_1/\gamma_1}-1}{\tau_1\gamma_1}$, inside the sign of the previous absolute value, and adding the same quantity, then applying inequality (3.10), we may readily show that $E_t^{(2,2)}(\beta) = o(A_{\mathbf{F}}(t))$, as $t \to \infty$, that we omit further details. Thereby $E_t^{(2,1)}(\beta) = (1+o(1))\eta_3$ leading to the result of assertion (ii). To show the second assertion (ii), is suffices to use Taylor's expansion to function f and similar arguments as used to first assertion (i), that we omit details. Let us now focus on the third one (iii). Multiplying, the equation of assertion (i) by $f(\gamma_1)$, we write

$$f(\gamma_1)E_t(\beta) = f(\gamma_1)\eta_2 + f(\gamma_1)\eta_3 A_{\mathbf{F}}(t)(1+o(1)), \tag{3.11}$$

and from assertion (ii) we have $f(\gamma_1) = f(L_t) - \eta_1 f'(\gamma_1) A_{\mathbf{F}}(t)(1+o(1))$. Substituting the previous expression into the first term of the right-side of equation (3.11), yields

$$f(\gamma_1) E_t(\beta) = \left(f(L_t) - \eta_1 f'(\gamma_1) A_{\mathbf{F}}(t)\right)\eta_2(1+o(1)) + f(\gamma_1)\eta_3 A_{\mathbf{F}}(t)(1+o(1)),$$

which gives

$$A_{\mathbf{F}}(t) = \frac{E_t(\beta) - \frac{f(L_t)}{f(\gamma_1)}\eta_2}{\eta_3 - \frac{f'(\gamma_1)}{f(\gamma_1)}\eta_2\eta_1}(1+o(1)).$$

In particular, for $f = f_*$, we have $f_*(\gamma_1) = \eta_2$, it follows that

$$A_{\mathbf{F}}(t) = \frac{E_t(\beta) - f_*(L_t)}{\eta_3 - f_*'(\gamma_1)\eta_1}(1+o(1)),$$

which gives the third assertion (iii). ■

3.2.2 Proof of the Theorem

To simplify the notation, let us set $\widehat{\gamma}_1^* := \widehat{\gamma}_{1,K}^*(\beta)$, $\widehat{\gamma}_1 := \widehat{\gamma}_{1,K}$, $\widehat{E}_k := \widehat{E}_{k,K}(\beta)$ and $\widehat{\rho}_k := \left(\widehat{\eta}_{3,K}/\widehat{\eta}_{1,K} - \widehat{\eta}_{4,K}\right)^{-1}$. Fix $\beta > 0$ and observe that from (3.7) we may write $\widehat{\gamma}_1^* - \gamma_1 = (\widehat{\gamma}_1 - \gamma_1) - \widehat{\rho}_k(\widehat{E}_k - \widehat{\eta}_2)$, which may be decomposed into the sum of $S_{1,k} := \widehat{\gamma}_1 - \gamma_1$, $S_{2,k} := -\widehat{\rho}_k(\widehat{E}_k - \eta_2)$ and $S_{3,k} := \widehat{\rho}_k(\widehat{\eta}_2 - \eta_2)$. We first consider the third term $S_{3,k}$. It is obvious that $S_{3,k} = \widehat{\rho}_k(f_*(\widehat{\gamma}_1) - f_*(\gamma_1))$, where f_* is as in (3.4). Applying Taylor's expansion, yields

$$S_{3,k} = \widehat{\rho}_k(\widehat{\gamma}_1 - \gamma_1) f_*'(\gamma_1) + \frac{1}{2}\widehat{\rho}_k(\widehat{\gamma}_1 - \gamma_1)^2 f_*''(\overline{\gamma}_1), \tag{3.12}$$

where $f_*''(x) := -\beta^2 \int_0^1 s^{\beta x - 1}(\log s)(\beta x \log s + 2) K^*(s) ds$ and $\overline{\gamma}_1$ is between $\widehat{\gamma}_1$ and γ_1. From asymptotic approximation (1.24), we deduce that $\sqrt{k}(\widehat{\gamma}_1 - \gamma_1) = O_{\mathbf{P}}(1)$ and therefore $\widehat{\gamma}_1 \xrightarrow{\mathbf{P}} \gamma_1$, it follows that $\overline{\gamma}_1 \xrightarrow{\mathbf{P}} \gamma_1$ too. On the other hand, since K is bounded on the real line then there is $M > 0$, such that $K^*(s) < Ms$, thus

$$\left|f_*''(x)\right| \leq M\beta^2 \int_0^1 s^{x\beta}(|\log s|)(\beta x |\log s| + 2) ds,$$

which equals $M\beta^3(4x\beta + 2)(x\beta + 1)^{-3}$. It is easy to check that for all $0 \leq x \leq 1$, the previous expression is less than $M\beta^3(4\beta + 2)$, which implies that $f_*''(\overline{\gamma}_1) = O_{\mathbf{P}}(1)$.

We noticed that $\widehat{\tau}_1$ is a consistent estimator for τ_1, then thanks to the consistency of $\widehat{\gamma}_1$, we have $\widehat{\rho}_k \xrightarrow{\mathbf{P}} \rho = \left(\eta_3/\eta_1 - \eta_4\right)^{-1}$. All the aforementioned arguments imply that the second term in (3.12), times $\sqrt{k}$, tends to zero in probability as $n \to \infty$, it follows that $\sqrt{k} S_{3,k} = \rho\eta_4\sqrt{k}\left(\widehat{\gamma}_1 - \gamma_1\right) + o_{\mathbf{P}}(1)$, where $\eta_4 = f_*'\left(\gamma_1\right)$. Thus

$$\sqrt{k}\left(S_{1,k} + S_{3,k}\right) = \left(1 + \rho\eta_4\right)\sqrt{k}\left(\widehat{\gamma}_1 - \gamma_1\right) + o_{\mathbf{P}}(1). \tag{3.13}$$

Let us now focus on the second term $S_{2,k}$. We have $\widehat{\rho}_k = \rho + o_{\mathbf{P}}(1)$, then

$$S_{2,k} = -\rho\left(\widehat{E}_k - \eta_2\right) + o_{\mathbf{P}}\left(\widehat{E}_k - \eta_2\right). \tag{3.14}$$

Observe that formula (3.5) may be rewritten into

$$\widehat{E}_k = 1 - \int_1^\infty K^*\left(\frac{\overline{\mathbf{F}}_n\left(xX_{n-k:n}\right)}{\overline{\mathbf{F}}_n\left(X_{n-k:n}\right)}\right)x^{-\beta-1}dx,$$

thereby the first term in (3.14) equals

$$\rho\int_1^\infty x^{-\beta-1}\left\{K^*\left(\frac{\overline{\mathbf{F}}_n\left(xX_{n-k:n}\right)}{\overline{\mathbf{F}}_n\left(X_{n-k:n}\right)}\right) - K^*\left(x^{-1/\gamma_1}\right)\right\}dx.$$

By applying Taylor's expansion to function K^*, we decompose $\sqrt{k}S_{2,k}$ into the sum of

$$\sqrt{k}S_{2,k}^{(1)} := \rho\int_1^\infty x^{-\beta-1}\mathbf{D}_n(x)K\left(x^{-1/\gamma_1}\right)dx,$$

$$\sqrt{k}S_{2,k}^{(2)} := k^{-1/2}\frac{\rho}{2}\int_1^\infty x^{-\beta-1}\mathbf{D}_n^2(x)K'\left(\ell_n(x)\right)dx,$$

and $\sqrt{k}S_{2,k}^{(3)} := \sqrt{k}\,o_{\mathbf{P}}\left(\widehat{E}_k - \eta_2\right)$, where $\ell_n(x)$ is a random sequence between x^{-1/γ_1} and $\overline{\mathbf{F}}_n\left(xX_{n-k:n}\right)/\overline{\mathbf{F}}_n\left(X_{n-k:n}\right)$, and

$$\mathbf{D}_n(x) := \sqrt{k}\left(\frac{\overline{\mathbf{F}}_n\left(xX_{n-k:n}\right)}{\overline{\mathbf{F}}_n\left(X_{n-k:n}\right)} - x^{-1/\gamma_1}\right),\ x \geq 1,$$

is the tail empirical process for randomly truncated data introduced by [6]. The authors showed that, there exists a standard Wiener process $\{W(x); x \geq 0\}$, defined on the probability space $(\Omega, \mathscr{A}, \mathbf{P})$, such that, for a sufficiently small $\xi > 0$ and $x_0 > 0$, one has

$$\sup_{x \geq x_0} x^{(1/2-\xi)/\gamma}\left|\mathbf{D}_n(x) - \Gamma(x;W) - x^{-1/\gamma_1}\frac{x^{\tau_1/\gamma_1} - 1}{\tau_1\gamma_1}A_{\mathbf{F}}(n/k)\right| \xrightarrow{\mathbf{P}} 0,\ \text{as } n \to \infty, \tag{3.15}$$

where

$$x^{1/\gamma_1}\Gamma(x;W) := \frac{\gamma}{\gamma_1}B(x;1) + \frac{\gamma}{\gamma_1+\gamma_2}\int_0^1 s^{-\gamma/\gamma_2-1}B(s;x)\,ds,$$

is a Gaussian process, with $B(s;x) := x^{1/\gamma}W\left(x^{-1/\gamma}s\right) - W(1)$. Applying weak approximation (4.23), yields

$$\sqrt{k}S_{2,k}^{(1)} = \sqrt{k}S_{2,k}^{(1,1)} + \rho\eta_3\sqrt{k}A_{\mathbf{F}}(n/k) + o_{\mathbf{P}}(v(x)),$$

where

$$\sqrt{k}S_{2,k}^{(1,1)} := \rho\int_1^\infty x^{-\beta-1}\Gamma(x;W)K\left(x^{-1/\gamma_1}\right)dx,$$

and $v(x) := \beta\int_1^\infty x^{-(1/2-\xi)/\gamma-\beta-1}K\left(x^{-1/\gamma_1}\right)dx$. Recall that K is bounded and for any sufficiently small $\xi > 0$, $\int_1^\infty x^{-(1/2-\xi)/\gamma-\beta-1}dx = \frac{2\gamma}{2\beta\gamma-2\xi+1}$, then $v(x)$ is finite, hence $o_{\mathbf{P}}(v(x)) = o_{\mathbf{P}}(1)$. Thereby

$$\sqrt{k}S_{2,k} = \rho\int_1^\infty x^{-\beta-1}\Gamma(x;W)K\left(x^{-1/\gamma_1}\right)dx + \rho\eta_3\sqrt{k}A_{\mathbf{F}}(n/k) + o_{\mathbf{P}}(1).$$

On the other hand, [6] showed that

$$\sqrt{k}\left(\widehat{\gamma}_1 - \gamma_1\right) = \int_1^\infty x^{-1}\Gamma(x;W)K\left(x^{-1/\gamma_1}\right)dx + \eta_1\sqrt{k}A_{\mathbf{F}}(n/k) + o_{\mathbf{P}}(1),$$

it follows from (3.13), that $\sqrt{k}\left(S_{1,k} + S_{3,k}\right)$ equals

$$\left(1+\rho\eta_4\right)\int_1^\infty x^{-1}\Gamma(x;W)K\left(x^{-1/\gamma_1}\right)dx + \left(1+\rho\eta_4\right)\eta_1\sqrt{k}A_{\mathbf{F}}(n/k) + o_{\mathbf{P}}(1).$$

To summarize, we showed that $\sqrt{k}\left(\widehat{\gamma}_1^* - \gamma_1\right)$ is asymptotically approximated by

$$\int_1^\infty x^{-1}\Gamma(x;W)u\left(x^{-1/\gamma_1}\right)dx + \left(\left(1+\rho\eta_4\right)\eta_1 + \rho\eta_3\right)\sqrt{k}A_{\mathbf{F}}(n/k) + o_{\mathbf{P}}(1),$$

where $u\left(x^{-1/\gamma_1}\right) := \left(1+\rho\eta_4+\rho x^{-\beta}\right)K\left(x^{-1/\gamma_1}\right)$. Substituting ρ by its expression $\left(\eta_3/\eta_1 - \eta_4\right)^{-1}$, we get $\left(1+\rho\eta_4\right)\eta_1 + \rho\eta_3 = 0$, therefore

$$\sqrt{k}\left(\widehat{\gamma}_1^* - \gamma_1\right) = \int_1^\infty x^{-1}\Gamma(x;W)u\left(x^{-1/\gamma_1}\right)dx + o_{\mathbf{P}}(1), \text{ as } n \to \infty.$$

An elementary calculation yields

$$\int_1^\infty x^{-1}\Gamma(x;W)u\left(x^{-1/\gamma_1}\right)dx = \left(\gamma^2/\gamma_1\right)\int_0^1 s^{-1}W(s)\,d\left\{sq_\beta(s)\right\} =: Z,$$

where

$$q_{\beta}(s) := s^{-1} \int_0^s t^{-\gamma/\gamma_2} \left\{ K_{\beta}\left(t^{\gamma/\gamma_1}\right) - \frac{\gamma_1}{\gamma_2} t^{-\gamma_2/\gamma_1} K_{\beta}\left(t^{\gamma/\gamma_1}\right) + t^{\gamma/\gamma_1} K'_{\beta}\left(t^{\gamma/\gamma_1}\right) \right\} dt,$$

where K_β is as in Theorem 3.1. Using the change of variables $u = t^{\gamma/\gamma_1}$, the previous expression meets that is (3.9). Making use of Lemma 8 in [13], we show that the variance of the centred Gaussian rv Z equals $\left(\gamma^2/\gamma_1\right)^2 \int_0^1 q_\beta^2(s)\,ds$. Next we show that $\sqrt{k}S_{2,k}^{(2)} = o_{\mathbf{P}}(1)$. Indeed, in view of assumption $[A4]$, the function g' is bounded, then there exists a constant $C > 0$, such that

$$\sqrt{k}S_{2,k}^{(2)} \leq Ck^{-1/2} \int_1^{\infty} x^{-\beta-1} \mathbf{D}_n^2(x)\,dx.$$

Note that $\mathbf{E}|W(s)| \leq s^{1/2}$, then it is easy to show that $\sup_{x\geq 1}|\Gamma(x;W)| = O_{\mathbf{P}}(1)$. Thanks to weak approximation (4.23), we infer that $\sup_{x\geq 1}|\mathbf{D}_n(x)| = O_{\mathbf{P}}(1)$ too, which implies that $\int_1^\infty x^{-\beta-1}\mathbf{D}_n^2(x)\,dx = O_{\mathbf{P}}(1)$ as well, therefore $\sqrt{k}S_{2,k}^{(2)} = O_{\mathbf{P}}\left(k^{-1/2}\right)$ which converges in probability to zero. Since $\sqrt{k}\left(\widehat{E}_k - \eta_2\right)$ is an asymptotically normal rv, hence it is bounded in probability, therefore $\sqrt{k}S_{2,k}^{(3)} = o_{\mathbf{P}}(1)$. The proof of Theorem 3.1 is now completed.

3.3 Simulation study

3.3.1 Graphical diagnostics

We will study the performance of the kernel bias-reduced estimator $\widehat{\gamma}_{1,K}^* := \widehat{\gamma}_{1,K}^*(\beta)$, given in (3.7), and compare it with $\overline{\gamma}_{1,K}^* := \overline{\gamma}_{1,K}^*(\beta)$ and $\widetilde{\gamma}_{1,K}^* := \widetilde{\gamma}_{1,K}^*(\beta)$ as well as the non bias-reduced ones $\widehat{\gamma}_{1,K}$ and $\widehat{\gamma}_1$. To this end, let us consider sets of truncated and truncation data drawn from Burr and Fréchet models:

- Burr $(\delta, \theta, \lambda)$ distribution with right-tail function:

$$\overline{F}(x) = \left(\frac{\delta}{\delta + x^\theta}\right)^\lambda, \; x > 0, \; \delta > 0, \; \theta > 0, \; \lambda > 0,$$

  with the tail index $\gamma = 1/(\lambda\theta)$ and the second-order parameter $\tau = -1/\lambda$.

- Fréchet (ξ) distribution with right-tail function:

$$\overline{F}(x) = 1 - \exp\left(-x^{-1/\xi}\right),\ x > 0,\ \xi > 0,$$

with the tail index $\gamma = \xi$ and the second-order parameter $\tau = -1$.

We first choose three different values for the tuning parameter $\beta = 0.5,\ 1,\ 2,$ consider the triweight kernel function K_3 and its corresponding bias-weight

$$K_3^*(s) := \int_0^s K_3(t)\,dt = \frac{1}{16}s\left(5s^6 + 21s^4 - 35s^2 + 35\right)\mathbf{1}\{[0,1)\}. \tag{3.16}$$

For each distribution, we generate 2000 random samples of length $N = 500$ and plot the absolute biases and RMSE's (root of the mean squared error) of the above-mentioned five estimators (y-axis) against different values of the number k of upper quantiles used in the estimation (x-axis). We consider two cases: a Burr distribution truncated by another Burr distribution (Figures 3.1 − 3.6) and a Fréchet distribution truncated by another Fréchet distribution (Figures 3.7−3.12). . Both tail indices γ_1 and γ_2 have to be chosen so that $p > 1/2$ which corresponds to average to strong truncation in the tail. For this, we take $\gamma_1 = 0.6,\ 0.8$ and $p = 0.55,\ 0.7,\ 0.9$ leading to the triplets $\left(\gamma_1, \gamma_2, p\right) = (0.6, 0.73, 0.55), (0.6, 1.4, 0.7), (0.6, 5.4, 0.9), (0.8, 0.97, 0.55), (0.8, 1.86, 0.7)$ and $(0.8, 7.2, 0.9)$. Our oveall results are taken as the empirical averages over the 2000 replicates. The situation $p < 1/2$ (strong truncation) is not considered in our simulation study since the proposed bias-reduced estimators are defined thanks to the asymptotic distribution (1.24) which is given provided that $p > 1/2$. The figures 3.1 − 3.6, corresponding to the Burr-Burr scheme, show that the three bias-reduced estimators perform better in terms of bias with reasonable RMSE for a large interval of $k-$values which is helpful to choose an appropriate value of k. As expected, the "full" kernel estimator $\widehat{\gamma}_{1,K}$ works well compared to $\overline{\gamma}_{1,K}^*$ and $\widetilde{\gamma}_{1,K}^*$. Note that for weak truncations, that is for large p, the three mentioned estimators behave well both in terms of bias and RMSE. In this scenario, we also observe that the tuning parameter β does not have any influence on the quality of the estimation, at least for the three values 0.5, 1 and 2, thus the choice of this parameter may be arbitrary. As far as the Fréchet-Fréchet scheme is concerned, the quality of the new estimators is not as good. Indeed, while the biases are low, the RMSE's are high mainly with the medium truncation rate, that is when p is close to 1/2 (see the top panels of Figures 3.7 – 3.12). This was somehow expectable because an improvement of the bias component may be made at the expense of the variance leading to a larger RMSE. On the other hand, unlike the first scheme, a change in the value of the tuning parameter has an impact on the estimation quality. At this stage we cannot say more about the optimal choice of β but it seems that the value 1 gives satisfactory results for both moderate and light truncation cases.

3.3.2 A heuristic procedure to estimate the tail Index γ_1

The choice of the number of top order statistics $\widehat{k}$ used in the computation of the tail index estimate is crucial. Several heuristic methods to select this latter are available in the extreme value theory literature, see for instance [10]. Here, we consider Reiss-Thomas's algorithm given in [43], page 137, in which the optimal sample fraction is defined by

$$\widehat{k} := \arg \min_{1<k<n} \frac{1}{k} \sum_{i=1}^{k} i^{\epsilon} \left| \widehat{\gamma}(i) - \text{median}\left\{ \widehat{\gamma}(1), ..., \widehat{\gamma}(k) \right\} \right|, \tag{3.17}$$

for a suitable constant $0 \leq \epsilon \leq 1/2$, where $\widehat{\gamma}(i)$ is an estimator of tail index γ, based on the i extreme values corresponding to a Pareto-type model. From our simulation study, we observed that $\epsilon = 0.3$ provides better results both in terms of bias and RMSE. We pointed out that this agrees with that early found by [39] in complete data case. We will opt for this procedure to select $\widehat{k}$ in the computation of aforementioned five tail index estimators. Here, we consider two sample sizes $N = 500$ and $N = 150$, keep the kernel K_3 and, for brevity, choose $\beta = 1$. The performances of the estimators, in terms of the biases and RMSE's, are summarized in Tables 3.2-3.4. The second columns of the tables contain the size n of the really observed sample which, due to the truncation, is expectedly less than the original size N. Actually, in real applications, one only has in hand datasets that are already truncated (i.e. of size n). We note that, in all instances, the estimators $\widehat{\gamma}_1$ and $\widehat{\gamma}_{1,K}$ overestimate the tail index while the estimators $\widetilde{\gamma}_1^*$ and $\overline{\gamma}_{1,K}^*$ underestimate it. However, the newly proposed estimator $\widehat{\gamma}_{1,K}^*$ alternate the sign of its (smaller) bias meaning that it can be considered as a good compromise between the first two categories of estimators.

$\gamma_1 = 0.6$, $N = 500$

	n	$\widehat{k}$	$\widehat{\gamma}^*_{1,K}$	$\widehat{k}$	$\widehat{\gamma}_1$	$\widehat{k}$	$\widehat{\gamma}_{1,K}$	$\widehat{k}$	$\widetilde{\gamma}^*_1$	$\widehat{k}$	$\overline{\gamma}^*_{1,K}$
$p = 0.55$	257	102	0.59	90	0.72	91	0.66	98	0.48	96	0.54
$p = 0.70$	278	99	0.59	93	0.68	93	0.63	96	0.52	95	0.55
$p = 0.90$	304	101	0.63	99	0.66	98	0.64	99	0.58	97	0.60

$\gamma_1 = 0.6$, $N = 150$

	n	$\widehat{k}$	$\widehat{\gamma}^*_{1,K}$	$\widehat{k}$	$\widehat{\gamma}_1$	$\widehat{k}$	$\widehat{\gamma}_{1,K}$	$\widehat{k}$	$\widetilde{\gamma}^*_1$	$\widehat{k}$	$\overline{\gamma}^*_{1,K}$
$p = 0.55$	76	37	0.71	27	0.69	28	0.71	30	0.36	29	0.53
$p = 0.70$	83	35	0.69	30	0.68	30	0.67	31	0.46	30	0.56
$p = 0.90$	91	37	0.70	31	0.64	33	0.65	32	0.55	32	0.61

Table 3.2: Optimal sample fractions $\widehat{k}$ and estimate values, through $\widehat{\gamma}^*_{1,K}$, $\widehat{\gamma}_1$, $\widehat{\gamma}_{1,K}$, $\widetilde{\gamma}^*_1$ and $\bar{\gamma}^*_{1,K}$, of the tail index $\gamma_1 = 0.6$ based on 2000 samples from a Fréchet distribution truncated by another Fréchet distribution with: $N = \{500, 150\}$, $\beta = 1$ and three truncating proportions.

$\gamma_1 = 0.8$, $N = 500$

	n	$\widehat{k}$	$\widehat{\gamma}^*_{1,K}$	$\widehat{k}$	$\widehat{\gamma}_1$	$\widehat{k}$	$\widehat{\gamma}_{1,K}$	$\widehat{k}$	$\widetilde{\gamma}^*_1$	$\widehat{k}$	$\overline{\gamma}^*_{1,K}$
$p = 0.55$	257	99	0.82	91	0.94	87	0.88	96	0.63	94	0.71
$p = 0.70$	279	98	0.82	94	0.90	95	0.86	99	0.74	95	0.77
$p = 0.90$	303	103	0.82	97	0.88	98	0.85	100	0.76	99	0.78

$\gamma_1 = 0.8$, $N = 150$

	n	$\widehat{k}$	$\widehat{\gamma}^*_{1,K}$	$\widehat{k}$	$\widehat{\gamma}_1$	$\widehat{k}$	$\widehat{\gamma}_{1,K}$	$\widehat{k}$	$\widetilde{\gamma}^*_1$	$\widehat{k}$	$\overline{\gamma}^*_{1,K}$
$p = 0.55$	77	35	0.77	28	0.84	27	0.80	29	0.42	28	0.58
$p = 0.70$	82	34	0.92	28	0.89	30	0.90	31	0.66	28	0.78
$p = 0.90$	90	35	0.93	32	0.87	31	0.86	31	0.71	30	0.79

Table 3.3: Optimal sample fractions $\widehat{k}$ and estimate values, through $\widehat{\gamma}^*_{1,K}$, $\widehat{\gamma}_1$, $\widehat{\gamma}_{1,K}$, $\widetilde{\gamma}^*_1$ and $\bar{\gamma}^*_{1,K}$, of the tail index $\gamma_1 = 0.8$ based on 2000 samples from a Fréchet distribution truncated by another Fréchet distribution with: $N = \{500, 150\}$, $\beta = 1$ and three truncating proportions.

$\gamma_1 = 0.6$, $N = 500$

	n	$\widehat{k}$	$\widehat{\gamma}^*_{1,K}$	$\widehat{k}$	$\widehat{\gamma}_1$	$\widehat{k}$	$\widehat{\gamma}_{1,K}$	$\widehat{k}$	$\widetilde{\gamma}^*_1$	$\widehat{k}$	$\overline{\gamma}^*_{1,K}$
$p = 0.55$	275	98	0.47	83	1.07	84	0.86	94	0.22	91	0.29
$p = 0.70$	350	103	0.56	86	0.94	89	0.80	96	0.44	94	0.47
$p = 0.90$	450	101	0.63	93	0.81	94	0.74	100	0.55	99	0.57
$\gamma_1 = 0.6$, $N = 150$											
	n	$\widehat{k}$	$\widehat{\gamma}^*_{1,K}$	$\widehat{k}$	$\widehat{\gamma}_1$	$\widehat{k}$	$\widehat{\gamma}_{1,K}$	$\widehat{k}$	$\widetilde{\gamma}^*_1$	$\widehat{k}$	$\overline{\gamma}^*_{1,K}$
$p = 0.55$	81	39	0.43	23	0.68	26	0.66	28	0.56	27	0.57
$p = 0.70$	105	37	0.53	33	0.68	35	0.66	36	0.57	36	0.57
$p = 0.90$	135	38	0.57	34	0.65	36	0.65	36	0.58	35	0.58

Table 3.4: Optimal sample fractions $\widehat{k}$ and estimate values, through $\widehat{\gamma}^*_{1,K}$, $\widehat{\gamma}_1$,$\widehat{\gamma}_{1,K}$,$\widetilde{\gamma}^*_1$ and $\bar{\gamma}^*_{1,K}$, of the tail index $\gamma_1 = 0.6$ based on 2000 samples from a Burr distribution truncated by another Burr distribution with: $N = \{500, 150\}$, $\beta = 1$ and three truncating proportions.

$\gamma_1 = 0.8$, $N = 500$

	n	$\widehat{k}$	$\widehat{\gamma}^*_{1,K}$	$\widehat{k}$	$\widehat{\gamma}_1$	$\widehat{k}$	$\widehat{\gamma}_{1,K}$	$\widehat{k}$	$\widetilde{\gamma}^*_1$	$\widehat{k}$	$\overline{\gamma}^*_{1,K}$
$p = 0.55$	275	99	0.81	83	1.25	88	1.08	93	0.54	91	0.64
$p = 0.70$	349	101	0.84	88	1.09	93	0.98	98	0.72	98	0.75
$p = 0.90$	449	105	0.81	92	0.96	96	0.89	99	0.74	99	0.76
$\gamma_1 = 0.8$, $N = 150$											
	n	$\widehat{k}$	$\widehat{\gamma}^*_{1,K}$	$\widehat{k}$	$\widehat{\gamma}_1$	$\widehat{k}$	$\widehat{\gamma}_{1,K}$	$\widehat{k}$	$\widetilde{\gamma}^*_1$	$\widehat{k}$	$\overline{\gamma}^*_{1,K}$
$p = 0.55$	82	37	0.70	32	0.89	33	0.86	34	0.40	34	0.44
$p = 0.70$	106	38	0.72	35	0.89	35	0.85	36	0.70	35	0.66
$p = 0.90$	135	38	0.75	36	0.88	35	0.85	36	0.72	35	0.74

Table 3.5: Optimal sample fractions $\widehat{k}$ and estimate values, through $\widehat{\gamma}^*_{1,K}$, $\widehat{\gamma}_1$,$\widehat{\gamma}_{1,K}$,$\widetilde{\gamma}^*_1$ and $\bar{\gamma}^*_{1,K}$, of the tail index $\gamma_1 = 0.8$ based on 2000 samples from a Burr distribution truncated by another Burr distribution with: $N = \{500, 150\}$, $\beta = 1$ and three truncating proportions.

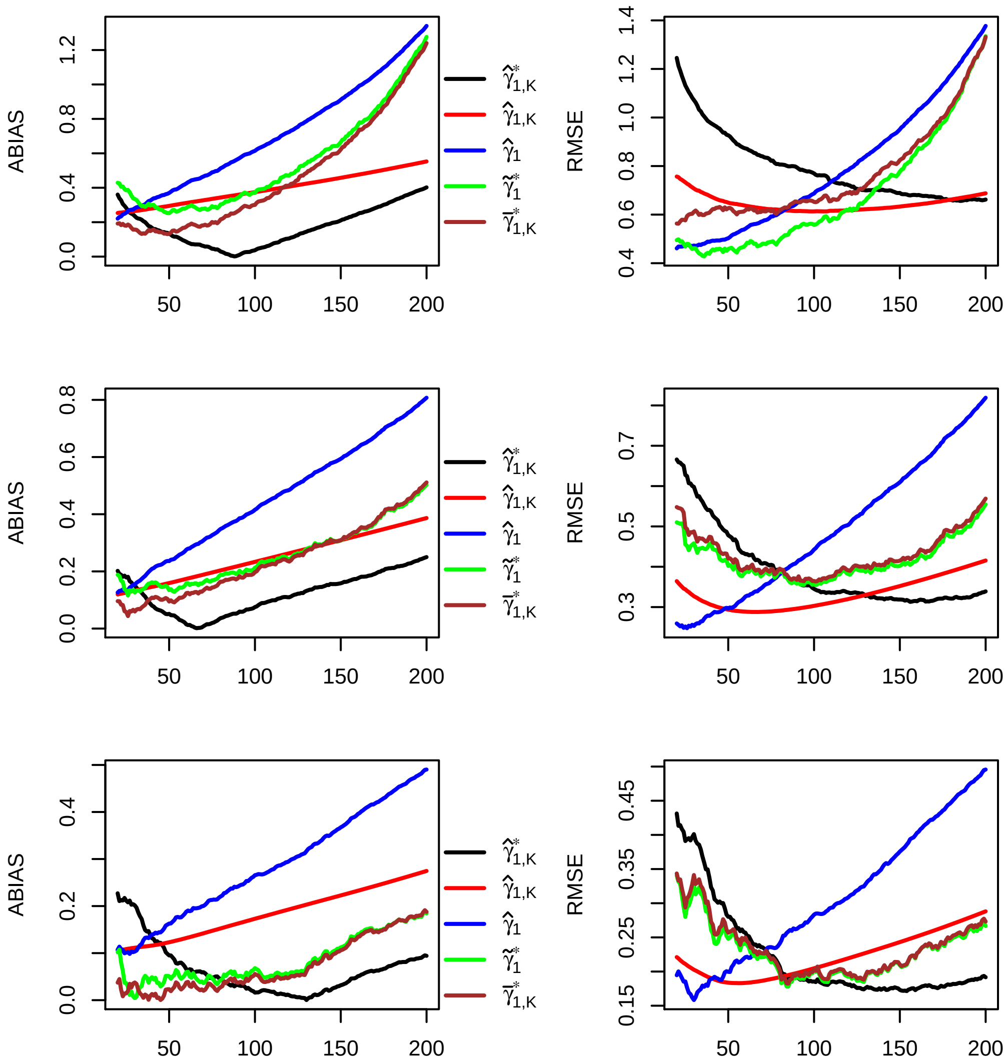

Figure 3.1: Absolute biases (left panel) and RMSE's (right panel) of $\widehat{\gamma}^*_{1,K}, \widehat{\gamma}_{1,K}, \widehat{\gamma}_1, \widetilde{\gamma}^*_1$ and $\overline{\gamma}^*_{1,K}$ for a Burr distribution truncated by another Burr distribution, with $\beta = 1$ and $\gamma_1 = 0.6$ under the following cases: $p = 0.55$ (top), $p = 0.7$ (middle) and $p = 0.9$ (bottom). The simulation is based on 2000 replicates of size 500.

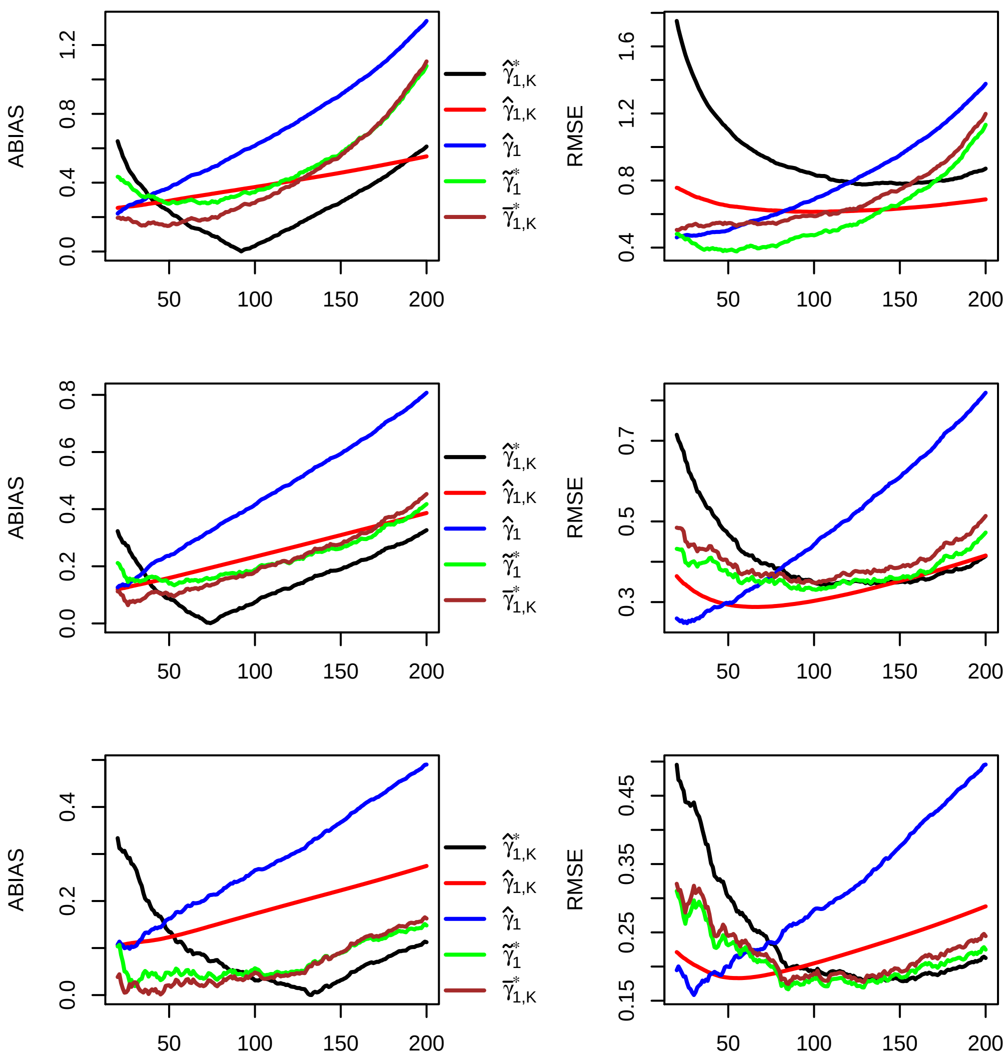

Figure 3.2: Absolute biases (left panel) and RMSE's (right panel) of $\widehat{\gamma}^*_{1,K}, \widehat{\gamma}_{1,K}, \widehat{\gamma}_1, \widetilde{\gamma}^*_1$ and $\overline{\gamma}^*_{1,K}$ for a Burr distribution truncated by another Burr distribution, with $\beta = 0.5$ and $\gamma_1 = 0.6$ under the following cases: $p = 0.55$ (top), $p = 0.7$ (middle) and $p = 0.9$ (bottom). The simulation is based on 2000 replicates of size 500.

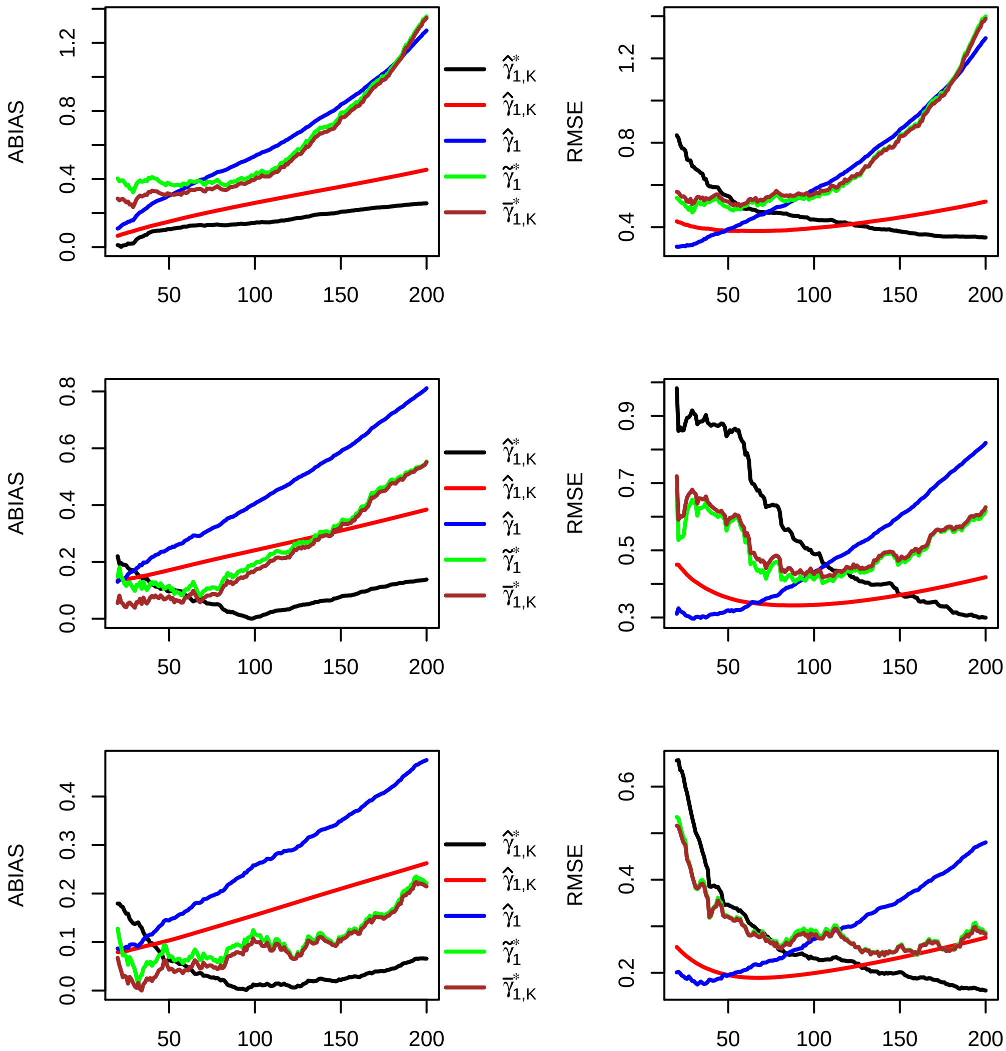

Figure 3.3: Absolute biases (left panel) and RMSE's (right panel) of $\widehat{\gamma}^*_{1,K}, \widehat{\gamma}_{1,K}, \widehat{\gamma}_1, \widetilde{\gamma}^*_1$ and $\overline{\gamma}^*_{1,K}$ for a Burr distribution truncated by another Burr distribution, with $\beta = 2$ and $\gamma_1 = 0.6$ under the following cases: $p = 0.55$ (top), $p = 0.7$ (middle) and $p = 0.9$ (bottom). The simulation is based on 2000 replicates of size 500.

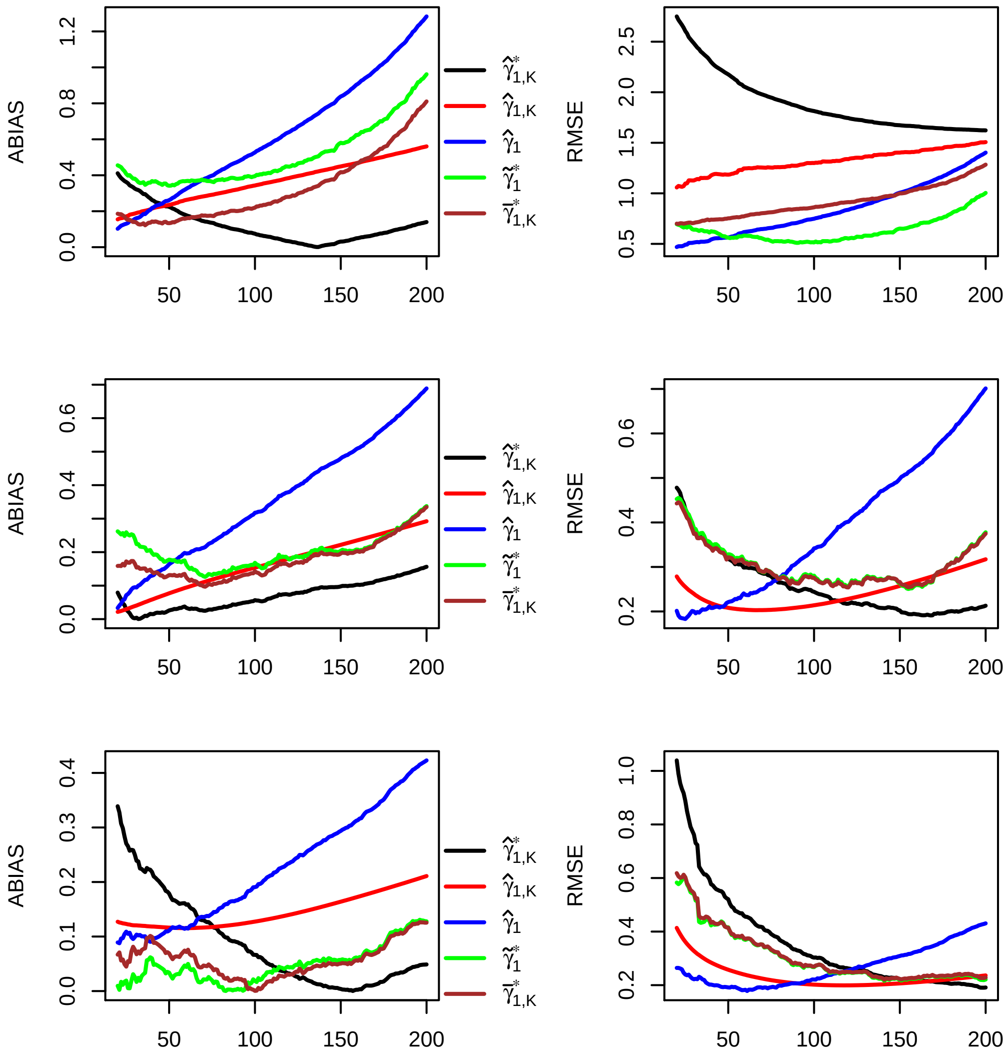

Figure 3.4: Absolute biases (left panel) and RMSE's (right panel) of $\widehat{\gamma}^{*}_{1,K}, \widehat{\gamma}_{1,K}, \widehat{\gamma}_{1}, \widetilde{\gamma}^{*}_{1}$ and $\overline{\gamma}^{*}_{1,K}$ for a Burr distribution truncated by another Burr distribution, with $\beta = 1$ and $\gamma_1 = 0.8$ under the following cases: $p = 0.55$ (top), $p = 0.7$ (middle) and $p = 0.9$ (bottom). The simulation is based on 2000 replicates of size 500.

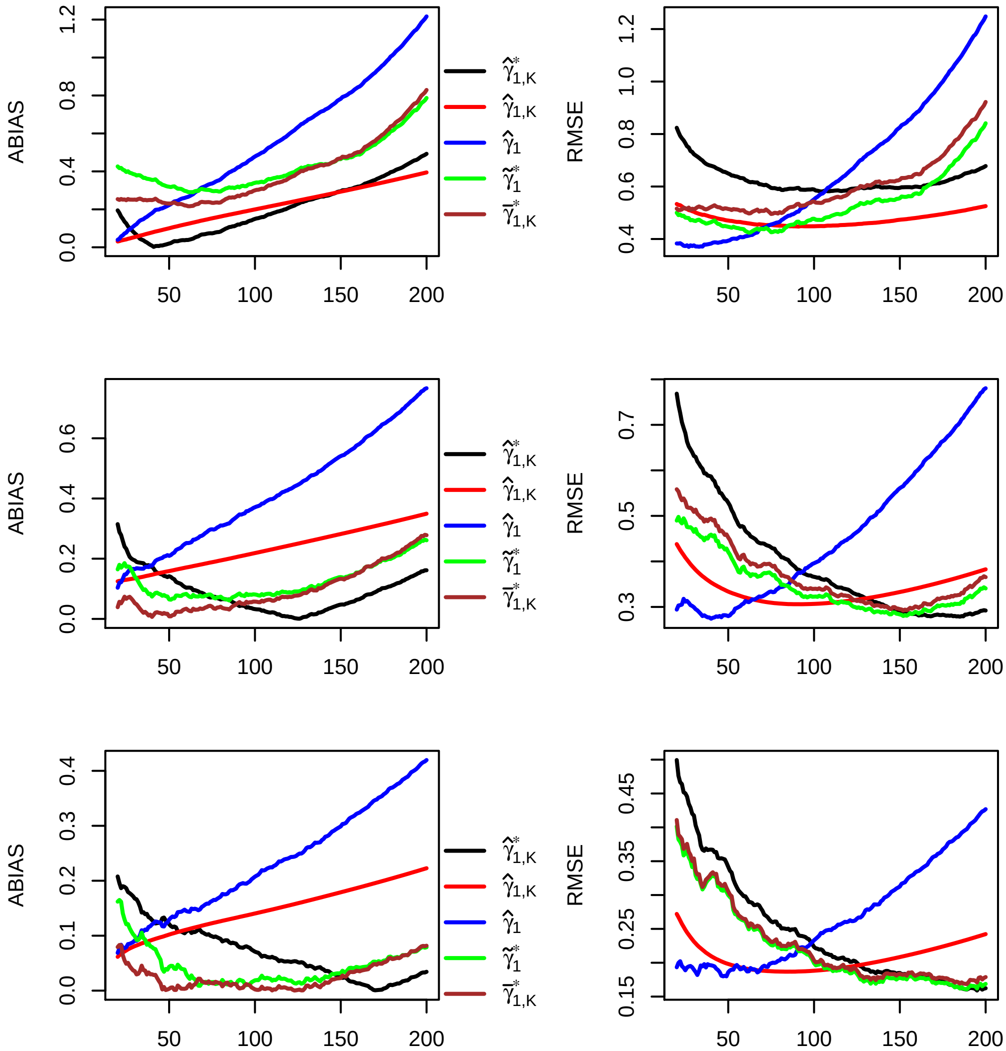

Figure 3.5: Absolute biases (left panel) and RMSE's (right panel) of $\widehat{\gamma}^*_{1,K}, \widehat{\gamma}_{1,K}, \widehat{\gamma}_1, \widetilde{\gamma}^*_1$ and $\overline{\gamma}^*_{1,K}$ for a Burr distribution truncated by another Burr distribution, with $\beta = 0.5$ and $\gamma_1 = 0.8$ under the following cases: $p = 0.55$ (top), $p = 0.7$ (middle) and $p = 0.9$ (bottom). The simulation is based on 2000 replicates of size 500.

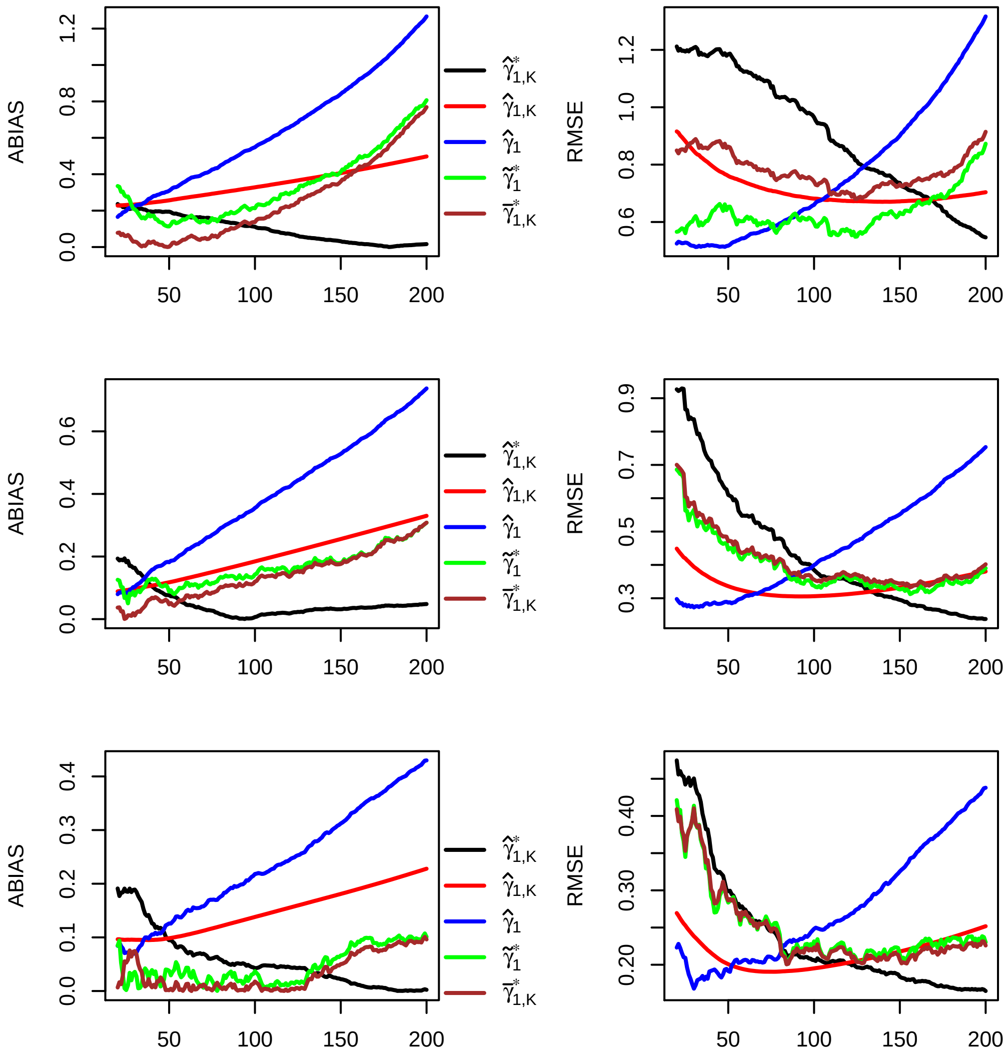

Figure 3.6: Absolute biases (left panel) and RMSE's (right panel) of $\widehat{\gamma}^*_{1,K}, \widehat{\gamma}_{1,K}, \widehat{\gamma}_1, \widetilde{\gamma}^*_1$ and $\overline{\gamma}^*_{1,K}$ for a Burr distribution truncated by another Burr distribution, with $\beta = 2$ and $\gamma_1 = 0.8$ under the following cases: $p = 0.55$ (top), $p = 0.7$ (middle) and $p = 0.9$ (bottom). The simulation is based on 2000 replicates of size 500.

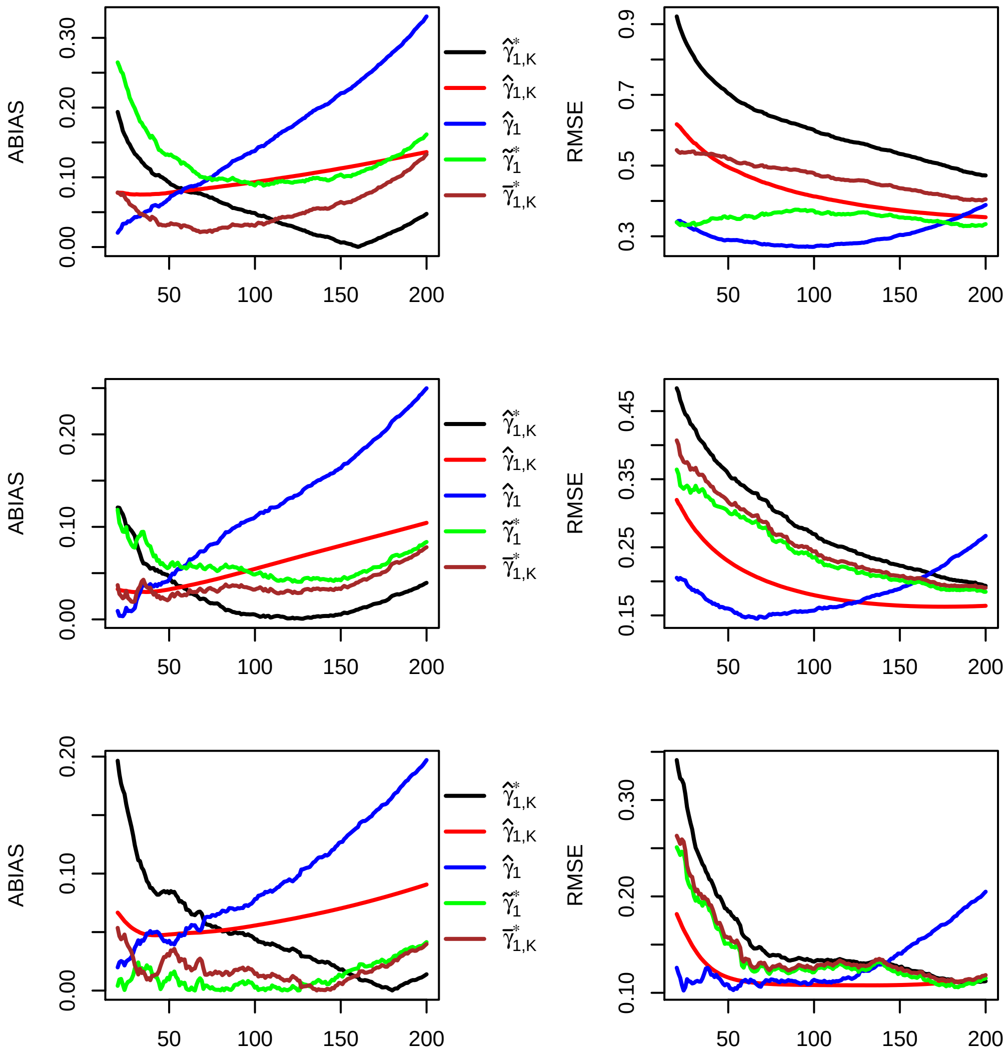

Figure 3.7: Absolute biases (left panel) and RMSE's (right panel) of $\widehat{\gamma}^*_{1,K}, \widehat{\gamma}_{1,K}, \widehat{\gamma}_1, \widetilde{\gamma}^*_1$ and $\overline{\gamma}^*_{1,K}$ for a Fréchet distribution truncated by another Fréchet distribution, with $\beta = 1$ and $\gamma_1 = 0.6$ under the following cases: $p = 0.55$ (top), $p = 0.7$ (middle) and $p = 0.9$ (bottom). The simulation is based on 2000 replicates of size 500.

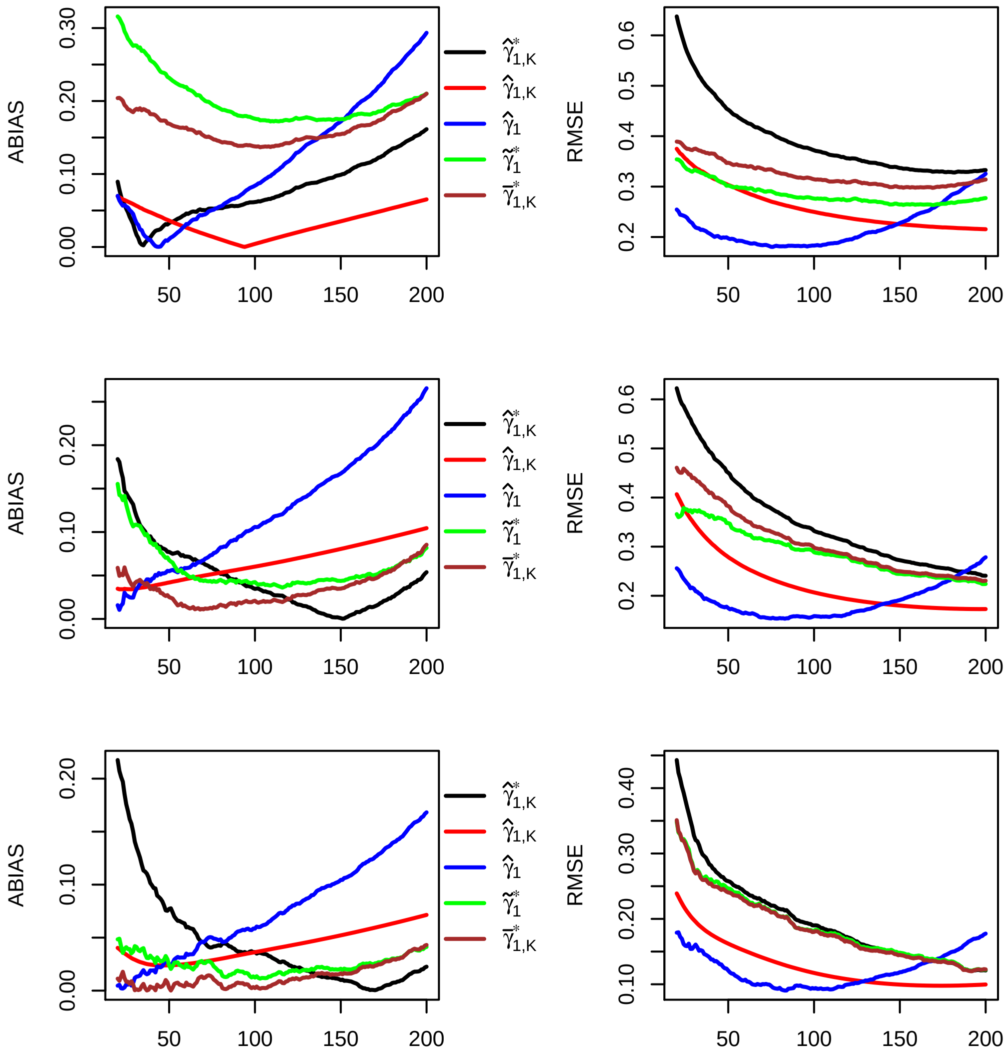

Figure 3.8: Absolute biases (left panel) and RMSE's (right panel) of $\widehat{\gamma}^*_{1,K}, \widehat{\gamma}_{1,K}, \widehat{\gamma}_1, \widetilde{\gamma}^*_1$ and $\overline{\gamma}^*_{1,K}$ for a Fréchet distribution truncated by another Fréchet distribution, with $\beta = 0.5$ and $\gamma_1 = 0.6$ under the following cases: $p = 0.55$ (top), $p = 0.7$ (middle) and $p = 0.9$ (bottom). The simulation is based on 2000 replicates of size 500.

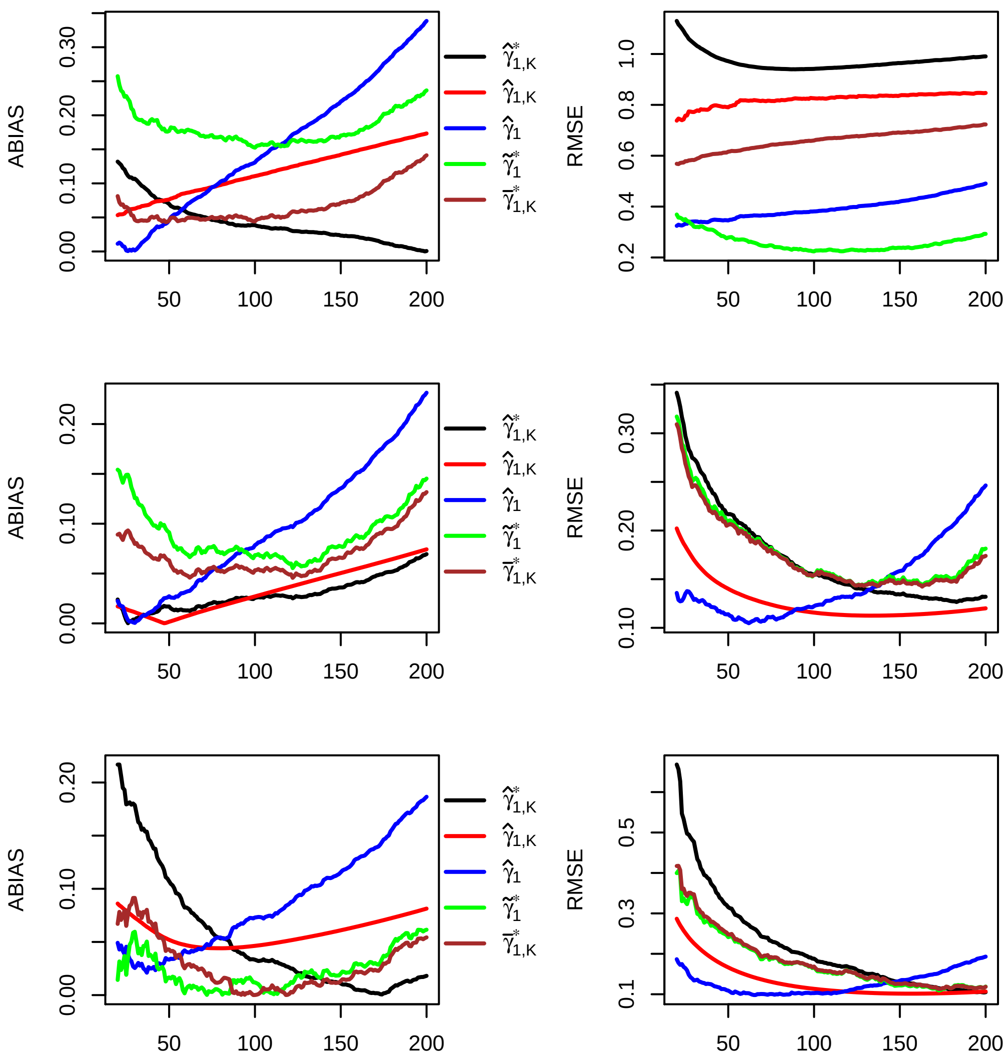

Figure 3.9: Absolute biases (left panel) and RMSE's (right panel) of $\widehat{\gamma}^*_{1,K}, \widehat{\gamma}_{1,K}, \widehat{\gamma}_1, \widetilde{\gamma}^*_1$ and $\overline{\gamma}^*_{1,K}$ for a Fréchet distribution truncated by another Fréchet distribution, with $\beta = 2$ and $\gamma_1 = 0.6$ under the following cases: $p = 0.55$ (top), $p = 0.7$ (middle) and $p = 0.9$ (bottom). The simulation is based on 2000 replicates of size 500.

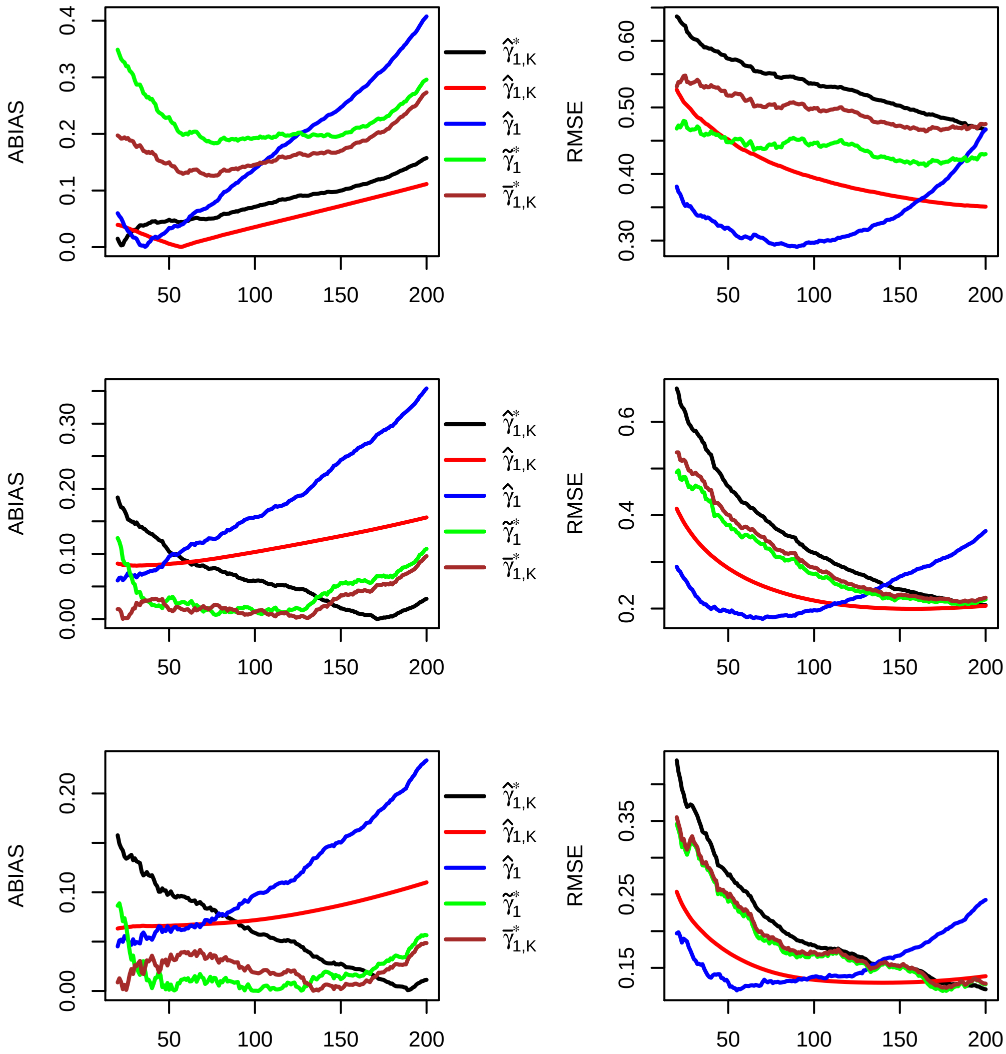

Figure 3.10: Absolute biases (left panel) and RMSE's (right panel) of $\widehat{\gamma}^*_{1,K}, \widehat{\gamma}_{1,K}, \widehat{\gamma}_1, \widetilde{\gamma}^*_1$ and $\overline{\gamma}^*_{1,K}$ for a Fréchet distribution truncated by another Fréchet distribution, with $\beta = 1$ and $\gamma_1 = 0.8$ under the following cases: $p = 0.55$ (top), $p = 0.7$ (middle) and $p = 0.9$ (bottom). The simulation is based on 2000 replicates of size 500.

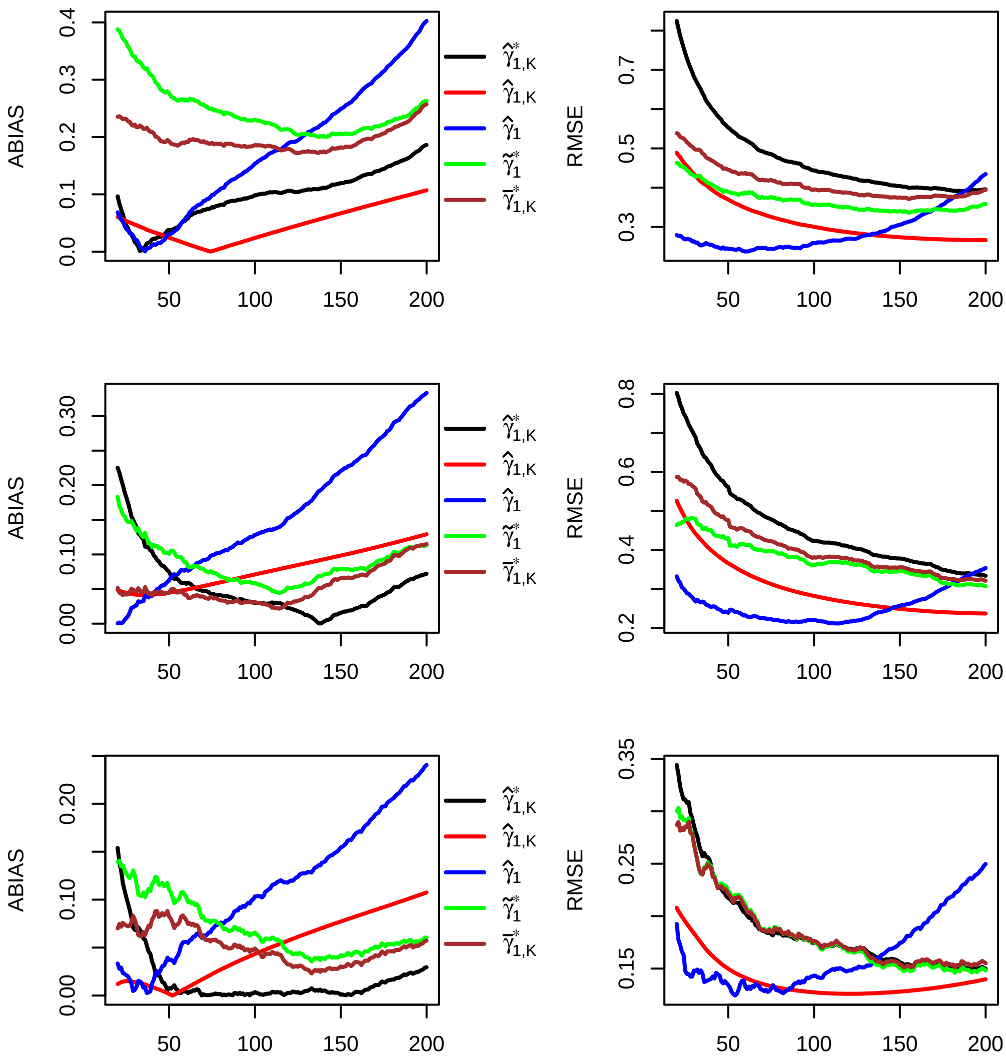

Figure 3.11: Absolute biases (left panel) and RMSE's (right panel) of $\widehat{\gamma}^*_{1,K}, \widehat{\gamma}_{1,K}, \widehat{\gamma}_1, \widetilde{\gamma}^*_1$ and $\overline{\gamma}^*_{1,K}$ for a Fréchet distribution truncated by another Fréchet distribution, with $\beta = 0.5$ and $\gamma_1 = 0.8$ under the following cases: $p = 0.55$ (top), $p = 0.7$ (middle) and $p = 0.9$ (bottom). The simulation is based on 2000 replicates of size 500.

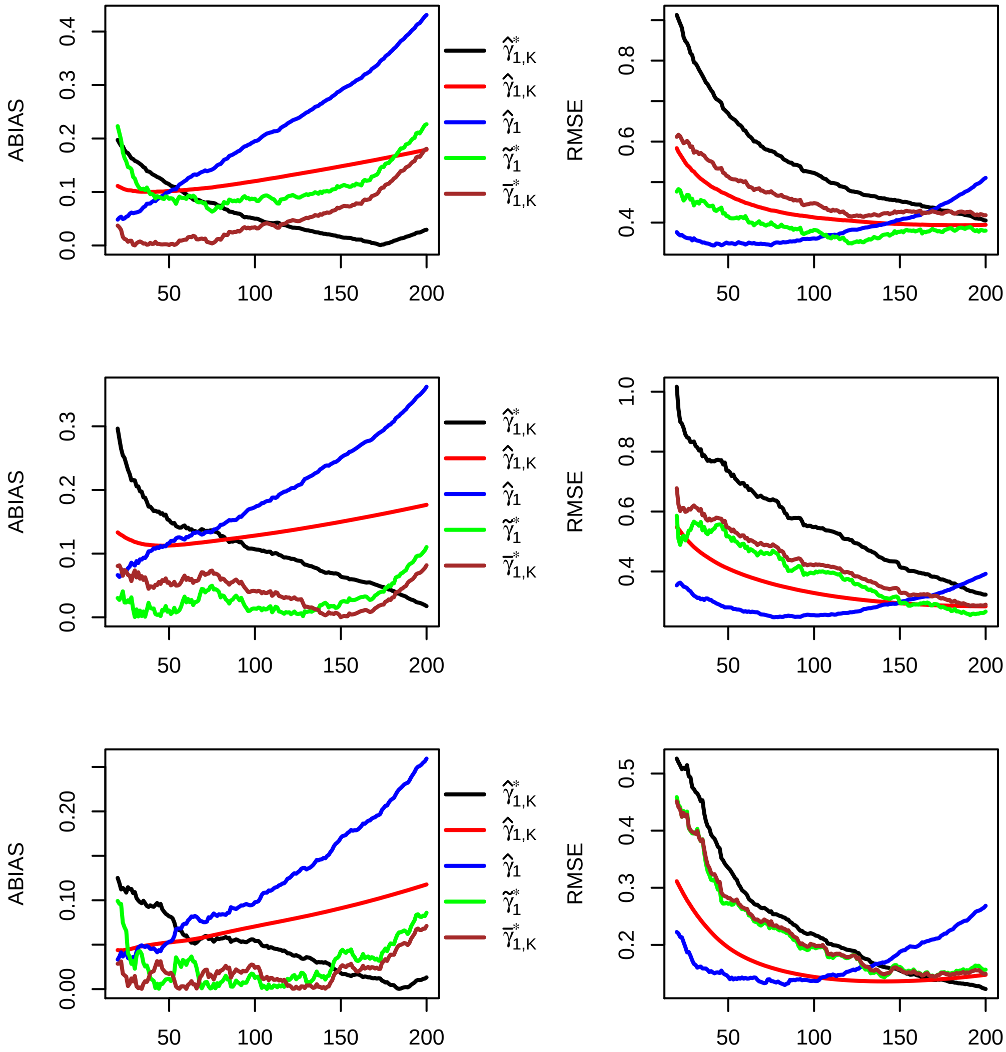

Figure 3.12: Absolute biases (left panel) and RMSE's (right panel) of $\widehat{\gamma}^*_{1,K}, \widehat{\gamma}_{1,K}, \widehat{\gamma}_1, \widetilde{\gamma}^*_1$ and $\overline{\gamma}^*_{1,K}$ for a Fréchet distribution truncated by a nother Fréchet distribution, with $\beta = 2$ and $\gamma_1 = 0.8$ under the following cases: $p = 0.55$ (top), $p = 0.7$ (middle) and $p = 0.9$ (bottom). The simulation is based on 2000 replicates of size 500.

For further illustration of the dispersion of the five estimators, we provide, in Figures 3.13-3.16, box-plot representations which show two things. The first one is that the three estimators $\widehat{\gamma}_{1,K}^{*}$, $\overline{\gamma}_{1,K}^{*}$, $\widetilde{\gamma}_{1}^{*}$ have, as expected and mentioned above, slight dispersions with respect to the remaining two. The second, is that all five estimators globally exhibit symmetric distributions.

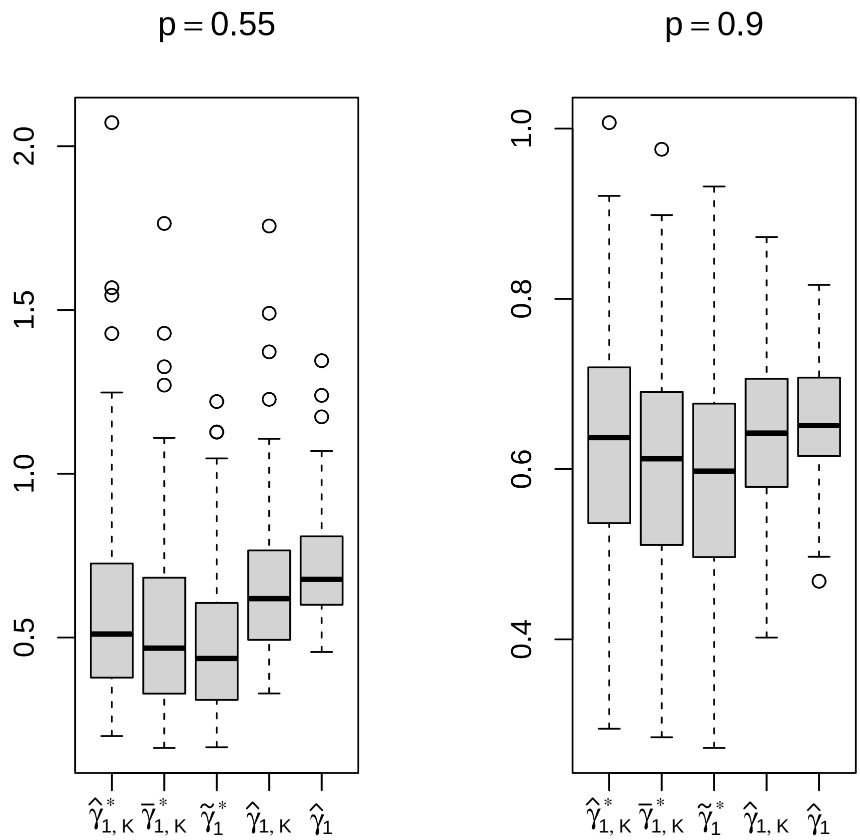

Figure 3.13: Box-plots corresponding to estimators $\widehat{\gamma}_{1,K}^{*}$, $\overline{\gamma}_{1,K}^{*}$, $\widetilde{\gamma}_{1}^{*}$, $\widehat{\gamma}_{1,K}$ and $\widehat{\gamma}_{1}$ for a Fréchet distribution truncated by another Fréchet distribution, with $\beta = 1$, $\gamma_1 = 0.6$, $p = 0.55$ and $p = 0.9$ based on 2000 replicates of size $N = 500$.

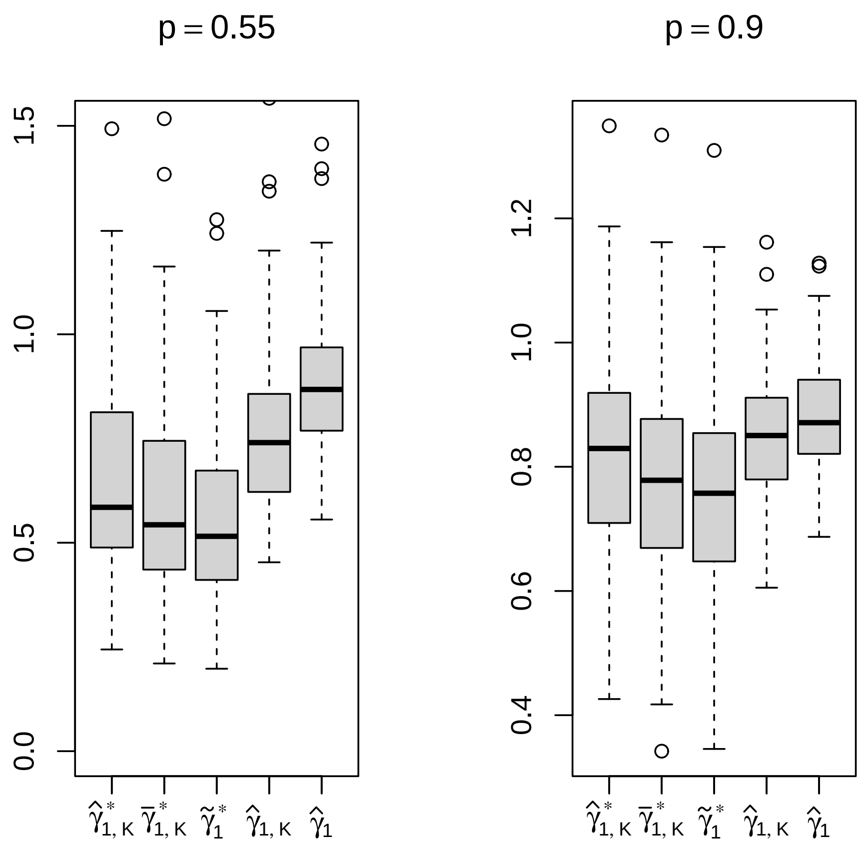

Figure 3.14: Box-plots corresponding to estimators $\widehat{\gamma}^*_{1,K}$, $\overline{\gamma}^*_{1,K}$, $\widetilde{\gamma}^*_1$, $\widehat{\gamma}_{1,K}$ and $\widehat{\gamma}_1$ for a Fréchet distribution truncated by another Fréchet distribution, with $\beta = 1$, $\gamma_1 = 0.8$, $p = 0.55$ and $p = 0.9$ based on 2000 replicates of size $N = 500$.

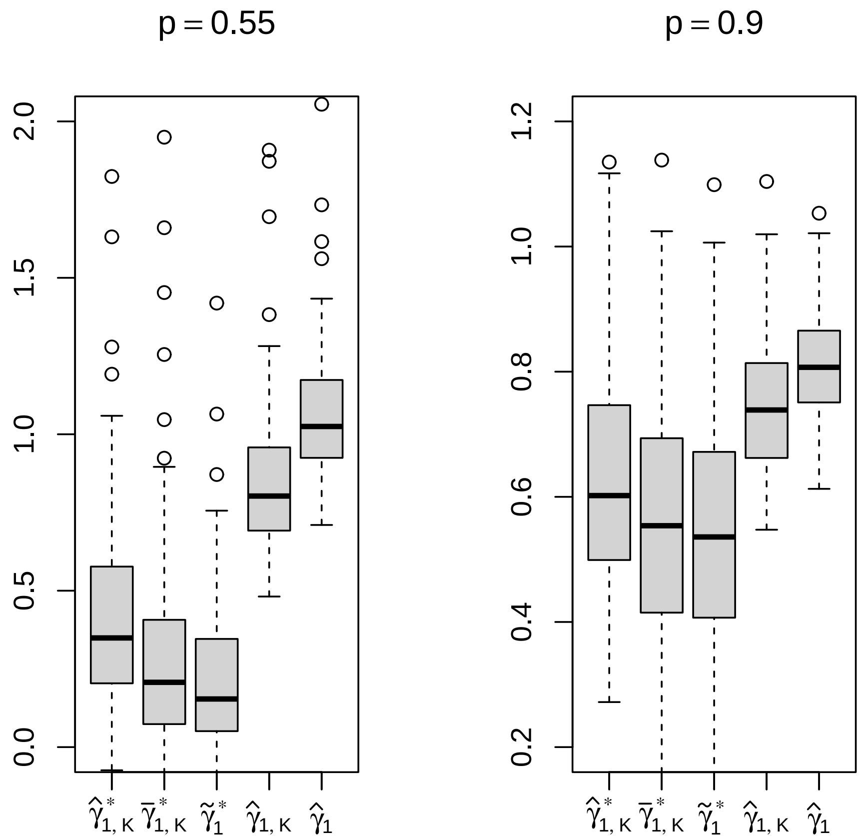

Figure 3.15: Box-plots corresponding to estimators $\widehat{\gamma}^*_{1,K}$, $\overline{\gamma}^*_{1,K}$, $\widetilde{\gamma}^*_1$, $\widehat{\gamma}_{1,K}$ and $\widehat{\gamma}_1$ for a Burr distribution truncated by another Burr distribution, with $\beta = 1$, $\gamma_1 = 0.6$, $p = 0.55$ and $p = 0.9$ based on 2000 replicates of size $N = 500$.

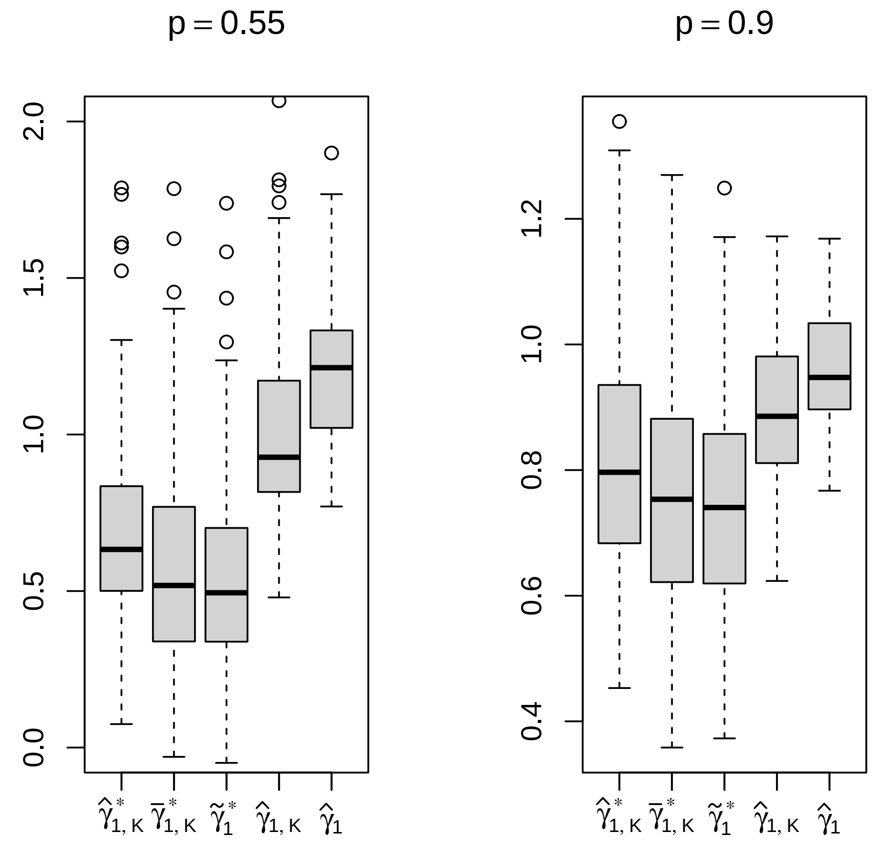

Figure 3.16: Box-plots corresponding to estimators $\widehat{\gamma}^*_{1,K}$, $\overline{\gamma}^*_{1,K}$, $\widetilde{\gamma}^*_1$, $\widehat{\gamma}_{1,K}$ and $\widehat{\gamma}_1$ for a Burr distribution truncated by another Burr distribution, with $\beta = 1$, $\gamma_1 = 0.8$, $p = 0.55$ and $p = 0.9$ based on 2000 replicates of size $N = 500$.

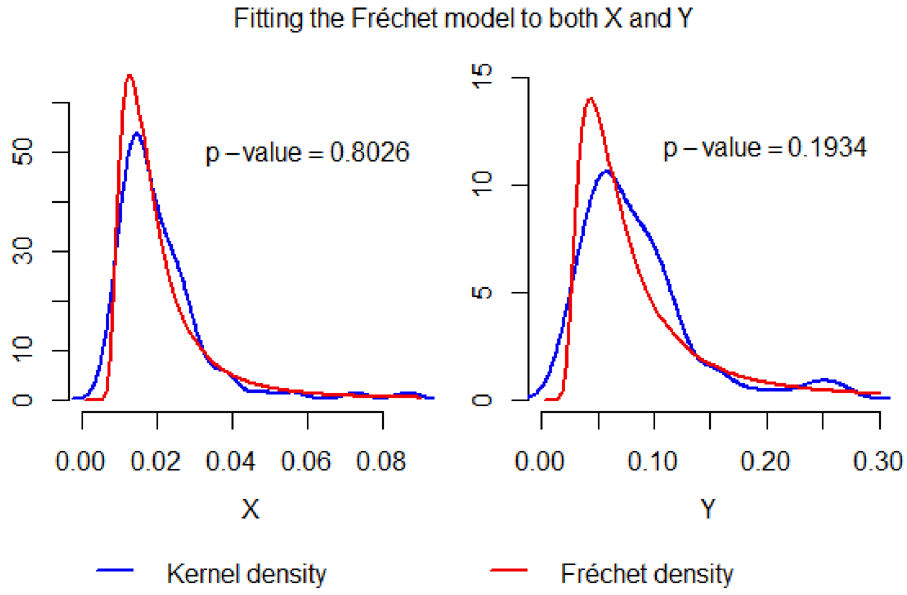

Figure 3.17: Fitting the Fréchet model to samples X and Y with respective Kolmogorov-Smirnov's p-values 0.8026 and 0.1934.

3.4 Real data example

We consider, as it is already used in [18], [6] and [22], the (originally left-truncated) lifetimes, denoted, L of car brake pads by the mileage, denoted M, given in [32], page 69. [18] gave a detailed description of these data and transformed them to X and Y in a right-truncation scheme. They also discussed the Pareto-like nature of their distributions. As far as we are concerned, we graphically checked the heavy-tailedness of X and Y, see Fig 3.17. Moreover, we applied Kolmogorov-Smirnov goodness-of-fit testing procedure to fit Fréchet models to both datasets with respective (large) p-values of 0.8026 and 0.1934.

High quantile estimation An extreme quantile of df $\mathbf{F}$ is a value q_ν defined in terms of the generalized inverse by $q_\nu := U_{\mathbf{F}}(1/\nu)$ for $v \downarrow 0$. In other words, it is an X-value which is sufficiently large so that the probability of exceeding it is very small. Also known as value-at-risk (VaR), this quantity is largely used, as a risk measure, in several fields such as in finance, insurance, hydrology and

$\overline{\nu}$	Transformed data	Original data
0.990	0.101	16.819
0.995	0.138	14.185
0.999	0.284	10.468

Table 3.6: Extreme quantiles for car brake pad lifetimes.

reliability. For asymptotic needs, we suppose that v is a function of the observed sample size n, denoted by $v = v_n$, and assumed to be much smaller than $1/n$. The estimation of high quantiles of heavy-tailed distributions, in the case of complete data, has been extensively studied in the literature (see, for instance, [24]). The well-known Weissman estimator [49] of high quantile q_ν adapted to our new tail index estimator $\widehat{\gamma}^*_{1,K}$ is given by

$$\widehat{q}_\nu := X_{n-k:n}\left(\frac{\nu}{\overline{\mathbf{F}}_n\left(X_{n-k:n}\right)}\right)^{-\widehat{\gamma}^*_{1,K}},$$

where $\mathbf{F}_n$ is Woodroofe's nonparametric estimator of df $\mathbf{F}$.

Using He-Wang's estimator [23] , we estimated proportion p of the observed sample, defined in (1.13), by $0.86 := p_0$ which means that the truncation is relatively weak. Then, as is noticed the simulation study, the tuning parameter β may be chosen equals 1. The sample size of this data set is $n = 98$ which almost equals 100, then relying to the results in Tables 3.2-3.4 (for $N = 150$), we suggest that the new estimator $\widehat{\gamma}^*_{1,K_3}$ is a better candidate to estimate the tail index γ_1.

Making use of Reiss-Thomas's algorithm, we select the optimal sample fraction $\widehat{k}$ and then compute the corresponding value of $\widehat{\gamma}^*_{1,K_3}$.

The result gives $\widehat{\gamma}^*_{1,K_3} = 0.46$, however [22] obtained the value 0.49, thereby, we compute, for three different high levels $\overline{\nu} = 1-\nu = 0.990$, 0.995 and 0.999 the corresponding extreme quantiles (see 3.4). Finally, via the aforementioned transformation, we obtain the pertaining extreme quantiles of the original dataset. The results are summarized in Table 3.6.

For instance, we may conclude that the estimated value of the brake pad lifetime is less than 16.819 km for 1% of the cars. However, only one out of a thousand brake pads lasts less than 10.468 km.

CHAPTER 4

A weighted minimum density power divergence estimator for the Pareto-tail index

Assigning a weight function to the density power divergence, we derive a new class of estimators for the tail index of a Pareto-type distribution. The proposed estimators may be considered as a robust generalization of the weighted least squares estimator and the kernel estimator of the tail index. The consistency and asymptotic normality of the proposed class of estimators are established. The study of finite sample behavior of the given estimator are done.

4.1 Minimum density power divergence

Given two probability densities ℓ and h, [2] introduced a new distance between them called the density power divergence

$$d_\alpha(h,\ell)=\begin{cases}\int_{\mathbb{R}}\left[\ell^{1+\alpha}(z)-\left(1+\frac{1}{\alpha}\right)\ell^{\alpha}(z)h(z)+\frac{1}{\alpha}h^{1+\alpha}(z)\right]dz, & \alpha>0\\ \int_{\mathbb{R}}h(z)\log\frac{h(z)}{\ell(z)}dz, & \alpha=0.\end{cases} \tag{4.1}$$

where α is a nonnegative tuning parameter. The case corresponding to $\alpha=0$ is obtained from the general case by letting $\alpha\to 0$ leading to the classical Kullback-Leibler divergence denoted $d_0(h,\ell)$. Let us consider a parametric model of densities $\{\ell_\theta:\Theta\subset\mathbb{R}^p\}$ and suppose that we consider the estimation of the parameter θ. Let H be the cdf corresponding to the density h. The minimum density power divergence (MDPD) is a functional $T_\alpha(H)$ defined by $d_\alpha\left(h,\ell_{T_\alpha(H)}\right)=\min_{\theta\in\Theta}d_\alpha(h,\ell_\theta)$. It is clear that the term $\int h^{1+\alpha}(z)dz$ in (4.1) does not contribute in the minimization of $d_\alpha(h,\ell_\theta)$ over $\theta\in\Theta$. Then minimization in the computation of the MDPD functional $T_\alpha(H)$ reduces to

$$\delta_\alpha(h;\theta):=\begin{cases}\int_{\mathbb{R}}\ell_\theta^{1+\alpha}(z)dz-\left(1+\frac{1}{\alpha}\right)\int_{\mathbb{R}}\ell_\theta^{\alpha}(z)dH(z), & \alpha>0,\\ -\int_{\mathbb{R}}\log\ell_\theta(z)dH(z), & \alpha=0.\end{cases} \tag{4.2}$$

Given a random sample $Z_1,\ldots,Z_n$ from the distribution H we may estimate the objective function h in (4.2) by substituting H with its empirical counterpart H_n. For a given tuning parameter α, the MDPD estimator $\widehat{\theta}_{n,\alpha}$ of θ may be obtained by minimizing (over $\theta\in\Theta$) the quantity

$$\delta^*_{n.\alpha}(\theta):=\begin{cases}\int_{\mathbb{R}}\ell_\theta^{1+\alpha}(z)dz-\left(1+\frac{1}{\alpha}\right)\int_{\mathbb{R}}\ell_\theta^{\alpha}(z)dH_n(z), & \alpha>0,\\ -\int_{\mathbb{R}}\log\ell_\theta(z)dH_n(z), & \alpha=0,\end{cases},$$

which in turn equals

$$\begin{cases}\int_{\mathbb{R}}\ell_\theta^{1+\alpha}(z)dz-\left(1+\frac{1}{\alpha}\right)\frac{1}{n}\sum_{i=1}^{n}\ell_\theta^{\alpha}(Z_i), & \alpha>0,\\ -\frac{1}{n}\sum_{i=1}^{n}\log\ell_\theta(Z_i), & \alpha=0.\end{cases}. \tag{4.3}$$

Thus the parameter θ minimizing $\delta^*_{n,\alpha}(\theta)$ will be a solution of the following equation

$$\begin{cases} \int_{\mathbb{R}} \frac{d}{d\theta}\ell_\theta^{1+\alpha}(z)dz - \left(1+\frac{1}{\alpha}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{d}{d\theta}\ell_\theta^{\alpha}(Z_i) = 0, & \alpha > 0, \\ \frac{1}{n}\sum_{i=1}^{n}\frac{d}{d\theta}\log \ell_\theta(Z_i) = 0, & \alpha = 0. \end{cases}$$

This may be rewritten into

$$\begin{cases} \int_{\mathbb{R}} u_\theta(z)\ell_\theta^{\alpha+1}(z)dz - \frac{1}{n}\sum_{i=1}^{n} u_\theta(z)\ell_\theta^{\alpha}(Z_i) = 0, & \alpha > 0, \\ \frac{1}{n}\sum_{i=1}^{n} u_\theta(z)(Z_i) = 0, & \alpha = 0. \end{cases}$$

where $u_\theta(z) := d\log \ell_\theta(z)/d\theta$ called the score function pertaining to density ℓ_θ. The role of the tuning parameter α is crucial in the sense that it offers a compromise between efficiency and robustness of the MDPD. In other terms when α is close to zero the estimator becomes more efficient however is less robust against outliers, while when α increases the robustness increases as well and the efficiency decreases. It is found that the estimators with small ac have strong robustness properties with little loss in asymptotic efficiency relative to maximum likelihood under model conditions.

4.2 Weighted MDPD

Since we are dealing with the upper extreme values, then it is convenient to assign a suitable weight to the right-tail of distributions. To this end, we consider the weighted densities instead of the original ones, namely $\ell_J := J\left(\overline{L}\right)\ell$ and $h_J = J\left(\overline{H}\right)h$, where $L(z) := \int_{-\infty}^{z}\ell(y)dy$ denotes the cdf corresponding to the density function ℓ and J be a nonnegative nonincreasing function such that $\int_0^1 J(s)ds = 1$, so that $\ell_J, h_J \geq 0$ and $\int_{\mathbb{R}} \ell_J(z)dz = \int_{\mathbb{R}} h_J(z)dz = 1$. We define the weighted minimum density power divergence (WMDPD), pertaining to a weight function J, between ℓ and h by $d_{\alpha,J}(h,\ell) := d_\alpha(h_J, \ell_J)$, that is

$$d_{\alpha,J}(h,\ell) := \begin{cases} \int_{\mathbb{R}} \left[\ell_J^{1+\alpha}(z) - \left(1+\frac{1}{\alpha}\right) \ell_J^\alpha(z) h_K(z) + \frac{1}{\alpha} h_J^{1+\alpha}(z) \right] dz, & \alpha > 0, \\ \int_{\mathbb{R}} h_J(z) \log \frac{h_J(z)}{\ell_J(z)} dz, & \alpha = 0. \end{cases} \tag{4.4}$$

Considering the indicator weight function $J_1 := \mathbf{1}_{[0,1)}$, the distance $d_{\alpha,J_1}(h,\ell)$ reduces to the original one $d_\alpha(h,\ell)$. The term $\int h_J^{1+\alpha}(z)dz$ in (4.4) remains has no role in the minimization of $d_{\alpha,J}(h,\ell_\theta)$ over $\theta \in \Theta$, therefore it suffices to minimize

$$d^*_{\alpha,J}(h,\ell) = \begin{cases} \int_{\mathbb{R}} \ell_J^{1+\alpha}(z)dz - \left(1+\frac{1}{\alpha}\right) \int_{\mathbb{R}} \ell_J^\alpha(z) J\left(\overline{H}\right) dH(z), & \alpha > 0 \\ \int_{\mathbb{R}} J\left(\overline{H}\right) \log \ell_J(z) dH(z), & \alpha = 0. \end{cases}$$

Following the above procedure, the corresponding estimator $\widehat{\theta}_{n,\alpha,J}$ of θ can be obtained by minimizing (over $\theta \in \Theta$) the quantity

$$\delta^*_{n,\alpha,J}(\theta) := \begin{cases} \int_{\mathbb{R}} \ell_{\theta,J}^{1+\alpha}(z)dz - \left(1+\frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^n J\left(\overline{H}(Z_i)\right) \ell_{\theta,J}^\alpha(Z_i), & \alpha > 0, \\ \frac{1}{n} \sum_{i=1}^n J\left(\overline{H}(Z_i)\right) \log \ell_{\theta,J}(Z_i), & \alpha = 0. \end{cases}.$$

4.3 WMDPD estimation of the tail index

Let us now consider the estimation the tail index γ by using the WMDPD. To this end, let us consider the relative excess rv $Z_u := X/u$ given $X > u$, with cdf $H_u(z) = 1 - \overline{F}(uz)/\overline{F}(uz)$ of corresponding density function h_u. In this case, the parametric model of densities is

$$\ell_\gamma(z) := \frac{d}{dz}\left(1 - z^{-1/\gamma}\right) = \gamma^{-1} z^{-1-1/\gamma},\ z \geq 1,\ \gamma > 0.$$

We are dealing to minimize the density weighted power divergence objective function the quantity

$$d^*_{u,\alpha,J}(\gamma) := \begin{cases} \int_1^\infty \ell_{\gamma,J}^{1+\alpha}(z)dz - \left(1+\frac{1}{\alpha}\right) \int_1^\infty \ell_{\gamma,J}^\alpha(z) J\left(\overline{H}_u(z)\right) dH_u(z), & \alpha > 0, \\ \int_1^\infty J\left(\overline{H}_u(z)\right) \log \ell_{\gamma,J}(z) dH_u(z), & \alpha = 0, \end{cases} \tag{4.5}$$

for sufficiently large u, which equals

$$\begin{cases} \int_1^\infty \ell_{\gamma,J}^{1+\alpha}(x)\,dx - \left(1+\dfrac{1}{\alpha}\right)\int_1^\infty \ell_{\gamma,J}^{\alpha}(x)\,J\left(\dfrac{\overline{F}(ux)}{\overline{F}(u)}\right) d\dfrac{F(ux)}{\overline{F}(u)}, & \alpha>0, \\ \int_1^\infty J\left(\dfrac{\overline{F}(ux)}{\overline{F}(u)}\right)\log \ell_{\gamma,J}(x)\,d\dfrac{F(ux)}{\overline{F}(u)}, & \alpha=0. \end{cases}$$

Substituting F by its empirical cdf $F_n(x) := n^{-1}\sum_{i=1}^n \mathbf{1}_{(X_i\leq x)}$ and letting $u = X_{n-k:n}$ in the previous functional, we end up with

$$\begin{cases} \dfrac{1}{k}\sum_{i=1}^{k} J\left(\dfrac{i}{k}\right)\dfrac{d}{d\gamma}\ell_{\gamma,J}^{\alpha}\left(\dfrac{X_{n-i+1:n}}{X_{n-k:n}}\right) = \dfrac{\alpha}{\alpha+1}\int_1^\infty \dfrac{d}{d\gamma}\ell_{\gamma,J}^{1+\alpha}(x)\,dx, & \alpha>0, \\ \dfrac{1}{k}\sum_{i=1}^{k} J\left(\dfrac{i}{k}\right)\dfrac{d}{d\gamma}\log \ell_{\gamma,J}\left(\dfrac{X_{n-i+1:n}}{X_{n-k:n}}\right) = 0, & \alpha=0. \end{cases} \tag{4.6}$$

This may be rewritten into

$$\begin{cases} \dfrac{1}{k}\sum_{i=1}^{k} J\left(\dfrac{i}{k}\right)u_{\gamma,J}\left(\dfrac{X_{n-i+1:n}}{X_{n-k:n}}\right)\ell_{\gamma,J}^{\alpha}\left(\dfrac{X_{n-i+1:n}}{X_{n-k:n}}\right) = \int_1^\infty u_{\gamma,J}(x)\,\ell_{\gamma,J}^{\alpha+1}(x)\,dx, & \alpha>0, \\ \dfrac{1}{k}\sum_{i=1}^{k} J\left(\dfrac{i}{k}\right)u_{\gamma,J}\left(\dfrac{X_{n-i+1:n}}{X_{n-k:n}}\right)\ell_{\gamma,J}\left(\dfrac{X_{n-i+1:n}}{X_{n-k:n}}\right) = 0, & \alpha=0, \end{cases} \tag{4.7}$$

where $u_\gamma(x) := d\log \ell_{\gamma,J}(x)/d\gamma$ is the weighted maximum likelihood score function.

	$J_{\log}$	J_0	J_1	J_2	J_3	J_4
$\mathscr{L}$	1	0	$s\frac{\log s}{s-1}$	$4s^2\frac{\log s}{s^2-1}$	$6s^2\frac{\log s}{s^2-1}$	$8s^2\frac{\log s}{s^2-1}$
ζ_1	1	0	1	2	3	4
ζ_2	0	0	0.644	0.934	1.40	1.86

Thereby, we define our newly WMDPD tail index estimator of γ, denote by $\widehat{\gamma}_{k,\alpha,J}$ as a solution of equation (4.6) (or (4.7)). It is worth noting that in the case $\alpha = 0$, the class of WMDPD estimators asymptotically meets that of WLSE ones, in other terms $\widehat{\gamma}_{k,0,J} = \widehat{\gamma}_{k,J} + o_{\mathbf{P}}(1)$, as $n\to\infty$. Indeed, let us introduce the class of weight functions J satisfying $[A1]$ below and further assume that

$$\zeta_1 := \sup_{0<s<1} \mathscr{L}(s) < \infty \text{ and } \zeta_2 := \int_0^1 \left|\mathscr{L}'(s)\right| ds < \infty, \tag{4.8}$$

where $\mathscr{L}(s) := \left(J'(s)/J(s)\right) s \log s$, for $s \in (0,1)$ and $\mathscr{L}'$ stands for the first derivative of $\mathscr{L}$. The function $\mathscr{L}$ is nonnegative because $\log s < 0$ and J is nonincreasing on $(0,1)$. The additional regularity assumptions (4.8) on J is somewhat not restrictive, in the sense that the common weight functions satisfy it. Indeed, in Table 4.3, we give the formula of function $\mathscr{L}$ and both the values of ζ_1 and ζ_2 for the aforementioned weight functions. Assuming (4.8), we showed in Lemma 4.1, that

$$\widehat{\gamma}_{k,0,J} := \left(\int_0^1 J(s) \log s^{-1} ds\right)^{-1} \frac{1}{k} \sum_{i=1}^{k} J\left(\frac{i}{k}\right) \log \frac{X_{n-i+1:n}}{X_{n-k:n}} + O_{\mathbf{P}}\left(k^{-1}\right),$$

which asymptotically meets to the weighted least squares estimator $\widehat{\gamma}_{k,J}$ sated in (1.10). Moreover, Theorem 2.1 and Theorem 2.2 in [28] concerning the consistency and asymptotic normality of $\widehat{\gamma}_{k,J}$ lead to that of $\widehat{\gamma}_{k,0,J}$. Thus, we may consider that $\widehat{\gamma}_{k,\alpha,J}$ for $\alpha > 0$ is a robust generalization of $\widehat{\gamma}_{k,J}$ and $\widehat{\gamma}_{k,\mathbb{K}}^{(CDM)}$. Anyway, the assumption (4.8) does not concern the estimator $\widehat{\gamma}_{k,\alpha,J}$ for $\alpha > 0$.

Next we study the asymptotic behavior of the solution to the estimating equation (4.6), for $\alpha > 0$. To this end we assume that the underlying distribution F satisfies the second-order condition of regular variation [20], that is: for any $x > 0$

$$\lim_{t\to\infty} \frac{U(tx)/U(t) - x^{\gamma}}{a(t)} = x^{\gamma} \frac{x^{\rho} - 1}{\rho}, \tag{4.9}$$

where $\rho < 0$ is the second-order parameters and a is a function tending to zero and not changing signs near infinity with regularly varying absolute values with index ρ. The notation $U(t) := F^{\leftarrow}(1 - 1/t)$, $t > 1$, where $F^{\leftarrow}(s) := \inf\{x : F(x) \geq s\}$, $0 < s < 1$, stands for the quantile function.

4.4 Main results

We state two theorems in which we establish existence and consistency of a sequence of solutions to the estimating equation (4.6), for $\alpha > 0$. We will consider the class of weight functions satisfying the following regularity assumptions:

- $[A1]$ J is nonincreasing nonnegative on $(0,1)$ with $\int_0^1 J(s) ds = 1$.
- $[A2]$ J and their first three derivatives J', J'' and J''' are bounded on $(0,1)$.

From now on we denote the true value of γ by γ_0.

Theorem 4.1. *(existence and consistency) Let $X_1,...,X_n$ be a sample of iid rv's from a cdf satisfying condition* (4.9)*. Let J be a continuous function fulfilling assumptions $[A1]-[A2]$ and let k be a sequence of integers such that $k\to\infty$ and $k/n\to 0$, as $n\to\infty$. Then with probability tending to* 1, *there exists solution $\widehat{\gamma}_{k,\alpha,J}$, for $\alpha>0$, of the estimating equation* (4.6) *such that $\widehat{\gamma}_{k,\alpha,J}\xrightarrow{\mathbf{P}}\gamma_0$, as $n\to\infty$.*

Proof. To show the existence and consistency of $\widehat{\gamma}_{k,\alpha,J}$ we adapt the proof of Theorem 1in [15] which in turns is an adaptation of the proof of Theorem 5.1 in Chapter 6 of [33], proving the existence and establishing the consistency of solutions of the likelihood equations, to the WMDPDE context. Let $d^*_{n,\alpha,J}(\gamma)$ denotes the empirical counterpart of the weighted density power divergence objective function $d^*_{u,\alpha,J}(\gamma)$, given in (4.5), namely

$$\widehat{d}^*_{n,\alpha,J}(\gamma):=\int_1^\infty \ell^{1+\alpha}_{\gamma,J}(x)\,dx-\left(1+\frac{1}{\alpha}\right)\frac{1}{k}\sum_{i=1}^{k}J\left(\frac{i}{k}\right)\ell^{\alpha}_{\gamma,J}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right),\ \alpha>0.$$

Next we show that

$$\mathbf{P}_{\gamma_0}\left(d^*_{n,\alpha,J}(\gamma_0)<d^*_{n,\alpha,J}(\gamma),\text{ for all }\gamma\in I_\epsilon\right)\to 1,\text{ as }\epsilon\downarrow 0, \tag{4.10}$$

where $I_\epsilon:=(\gamma_0-\epsilon,\gamma_0+\epsilon)$, for $0<\epsilon<\gamma_0$. By applying Taylor's expansion near γ_0 to function $\gamma\to d^*_{n,\alpha,J}(\gamma)$, we decompose $d^*_{n,\alpha,J}(\gamma)-d^*_{n,\alpha,J}(\gamma_0)$ into

$$\begin{aligned}&\pi^{(1)}_k(\gamma_0)(\gamma-\gamma_0)+2^{-1}\pi^{(2)}_k(\gamma_0)(\gamma-\gamma_0)^2+6^{-1}\pi^{(3)}_k(\widetilde{\gamma})(\gamma-\gamma_0)^3\\ =:\ &S_{1,k}+S_{2,k}+S_{3,k},\end{aligned}$$

where

$$\pi^{(m)}_k(\gamma_0):=\int_1^\infty\frac{d^m}{d\gamma^m}\ell^{\alpha+1}_{\gamma_0,J}(x)\,dx-\left(1+\frac{1}{\alpha}\right)A^{(m)}_k(\gamma_0), \tag{4.11}$$

$$A^{(m)}_k(\gamma_0):=\frac{1}{k}\sum_{i=1}^{k}J\left(\frac{i}{k}\right)\frac{d^m}{d\gamma^m}\ell^{\alpha}_{\gamma_0,J}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right),\ m=1,2 \tag{4.12}$$

and $\widetilde{\gamma}$ is between γ and γ_0. The result (4.21) of Lemma 4.2 gives $\pi^{(1)}_k(\gamma_0)\xrightarrow{\mathbf{P}}0$, as $n\to\infty$. In the other terms, for any $\epsilon>0$ sufficiently small $\left|\pi^{(1)}_k(\gamma_0)\right|<\epsilon^2$, which entails that $|S_{1,k}|<\epsilon^3$, for ever $\gamma\in I_\epsilon$, with probability tending to 1. Once again using assertion (4.21) of Lemma 4.2, we deduce that $\pi_k^{(2)}(\gamma_0) \xrightarrow{\mathbf{P}} \eta_{\gamma_0}$ as $n \to \infty$, where η_{γ_0} is as in (4.13), therefore $S_{2,k} = (1 + o_{\mathbf{P}}(1)) 2^{-1} \eta_{\gamma_0} (\gamma - \gamma_0)^2$. Observe that $\eta_{\gamma_0} > 0$, then there exists $c > 0$ and $\epsilon_0 > 0$ such that for $\epsilon < \epsilon_0$ such that $S_{2,k} > c\epsilon^2$. Note that $\widetilde{\gamma}$ is a consistent estimator for γ_0, then by Lemma 4.6 $\pi_k^{(3)}(\widetilde{\gamma}) = O_{\mathbf{P}}(1)$. This means that with probability tending to 1 there exists a constant $d > 0$ such that $|S_{3,k}| < d\epsilon^3$. Combining the above we find that

$$\min_{\gamma \in I_\epsilon} \left(S_{1,k} + S_{2,k} + S_{3,k} \right) > c\epsilon^2 - (d+1)\epsilon^3,$$

with probability tending to 1. Choosing $0 < \epsilon < c/(d+1)$ gives $c\epsilon^2 - (d+1)\epsilon^3 > 0$ leading to inequality (4.10). To complete the proof of existence and consistency, we follow the same steps to those used in the proof of Theorem 3.7 in Chapter 6 of [33]. Let $\epsilon > 0$ be small so that $0 < \epsilon < c/(d+1)$ and $I_\epsilon \subset (0, \infty)$, then consider the set

$$S_n(\epsilon) := \left\{ \gamma : d_{n,\alpha,J}^*(\gamma_0) < d_{n,\alpha,J}^*(\gamma) \text{ for all } \gamma \in I_\epsilon \right\}.$$

We already showed that $\mathbf{P}_{\gamma_0}\{S_n(\epsilon)\} \to 1$ for any such ϵ, then there exists a sequence $\epsilon_n \downarrow 0$ such that $\mathbf{P}_{\gamma_0}\{S_n(\epsilon_n)\} \to 1$ as $n \to \infty$. Note that $\gamma \to d_{n,\alpha,J}^*(\gamma)$ being differentiable on $(0, \infty)$, then given $\gamma \in S_n(\epsilon_n)$ there exists a point $\widehat{\gamma}_{k,\alpha,J}(\epsilon_n) \in I_{\epsilon_n}$ for which $d_{n,\alpha,J}^*(\gamma)$ attains a local minimum, thereby $\pi_k^{(1)}(\widehat{\gamma}_{k,\alpha,J}(\epsilon_n)) = 0$. Let us set $\widehat{\gamma}_{k,\alpha,J}^* = \widehat{\gamma}_{k,\alpha,J}(\epsilon_n)$ for $\gamma \in S_n(\epsilon_n)$ and arbitrary otherwise. Obviously

$$\mathbf{P}_{\gamma_0}\left\{ \pi_k^{(1)}\left(\widehat{\gamma}_{k,\alpha,J}^*\right) = 0 \right\} \geq \mathbf{P}_{\gamma_0}\{S_n(\epsilon_n)\} \to 1, \text{ as } n \to \infty,$$

thus with probability tending to 1 there exists a sequence of solutions to estimating Equation (4.6). Observe now that for any fixed $\epsilon > 0$ and n sufficiently large $\mathbf{P}_{\gamma_0}\left\{ \left| \widehat{\gamma}_{k,\alpha,J}^* - \gamma_0 \right| < \epsilon \right\} \geq \mathbf{P}_{\gamma_0}\left\{ \left| \widehat{\gamma}_{k,\alpha,J}^* - \gamma_0 \right| < \epsilon_n \right\} \to 1$, as $n \to \infty$, which establishes the consistency of $\widehat{\gamma}_{k,\alpha,J}^*$, as sought. ∎

Theorem 4.2. *(asymptotic normality) Let $X_1, ..., X_n$ be a sample of iid rv's from a cdf satisfying the condition* (4.9) *and assume that $\widehat{\gamma}_{k,\alpha,J}$, for $\alpha > 0$, is a consistent estimator for γ_0 satisfying* (4.6). *Let J be a continuous function fulfilling assumptions* $[A1]-[A2]$ *and let k be a sequence of integers such that $k \to \infty$, $k/n \to 0$ and $\sqrt{k}\,a(n/k) \to \lambda \in \mathbb{R}$. Then, in the probability space $(\Omega, \mathscr{A}, \mathbf{P})$ there exists a standard*

Wiener process $\{W(x),\ x \geq 0\}$, *such that*

$$\left(1+\frac{1}{\alpha}\right)\eta_{\gamma_0}\sqrt{k}\left(\widehat{\gamma}_{k,\alpha,J}-\gamma_0\right)$$
$$=\int_1^{\infty}\left(W\left(x^{-1/\gamma_0}\right)-x^{-1/\gamma_0}W(1)\right)J\left(x^{-1/\gamma_0}\right)d\Psi_{\gamma_0}^{(1)}(x)+\lambda B_{\gamma_0}^{(1)}+o_{\mathbf{P}}(1),$$

where

$$\eta_{\gamma_0}:=(1+\alpha)\int_1^{\infty}\left(\Psi_{\alpha,\gamma_0}^{(1)}(x)\right)^2\ell_{\gamma_0,J}^{\alpha-1}(x)\,dx,\ \Psi_{\alpha,\gamma_0}^{(1)}(x):=d\ell_{\gamma_0,J}^{\alpha}(x)/d\gamma \tag{4.13}$$

and

$$B_{\gamma_0}^{(1)}:=\frac{1}{\rho\gamma_0}\int_0^1 s\left(1-s^{-\rho}\right)J(s)\,d\Psi_{\alpha,\gamma_0}^{(1)}\left(s^{-\gamma_0}\right).$$

Thus $\sqrt{k}\left(1+\frac{1}{\alpha}\right)\eta_{\gamma_0}\left(\widehat{\gamma}_{k,\alpha,J}-\gamma_0\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(\lambda B_{\gamma_0}^{(1)},\sigma_{\gamma_0}^2\right)$, *as* $n\to\infty$, *where*

$$\sigma_{\gamma_0}^2:=\int_0^1\int_0^1(\min(s,t)-st)J(s)J(t)\,d\Psi_{\alpha,\gamma_0}^{(1)}\left(s^{-\gamma_0}\right)d\Psi_{\alpha,\gamma_0}^{(1)}\left(t^{-\gamma_0}\right).$$

Proof. Applying Taylor's expansion of the estimating equation $\pi_k^{(1)}\left(\widehat{\gamma}_{k,\alpha,J}\right)=0$, yields

$$0=\pi_k^{(1)}(\gamma_0)+\pi_k^{(2)}(\gamma_0)\left(\widehat{\gamma}_{k,\alpha,J}-\gamma_0\right)+\frac{1}{2}\pi_k^{(3)}\left(\widehat{\gamma}_0\right)\left(\widehat{\gamma}_{k,\alpha,J}-\gamma_0\right)^2,$$

where

$$\pi_k^{(3)}(\gamma_0):=\int_1^{\infty}\frac{d^3}{d\gamma^3}\ell_{\gamma_0,J}^{\alpha+1}(x)\,dx-\left(1+\frac{1}{\alpha}\right)A_k^{(3)}(\gamma_0) \tag{4.14}$$

and

$$A_k^{(3)}(\gamma_0):=\frac{1}{k}\sum_{i=1}^{k}J\left(\frac{i}{k}\right)\frac{d^3}{d\gamma^3}\ell_{\gamma_0,J}^{\alpha}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right), \tag{4.15}$$

with $\widehat{\gamma}_0$ is between γ_0 and $\widehat{\gamma}_{k,\alpha,J}$. We have $\widehat{\gamma}_0\xrightarrow{\mathbf{P}}\gamma_0$ as $n\to\infty$, then by Lemma 4.6 we get $\pi_k^{(3)}\left(\widehat{\gamma}_0\right)=O_{\mathbf{P}}(1)$. On the other hand $\widehat{\gamma}_{k,\alpha,J}\xrightarrow{\mathbf{P}}\gamma_0$, it follows that

$$2^{-1}\pi_k^{(3)}\left(\widehat{\gamma}_0\right)\left(\widehat{\gamma}_{k,\alpha,J}-\gamma_0\right)^2=o_{\mathbf{P}}(1)\left(\widehat{\gamma}_{k,\alpha,J}-\gamma_0\right),$$

therefore

$$\pi_k^{(2)}(\gamma_0)\sqrt{k}\left(\widehat{\gamma}_{k,\alpha,J}-\gamma_0\right)(1+o_{\mathbf{P}}(1))=-\sqrt{k}\,\pi_k^{(1)}(\gamma_0),\ \text{as } n\to\infty.$$

Once again using Gaussian approximation (4.18) of Lemma 4.2, we may write

$$\begin{aligned}&\left(1+\frac{1}{\alpha}\right)\pi_k^{(2)}(\gamma_0)\sqrt{k}\left(\widehat{\gamma}_{k,\alpha,J}-\gamma_0\right)\\&=\int_1^\infty\left(W\left(x^{-1/\gamma_0}\right)-x^{-1/\gamma_0}W(1)\right)J\left(x^{-1/\gamma_0}\right)d\Psi_{\gamma_0}(x)+\lambda B_{\gamma_0}^{(1)}+o_{\mathbf{P}}(1),\end{aligned}$$

where $\{W(x),\ x\geq 0\}$ is a standard Wiener process, $B_{\gamma_0}^{(1)}$ is as in (4.20) and $\Psi_{\alpha,\gamma_0}^{(1)}(x)=d\ell_{\gamma_0,J}^{\alpha}(x)/d\gamma$. From Assertion (4.21) of Lemma 4.2, we have $\pi_k^{(2)}(\gamma_0)\xrightarrow{\mathbf{P}}\eta_{\gamma_0}$ therefore

$$\left(1+\frac{1}{\alpha}\right)\eta_{\gamma_0}\sqrt{k}\left(\widehat{\gamma}_{k,\alpha,J}-\gamma_0\right)\to\mathcal{N}\left(\lambda B_{\gamma_0}^{(1)},\sigma_{\gamma_0}^2\right),$$

as $n\to\infty$, where

$$\begin{aligned}\sigma_{\gamma_0}^2:=\int_1^\infty\int_1^\infty&\left(\min\left(x^{-1/\gamma_0},y^{-1/\gamma_0}\right)-x^{-1/\gamma_0}y^{-1/\gamma_0}\right)\\&\times J\left(x^{-1/\gamma_0}\right)J\left(y^{-1/\gamma_0}\right)d\Psi_{\alpha,\gamma_0}^{(1)}(x)\,d\Psi_{\alpha,\gamma_0}^{(1)}(y),\end{aligned}$$

which equals $\int_0^1\int_0^1(\min(s,t)-st)J(s)J(t)\,d\Psi_{\alpha,\gamma_0}^{(1)}(s^{-\gamma_0})\,d\Psi_{\alpha,\gamma_0}^{(1)}(t^{-\gamma_0})$. This completes the proof of Theorem 4.2. ■

4.5 Influence function

The infulence function

$$\mathbf{IF}_t\left(x,T,L_\gamma\right)=\frac{\mathbf{1}(x>t)\,J\left(\dfrac{\overline{F}(xt)}{\overline{F}(t)}\right)u_{\gamma,J}(x)\,\ell_{\gamma,J}^{\alpha}(x)-\int_1^\infty u_{\gamma,J}(z)\,\ell_{\gamma,J}^{1+\alpha}(z)\,dz}{\int_1^\infty u_{\gamma,J}^2(z)\,\ell_{\gamma,J}^{1+\alpha}(z)\,dz}.$$

First its worth noting that

$$\int_1^\infty u_{\gamma,J}(z)\,\ell_{\gamma,J}^{1+\alpha}(z)\,dz<\infty\text{ and }0\neq\int_1^\infty u_{\gamma,J}^2(z)\,\ell_{\gamma,J}^{1+\alpha}(z)\,dz<\infty.$$

For all large t, we have

$$\mathbf{IF}_t\left(x,T,L_\gamma\right)=\frac{\mathbf{1}(x>t)\,J\left(x^{-1/\gamma}\right)u_{\gamma,J}(x)\,\ell_{\gamma,J}^{\alpha}(x)-\int_1^\infty u_{\gamma,J}(z)\,\ell_{\gamma,J}^{1+\alpha}(z)\,dz}{\int_1^\infty u_{\gamma,J}^2(z)\,\ell_{\gamma,J}^{1+\alpha}(z)\,dz}+o(1),$$

uniformly on $x\geq 1$. We show that

$$\mathbf{1}(x>t)\,J\left(x^{-1/\gamma}\right)u_{\gamma,J}(x)\,\ell^{\alpha}_{\gamma,J}(x)\to 0,\text{ as } x\to\infty,$$

then

$$\lim_{x\to\infty}\lim_{t\to\infty}\mathbf{IF}_t\left(x,T,L_\gamma\right)=-\frac{\int_1^\infty u_{\gamma,J}(z)\,\ell^{1+\alpha}_{\gamma,J}(z)\,dz}{\int_1^\infty u^2_{\gamma,J}(z)\,\ell^{1+\alpha}_{\gamma,J}(z)\,dz},$$

this means that $\mathbf{IF}_t\left(x,T,L_\gamma\right)$ is asymptotically bounded.

4.6 Simulation study

We consider three distributions in the Fréchet domain of attraction namely the Fréchet and Burr. For each distribution F, we generated samples from a mixture contaminated model: 1 E F EG where G is a nuisance distribution. Specifically, G is chosen in two ways: from the same distribution as F but with different parameters and a different distribution from F. In each case, we assess the robustness of the estimators under different contamination scenarios with E 0:05 and E 0:15: Furthermore, to assess the effect of the robustness parameter, we take three values of a, at 0.1, 0.5 and 1 representing levels for increased robustness.

4.7 Important lemma

Lemma 4.1. *Assume that $\overline{F}$ satisfies the condition* (1.8) *and let J be a weight function fulfilling assumptions* $[A1]$ *and* (4.8), *then*

$$\begin{aligned}0&=\frac{1}{k}\sum_{i=1}^{k}J\left(\frac{i}{k}\right)\frac{d}{d\gamma}\log\ell_{\gamma,J}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)\\&=\gamma^{-1}\int_0^1 J(s)\log s\,ds+\gamma^{-2}\frac{1}{k}\sum_{i=1}^{k}J\left(\frac{i}{k}\right)\log\frac{X_{n-i+1:n}}{X_{n-k:n}}+O_{\mathbf{P}}\left(k^{-1}\right),\ \text{as } n\to\infty,\end{aligned}$$

and therefore

$$\widehat{\gamma}_{k,0,J}=\left(\int_0^1 J(s)\log s^{-1}ds\right)^{-1}\frac{1}{k}\sum_{i=1}^{k}J\left(\frac{i}{k}\right)\log\frac{X_{n-i+1:n}}{X_{n-k:n}}+O_{\mathbf{P}}\left(k^{-1}\right).$$

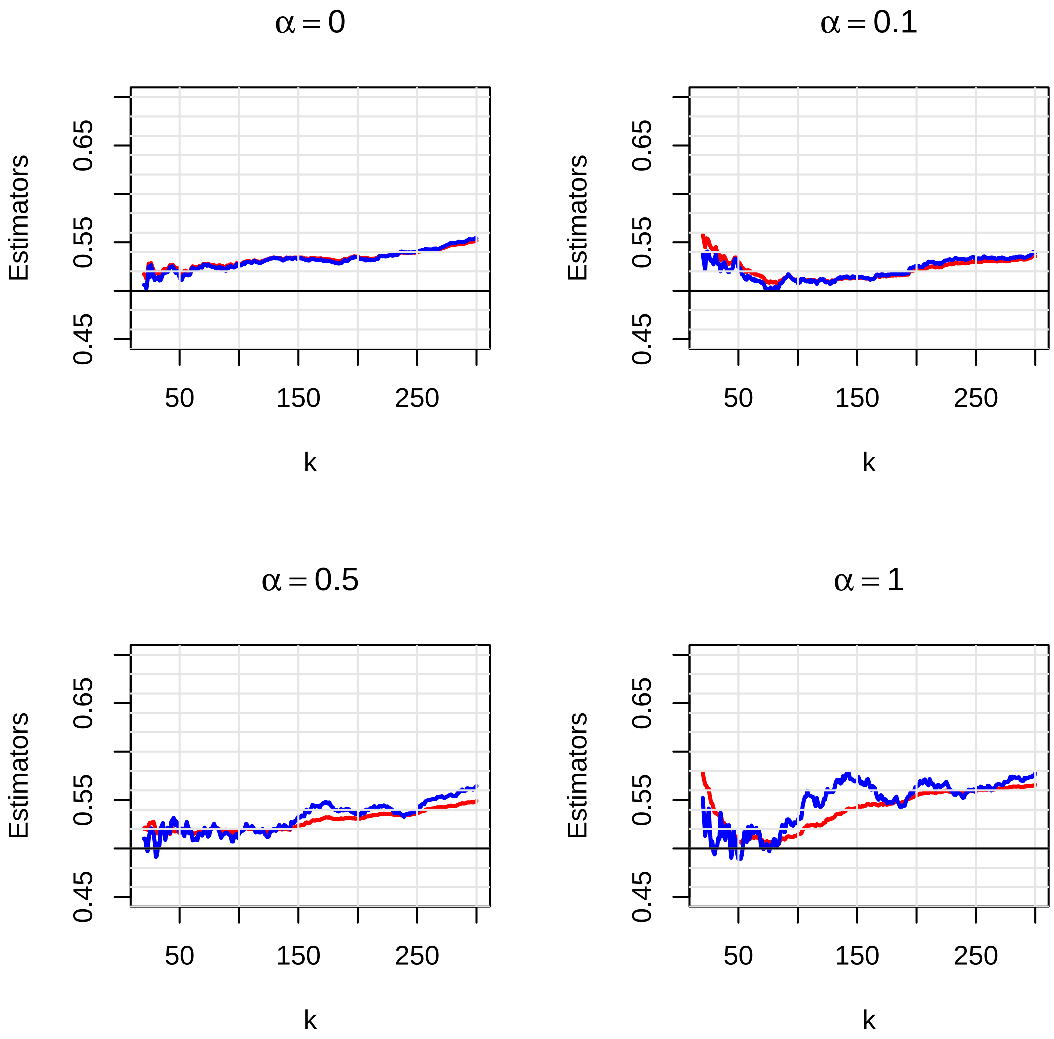

Figure 4.1: Plotting the estimators $\widehat{\gamma}_{\alpha,k,J}$ (red line) and $\widehat{\gamma}_{\alpha,k,\mathbf{1}}$ (blue line) for a Fréchet distribution with tail index: $\gamma = 0.5$ and different values of α, based on 20 samples of size 1000.

Proof. We have

$$\begin{aligned}
&\frac{1}{k}\sum_{i=1}^{k} J\left(\frac{i}{k}\right)\frac{d}{d\gamma}\log \ell_{\gamma,J}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)\\
&= \frac{n}{k}\int_1^\infty J\left(\frac{n}{k}\overline{F}_n(xX_{n-k:n})\right)\frac{d}{d\gamma}\log \ell_{\gamma,J}(x)\,dF_n(xX_{n-k:n}) =: I_k,
\end{aligned}$$

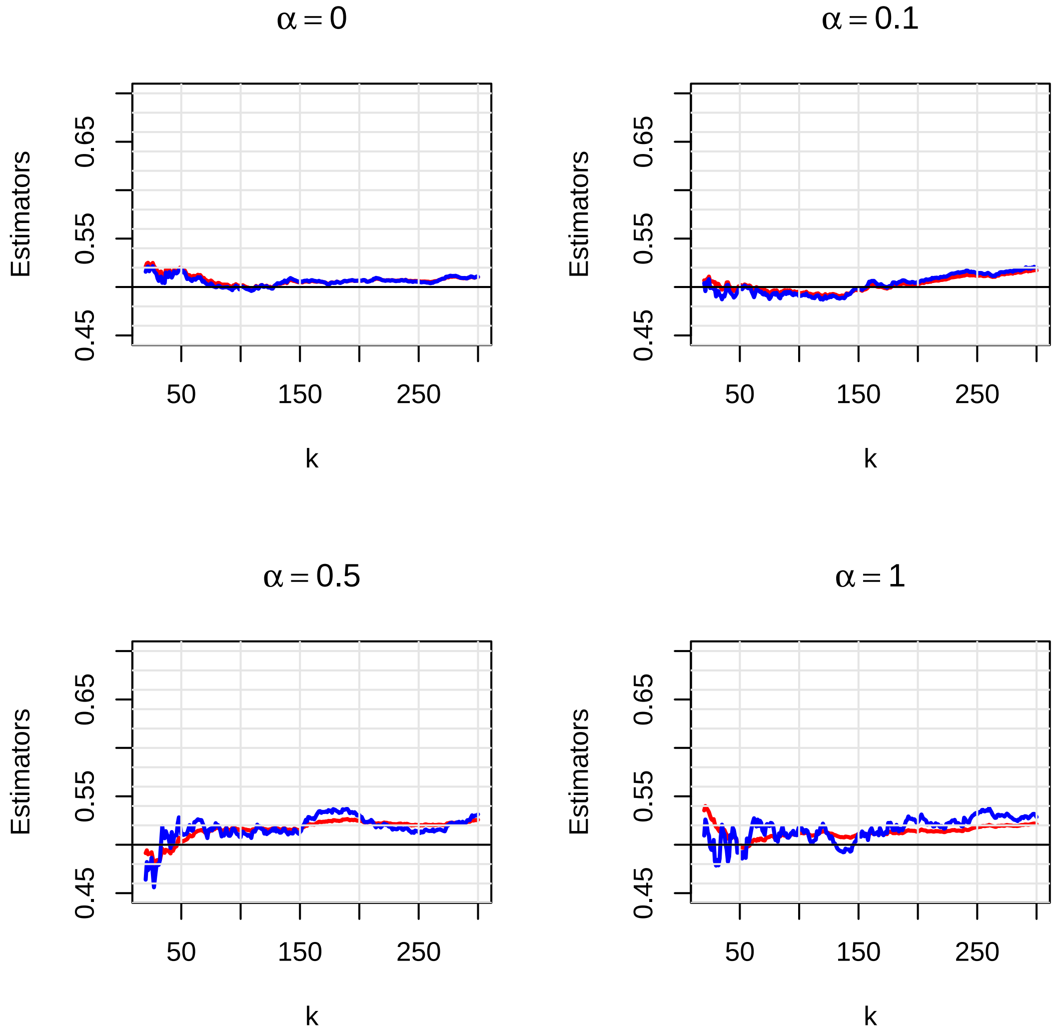

Figure 4.2: Plotting the estimators $\widehat{\gamma}_{\alpha,k,J}$ (red line) and $\widehat{\gamma}_{\alpha,k,\mathbf{1}}$ (blue line) for a Burr distribution with tail index: $\gamma = 0.5$ and diffrent values of α, based on 20 samples of size 1000.

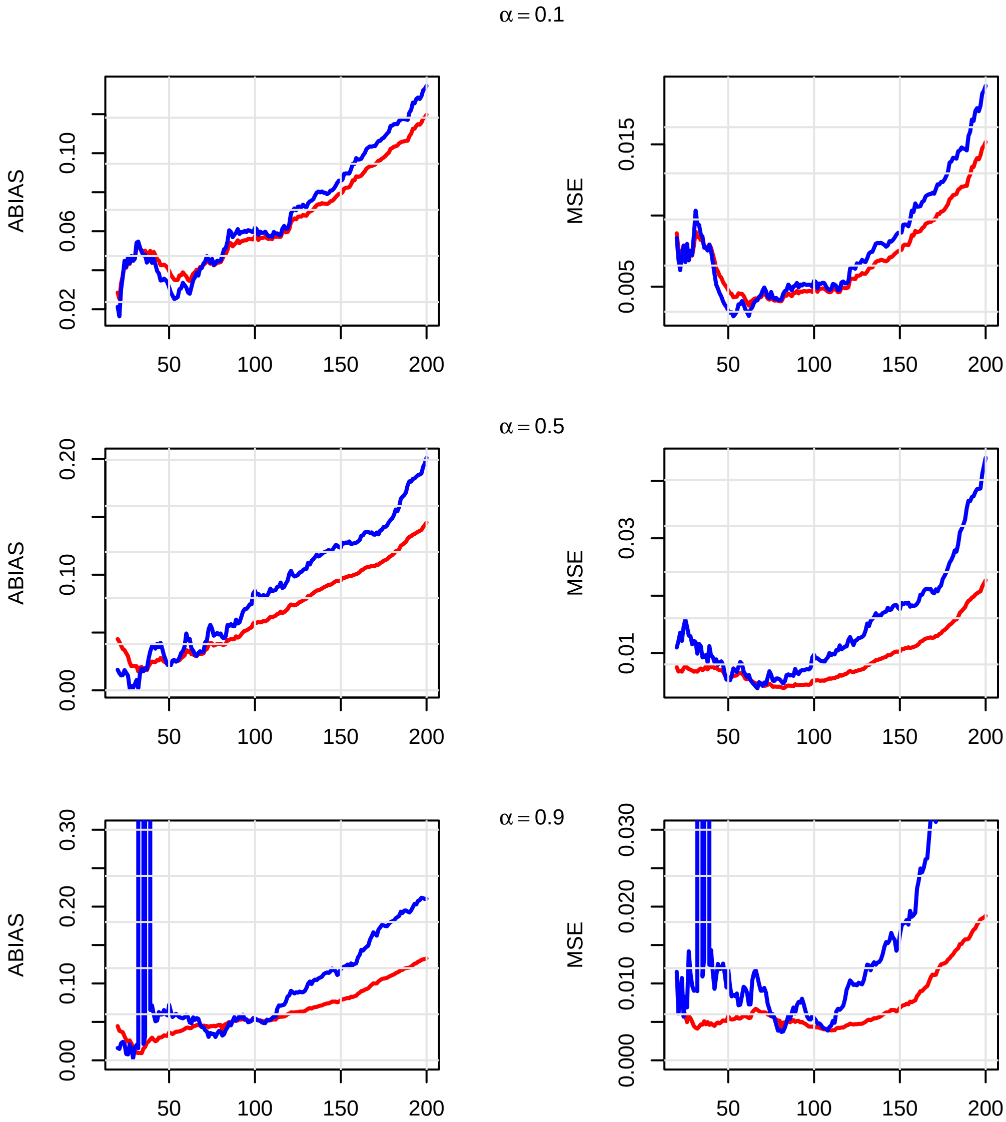

Figure 4.3: Absolute bias (left panel) and MSE (right panel) of $\widehat{\gamma}_{\alpha,k,J}$ (red) and $\widehat{\gamma}_{\alpha,k,\mathbf{1}}$ (blue), corresponding to Frechet distribution with tail index: $\gamma = 0.4$ and diffrent values of α ,based on 20 samples of size 300.

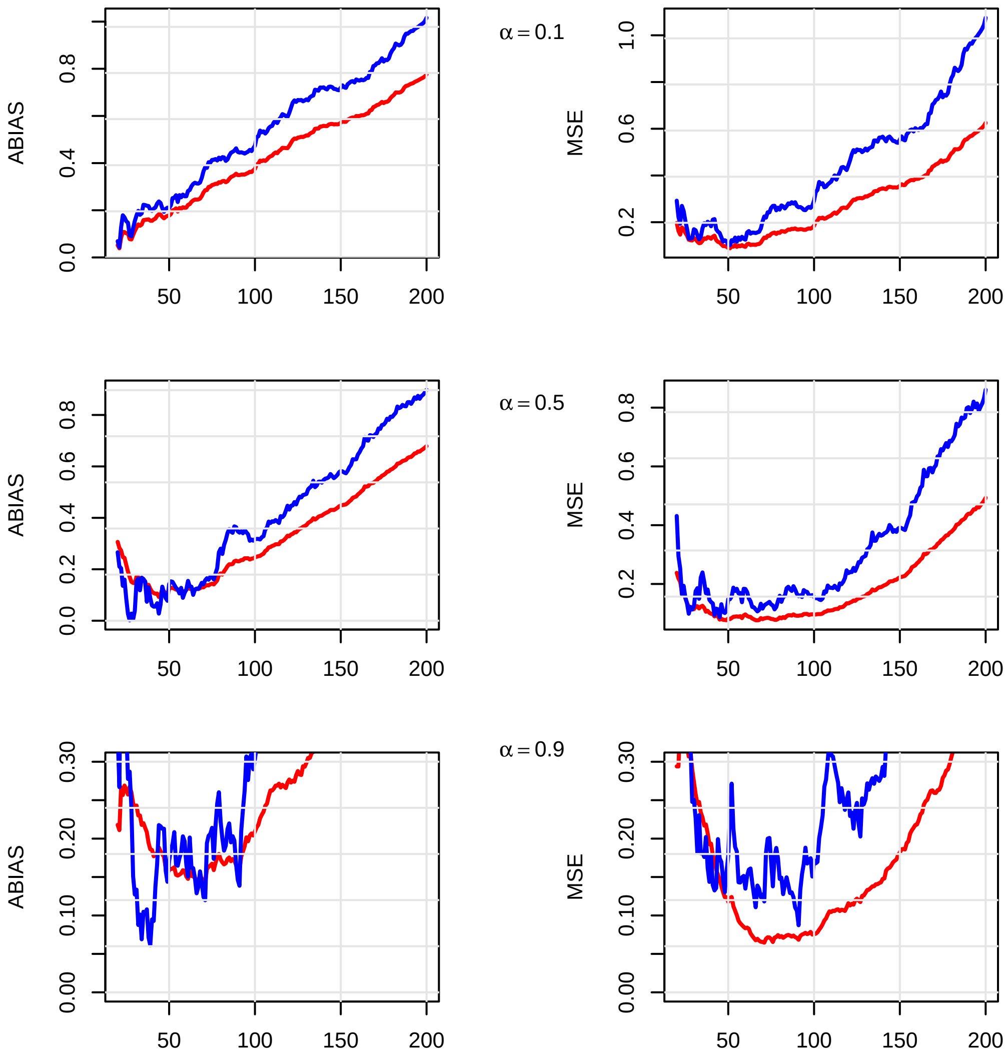

Figure 4.4: Absolute bias (left panel) and MSE (right panel) of $\widehat{\gamma}_{\alpha,k,J}$ (red) and $\widehat{\gamma}_{\alpha,k,\mathbf{1}}$ (blue), corresponding to Fréchet distribution with tail index: $\gamma = 1.5$ and diffrent values of α ,based on 20 samples of size 300.

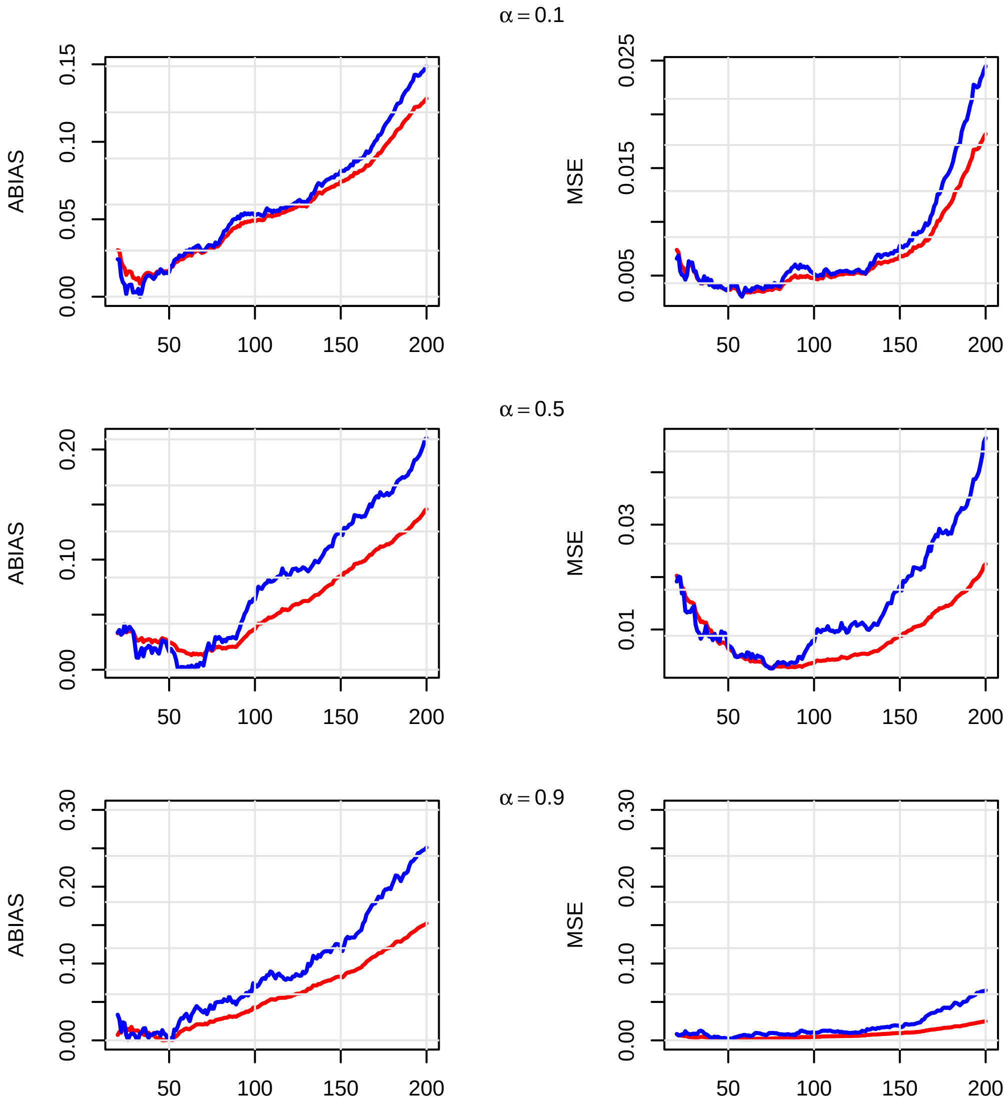

Figure 4.5: Absolute bias (left panel) and MSE (right panel) of $\widehat{\gamma}_{\alpha,k,J}$ (red) and $\widehat{\gamma}_{\alpha,k,\mathbf{1}}$ (blue), corresponding to Burr distribution with tail index: $\gamma = 0.4$ and diffrent values of α ,based on 20 samples of size 300.

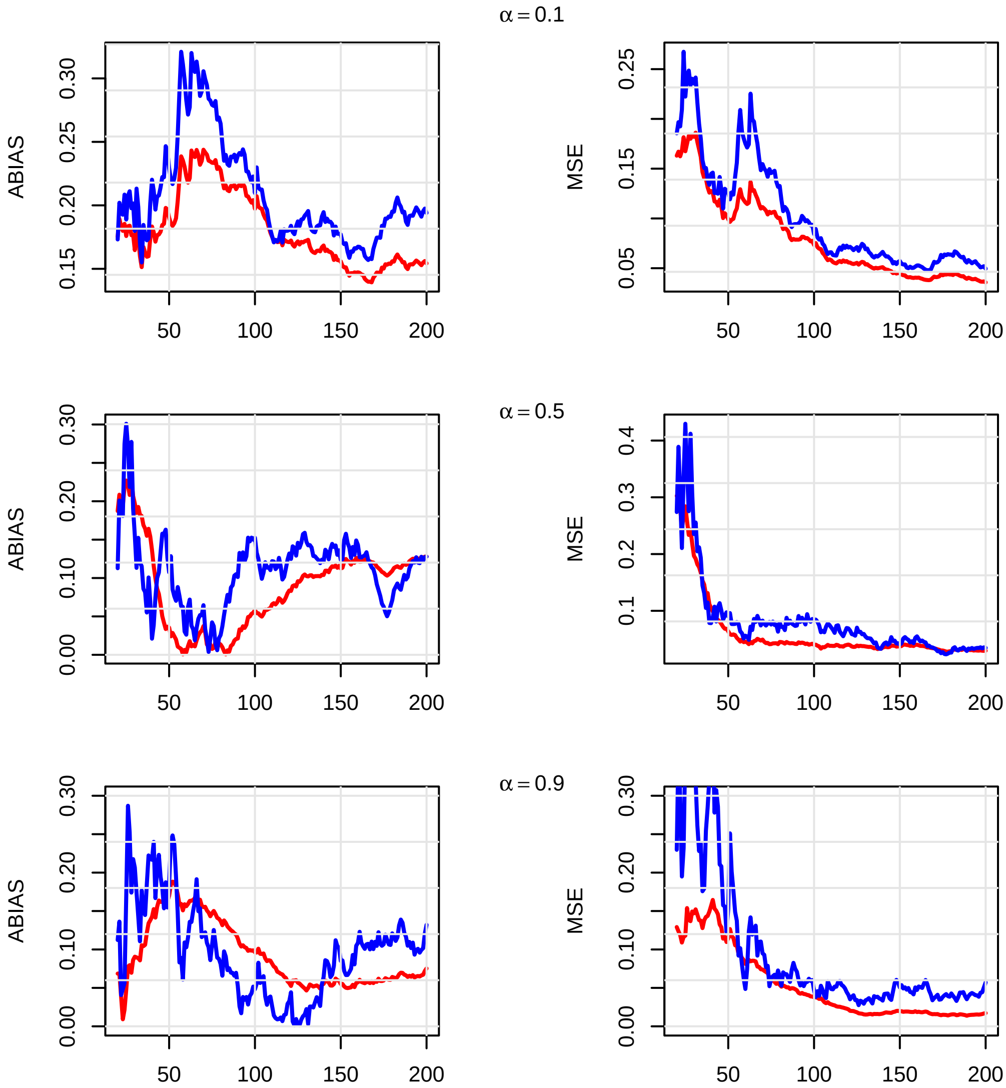

Figure 4.6: Absolute bias (left panel) and MSE (right panel) of $\widehat{\gamma}_{\alpha,k,J}$ (red) and $\widehat{\gamma}_{\alpha,k,\mathbf{1}}$ (blue), corresponding to Burr distribution with tail index: $\gamma = 1.5$ and diffrent values of α ,based on 20 samples of size 300.

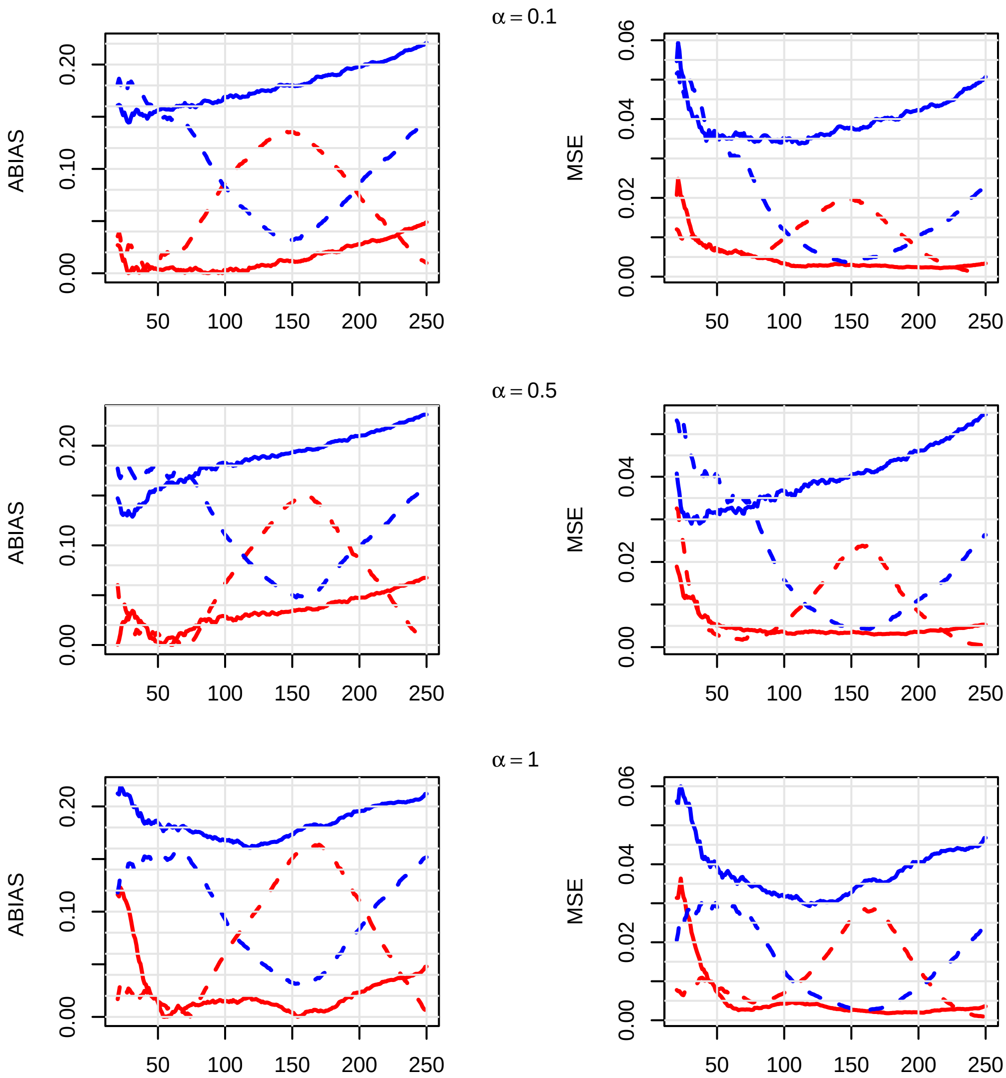

Figure 4.7: Comparaison in terms of absolute bias (left panel) and MSE (right panel) of the two estimators $\widehat{\gamma}_{k,\alpha,J}$ (red) and $\widehat{\gamma}_{k,J}$ (blue) in the both cases when the estimators are pure (solid line) and contaminated (dashed line), corresponding to Burr distribution with tail index $\gamma = 0.5$ and different values of α, based on 20 samples of size 500

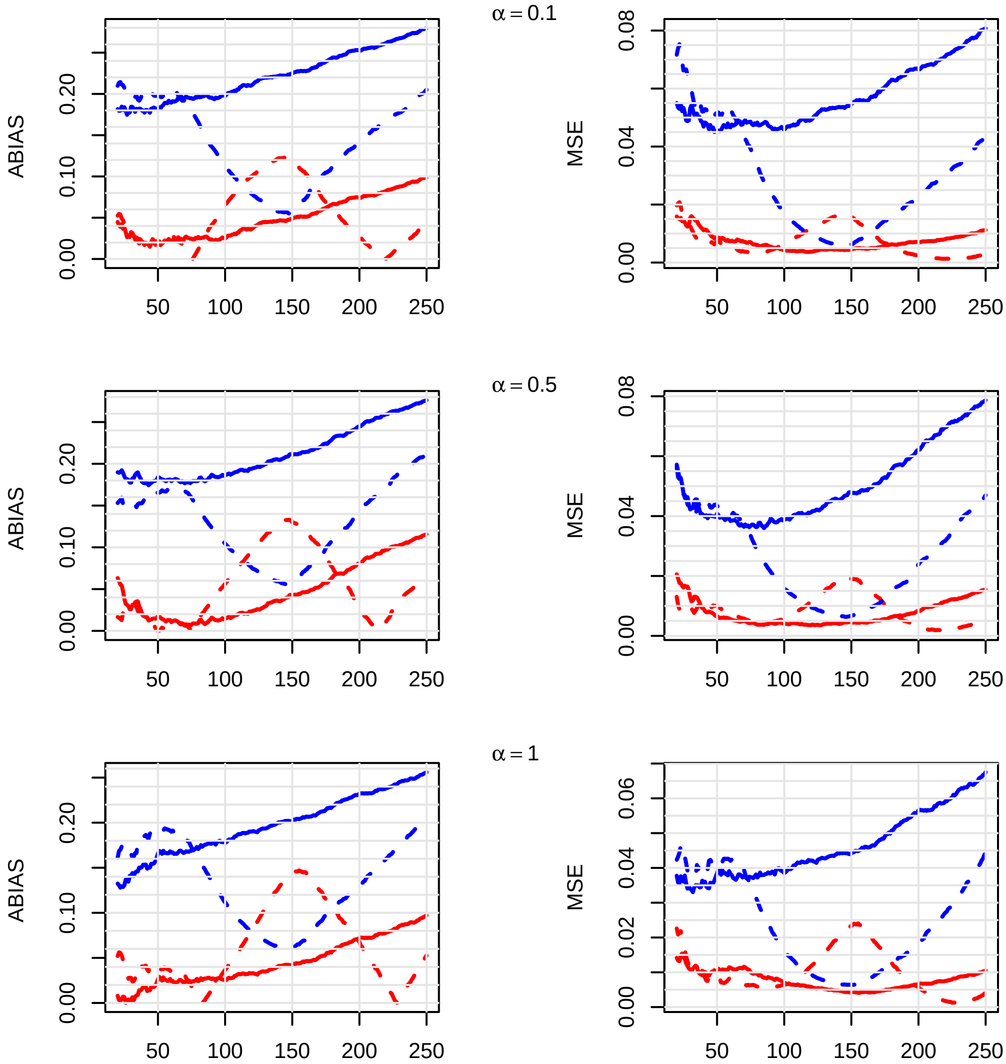

Figure 4.8: Comparaison in terms of absolute bias (left panel) and MSE (right panel) of the two estimators $\widehat{\gamma}_{k,\alpha,J}$ (red) and $\widehat{\gamma}_{k,J}$ (blue) in the both cases when the estimators are pure (solid line) and cntaminated (dashed line), corresponding to Frechet distribution with tail index $\gamma = 0.5$ and different values of α, based on 20 samples of size 500.

where F_n being the empirical cdf pertaining to the sample $X_1,...,X_n$. Recall that $\ell_{\gamma,J}(x)=J\left(x^{-1/\gamma}\right)\ell_\gamma(x)$ and let us decompose I_k into the sum of

$$I_k^{(1)}:=\frac{n}{k}\int_1^\infty J\left(\frac{n}{k}\overline{F}_n(xX_{n-k:n})\right)\frac{d}{d\gamma}\log J\left(x^{-1/\gamma}\right)dF_n(xX_{n-k:n})$$

and

$$I_k^{(2)}:=\frac{n}{k}\int_1^\infty J\left(\frac{n}{k}\overline{F}_n(xX_{n-k:n})\right)\frac{d}{d\gamma}\log\ell_\gamma(x)\,dF_n(xX_{n-k:n}).$$

Next we show that $I_k^{(1)}\xrightarrow{\mathbf{P}}\gamma^{-1}\int_0^1 J(s)\log s\,ds+\gamma^{-1}$, as $n\to\infty$. It is clear that

$$I_k^{(1)}=\int_1^\infty\frac{d}{d\gamma}\log J\left(x^{-1/\gamma}\right)d\varphi\left(\frac{n}{k}\overline{F}_n(xX_{n-k:n})\right),$$

where $\varphi(s):=\int_s^1 J(s)\,dt$. Observe that

$$\frac{d}{d\gamma}\log J\left(x^{-1/\gamma}\right)=\frac{x^{-1/\gamma}}{\gamma}\frac{J'\left(x^{-1/\gamma}\right)}{J\left(x^{-1/\gamma}\right)}\log x=-\frac{1}{\gamma}\mathscr{L}\left(x^{-1/\gamma}\right),$$

where $\mathscr{L}(s):=\left(J'(s)/J(s)\right)s\log s$, therefore

$$I_k^{(1)}=-\frac{1}{\gamma}\int_1^\infty\mathscr{L}\left(x^{-1/\gamma}\right)d\varphi\left(\frac{n}{k}\overline{F}_n(xX_{n-k:n})\right).$$

By assumption (4.8) we have $\sup_{0<s<1}|\mathscr{L}(s)|<\infty$, this implies that $\mathscr{L}(0+)<\infty$. Then using an integration by parts, yields

$$\gamma I_k^{(1)}=-\mathscr{L}(0+)+\int_1^\infty\varphi\left(\frac{n}{k}\overline{F}_n(xX_{n-k:n})\right)d\mathscr{L}\left(x^{-1/\gamma}\right).$$

By a change of variables, we write

$$\gamma I_k^{(1)}=-\mathscr{L}(0+)-\int_0^1\varphi\left(\frac{n}{k}\overline{F}_n\left(s^{-\gamma}X_{n-k:n}\right)\right)\left(\frac{d}{ds}\mathscr{L}(s)\right)ds.$$

Observe that

$$\begin{aligned}\varphi\left(\frac{n}{k}\overline{F}_n(s^{-\gamma}X_{n-k:n})\right)&=\left\{\varphi\left(\frac{n}{k}\overline{F}_n(s^{-\gamma}X_{n-k:n})\right)-\varphi(s)\right\}+\varphi(s)\\&=:\ \omega_n(s)+\varphi(s).\end{aligned}$$

Note that $d\varphi(s)/ds=-J(s)$, then by the mean value theorem, yields

$$\omega_n(s)=-\left\{\frac{n}{k}\overline{F}_n\left(s^{-\gamma}X_{n-k:n}\right)-s\right\}J(\xi_n(s)),$$

where $\xi_n(s)$ is between s and $\frac{n}{k}\overline{F}_n(s^{-\gamma}X_{n-k:n})$. The function J being bounded, then there exists a constant $M_1>0$, such that $\omega_n(s)=O_{\mathbf{P}}(1)\left(\frac{n}{k}\overline{F}_n(s^{-\gamma}X_{n-k:n})-s\right)$. From weak approximation (4.23), we infer that for a sufficiently small $\epsilon>0$, $\frac{n}{k}\overline{F}_n(s^{-\gamma}X_{n-k:n})-s=s^{\epsilon}O_{\mathbf{P}}\left(k^{-1}\right)$, uniformly over $0<s<1$, thus $\omega_n(s)=o_{\mathbf{P}}(s^{\epsilon})$, leading to

$$\varphi\left(\frac{n}{k}\overline{F}_n\left(s^{-\gamma}X_{n-k:n}\right)\right)=\varphi(s)+s^{\epsilon}o_{\mathbf{P}}\left(k^{-1}\right),$$

uniformly over $0<s<1$. Then we showed that

$$\gamma I_k^{(1)}=-\mathscr{L}(0+)-\int_0^1\varphi(s)\frac{d}{ds}\mathscr{L}(s)ds+O_{\mathbf{P}}\left(k^{-1}\right)\int_0^1 s^{\epsilon}\left|\mathscr{L}'(s)\right|ds,$$

where $\mathscr{L}(0+):=\lim_{\epsilon\downarrow 0}\mathscr{L}(\epsilon)$. From assumption (4.8), we have $\int_0^1\left|\mathscr{L}'(s)\right|ds<\infty$, thus

$$\gamma I_k^{(1)}=-\mathscr{L}(0+)-\int_0^1\varphi(s)d\mathscr{L}(s)+O_{\mathbf{P}}\left(k^{-1}\right).$$

By using an integration by parts, we get

$$\gamma I_k^{(1)}=-\int_0^1 J(s)\mathscr{L}(s)ds+o_{\mathbf{P}}(1)=:r+O_{\mathbf{P}}\left(k^{-1}\right).$$

We have $\mathscr{L}(s)=\left(J'(s)/J(s)\right)s\log s$, then $r=-\int_0^1 J'(s)s\log s ds$, which by an integration by parts becomes $\int_0^1 J(s)\log s ds+1$, thereby

$$I_k^{(1)}=\frac{1}{\gamma}\int_0^1 J(s)\log s ds+\frac{1}{\gamma}+O_{\mathbf{P}}\left(k^{-1}\right). \tag{4.16}$$

Let us now consider the term $I_k^{(2)}$, which equals

$$\begin{aligned}
&\int_1^{\infty}J\left(\frac{\overline{F}_n(xX_{n-k:n})}{\overline{F}_n(xX_{n-k:n})}\right)\left\{-\frac{1}{\gamma^2}\left(\gamma-\log x\right)\right\}d\frac{F_n(xX_{n-k:n})}{\overline{F}_n(xX_{n-k:n})}\\
&=-\gamma^{-1}\int_1^{\infty}J\left(\frac{\overline{F}_n(xX_{n-k:n})}{\overline{F}_n(xX_{n-k:n})}\right)d\frac{F_n(xX_{n-k:n})}{\overline{F}_n(xX_{n-k:n})}\\
&+\gamma^{-2}\int_1^{\infty}J\left(\frac{\overline{F}_n(xX_{n-k:n})}{\overline{F}_n(xX_{n-k:n})}\right)\log x d\frac{F_n(xX_{n-k:n})}{\overline{F}_n(xX_{n-k:n})}.
\end{aligned}$$

Observe that

$$-\gamma^{-1}\int_1^{\infty}J\left(\frac{\overline{F}_n(xX_{n-k:n})}{\overline{F}_n(xX_{n-k:n})}\right)d\frac{F_n(xX_{n-k:n})}{\overline{F}_n(xX_{n-k:n})}=-\gamma^{-1}\int_0^1 J(s)ds=-\gamma^{-1},$$

it follows that

$$I_k^{(2)} = -\gamma^{-1} + \gamma^{-2} \frac{1}{k} \sum_{i=1}^{k} J\left(\frac{i}{k}\right) \log \frac{X_{n-i+1:n}}{X_{n-k:n}}. \tag{4.17}$$

Combining two equations (4.16) and (4.17) together yields

$$\begin{aligned} I_k &= I_k^{(1)} + I_k^{(2)} \\ &= \gamma^{-1} \int_0^1 J(s) \log s ds + \gamma^{-2} \frac{1}{k} \sum_{i=1}^{k} J\left(\frac{i}{k}\right) \log \frac{X_{n-i+1:n}}{X_{n-k:n}} + O_{\mathbf{P}}\left(k^{-1}\right), \end{aligned}$$

as $n \to \infty$, as sought. ∎

Lemma 4.2. *Let $X_1, ..., X_n$ be a sample of iid random variables from a distribution function satisfying* (4.9)*. Let J be a continuous function fulfilling assumptions* $[A1]-[A2]$ *and k be a sequence of integers such that $k \to \infty$, $k/n \to 0$ and $\sqrt{k}\, a(n/k) \to \lambda \in \mathbb{R}$. Then, in the probability space $(\Omega, \mathscr{A}, \mathbf{P})$ there exists a standard Wiener process $\{W(x),\ x \geq 0\}$, such that*

$$\begin{aligned} &-\left(1+\frac{1}{\alpha}\right)^{-1} \sqrt{k}\, \pi_k^{(1)}(\gamma_0) \\ &= \int_1^\infty \left(W\left(x^{-1/\gamma_0}\right) - x^{-1/\gamma_0} W(1)\right) J\left(x^{-1/\gamma_0}\right) d\Psi_{\gamma_0,\alpha}^{(1)}(x) + \lambda B_{\gamma_0}^{(1)} + o_{\mathbf{P}}(1), \end{aligned} \tag{4.18}$$

and

$$\begin{aligned} &-\left(1+\frac{1}{\alpha}\right)^{-1} \sqrt{k}\left(\pi_k^{(2)}(\gamma_0) - \eta_{\gamma_0}\right) \\ &= \int_1^\infty \left(W\left(x^{-1/\gamma_0}\right) - x^{-1/\gamma_0} W(1)\right) J\left(x^{-1/\gamma_0}\right) d\Psi_{\gamma_0,\alpha}^{(2)}(x) + \lambda B_{\gamma_0}^{(2)} + o_{\mathbf{P}}(1), \end{aligned} \tag{4.19}$$

as $n \to \infty$, where η_{γ_0} is as in (4.13)*,*

$$B_{\gamma_0}^{(m)} := \int_1^\infty x^{-1/\gamma_0} \frac{x^{\rho/\gamma_0} - 1}{\rho\gamma_0} J\left(x^{-1/\gamma_0}\right) d\Psi_{\gamma_0,\alpha}^{(m)}(x),\ m = 1, 2. \tag{4.20}$$

and $\Psi_{\gamma_0,\alpha}^{(m)}(x) := d^m \ell_{\gamma_0,J}^{\alpha}(x)/d\gamma^m$ denotes the m-th derivative of function $\gamma \to \ell_{\gamma,J}^{\alpha}$ in $\gamma = \gamma_0$. Moreover,

$$\pi_k^{(1)}(\gamma_0) \xrightarrow{\mathbf{P}} 0 \text{ and } \pi_k^{(2)}(\gamma_0) \xrightarrow{\mathbf{P}} \eta_{\gamma_0}, \text{ as } n \to \infty. \tag{4.21}$$

Proof. Recall that

$$\pi_k^{(m)}(\gamma_0) := \int_1^\infty \Psi_{\gamma_0,\alpha+1}^{(m)}(x)\,dx - \left(1+\frac{1}{\alpha}\right) A_k^{(m)}(\gamma_0)$$

and

$$A_k^{(m)}(\gamma_0) := \frac{1}{k}\sum_{i=1}^{k} J\left(\frac{i}{k}\right)\Psi_{\gamma_0,\alpha}^{(m)}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right), \quad m=1,2. \tag{4.22}$$

Let us rewrite $A_k^{(1)}(\gamma_0)$ into

$$\int_1^\infty J\left(\frac{\overline{F}_n(xX_{n-k:n})}{\overline{F}_n(X_{n-k:n})}\right)\Psi_{\gamma_0,\alpha}^{(1)}(x)\,d\frac{F_n(xX_{n-k:n})}{\overline{F}_n(X_{n-k:n})},$$

where F_n being the empirical cdf pertaining to the sample $X_1,...,X_n$. It is clear that

$$A_k^{(m)}(\gamma_0) = -\int_1^\infty \Psi_{\gamma_0,\alpha}^{(1)}(x)\,d\varphi\left(\frac{\overline{F}_n(xX_{n-k:n})}{\overline{F}_n(X_{n-k:n})}\right),$$

where $\varphi(v) := \int_0^v J(t)\,dt$. From Lemma 4.6 we have $\Psi_{\gamma_0,\alpha}^{(1)}(1) = \Psi_{\gamma_0,\alpha}^{(1)}(\infty) = 0$, then using an integration by parts yields

$$A_k^{(m)}(\gamma_0) = \int_1^\infty \varphi\left(\frac{\overline{F}_n(xX_{n-k:n})}{\overline{F}_n(X_{n-k:n})}\right) d\Psi_{\gamma_0,\alpha}^{(1)}(x).$$

By Lemma 4.3, we have

$$\int_1^\infty \frac{d}{d\gamma}\ell_{\gamma_0,J}^{\alpha+1}(x)\,dx = \left(1+\frac{1}{\alpha}\right)\int_1^\infty \ell_{\gamma_0,J}(x)\frac{d}{d\gamma}\ell_{\gamma_0,J}^{\alpha}(x)\,dx.$$

Recall that $\ell_{\gamma_0,J}(x) = J\left(x^{-1/\gamma_0}\right)\ell_{\gamma_0}(x) = \gamma_0^{-1}x^{-1/\gamma_0-1}J\left(x^{-1/\gamma_0}\right)$, it follows that

$$\int_1^\infty \frac{d}{d\gamma}\ell_{\gamma_0,J}^{\alpha+1}(x)\,dx = -\left(1+\frac{1}{\alpha}\right)\int_1^\infty J\left(x^{-1/\gamma_0}\right)\frac{d}{d\gamma}\ell_{\gamma_0,J}^{\alpha}(x)\,dx^{-1/\gamma_0},$$

which equals $-\left(1+\frac{1}{\alpha}\right)\int_1^\infty \Psi_{\gamma_0,\alpha}^{(1)}(x)\,d\varphi\left(x^{-1/\gamma_0}\right)$. By using an integration by parts yields

$$\int_1^\infty \frac{d}{d\gamma}\ell_{\gamma_0,J}^{\alpha+1}(x)\,dx = \left(1+\frac{1}{\alpha}\right)\int_1^\infty \varphi\left(x^{-1/\gamma_0}\right)d\Psi_{\gamma_0,\alpha}^{(1)}(x).$$

Let us now write

$$-\left(1+\frac{1}{\alpha}\right)^{-1}\pi_k^{(1)}(\gamma_0) = \int_1^\infty \left\{\varphi\left(\frac{n}{k}\overline{F}_n(xX_{n-k:n})\right) - \varphi\left(x^{-1/\gamma}\right)\right\} d\Psi_{\gamma_0,\alpha}^{(1)}(x).$$

Applying Taylor's expansion, we may decompose the later quantity into the sum of

$$T_k^{(1)} := \int_1^\infty \left(\frac{n}{k}\overline{F}_n(xX_{n-k:n}) - x^{-1/\gamma_0}\right) J\left(x^{-1/\gamma_0}\right) d\Psi^{(1)}_{\gamma_0,\alpha}(x)$$

and

$$R_k^{(1)} := \frac{1}{2}\int_1^\infty \left(\frac{n}{k}\overline{F}_n(xX_{n-k:n}) - x^{-1/\gamma_0}\right)^2 J'(d_n(x))\, d\Psi^{(1)}_{\gamma_0,\alpha}(x),$$

where $d_n(x)$ is between x^{-1/γ_0} and $\frac{n}{k}\overline{F}_n(xX_{n-k:n})$. It is clear that

$$\sqrt{k}\,T_k^{(1)} = \int_1^\infty D_k(x) J\left(x^{-1/\gamma_0}\right) d\Psi^{(1)}_{\gamma_0,\alpha}(x),$$

where $D_k(x) := \sqrt{k}\left(\frac{n}{k}\overline{F}_n(xX_{n-k:n}) - x^{-1/\gamma_0}\right), x \geq 1$. In Proposition 4.1 we showed that, on the probability space $(\Omega, \mathscr{A}, \mathbf{P})$, there exists a standard Wiener process $\{W(x),\ x \geq 0\}$, such that for $x \geq 1$ and $0 < \epsilon < 1/2$,

$$\sup_{x\geq 1} x^{(1/2-\epsilon)} \left| D_k(x) - \Gamma(x;W) - x^{-1/\gamma_0}\frac{x^{\rho/\gamma_0}-1}{\rho\gamma_0}\sqrt{k}A_0(n/k)\right| \xrightarrow{\mathbf{P}} 0, \tag{4.23}$$

as $n \to \infty$, where $\Gamma(x;W) := W\left(x^{-1/\gamma_0}\right) - x^{-1/\gamma_0}W(1)$ and $A_0(t) \sim a(t)$, as $t \to \infty$. Let us now decompose $\sqrt{k}\,T_k^{(1)}$ into the sum of

$$N^{(1)} := \int_1^\infty \Gamma(x;W) J\left(x^{-1/\gamma_0}\right) d\Psi^{(1)}_{\gamma_0,\alpha}(x),$$

$$\mathscr{B}_k^{(1)} := \sqrt{k}A_0(n/k)\int_1^\infty x^{-1/\gamma_0}\frac{x^{\rho/\gamma_0}-1}{\rho\gamma_0} J\left(x^{-1/\gamma_0}\right) d\Psi^{(1)}_{\gamma_0,\alpha}(x)$$

and

$$\begin{aligned} S_k^{(1)} \;:=\; & \int_1^\infty \left(D_k(x) - \Gamma(x;W) - \int_1^\infty x^{-1/\gamma_0}\frac{x^{\rho/\gamma_0}-1}{\rho\gamma_0}\sqrt{k}A_0(n/k)\right) \\ & \qquad\qquad\qquad\qquad \times J\left(x^{-1/\gamma_0}\right) d\Psi^{(1)}_{\gamma_0,\alpha}(x). \end{aligned}$$

Next we show that $S_k^{(1)} \xrightarrow{\mathbf{P}} 0$ as $n \to \infty$. Applying Gaussian approximation (4.23) yields

$$S_k^{(1)} = o_{\mathbf{P}}(1)\int_1^\infty x^{-(1/2-\epsilon)} J\left(x^{-1/\gamma_0}\right) \left|\frac{d}{dx}\Psi^{(1)}_{\gamma_0,\alpha}(x)\right| dx.$$

From Lemma 4.5, we have $\sup_{x\geq 1}\left|\frac{d}{dx}\Psi^{(1)}_{\gamma_0,\alpha}(x)\right| < \infty$, it follows that

$$S_k^{(1)} = o_{\mathbf{P}}(1)\int_1^\infty x^{-(1/2-\epsilon)} J\left(x^{-1/\gamma_0}\right) dx.$$

The function J is bounded then the last integral is finite and therefore $S_k^{(1)} = o_{\mathbf{P}}(1)$. Let us now consider the term $R_k^{(1)}$. It is clear that there exists a constant $c > 0$, such that

$$\sqrt{k}\left|R_k^{(1)}\right| \leq ck^{-1/2}\int_1^\infty D_k^2(x)\left|\frac{d}{dx}\Psi_{\gamma_0,\alpha}^{(1)}(x)\right|dx.$$

From Lemma 4.7, we have $\sup_{x\geq 1} x^{(1-2\epsilon)}D_k^2(x) = O_{\mathbf{P}}(1)$, therefore

$$\sqrt{k}R^{(1)} = O_{\mathbf{P}}\left(k^{-1/2}\right)\int_1^\infty x^{-1+2\epsilon}\left|\frac{d}{dx}\Psi_{\gamma_0,\alpha}^{(1)}(x)\right|dx.$$

Since $\sup_{x\geq 1} x^{-\epsilon}\left|\frac{d}{dx}\Psi^{(1)}(x)\right| < \infty$, hence $\sqrt{k}R_k^{(1)} = o_{\mathbf{P}}(1)$, as $n \to \infty$. Let us now consider the asymptotic bias

$$\mathscr{B}_k^{(1)} = \sqrt{k}A_0(n/k)\int_1^\infty x^{-1/\gamma_0}\frac{x^{\rho/\gamma_0}-1}{\rho\gamma_0}J\left(x^{-1/\gamma_0}\right)d\Psi_{\gamma_0,\alpha}^{(1)}(x).$$

Since $\sqrt{k}A_0(n/k) \to \lambda < \infty$, then $\mathscr{B}_k^{(1)} = (1+o(1))\lambda B_{\gamma_0}^{(1)}$, as $n \to \infty$, where $B_{\gamma_0}^{(1)}$ is as in (4.20). Since J is bounded, then is easy to verify that the previous integral is finite, therefore $\mathscr{B}_k^{(1)} = \lambda B_{\gamma_0}^{(1)} + o(1)$. In summary, we showed that

$$\begin{aligned}&-\left(1+\frac{1}{\alpha}\right)^{-1}\sqrt{k}\pi_k^{(1)}(\gamma_0)\\&=\int_1^\infty \left(W\left(x^{-1/\gamma_0}\right) - x^{-1/\gamma_0}W(1)\right)J\left(x^{-1/\gamma_0}\right)d\Psi_{\gamma_0,\alpha}^{(1)}(x) + \lambda B_{\gamma_0}^{(1)} + o_{\mathbf{P}}(1),\end{aligned}$$

thus (4.18) holds. Let us now prove assertion (4.19). From Lemma 4.3 we have

$$\int_1^\infty \frac{d^2}{d\gamma^2}\ell_{\gamma_0,J}^{\alpha+1}(x)dx = \left(1+\frac{1}{\alpha}\right)\int_1^\infty \ell_{\gamma_0}(x)\Psi_{\gamma_0,\alpha}^{(2)}(x)dx + \eta_{\gamma_0},$$

it follows that

$$\left(1+\frac{1}{\alpha}\right)^{-1}\left(\pi_k^{(2)}(\gamma_0) - \eta_{\gamma_0}\right) = \int_1^\infty \ell_{\gamma_0}(x)\Psi_{\gamma_0,\alpha}^{(2)}(x)dx - A_k^{(2)}(\gamma_0).$$

Using similar arguments as used in the proof of assertion (4.18), we also show that the right side of the previous equation equals

$$\int_1^\infty \left(W\left(x^{-1/\gamma_0}\right) - x^{-1/\gamma_0}W(1)\right)J\left(x^{-1/\gamma_0}\right)d\Psi_{\gamma_0,\alpha}^{(2)}(x) + \lambda B_{\gamma_0}^{(2)} + o_{\mathbf{P}}(1), \text{ as } n \to \infty,$$

thus (4.19) holds too. To show (4.20), let us first note that

$$\begin{aligned}&\mathbf{E}\left|\int_1^\infty \left(W\left(x^{-1/\gamma_0}\right)-x^{-1/\gamma_0}W(1)\right)J\left(x^{-1/\gamma_0}\right)d\Psi_{\gamma_0,\alpha}^{(m)}(x)\right|\\&\leq \int_1^\infty \left|B\left(x^{-1/\gamma_0}\right)\right|J\left(x^{-1/\gamma_0}\right)\left|\frac{d\Psi_{\gamma_0,\alpha}^{(m)}(x)}{dx}\right|dx,\end{aligned}$$

where $B(s):=W(s)-sW(1)$ is a Brownian bridge. Note that

$$\mathbf{E}\left|B\left(x^{-1/\gamma_0}\right)\right|\leq x^{-1/(2\gamma_0)}\leq 1,\ \text{for } x\geq 1,$$

and J is bounded, then the right-side of the previous inequality is less than or equal to This implies that there exist a constant $0<M<\infty$, such that $M\int_1^\infty \left|d\Psi_{\gamma_0,\alpha}^{(m)}(x)/dx\right|dx$, which from Lemma 4.4 is finite. By the last argument, it is obvious that $\left|B_{\gamma_0}^{(m)}\right|<\infty$, thereby

$$\sqrt{k}\pi_k^{(1)}(\gamma_0)=O_{\mathbf{P}}(1)\sqrt{k}\left(\pi_k^{(2)}(\gamma_0)-\eta_{\gamma_0}\right),$$

therefore $\pi_k^{(1)}(\gamma_0)=o_{\mathbf{P}}(1)=\left(\pi_k^{(2)}(\gamma_0)-\eta_{\gamma_0}\right)$, because $k^{-1}\to 0$, as $n\to\infty$ as sought. ∎

Lemma 4.3. *For any $\alpha>0$, we have*

$$\frac{d}{d\gamma}\ell_{\gamma,J}^{\alpha+1}(x)=\left(1+\frac{1}{\alpha}\right)\ell_{\gamma,J}(x)\frac{d}{d\gamma}\ell_{\gamma,J}^{\alpha}(x)$$

and

$$\frac{d^2}{d\gamma^2}\ell_\gamma^{\alpha+1}(x)=\left(1+\frac{1}{\alpha}\right)\ell_\gamma(x)\frac{d^2}{d\gamma^2}\ell_\gamma^{\alpha}(x)+(1+\alpha)\left(\frac{d}{d\gamma}\ell_\gamma(x)\right)^2\ell_\gamma^{\alpha-1}(x).$$

Proof. The proof of first equation is obvious. Indeed

$$\frac{d}{d\gamma}\ell_{\gamma,J}^{\alpha+1}(x)=(1+\alpha)\ell_{\gamma,J}^{\alpha}(x)\frac{d}{d\gamma}\ell_{\gamma,J}(x)=\left(1+\frac{1}{\alpha}\right)\ell_{\gamma,J}(x)\frac{d}{d\gamma}\ell_{\gamma,J}^{\alpha}(x).$$

For the second, we write

$$\begin{aligned}\frac{d^2}{d\gamma^2}\ell_\gamma^{\alpha+1}(x)&=\frac{d}{d\gamma}\left[(1+\alpha)\frac{d}{d\gamma}\ell_\gamma(x)\ell_\gamma^{\alpha}(x)\right]\\&=(1+\alpha)\left[\frac{d^2}{d\gamma^2}\ell_\gamma(x)\ell_\gamma^{\alpha}(x)+\alpha\left(\frac{d}{d\gamma}\ell_\gamma(x)\right)^2\ell_\gamma^{\alpha-1}(x)\right].\end{aligned}$$

On the other hand

$$\ell_\gamma(x)\frac{d^2}{d\gamma^2}\ell_\gamma^\alpha(x)=\alpha\ell_\gamma(x)\left[\frac{d^2}{d\gamma^2}\ell_\gamma(x)\ell_\gamma^{\alpha-1}(x)+(\alpha-1)\left(\frac{d}{d\gamma}\ell_\gamma(x)\right)^2\ell_\gamma^{\alpha-2}(x)\right],$$

it follows that

$$\begin{aligned}&\left(1+\frac{1}{\alpha}\right)\ell_\gamma(x)\frac{d^2}{d\gamma^2}\ell_\gamma^\alpha(x)\\&=(\alpha+1)\left[\ell_\gamma(x)\frac{d^2}{d\gamma^2}\ell_\gamma(x)\ell_\gamma^{\alpha-1}(x)+(\alpha-1)\left(\frac{d}{d\gamma}\ell_\gamma(x)\right)^2\ell_\gamma^{\alpha-1}(x)\right].\end{aligned}$$

Thus

$$\frac{d^2}{d\gamma^2}\ell_\gamma^{\alpha+1}(x)=\left(1+\frac{1}{\alpha}\right)\ell_\gamma(x)\frac{d^2}{d\gamma^2}\ell_\gamma^\alpha(x)+(1+\alpha)\left(\frac{d}{d\gamma}\ell_\gamma(x)\right)^2\ell_\gamma^{\alpha-1}(x).$$

■

Lemma 4.4. *There exists a constant* $0<M_\gamma<\infty$, *such that for every* $z\ge 1$

$$\left|\frac{d^m}{d\gamma^m}\ell_{\gamma,J}^\alpha(z)\right|\le M_\gamma z^{-\alpha(1+1/\gamma)}\sum_{j=0}^m\log^j z,\ \alpha>0.$$

Moreover

$$\int_1^\infty\frac{d^m}{d\gamma^m}\ell_{\gamma,J}^\alpha(z)dz<\infty,\ \sup_{z\ge1}\left|\frac{d}{dz}\frac{d^m}{d\gamma^m}\ell_{\gamma,J}^\alpha(z)\right|<\infty,$$

and $\int_1^\infty\left|\frac{d}{dz}\frac{d^m}{d\gamma^m}\ell_{\gamma,J}^\alpha(z)\right|dz<\infty$.

Proof. Observe that the first two derivatives of ℓ_γ^α are

$$\frac{d}{d\gamma}\ell_\gamma^\alpha(z)=\alpha\gamma^{-3-\alpha}z^{-\alpha(1+1/\gamma)}\left(\log z-\gamma\right),$$

and

$$\frac{d^2}{d\gamma^2}\ell_\gamma^\alpha(z)=\alpha\gamma^{-5-\alpha}z^{-\alpha(1+1/\gamma)}\left((2+\alpha)\gamma^2-3\gamma\log z-\alpha\log^2 z\right).$$

Thus we may easily shown that or every $z\ge1$

$$\frac{d^m}{d\gamma^m}\ell_\gamma^\alpha(z)=z^{-\alpha(1+1/\gamma)}\sum_{j=0}^m a_{j,\gamma,\alpha}\log^j z,\ \alpha>0.$$

for some constants $0 < a_{j,\gamma,\alpha} < \infty$. We have $\ell_{\gamma,J}(z) = J\left(x^{-1/\gamma}\right)\ell_\gamma(z)$, so $\frac{d^m}{d\gamma^m}\ell^\alpha_{\gamma,J}(z)$ will be linear combinations of $\left\{z^{-\alpha(1+1/\gamma)-\beta_m}\log^m z\right\}_{0\le m\le 3}$ and derivatives $\left\{\frac{d^m}{dz^m}J\left(x^{-1/\gamma}\right)\right\}_{0\le m\le 3}$, for some sequence of constants $\beta_m > 0$.

In other terms

$$\frac{d^m}{d\gamma^m}\ell^\alpha_{\gamma,J}(z) = z^{-\alpha(1+1/\gamma)}\sum_{j=0}^m b_{j,\gamma,\alpha,J}\log^j z,\ \alpha > 0, \tag{4.24}$$

for some constants $0 < b_{j,\gamma,\alpha} < \infty$. Since $z^{-\beta_m} \le 1$, for every $z \ge 1$ and J and their four three derivatives are bounded, then using equation (4.24), we show that there exists a constant $0 < M_\gamma < \infty$ such that for every $z \ge 1$

$$\left|\frac{d^m}{d\gamma^m}\ell^\alpha_{\gamma,J}(z)\right| \le Mz^{-\alpha(1+1/\gamma)}\sum_{j=0}^m \log^j z,\ m = 1,2,3,4,$$

It is obvious that $\sup_{z\ge1} z^{-\alpha(1+1/\gamma)+\epsilon}\log z^j$, for every sufficiently small $0 < \epsilon < \alpha\left(1+1/\gamma\right)$, which implies that

$$\left|\int_1^\infty \frac{d^m}{d\gamma^m}\ell^\alpha_{\gamma,J}(z)dz\right| \le M_\gamma\int_1^\infty z^{-\epsilon}dz < \infty.$$

Using equation (4.24) and an elementary calculus yields

$$\frac{d}{dz}\frac{d^m}{d\gamma^m}\ell^\alpha_{\gamma,J}(z) = \gamma^{-1}z^{-\alpha(1+1/\gamma)-1}\sum_{j=1}^m b_{j,\gamma,\alpha}\left(\gamma j\log^{j-1}z - \alpha\left(1+\gamma\right)\log^j z\right).$$

Since $\sup_{z\ge1} z^{-\alpha(1+1/\gamma)-1}\log^j z$ is finite, for $j = 0,\dots,m$, then $\sup_{z\ge1}\left|\frac{d}{dz}\frac{d^m}{d\gamma^m}\ell^\alpha_{\gamma,J}(z)\right|$ is finite as well. ■

Lemma 4.5. *We have*

$$\frac{d^3}{d\gamma^3}\ell^{\alpha+1}_{\gamma,J}(z) = \left(1+\frac{1}{\alpha}\right)\ell_{\gamma,J}(z)\frac{d^3}{d\gamma^3}\ell^\alpha_{\gamma,J}(z) + g_\gamma(z),\ \alpha > 0,$$

where $z \to g_\gamma(z)$ is a function such that $\int_1^\infty g_\gamma(z)dz < \infty$.

Proof. It is easy to check that

$$\frac{d^3}{d\gamma^3}\ell^{\alpha+1}_{\gamma,J}(z) = \left(1+\frac{1}{\alpha}\right)\ell_{\gamma,J}(z)\frac{d^3}{d\gamma^3}\ell^\alpha_{\gamma,J}(z) + g_\gamma(z),$$

where

$$g_\gamma(z) := \left(1+\frac{1}{\alpha}\right)\frac{d}{d\gamma}\ell_{\gamma,J}(z)\frac{d^2}{d\gamma^2}\ell^{\alpha}_{\gamma,J}(z)$$
$$+(1+\alpha)\ell^{\alpha-2}_{\gamma,J}(z)\frac{d}{d\gamma}\ell_{\gamma,J}(z)\left\{2\ell_{\gamma,J}(z)\tfrac{d^2}{d\gamma^2}\ell^{\alpha}_{\gamma,J}(z)+(\alpha-1)\left(\frac{d}{d\gamma}\ell_{\gamma,J}(z)\right)^2\right\}.$$

Using Lemma 4.6, we show that indeed $\int_1^\infty g_\gamma(z)dz<\infty$ that we omit further details. ■

Lemma 4.6. *Given a consistent estimator* $\widehat{\gamma}$ *of* γ_0, *we have* $\pi_k^{(3)}(\widehat{\gamma}) = O_{\mathbf{P}}(1)$.

Proof. Let us first show that $\pi_k^{(3)}(\gamma_0) = O_{\mathbf{P}}(1)$. Recall that

$$\pi_k^{(3)}(\gamma_0) = \int_1^\infty \frac{d^3}{d\gamma^3}\ell^{\alpha+1}_{\gamma_0,J}(x)dx - \left(1+\frac{1}{\alpha}\right)\frac{1}{k}\sum_{i=1}^k J\left(\frac{i}{k}\right)\frac{d^3}{d\gamma^3}\ell^{\alpha}_{\gamma_0,J}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right).$$

Making use of Lemma 4.5, we may rewrite $\pi_k^{(3)}(\gamma_0)$ into the sum of $\int_1^\infty g_{\gamma_0}(x)dx$ and $\left(1+\frac{1}{\alpha}\right)\mathbb{A}^{(3)}(\gamma_0)$, where

$$\mathbb{A}^{(3)}(\gamma_0) := \int_1^\infty \frac{d}{d\gamma}\ell_{\gamma_0,J}(x)\frac{d^3}{d\gamma^3}\ell^{\alpha}_{\gamma_0,J}(x)dx - \frac{1}{k}\sum_{i=1}^k J\left(\frac{i}{k}\right)\frac{d^3}{d\gamma^3}\ell^{\alpha}_{\gamma_0,J}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right).$$

By similar arguments as those used in the proof of Lemma 4.2, we also show that $\mathbb{A}^{(3)}(\gamma_0) \xrightarrow{\mathbf{P}} 0$, that we omit further details. So $\pi_k^{(3)}(\gamma_0) \xrightarrow{\mathbf{P}} \int_1^\infty g_{\gamma_0}(x)dx<\infty$, thus $\pi_k^{(3)}(\gamma_0) = O_{\mathbf{P}}(1)$. Next we prove that $\pi_k^{(3)}(\widehat{\gamma}_0)-\pi_k^{(3)}(\gamma_0) = o_{\mathbf{P}}(1)$, as $n\to\infty$. Indeed, let us write

$$\pi_k^{(3)}(\widehat{\gamma}) - \pi_k^{(3)}(\gamma_0)$$
$$= \int_1^\infty \left\{\frac{d^3}{d\gamma^3}\ell^{\alpha+1}_{\widehat{\gamma},J}(x) - \frac{d^3}{d\gamma^3}\ell^{\alpha+1}_{\gamma_0,J}(x)\right\}dx$$
$$-\left(1+\frac{1}{\alpha}\right)\frac{1}{k}\sum_{i=1}^k J\left(\frac{i}{k}\right)\left\{\frac{d^3}{d\gamma^3}\ell^{\alpha}_{\widehat{\gamma},J}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right) - \frac{d^3}{d\gamma^3}\ell^{\alpha}_{\gamma_0,J}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)\right\}.$$

Using the mean value theorem, yields

$$\int_1^\infty \left\{\frac{d^3}{d\gamma^3}\ell^{\alpha+1}_{\widehat{\gamma},J}(x) - \frac{d^3}{d\gamma^3}\ell^{\alpha+1}_{\gamma_0,J}(x)\right\}dx = (\widehat{\gamma}-\gamma_0)\int_1^\infty \frac{d^4}{d\gamma^4}\ell^{\alpha+1}_{\widehat{\gamma}_0^*,J}(x)dx,$$

where $\widehat{\gamma}_0^*$ is between $\widehat{\gamma}$ and γ. Recall that $\widehat{\gamma} \xrightarrow{\mathbf{P}} \gamma_0$ and by Lemma 4.4 we have $\int_1^\infty \frac{d^4}{d\gamma^4} \ell^{\alpha+1}_{\widehat{\gamma}_0^*,J}(x)\,dx = O_{\mathbf{P}}(1)$, thus

$$\int_1^\infty \left\{ \frac{d^3}{d\gamma^3} \ell^{\alpha+1}_{\widehat{\gamma},J}(x) - \frac{d^3}{d\gamma^3} \ell^{\alpha+1}_{\gamma_0,J}(x) \right\} dx = o_{\mathbf{P}}(1).$$

On the other hand

$$\begin{aligned} &\frac{1}{k}\sum_{i=1}^{k} J\left(\frac{i}{k}\right)\left\{\frac{d^3}{d\gamma^3}\ell^{\alpha}_{\widehat{\gamma},J}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right) - \frac{d^3}{d\gamma^3}\ell^{\alpha}_{\gamma_0,J}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)\right\} \\ = &\int_1^\infty J\left(\frac{\overline{F}_n(xX_{n-k:n})}{\overline{F}_n(X_{n-k:n})}\right)\left\{\frac{d^3}{d\gamma^3}\ell^{\alpha}_{\widehat{\gamma},J}(x) - \frac{d^3}{d\gamma^3}\ell^{\alpha}_{\gamma_0,J}(x)\right\} d\frac{F_n(xX_{n-k:n})}{\overline{F}_n(X_{n-k:n})}. \end{aligned}$$

Once again making use of the mean value theorem, we write

$$\left(\widehat{\gamma} - \gamma_0\right)\int_1^\infty J\left(\frac{\overline{F}_n(xX_{n-k:n})}{\overline{F}_n(X_{n-k:n})}\right)\frac{d^4}{d\gamma^4}\ell^{\alpha}_{\overline{\gamma}_0^*,J}(x)\, d\frac{F_n(xX_{n-k:n})}{\overline{F}_n(X_{n-k:n})},$$

where $\overline{\gamma}_0^*$ is between $\widehat{\gamma}$ and γ_0. Using similar arguments as above, we show that the previous quantity equals $o_{\mathbf{P}}(1)\int_0^1 J(s)\,ds = o_{\mathbf{P}}(1)$. Thus we showed that

$$\pi_k^{(3)}\left(\widehat{\gamma}\right) - \pi_k^{(3)}\left(\gamma_0\right) = o_{\mathbf{P}}(1),$$

leading to $\pi_k^{(3)}\left(\widehat{\gamma}\right) = O_{\mathbf{P}}(1)$ as well. ■

Lemma 4.7. *For sufficiently small* $0 < \epsilon < 1/2$, *we have*

$$\sup_{x\geq 1} x^{1/2-\epsilon}\left|D_k(x)\right| = O_{\mathbf{P}}(1), \text{ as } n \to \infty.$$

Proof. Let $0 < \epsilon < 1/2$ be sufficiently small so that $0 < \delta := \gamma_0(1/2 - \epsilon) < 1/2$ and rewrite $D_k(x)$ into the sum of $\Gamma(x;W)$ and

$$\begin{aligned} &\left[D_k(x) - \left\{W\left(x^{-1/\gamma_0}\right) - x^{-1/\gamma_0}W(1)\right\} - x^{-1/\gamma_0}\frac{x^{\rho/\gamma_0}-1}{\rho\gamma_0}\sqrt{k}A_0(n/k)\right] \\ &\qquad + \left[x^{-1/\gamma_0}\frac{x^{\rho/\gamma_0}-1}{\rho\gamma_0}\sqrt{k}A_0(n/k)\right] \\ &=: V_{k,1}(x) + V_{k,2}(x). \end{aligned}$$

From weak approximation (4.23), we have $\sup_{x\geq 1} x^{(1/2-\epsilon)} \left|V_{k,1}(x)\right| = o_{\mathbf{P}}(1)$. On the other hand $\sqrt{k}A_0(n/k) \to \lambda$ as $n \to \infty$, then $\sup_{x\geq 1} x^{(1/2-\epsilon)} \left|V_{k,2}(x)\right| = O_{\mathbf{P}}(1)$, therefore

$$\sup_{x\geq 1} x^{(1/2-\epsilon)} \left|D_k(x)\right| \leq \sup_{x\geq 1} x^{(1/2-\epsilon)} \left|W\left(x^{-1/\gamma_0}\right) - x^{-1/\gamma_0} W(1)\right| + O_{\mathbf{P}}(1).$$

Note that $W(1) = O_{\mathbf{P}}(1)$, it follows that

$$\sup_{x\geq 1} x^{1/2-\epsilon} \left|\Gamma(x;W)\right| \leq \sup_{0<s\leq 1} s^{-\delta} \left|W(s)\right| + O_{\mathbf{P}}(1).$$

In view of Lemma 3.2 in [**?**], we infer that $\sup_{0<s\leq 1} s^{-\delta} \left|W(s)\right|$ is stochastically bounded which entails

$$\sup_{x\geq 1} x^{(1/2-\epsilon)} \left|\Gamma(x;W)\right| = O_{\mathbf{P}}(1) = \sup_{x\geq 1} x^{(1/2-\epsilon)} \left|D_k(x)\right|.$$

■

Proposition 4.1. *On the probability space* $(\Omega, \mathscr{A}, \mathbf{P})$ *there exists a standard Wiener process* $\{W(x),\ x \geq 0\}$, *such that for* $x \geq 1$ *and* $0 < \epsilon < 1/2$,

$$\sup_{x\geq 1} x^{(1/2-\epsilon)} \left|D_k(x) - \Gamma(x;W) - x^{-1/\gamma_0} \frac{x^{\rho/\gamma_0} - 1}{\rho\gamma_0} \sqrt{k} A_0(n/k)\right| \xrightarrow{\mathbf{P}} 0, \tag{4.25}$$

as $n \to \infty$, *where* $\Gamma(x;W) := W\left(x^{-1/\gamma_0}\right) - x^{-1/\gamma_0} W(1)$ *and* $A_0(t) \sim a(t)$, *as* $t \to \infty$.

Proof. Let us set $\widetilde{D}_k(z) := \sqrt{k}\left(\frac{n}{k}\overline{F}_n(za_k) - z^{-1/\gamma}\right)$, for $z \geq 1$, where $a_k := U(n/k)$ and let us decompose $D_k(x)$ into the sum of

$$E_{n1}(x) := \widetilde{D}_k\left(x \frac{X_{n-k:n}}{a_k}\right) \text{ and } E_{n2}(x) := x^{-1/\gamma} \sqrt{k}\left(\left(\frac{X_{n-k:n}}{a_k}\right)^{-1/\gamma} - 1\right).$$

From Theorems 5.1.4 page 161 in [24], we have, on the probability space $(\Omega, \mathscr{A}, \mathbf{P})$ there exists a standard Wiener process $\{W(z),\ z \geq 0\}$, such that for $x \geq 1$ and $0 < \epsilon < 1/2$,

$$\sup_{z\geq 1} z^{(1/2-\epsilon)} \left|\widetilde{D}_k(z) - W\left(z^{-1/\gamma_0}\right) - z^{-1/\gamma_0} \frac{z^{\rho/\gamma_0} - 1}{\rho\gamma_0} \sqrt{k} A_0(n/k)\right| \xrightarrow{\mathbf{P}} 0, \tag{4.26}$$

as $n \to \infty$, where $A_0(t) \sim a(t)$, as $t \to \infty$. It is clear that

$$\sup_{x \geq 1} x^{(1/2-\epsilon)} \Big| E_{n1}(z_n(x)) - W\left((z_n(x))^{-1/\gamma_0}\right) - (z_n(x))^{-1/\gamma_0} \frac{(z_n(x))^{\rho/\gamma_0} - 1}{\rho\gamma_0} \sqrt{k} A_0(n/k) \Big|,$$

$$= \left(\frac{a_k}{X_{n-k:n}}\right)^{(1/2-\epsilon)} \sup_{z_n \geq \frac{X_{n-k:n}}{a_k}} (z_n(x))^{(1/2-\epsilon)} \Big| \widetilde{D}_k(z_n(x)) - W\left((z_n(x))^{-1/\gamma_0}\right) - (z_n(x))^{-1/\gamma_0} \frac{(z_n(x))^{\rho/\gamma_0} - 1}{\rho\gamma_0} \sqrt{k} A_0(n/k) \Big|,$$

where $z_n(x) := x \frac{X_{n-k:n}}{a_k}$. We have $X_{n-k:n}/a_k \xrightarrow{\mathbf{P}} 1$, this means that the probability of $A_{n,\epsilon} := \{|X_{n-k:n}/a_k - 1| < \epsilon\}$ tends to 1 as $n \to \infty$, for any small $\epsilon > 0$. Then in the set $A_{n,\epsilon}$ the right-side of the previous equation is less than or equal to

$$\sup_{z_n(x) \geq 1-\epsilon} (z_n(x))^{(1/2-\epsilon)} \Big| \widetilde{D}_k(z_n(x)) - W\left((z_n(x))^{-1/\gamma_0}\right) - (z_n(x))^{-1/\gamma_0} \frac{(z_n(x))^{\rho/\gamma_0} - 1}{\rho\gamma_0} \sqrt{k} A_0(n/k) \Big|,$$

which by equals $o_{\mathbf{P}}(1)$ as $n \to \infty$. This means that

$$\sup_{x \geq 1} x^{(1/2-\epsilon)} \Big| E_{n1}(x) - W\left((z_n(x))^{-1/\gamma_0}\right) - (z_n(x))^{-1/\gamma_0} \frac{(z_n(x))^{\rho/\gamma_0} - 1}{\rho\gamma_0} \sqrt{k} A_0(n/k) \Big| = o_{\mathbf{P}}(1).$$

Since $z_n(x) = (1 + o_{\mathbf{P}}(1))x$, uniformly over $x \geq 1$, $\sqrt{k} A_0(n/k) = O_{\mathbf{P}}(1)$ and $\rho < 0$, then it is easy to check that

$$(z_n(x))^{(1/2-\varepsilon)-1/\gamma_0} \frac{(z_n(x))^{\rho/\gamma_0} - 1}{\rho\gamma_0} = x^{(1/2-\varepsilon)-1/\gamma_0} \frac{x^{\rho/\gamma_0} - 1}{\rho\gamma_0} + o_{\mathbf{P}}(1),$$

uniformly over $x \geq 1$. [6] (page 235) showed that with probability one

$$\left| W\left((z_n(x))^{-1/\gamma_0}\right) - W\left(x^{-1/\gamma_0}\right) \right| \leq 2\epsilon x^{-(1-\epsilon)/(2\gamma_0)},$$

uniformly over $x \geq 1$, for any small $\epsilon > 0$, as $n \to \infty$. Note that for sufficiently small $\epsilon > 0$, $x^{-(1-\epsilon)/(2\gamma_0)+(1/2-\epsilon)} < 1$, then (almost surely) uniformly over $x \geq 1$

$$x^{(1/2-\varepsilon)} \left| W\left((z_n(x))^{-1/\gamma_0}\right) - W\left(x^{-1/\gamma_0}\right) \right| \leq 2\epsilon,$$

Therefore

$$\sup_{x\geq 1} x^{(1/2-\epsilon)} \left| E_{n1}(x) - W\left(x^{-1/\gamma_0}\right) \right.$$
$$\left. - x^{-1/\gamma_0} \frac{x^{\rho/\gamma_0} - 1}{\rho\gamma_0} \sqrt{k} A_0(n/k) \right| = o_{\mathbf{P}}(1). \tag{4.27}$$

For the term $E_{n2}(x)$ note that by Theorem 2.4.8 in [24],

$$\sqrt{k}\,(X_{n-k:n}/a_k - 1) - \gamma_0 W(1) = o_{\mathbf{P}}(1),$$

it follows that

$$\sqrt{k}\left((X_{n-k:n}/a_k)^{-1/\gamma_0} - 1\right) + W(1) = o_{\mathbf{P}}(1),$$

thus $\sup_{x\geq 1} x^{(1/2-\epsilon)} \left| E_{n1}(x) + x^{-1/\gamma_0} W(1) \right| = o_{\mathbf{P}}(1)$. Combining the last statement and (4.27), the weak approximation (4.25) comes. ■

Conclusion

As we have seen throughout this thesis, we proposed a new method to estimate the tail index under randomly right truncated data, which is the so-called semi-parametric estimation that gives us an estimator with more efficiency than the existing ones. Also, we proved its consistency and asymptotic normality by using a weak approximation of the corresponding tail empirical process, and to guarantee the best performance of the proposed estimator we made the simulation study in terms of bias and rmse.

The other important point is the introduction of a bias reduction to a kernel estimator of the tail index of randomly right-truncated Pareto-type distributions. The asymptotic normality of the derived estimator is established by assuming the second-order condition of regular variation. A simulation study is carried out to evaluate the finite sample behavior of the proposed estimator and compare it to those with non-reduced bias. An application to a real dataset of the lifetimes of automobile brake pads is done.

While we concluded this thesis with our interest in the estimation of the tail index for complete data and this is by presenting a robust estimator which is not affected by the outliers in the data. We also established its consistency and asymptotic normality, and the study of the finite sample behavior of the given estimator is done.

Nevertheless, this does not imply that the novel estimators which we presented in our research are without flaws. For instance, only if the distribution function $\mathbf{G}$ is known can the first estimator be used. The second estimator, although it reduces the bias, it increases the RMSE and that is the price to pay. Whereas, the third estimator has lower efficiency than the original estimator.

But this does not prevent the continuation of research to always find the best.

BIBLIOGRAPHY

[1] Andersen, E. B., 1970. Asymptotic properties of conditional maximum-likelihood estimators. *J. Roy. Statist. Soc.* Ser. B **32**, 283-301.

[2] Basu, A., Harris, I. R., Hjort, N. L., & Jones, M. C., 1998. Robust and efficient estimation by minimising a density power divergence. Biometrika 85, no. 3, 549-559.

[3] Beirlant, J., Bardoutsos, A., de Wet, T., Gijbels, I., 2016. Bias reduced tail estimation for censored Pareto type distributions. *Statist. Probab. Lett.* **109**, 78-88.

[4] Beirlant, J., Maribe, G., Verster, A., 2018. Penalized bias reduction in extreme value estimation for censored Pareto-type data, and long-tailed insurance applications. *Insurance Math. Econom.* **78**, 114–122.

[5] Benchaira, S., Meraghni, D., Necir, A., 2015. On the asymptotic normality of the extreme value index for right-truncated data. *Statist. Probab. Lett.* **107,** 378-384.

[6] Benchaira, S., Meraghni, D., Necir, A., 2016a. Tail product-limit process for truncated data with application to extreme value index estimation. *Extremes* **19**, 219-251.

[7] Benchaira, S., Meraghni, D., Necir, A., 2016b. Kernel estimation of the tail index of a right-truncated Pareto-type distribution. *Statist. Probab. Lett.* **119,** 186-193.

[8] Bilker W. B.,Wang, M. C., 1996.Asemiparametric extension of theMann–Whitney test for randomly truncated data. *Biometrics* **52**, 10-20

[9] Bingham, Nicholas H., Charles M. Goldie, and Jef L. Teugels. Regular variation. Vol. 27. Cambridge university press, 1989.

[10] Caeiro, F., Gomes, M.I, 2015. Threshold Selection in Extreme Value Analysis. Chapter in: Dipak Dey and Jun Yan, *Extreme Value Modeling and Risk Analysis: Methods and Applications*, Chapman-Hall/CRC, ISBN 9781498701297, pp. 69-87.

[11] Caeiro, F., Mateus, A., & Soltane, L., 2021. A class of weighted Hill estimators. Comput. Math. Methods 3, Paper No. e1167, 12 pp.

[12] Ciuperca, G., Mercadier, C., 2010. Semi-parametric estimation for heavy tailed distributions. *Extremes* 13, 55–87

[13] Csörgő, S., Deheuvels, P., Mason, D., 1985. Kernel estimates of the tail index of a distribution. *Ann. Statist.* **13**, 1050–1077.

[14] Csörgő,M., Révész, P., 1981. Strong Approximations in Probability and Statistics. Probability andMathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London.

[15] Dierckx, G., Goegebeur, Y., & Guillou, A., 2013. An asymptotically unbiased minimum density power divergence estimator for the Pareto-tail index. *J. Multivariate Anal.* 121, 70–86.

[16] Drees, H., de Haan, L., Li, D., 2006. Approximations to the tail empirical distribution function with application to testing extreme value conditions. *J. Statist. Plann. Inference* **136**, 3498-3538.

[17] Einmahl, J.H.J., de Haan, L., Li, D., 2006. Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *Ann. Statist.* **34**, 1987–2014.

[18] Gardes, L., Stupfler, G., 2015. Estimating extreme quantiles under random truncation. *TEST* **24**, 207-227.

[19] Groeneboom, P., Lopuhaä, H. P., de Wolf, P. P., 2003. Kernel-type estimators for the extreme value index. *Ann. Statist.* 31, 1956–1995.

[20] de Haan, L., Stadtmüller, U., 1996. Generalized regular variation of second order. *J. Australian Math. Soc.* (Series A) **61**, 381-395.

[21] Haouas, N., Necir, A, Meraghni, D, Brahimi, B., 2018. A Lynden-Bell integral estimator for the tail index of right-truncated data with a random threshold. *Afr. Stat.* **12**, 1159-1170.

[22] Haouas, N., Necir, A., Brahimi, B., 2019. Estimating the second-order parameter of regular variation and bias reduction in tail index estimation under random truncation. *J. Stat. Theory Pract.* **13**.

[23] He, S., Yang, G. L., 1998. Estimation of the truncation probability in the random truncation model. *Ann. Statist*. **26**, 1011-027.

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

[25] Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163-1174.

[26] Huber, P. J. & Ronchetti, E. M, 2009. *Robust Statistics*. John Wiley&Sons, second edition.

[27] Hua, L., Joe, H., 2011. Second order regular variation and conditional tail expectation of multiple risks. *Insurance Math. Econom.* **49,** 537-546.

[28] Hüsler, J., Li, D., & Müller, S. (2006). Weighted least squares estimation of the extreme value index. *Stat Probab Lett.* 76, 920-930.

[29] Klein, J.P., Moeschberger, S., 1997. Survival Analysis: Techniques for Censored and Truncated Data. Springer, Berlin.https://doi.org/10.1007/978-1-4757-2728-9.

[30] Koning, A. J., Peng, L., 2008. Goodness-of-fit tests for a heavy tailed distribution. *J. Statist. Plann. Inference* **138**, 3960-3981

[31] Lagakos, S. W., Barraj, L. M., De Gruttola, V., 1988. Nonparametric analysis of truncated survival data, with applications to AIDS. Biometrika **75**, 515–523.

[32] Lawless, J.F., 2002. Statistical Models and Methods for Lifetime Data, Second Edition. Wiley Series in Probability and Statistics.

[33] Lehmann, E.L. & Casella, G., 1998. Theory of Point Estimation. Springer.

[34] Li, G., Qin, J., Tiwari, R. C., 1997. Semiparametric likelihood ratio-based inferences for truncated data. *J. Amer. Statist. Assoc*. **92**, 236–245.

[35] Lui, K. J., Lawrence, D. N., Morgan, W. M., Peterman, T. A., Haverkos, H. H., Breakman, D. J., 1986. A model-based approach for estimating the mean incubation period of transfusion-associated acquired immunodeficiency syndrome. *Proc. Nat. Acad. Sc.* **83**, 2913-7.

[36] Lynden-Bell, D.,1971. A method of allowing for known observational selection in small samples applied to 3CR quasars. Monthly Notices Roy. Astron. Soc. *155*, 95-118.

[37] Mancer, S., Necir, A., Benchaira, S. (2022). Semiparametric tail-index estimation for randomly right-truncated heavy-tailed data. Arab Journal of Mathematical Sciences, (ahead-of-print).

[38] Mancer, S., Necir, A., & Benchaira, S. (2022). Bias Reduction in Kernel Tail Index Estimation for Randomly Truncated Pareto-Type Data. Sankhya A, 1-38.

[39] Neves, C., Fraga Alves, M.I., 2004. Reiss and Thomas' automatic selection of the number of extremes. *Comput. Statist. Data Anal*. **47**, 689-704.

[40] Moreira, C., de Uña-Álvarez, J., 2010. A semiparametric estimator of survival for doubly truncated data. *Stat. Med.* **29**, 3147–3159

[41] Moreira, C., de Uña-Álvarez, J., Van Keilegom, I., 2014. Goodness-of-fit tests for a semiparametric model under random double truncation. *Comput. Statist.* **29**, 1365–137.

[42] Qin, J., Wang, M-C., 2001, Semiparametric analysis of truncated data. Lifetime Data Anal. **7**, no. 3, 225–242.

[43] Reiss, R.D., Thomas, M., 2007. *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, 3rd ed. Birkhäuser Verlag, Basel, Boston, Berlin.

[44] Shen, P-S., 2010. Semiparametric analysis of doubly truncated data. *Comm. Statist. Theory Methods* **39**, 3178–3190.

[45] Shen, P-S, Hsu, H., 2020. Conditional maximum likelihood estimation for semiparametric transformation models with doubly truncated data. Comput. Statist. Data Anal. **144**, 106862, 15 pp.

[46] Shorack, G.R.,Wellner, J.A., 1986. Empirical Processes with Applications to Statistics.Wiley, New York.

[47] Strzalkowska-Kominiak, E., Stute, W., 2009. Martingale representations of the Lynden-Bell estimator with applications. *Statist. Probab. Lett.* **79**, 814–820.

[48] Wang, M.-C., 1989. A semiparametric model for randomly truncated data. J. Amer. Statist. Assoc. **84**, 742–748.

[49] Weissman, I., 1978. Estimation of parameters and large quantiles based on the k largest observations. J. Am. Statist. Assoc. **73**, 812-815.

[50] Worms, J., Worms, R., 2016. A Lynden-Bell integral estimator for extremes of randomly truncated data. *Statist. Probab. Lett.* **109**, 106-117.

[51] Woodroofe, M., 1985. Estimating a distribution function with truncated data. *Ann. Statist.* **13**, 163-177.