

# Flexible Univariate Extreme Value Modelling with Applications in Insurance

Gaonyalelwe Maribe  
(2007052190)

**Promoters:**

Dr. Andrehette Verster & Prof. Jan Beirlant

A Thesis submitted in fulfillment of the degree of  
Doctor of Philosophy



UNIVERSITY OF THE FREE STATE  
Faculty of Natural and Agricultural Sciences  
Department of Mathematical Statistics and Actuarial Science  
December 2018

*In dedication to a dream deferred*

*What happens to a dream deferred?*

*Does it dry up*

*like a raisin in the sun?*

*Or fester like a sore—*

*And then run?*

*Does it stink like rotten meat?*

*Or crust and sugar over—*

*like a syrupy sweet?*

*Maybe it just sags like a heavy load.*

*Or does it explode?*

*—by Langston Hughes*

# **Flexible Univariate Extreme Value Modelling with Applications in Insurance**

**Gaonyalelwe Maribe**  
**(2007052190)**

Submitted for the degree of Doctor of Philosophy  
December 2018

## **Abstract**

Bias reduction in tail estimation has received considerable interest in extreme value analysis. Estimation methods that minimise the bias while keeping the mean squared error (MSE) under control, are especially useful when applying classical methods such as the Hill (1975) estimator. In the case of heavy tailed distributions, Caeiro et al. (2005) proposed minimum variance reduced bias estimators of the extreme value index, where the bias is reduced without increasing the variance with respect to the Hill estimator. This method is based on adequate external estimation of a pair of parameters of second order slow variation under a third order condition.

In this thesis we revisit this problem and exploit the mathematical fact that the bias of the Hill estimator tends to zero as the threshold increases. This leads to a shrinkage estimation of the extreme value index, which allows for a penalised likelihood and a Bayesian implementation of the extended Pareto distribution, as developed in Beirlant et al. (2009). Asymptotic results for the resulting shrinkage penalised likelihood estimator of the extreme value index are presented. Finite sample simulation results are proposed both for the penalised likelihood and Bayesian implementation. We then compare with the minimum variance reduced bias estimators of Caeiro et al. (2005).

---

The subject of tail estimation for randomly censored data from a heavy tailed distribution has received growing attention, motivated by applications for instance in actuarial statistics. The bias of the available estimators of the extreme value index can be substantial and depends strongly on the amount of censoring. We review the available estimators, propose a bias reduced estimator, and show again how shrinkage estimation can help to keep the MSE under control. A bootstrap algorithm is proposed to construct confidence intervals. These new proposals are compared with the existing estimators through simulation results. A detailed study of a long-tailed car insurance portfolio is made, which typically exhibit heavy censoring.

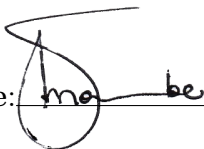
In recent years, several attempts have been made to extend tail modelling towards the modal part of the data. Great efforts have been made in the development of extreme value models which allow for simultaneous estimation of the threshold, quantification of threshold uncertainties and inference based on all the data.

When fitting second order models it turns out that the bias of the resulting extreme value estimators is significantly reduced compared to the classical tail fits using only the first order tail component based on the Pareto or generalised Pareto (GP) fits to peaks over threshold (POT) distributions.

This thesis proposes a novel bias reduced tail fitting technique, improving upon the classical GP approximation for POT using the flexible semiparametric GP modelling introduced in Tencaliec et al. (2018). We also revisit and extend the second order refined POT approach started in Beirlant et al. (2009) to all max-domains of attraction using flexible semiparametric modelling of the second order component. In this way we relax the classical second order regular variation assumptions.

# Declaration

I, Gaonyalelwe Maribe, hereby declare that this work, submitted for the degree Ph.D (Mathematical Statistics) at the University of the Free State is my own original work, and has not been submitted at another university/faculty for any reward, recognition, or degree. I further declare that all sources cited or quoted are indicated and acknowledged by means of a comprehensive list of references. Copyright hereby cedes to the University of the Free State.

Signature:  \_\_\_\_\_

Date: March 17, 2019

# Acknowledgements

I would firstly like to thank **God** for all the blessings he has showered upon me and my family, for the daily strength he has given me and for all the people he has placed in my path to make life more meaningful.

To my family for all the sacrifices they had to make for my education. Again to my family and friends for being understanding and having to deal with my “research moods”, most especially **Kgabo Mphahlele**.

Towards the end of the term in November of 2014, I found myself in **Andréhette Verster’s** office, seeking mentorship for my Honours research project. Not knowing she would continue to mentor me throughout my masters and PhD studies. This visit to her office has been one of the most fruitful endeavours of my life. I am truly grateful to have been called your student.

To **Jan Beirlant** from *KU Leuven, Belgium*, a great teacher and mentor. From you, I have learned a lot about myself. I would like to thank you for taking out your time to work with me. I jokingly like to refer to you as my grandfather ☺, this is because no one has ever had so much patience in explaining difficult concepts to me as you have. I am truly grateful for the life long contribution you have made in my life.

To **Philippe Naveau** from *Laboratoire des Sciences du Climat et de l’Environnement*,

---

*CNRS, Université Paris-Saclay*, thank you for our collaborative work described in Chapter 5, and for being a fantastic host on our visit to Paris.

I would also like to take pleasure in thanking **Sean van der Merwe** for his valuable advice concerning the Bayesian implementations in Chapter 3.

A special thanks goes out to **Rym and Julien Worms** for their helpful discussions and suggestions on the topic of tail estimation for randomly censored data.

I would also like to thank and acknowledge the UFS High-Performance Computing (HPC) unit for providing me with the massive amount of computing resources I used to conduct simulation studies. More especially **Stephanus Riekert**, I am truly grateful for all the technical support you provided me.

Thanks to the REVSTAT as well as Insurance: Mathematics & Economics referees for their constructive comments which improved the presentation of the accepted papers significantly. I extend this thanks to the assessors taking out their invaluable time to assess this PhD thesis.

I would lastly like to thank the “*SASA-NRF GRANT: Academic statistics bursary fund*” for sponsoring this PhD research.

*This work is based on the research supported wholly/in part by the National Research Foundation of South Africa (Grant Number 102628). The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research is that of the author(s), and that the NRF accepts no liability whatsoever in this regard.*

# List of acronyms

<b>2ERV</b>	Second order extended regular variation
<b>CH</b>	Corrected Hill
<b>CLT</b>	Central limit theorem
<b>CDF</b>	Cumulative distribution function
<b>IID</b>	Independent, identically distributed
<b>EP</b>	Extended Pareto
<b>EPD</b>	Extended Pareto distribution
<b>ERV</b>	Extended regular variation
<b>EVI</b>	Extreme value index
<b>EVT</b>	Extreme value theory
<b>GEV</b>	Generalised extreme value
<b>GP</b>	Generalised Pareto
<b>GPD</b>	Generalised Pareto distribution
<b>HTE</b>	Heavier than exponential
<b>KM</b>	Kaplan-Meier
<b>LTE</b>	Lighter than exponential
<b>ML</b>	Maximum likelihood
<b>MLE</b>	Maximum likelihood estimate
<b>MVBR</b>	Minimum variance reduced bias
<b>PML</b>	Penalised maximum likelihood
<b>POT</b>	Peaks over threshold



# Mathematical notation

$\gamma$	Extreme value index.
$\mathcal{D}(G_\gamma)$	Domain of attraction of $G_\gamma$ .
$\mathcal{F}$	Class of distributions in the Fréchet domain of attraction.
$F^\leftarrow$	left continuous inverse of the distribution function $F$ .
$f \in RV_\rho$	$f$ is regular varying with index $\rho$ .
$\Delta$	Censoring indicator, 0 if the observation is right-censored and 1 if not.
$l_{pen}$	Penalised likelihood.
▲	Indicates the end of a definition, proof, lemma or proposition.
$\ell_a, \ell_\delta, \ell_U$ & $\ell_U$	Slowly varying functions.
$f \in 2ERV_{\gamma,\rho}$	$f$ is of second order extended regular variation.
$Q(p)$	The quantile function denoted by $Q(p) = \inf\{x F(x) \geq p\}$ .
$H_{k,n}$	The Hill estimator based on $k$ excesses above the threshold.
$t \uparrow \infty$	$t$ approaches $\infty$ from below, in an increasing manner.
$\rightarrow^d \mathcal{N}_p$	Converges in distribution to a $p$ dimensional Normal.
$\omega$	Tuning constant regulating the amount of penalty.
$E_\infty, Var_\infty$ & $MSE_\infty$	Asymptotic bias, variance and mean squared error.
$\mathcal{T}$ and $\mathcal{E}$	Transformed and extended models.

# Contents

Abstract	iii
Declaration	v
Acknowledgements	vi
List of acronyms	viii
Mathematical notation	ix
<b>1 Introduction</b>	<b>1</b>
1.1 Purpose of study and contributions . . . . .	2
1.2 Thesis Objectives . . . . .	3
1.3 Outline of Thesis . . . . .	3
1.4 Authorship and publication . . . . .	5
<b>2 Background</b>	<b>6</b>
2.1 Introduction . . . . .	6
2.2 Motivating data set . . . . .	11
2.2.1 Reinsurance: Motor Liability Data . . . . .	12
2.2.2 Mean excess . . . . .	16
2.2.3 QQ-plots . . . . .	17
2.3 Tail Estimation . . . . .	21

2.3.1	Block maxima . . . . .	22
2.3.2	Peaks over threshold approach . . . . .	23
2.3.3	Fitting Pareto type data . . . . .	24
2.3.4	Threshold selection . . . . .	29
2.3.5	Bias reduction . . . . .	31
2.3.6	Mixture modelling in EVT . . . . .	32
2.4	Data sets and Software . . . . .	33
2.5	Chapter conclusion . . . . .	34
<b>I</b>	<b>Second-order Refinements: Heavy Tails (<math>\gamma &gt; 0</math>)</b>	<b>36</b>
<b>3</b>	<b>Using Shrinkage Estimators to Reduce Bias and MSE in Estimation of Heavy Tails</b>	<b>37</b>
3.1	INTRODUCTION . . . . .	37
3.2	Shrinkage estimators of the EPD parameters . . . . .	41
3.2.1	penalised likelihood and Bayesian interpretation . . . . .	41
3.2.2	Asymptotic results for the penalised ML estimator $\hat{\gamma}_k^P$ . . . . .	42
3.3	Simulations and practical case studies . . . . .	44
3.4	Chapter conclusion . . . . .	47
	<b>Appendices</b>	<b>54</b>
3.A	Derivation of the expressions of $(\hat{\gamma}_k^P, \hat{\delta}_k^P)$ . . . . .	54
3.B	Optimal $\omega$ minimising the $MSE_\infty$ . . . . .	55
3.C	Algorithm to estimate $\rho$ . . . . .	58
<b>4</b>	<b>Penalised Bias Reduction in Extreme Value Estimation for Censored Pareto-type Data</b>	<b>59</b>
4.1	Introduction . . . . .	59
4.2	A review of estimators of $\gamma_1$ . . . . .	61

4.3	Bias reduction of the Worms & Worms estimator and penalised estimation of bias . . . . .	65
4.4	Bootstrap confidence intervals for $\gamma_1$ . . . . .	67
4.5	Finite sample simulations . . . . .	68
4.6	A case study from car insurance . . . . .	71
4.7	Chapter conclusion . . . . .	72
<b>II</b>	<b>Semi-parametric Refinements: General Tails (<math>\gamma \in \mathbb{R}</math>)</b>	<b>81</b>
<b>5</b>	<b>Bias Reduced Peaks over Threshold Tail Estimation</b>	<b>82</b>
5.1	Introduction . . . . .	82
5.2	Transformed and extended GPD models . . . . .	86
5.3	Basic asymptotics under model $(\mathcal{E})$ . . . . .	93
5.4	Simulations and case studies . . . . .	97
5.5	Chapter conclusion . . . . .	101
	<b>Appendices</b>	<b>112</b>
5.A	Proof of Theorem 5.1 . . . . .	112
5.B	Shiny application . . . . .	116
<b>6</b>	<b>Thesis conclusion</b>	<b>122</b>
6.1	Outlook and future work . . . . .	123
	<b>References</b>	<b>125</b>

# List of Figures

2.1	Densities of the three standard extreme value distributions. With $\alpha = 1$ (left) and $\alpha = 2$ (right) . . . . .	7
2.2	Densities of the $GEV(\gamma)$ with $\gamma$ choices: 0,0.5,-0.5 (left) and 0,0.75,-0.75 (right) . . . . .	8
2.3	Scatter plot of ultimates loss data (left) and the corresponding empirical density plot (right) . . . . .	13
2.4	Diagnostic plots for the ultimates data . . . . .	21
2.5	Mean Excess (left) and Hill estimates of the EVI (right) of the ultimates . . . . .	25
2.6	Development patterns for four particular claims . . . . .	27
3.1	Bias (left) and root mean squared error (right) in case of the <b>EV distribution</b> with $\gamma = 0.25$ for sample sizes $n = 200$ (top), $n = 500$ (middle) and $n = 1000$ (bottom) for the Hill estimator (H), the EPD-ML estimator $\hat{\gamma}_k^{ML}$ (ML), the penalised ML estimator $\hat{\gamma}_k^P(1)$ with $\omega = 1$ (PML), the Bayesian estimator $\hat{\gamma}_k^B(1)$ with $\omega = 1$ (B), and the minimum variance reduced bias estimator $CH_k$ (CH). . . . .	48
3.2	Bias (left) and root mean squared error (right) in case of the <b>Fréchet distribution</b> with $\gamma = 0.5$ for sample sizes $n = 200$ (top), $n = 500$ (middle) and $n = 1000$ (bottom) for the Hill estimator (H), the EPD-ML estimator $\hat{\gamma}_k^{ML}$ (ML), the penalised ML estimator $\hat{\gamma}_k^P(1)$ with $\omega = 1$ (PML), the Bayesian estimator $\hat{\gamma}_k^B(1)$ with $\omega = 1$ (B), and the minimum variance reduced bias estimator $CH_k$ (CH). . . . .	49

3.3	Bias (left) and root mean squared error (right) in case of the <b>Burr distribution</b> with $\gamma = 0.75$ for sample sizes $n = 200$ (top), $n = 500$ (middle) and $n = 1000$ (bottom) for the Hill estimator (H), the EPD-ML estimator $\hat{\gamma}_k^{ML}$ (ML), the penalised ML estimator $\hat{\gamma}_k^P(1)$ with $\omega = 1$ (PML), the Bayesian estimator $\hat{\gamma}_k^B(1)$ with $\omega = 1$ (B), and the minimum variance reduced bias estimator $CH_k$ (CH). . . . .	50
3.4	Bias (left) and root mean squared error (right) in case of the <b>loggamma distribution</b> with $\gamma = 0.5$ for sample sizes $n = 200$ (top), $n = 500$ (middle) and $n = 1000$ (bottom) for the Hill estimator (H), the EPD-ML estimator $\hat{\gamma}_k^{ML}$ (ML), the penalised ML estimator $\hat{\gamma}_k^P(1)$ with $\omega = 1$ (PML), the Bayesian estimator $\hat{\gamma}_k^B(1)$ with $\omega = 1$ (B), and the minimum variance reduced bias estimator $CH_k$ (CH). . . . .	51
3.5	Bias (left) and root mean squared error (right) in case of the <b>Fréchet distribution</b> with $\gamma = 0.5$ (top) and Burr distribution with $\gamma = 0.75$ (bottom) for sample size $n = 200$ comparing the penalised ML estimator $\hat{\gamma}_k^P(1)$ with $\omega = 1$ , $\omega = \omega_{mv}$ from (3.2.18), and the optimal asymptotic RMSE from (3.2.17) replacing $\lambda$ by $C_a\sqrt{k}(k/n)^{-\rho}$ . . . . .	52
3.6	Estimates of $\gamma$ for Secura Belgian Re data set: results for the Hill estimator (H), the EPD-ML estimator $\hat{\gamma}_k^{ML}$ (ML), the penalised ML estimator $\hat{\gamma}_k^P(1)$ with $\omega = 1$ (PML), the Bayesian estimator $\hat{\gamma}_k^B(1)$ with $\omega = 1$ (B), and the minimum variance reduced bias estimator $CH_k$ (CH) (left), focused plot for $k = 1, \dots, 100$ (right). . .	53
4.1	Bias and RMSE for <b>Burr(10,2,2)</b> censored by <b>Burr(10,5,2)</b> . . . . .	73
4.2	Bias and RMSE for <b>Burr(10,2,1)</b> censored by <b>Burr(10,2,1)</b> . . . . .	74
4.3	Bias and RMSE for <b>Burr(10,5,2)</b> censored by <b>Burr(10,2,2)</b> . . . . .	75
4.4	Bias and RMSE for <b>Fréchet(2)</b> censored by <b>Fréchet(1)</b> . . . . .	76

4.5	Bias (top) and RMSE (bottom) for $\hat{\gamma}_{1,k}^{(W)}$ , $\hat{\gamma}_{1,k}^{(BR,W)}(-1.5)$ , $\hat{\gamma}_{1,k}^{(BR,W)}(-2)$ , $\hat{\gamma}_{1,k}^{(s,W)}(-1.5)$ , $\hat{\gamma}_{1,k}^{(s,W)}(-2)$ in case of <b>Burr(10,2,2)</b> censored by <b>Burr(10,5,2)</b> (left); <b>Burr(10,2,1)</b> censored by <b>Burr(10,2,1)</b> (middle); and <b>Burr(10,5,2)</b> censored by <b>Burr(10,2,2)</b> (right) . . . . .	77
4.6	Simulated bootstrap 95% confidence intervals with adaptive choice of $k_1$ and $k_2$ using $\epsilon = 0.01$ (4 left frames) and using $k_1 = k_2 = 25 = 5\%n$ for every sample (4 right frames) . . . . .	78
4.7	Car liability data: plot of $\hat{p}_k$ as a function of $k/n$ for all claims jointly (thick full line) and for claims that arrived in different time intervals. .	79
4.8	Car liability data: $\hat{\gamma}_{1,k}^{(W)}$ (left) and $\hat{\gamma}_{1,k}^{(BR,W)}(-3)$ (right) based on indexed cumulative payments at end of 2010, as a function of $k/n$ for all claims together (thick full line) and for claims arriving in different time periods	79
4.9	Car liability data: $\hat{\gamma}_{1,k}^{(W)}$ (left) and $\hat{\gamma}_{1,k}^{(BR,W)}(-3)$ (right) for all claims arriving in 1995-1999 and based on indexed cumulative payments at end of 2000(02)2010, as a function of $k/n$ . . . . .	80
4.10	Car liability data, using all claims: $\hat{\gamma}_{1,k}^{(W)}$ , $\hat{\gamma}_{1,k}^{(s,W)}(-3)$ , $\hat{\gamma}_{2,k}^{(W)}$ , $\hat{\gamma}_{2,k}^{(s,W)}(-3)$ (left); 95% bootstrap confidence intervals (right). . . . .	80
5.1	Ultimates of Belgian car insurance claims: bias reduction of Hill estimator (full line) using $\bar{H}_{\gamma,\delta}^{EP}$ with $\rho = -0.25$ (dashed line), $\rho = -0.5$ (dotted line) and $\rho = -1$ (dash-dotted line). . . . .	85
5.1	Burr distribution with $\gamma = 0.5$ and $\rho = -0.5$ . Estimation of $\gamma$ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line), $T\bar{p}$ (dotted), $Ep$ (dash-dotted) and $E\bar{p}$ (dashed). . . . .	102
5.2	Burr distribution with $\gamma = 0.5$ and $\rho = -0.5$ . Estimation of $\gamma$ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): Pareto-ML (full line), $T\bar{p}^+$ (dotted), $Ep^+$ (dash-dotted) and $E\bar{p}^+$ (dashed). . . . .	103

5.3	Fréchet distribution with $\gamma = 0.5$ . Estimation of $\gamma$ (top) and tail probability (bottom), bias (left), RMSE (right): GPD-ML (full line), $T\bar{p}$ (dotted, using minimum variance principle), $Ep$ with $\rho = -2$ (dash-dotted) and $E\bar{p}$ with $(k_*, m) = (190, 150)$ (dashed). . . . .	104
5.4	Fréchet distribution with $\gamma = 0.5$ . Estimation of $\gamma$ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): Pareto-ML (full line), $T\bar{p}^+$ (dotted), $Ep^+$ (dash-dotted) and $E\bar{p}^+$ (dashed). . . . .	105
5.5	Standard normal distribution ( $\gamma = 0$ and $\tilde{\rho} = 0$ ). Estimation of $\gamma$ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line), $T\bar{p}$ (dotted), $Ep$ (dash-dotted) and $E\bar{p}$ (dashed). . . . .	106
5.6	The exponential distribution ( $\gamma = 0$ and $\tilde{\rho} = 0$ ). Estimation of $\gamma$ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line), $T\bar{p}$ (dotted), $Ep$ (dash-dotted) and $E\bar{p}$ (dashed). . . . .	107
5.7	Reversed Burr distribution ( $\gamma = -0.2$ and $\tilde{\rho} = -1$ ). Estimation of $\gamma$ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line), $T\bar{p}$ (dotted), $Ep$ (dash-dotted) and $E\bar{p}$ (dashed). . . . .	108
5.8	Extreme value Weibull distribution ( $\gamma = -0.25$ and $\tilde{\rho} = -1$ ). Estimation of $\gamma$ (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line), $T\bar{p}$ (dotted), $Ep$ (dash-dotted) and $E\bar{p}$ (dashed). . . . .	109
5.9	Ultimates of Belgian car insurance claims: estimation of $\gamma$ (top left), tail probability at maximum observation (top right): Pareto-ML (full line), $T\bar{p}^+$ (dotted), $Ep^+$ (dashed) and $E\bar{p}^+$ (dash-dotted). Goodness-of-fit plot (bottom). . . . .	110



5.10 Winter rain data at Mont-Aigoual: estimation of $\gamma$ and $\sigma$ (top) and tail probability (bottom left) using minimum variance principle: GPD-ML (full line), $T\bar{p}$ (dotted), $Ep$ (dashed) and $E\bar{p}$ (dash-dotted). Goodness-of-fit plot (bottom right).	111
5.B.1 Window 1	119
5.B.2 Window 2	120
5.B.3 Window 3	121

# List of Tables

2.1	A list of some common distributions and their respective domains of attraction . . . . .	12
2.2	Summary of the ultimates data . . . . .	13
2.3	Q-Q plot and derivatives plot coordinates. . . . .	18

# Chapter 1

## Introduction

*“Nothing is more self-limiting than going to extremes.”*

— Marty Rubin

At the forefront of philosophical discourses in statistics, was the measure of central tendency of a population (the average). The inquiry of the “limiting” distribution of a sequence of random variables would later lead up to the first version of what is known as the Central Limit Theorem (CLT) which appeared in 1733, by the French-born mathematician Abraham de Moivre.

The CLT is perhaps the most fundamental result in all of statistics. It provides a theoretical framework that allows for justification of inference to extrapolate from a sample and make good approximations and predictions. Under some conditions the CLT states that the normal distribution (Gaussian) appears as a limiting distribution of the sample averages when the number of observations is sufficiently large.

A corresponding discourse which emerged in the 20th century, was that of the maxima

rather than the average. *What appears as the limiting distribution of the partial maxima?*

This question has led to a very fruitful and path-breaking intellectual pursuit, the development of a theoretical framework known as Extreme Value Theory (EVT). Most concur that it made a first appearance in a paper by Dodd (1923), followed quickly by the papers of Fréchet (1927) and Fisher and Tippett (1928).

EVT has a rich mathematical theory which helps us answer non-trivial inferential questions about extreme observations and is applicable to a vast number of disciplines. The goal is often motivated by the need to predict the upper quantile or tail probability. Areas of application include but is not limited to: climatology, finance, insurance, survival analysis, hydrology and geology.

## 1.1 Purpose of study and contributions

The main research focus in EVT has over the past half-century taken a dramatic change. Peaks over threshold (POT) methods based on large observations (herein referred to as excesses) above some suitably high threshold have become more popular since the pioneering work of Balkema and de Haan (1974) and Pickands (1975). The use of POT methods however prompt the question: “what is a suitable threshold?” Issues with threshold selection are discussed in Section 2.3.4. Work presented in this thesis is meant to subjugate the sensitivity inherent in threshold selection.

A second-order refined model makes use of additional information up to the second-order about the slowly varying function  $\ell$ . Beirlant et al. (2004) proposed a quantile and probabilistic view of second-order refinements for  $\gamma > 0$  estimators. In the first part of this PhD thesis, we propose some improvements to second order refinements made by Beirlant et al. (2009) by using the theoretical fact that, for large thresholds the distribution of excesses is strictly Pareto, leading to a shrinkage estimation technique

of the extreme value index (EVI).

Motivated by applications for instance in actuarial statistics, the subject of censoring in heavy-tailed data has received growing attention in EVT. This thesis reviews some existing tail estimation methods and proposes a second order improved version of the Worms and Worms (2014) estimator.

Second order models naturally add some restrictions. This thesis further proposes a bias reduced tail fitting technique that allows for these restrictions to be relaxed. The second order refined POT approach started in Beirlant et al. (2009) is revisited and extended to all max-domains of attraction by applying a flexible semi-parametric modelling of the second order component.

## 1.2 Thesis Objectives

- Propose a shrinkage estimation technique to reduce the mean squared error (MSE) of the Extended Pareto distribution (EPD) of Beirlant et al. (2009);
- Propose a bias reduced version of the Worms and Worms (2014) estimator;
- Construct confidence intervals of the bias reduced Worms and Worms (2014) using a bootstrap algorithm;
- Propose a novel bias reduced tail fitting technique for all max-domains of attractions.

## 1.3 Outline of Thesis

This thesis is divided into two parts, Part I focuses on second order refinements for censored and uncensored heavy-tailed distributions. Part II reviews attempts made to

build flexible full models capable of capturing the bulk and tail part of the data. A novel semi-parametric refinement is proposed in this part.

Chapter 2 gives an overview of EVT, introduces some key concepts and mentions historical developments in the field. In Chapter 3 we introduce a shrinkage estimation technique to improve an unbiased estimator of the EVI and reduce the overall MSE. The overall reduction in MSE is shown both mathematically and through a finite simulation study. A short case study is conducted on a Secura Belgian Reinsurance data set.

In Chapter 4 we focus on random right censored heavy tailed data, and propose a bias reduced version of the Worms and Worms (2014) estimator. We then apply the shrinkage technique introduced in Chapter 3 to improve MSE of this newly proposed estimator. A finite simulation study and a long tailed insurance application are conducted.

In Chapter 5, the second order bias reduction technique is extended to all domains of attraction. A general form of the second order refined model is proposed. Merits of the proposed estimators are illustrated with a case study on a rain fall data set. An interactive simulation study is conducted and hosted on a shiny-website provided at the end of the chapter.

In Chapter 6, a conclusion of the study is presented, and some areas which allow for further research are identified.

## 1.4 Authorship and publication

Some parts of this thesis are based on work by the author already published in international peer reviewed journals. Below is the list of published and publishable journal papers listed in chronological order as presented in chapters to follow.

1. Beirlant, J., Maribe, G. and Verster, A., 2019. Using shrinkage estimators to reduce bias and MSE in estimation of heavy tails. *REVSTAT-Statistical Journal*, 17(1), pp.91-108.
2. Beirlant, J., Maribe, G. and Verster, A., 2018. Penalized bias reduction in extreme value estimation for censored Pareto-type data, and long-tailed insurance applications. *Insurance: Mathematics and Economics*, 78, pp.114-122.
3. Beirlant, J., Maribe, G., Naveau, P. and Verster, A., 2018. Bias Reduced Peaks over Threshold Tail Estimation. *arXiv preprint arXiv:1810.01296*.

# Chapter 2

## Background

### 2.1 Introduction

EVT is primarily concerned with the limiting distribution of the maxima or minima of independent, identically distributed (IID) random variables.

Let  $X$  be an arbitrary random variable with a cumulative distribution function (CDF)  $F$ . Moreover, let  $X_1, X_2, \dots, X_n$  be IID random variables. The ordered set is then denoted as  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ . The limit problem is that of finding a potential limiting distributions of the maximum  $X_{n,n} = \max(X_1, X_2, \dots, X_n)$ . The distribution of this sample maximum can be written as  $P(X_{n,n} \leq x) = F^n(x)$ , non trivially for some constants  $a_n > 0$  and  $b_n$ :

$$F^n(a_n x + b_n) \rightarrow^d G(x), \text{ weakly as } n \rightarrow \infty. \quad (2.1.1)$$

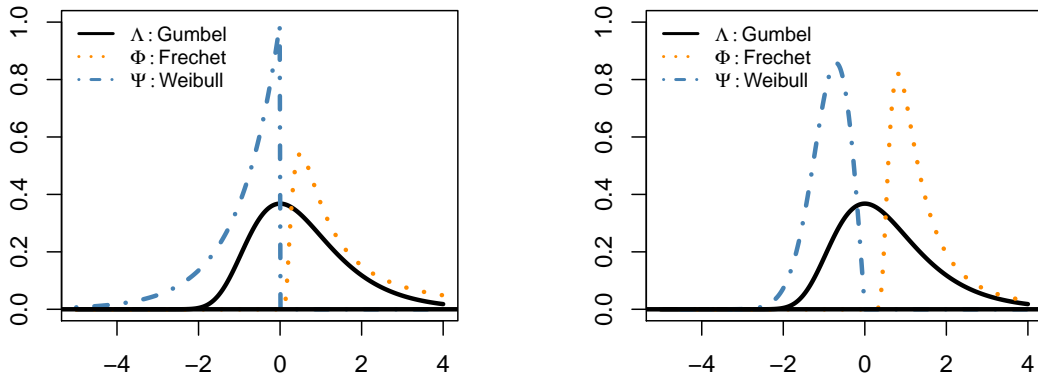
thereby reducing the limit problem to a location and scale problem where  $G(x)$  is one of the three possible limiting distribution functions.

This problem was initially solved by Fisher and Tippett (1928), and later derived rigorously by Gnedenko (1943) and further streamlined by de Haan (1970). The extremal



types theorem asserts that if a non-degenerate  $G$  exists it must be one of three types:

$$\begin{aligned}
 I &: \Lambda(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R} : \\
 II &: \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases} \\
 &\quad \text{for some } \alpha > 0 \\
 III &: \Psi(x) = \begin{cases} \exp\{-(-x)^\alpha\}, & x \leq 0 \\ 1, & x > 0 \end{cases} \\
 &\quad \text{for some } \alpha > 0
 \end{aligned}$$

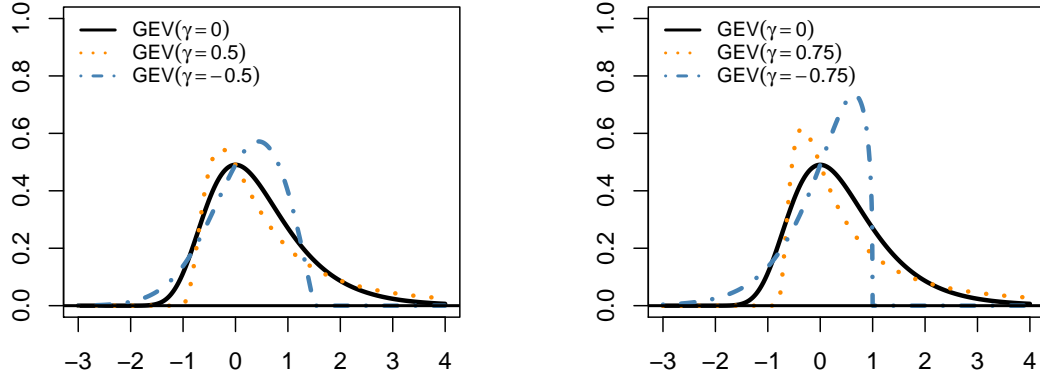


**Figure 2.1:** Densities of the three standard extreme value distributions. With  $\alpha = 1$  (left) and  $\alpha = 2$  (right)

These types of distributions are often referred to as the Gumbel, Fréchet and Weibull types, respectively. The three types of distributions can be joined into a single distribution function with the following cumulative distribution function (CDF):

$$G_\gamma(x) = \begin{cases} \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\}, & \text{if } \gamma \neq 0 \\ \exp\{-e^{-x}\}, & \text{if } \gamma = 0 \end{cases} \quad (2.1.2)$$

defined on  $\{x : 1 + \gamma x > 0\}$  and  $\gamma \in \mathbb{R}$ . This CDF is known as the Generalised Extreme value (GEV) distribution. The parameter  $\gamma$  is known as the shape parameter, also known as the EVI. This parameter is of primordial importance in EVT.



**Figure 2.2:** Densities of the  $GEV(\gamma)$  with  $\gamma$  choices: 0,0.5,-0.5 (left) and 0,0.75,-0.75 (right)

It becomes clear from Figure 2.2 that the EVI affects the tail behaviour. If  $\gamma = 0$ , the GEV distribution decays exponentially and belongs to the Gumbel family. If  $\gamma > 0$ , the GEV distribution has a much heavier tail and belongs to the Fréchet family. Lastly if  $\gamma < 0$ , the GEV has a finite upper endpoint and belongs to the Weibull family.

If (2.1.1) holds, we can say that the underlying distribution  $F$  is in the domain of attraction of  $G_\gamma$  and write  $F \in \mathcal{D}(G_\gamma)$ . This is the set of distributions attracted to  $G_\gamma$ . For example it can be shown that the Loggamma distribution is in the domain of attraction of  $G$  with  $\gamma > 0$ .

The rigorous mathematical framework provided by EVT necessitates some assumptions about regularity conditions for the underlying distributions. Throughout the different parts of this thesis we refer to the well known, necessary and sufficient conditions for  $F \in \mathcal{D}(G_\gamma)$ . Without going into much detail we therefore make a brief overview of the necessary and sufficient conditions in order to give a preliminary context.

Define the support of a distribution function  $F$  by the lower bound  ${}_+x := \sup\{x : F(x) = 0\}$  and the upper bound  $x_+ := \inf\{x : F(x) = 1\}$ . The quantile function  $Q(y) = F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$  and the tail quantile function  $U(y) = F^{\leftarrow}\left(1 - \frac{1}{y}\right)$ .

## First-Order Condition

Consider the underlying distribution function  $F$  underlying in the extremal domain of attraction of the GEV distribution,  $F \in \mathcal{D}(G_\gamma)$ . There exists a positive measurable function  $a$  for which (de Haan, 1984):

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D(x) = \begin{cases} \frac{x^\gamma - 1}{\gamma} & \text{if } \gamma \neq 0 \\ \log(x) & \text{if } \gamma = 0 \end{cases} \quad (2.1.3)$$

for all  $x > 0$ .

de Haan (1984) shows that Equation (2.1.3) is equivalent to the following statement:

*There is a positive function  $a$  such that:*

$$\lim_{t \rightarrow \infty} t(1 - F(a(t)x + U(t))) = (1 + \gamma x)^{-1/\gamma} \quad (2.1.4)$$

for all  $x$  with  $1 + \gamma x > 0$ .

▲

### Definition 2.1. Regular variation

*All functions ultimately zero are said to belong to a class of regular varying functions if:*

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\rho, \quad \text{for } x > 0, \text{ with some } \rho \in \mathbb{R}, \quad (2.1.5)$$

*$f$  is said to be regular varying with index  $\rho$ ,  $f \in RV_\rho$ . Furthermore a function  $f$  is said to be slowly varying if for  $x > 0$ :*

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = 1. \quad (2.1.6)$$

▲

## Second-Order Condition

### A general tail ( $\gamma \in \mathbb{R}$ )

$F$  satisfies the second order condition if there exists a function  $A(t)$ , tending to zero as  $t \rightarrow \infty$ , such that:

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a_0(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = H_{\gamma, \rho}(x) := \frac{1}{\rho} \left( \frac{x^{\gamma+\rho} - 1}{\gamma + \rho} - \frac{x^\gamma - 1}{\gamma} \right) \quad (2.1.7)$$

for all  $x > 0$  and  $a_0 > 0$ , with  $\rho < 0$ , the second order parameter controlling the speed of convergence of the partial maxima towards the limit law. It is then said that the function  $U$  is of second order extended regular variation, and write  $U \in 2ERV_{\gamma, \rho}$ , noting also that,  $|A| \in RV_\rho$  and tends to zero, hence larger absolute values of  $\rho$  correspond to a higher rate of convergence.

### A heavy tail ( $\gamma > 0$ )

In case of heavy tails the second order condition reads as follows. The existence of the function  $\tilde{A}(t)$  is such that:

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{\tilde{A}(t)} = x^\gamma \frac{x^{\tilde{\rho}} - 1}{\tilde{\rho}} &\iff \\ \lim_{t \rightarrow \infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{\tilde{A}(t)} = \frac{x^{\tilde{\rho}} - 1}{\tilde{\rho}} \end{aligned} \quad (2.1.8)$$

for all  $x > 0$ , and  $\tilde{\rho} \leq 0$ , the second order parameter. As in the general case, we then say  $U$  is of second order regular variation and write  $U \in 2RV_{\gamma, \tilde{\rho}}$ .  $F$  is said to satisfy the second order condition (Fraga Alves et al., 2007). ▲

There is a special class of Pareto type distributions, known as the Hall class. This is a wide class of models containing most useful heavy-tailed parents such as those used in this thesis, namely: the Fréchet and the Burr XII distribution.

### Definition 2.2. Hall class of distributions

Assuming the underlying distribution  $F$  satisfies  $F(0) = 0$ , the survival function of a distribution in the Hall class (Hall, 1982; Hall and Welsh, 1985) can be written as:

$$\bar{F}(t) = C^{1/\gamma} t^{-1/\gamma} [1 + \gamma^{-1} D C^{\rho/\gamma} t^{-\rho/\gamma} + o(1)], \quad t \uparrow \infty, \quad (2.1.9)$$

where  $\gamma > 0, C > 0, \rho < 0$  and  $D \in \mathbb{R}$ .

▲

For any  $\tau = \rho/\gamma$  and  $C^{1/\gamma} \propto C$ , we can write (2.1.9) as:

$$1 - F(x) = C x^{-\alpha} [1 + D x^\tau + o(x^\tau)], \quad x \uparrow \infty, \quad (2.1.10)$$

where  $\alpha > 0, C > 0, \tau < 0$  and  $D \in \mathbb{R}$ . Equation 2.1.10 can also be written as a tail quantile function:

$$U(t) = C t^\gamma \left( 1 + \frac{A(t)}{\rho} + o(t^\rho) \right), \quad A(t) = \gamma \tau t^\rho. \quad (2.1.11)$$

Much more insightful readings on second order conditions and regular variation can be found in the famous work of de Haan and Stadtmüller (1996). Fraga Alves et al. (2007) also provides a very comprehensive overview of the first, second and third order conditions. In this thesis we do not make use of this third order frame work. Furthermore a detailed explanation of the characterisation of normalising constants  $a_n$  and  $b_n$  more specific to the first order condition, can be found in Leadbetter et al. (2012). Table 2.1<sup>1</sup> provides a short list of some of the very common distribution functions classified according to the different domains.

## 2.2 Motivating data set

In order to demonstrate the need for EVT and methods proposed in this thesis, a simplified case study in reinsurance is considered.

---

<sup>1</sup>Table 2.1 is adapted from Albrecher et al. (2017)

Domain	Distribution	$1 - F(x)$	$(+x, x_+)$
$\gamma < 0$	Weibull		
	Reversed Burr	$\beta^\alpha(\beta + x_* - x)^{-\tau})^{-\alpha}$	$(0, x_*)$
	Extreme Weibull	$1 - e^{-(1-x)^\alpha}$	$(0, x_*)$
$\gamma = 0$	Beta	$\frac{1}{B(p, q)} \int_x^1 u^{p-1}(1-u)^{q-1}$	$(0, 1)$
	Gumbel	$\frac{\lambda^\alpha}{\Gamma(\alpha)} \int_x^\infty e^{-\lambda u} du$	$(0, \infty)$
	Exponential	$e^{-\lambda x}$	$(0, \infty)$
	Log Normal	$\int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\log u - t)^2\right) du$	$(0, \infty)$
$\gamma > 0$	Fréchet		
	Strict Pareto	$(x/x_0)^{-\alpha}$	$(x_0, \infty)$
	Burr (XII)	$\beta^\alpha(\beta + x^\tau)^{-\alpha}$	$(0, \infty)$
	Log-gamma	$\int_x^\infty \frac{\lambda^\alpha}{\gamma(\alpha)} u^{-\lambda-1} \log(u)^{\alpha-1} du$	$(1, \infty)$

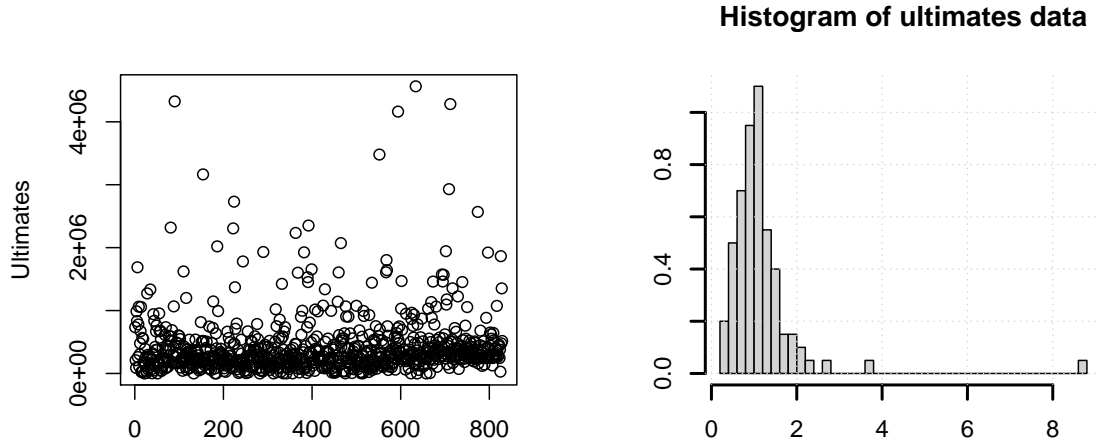
**Table 2.1:** A list of some common distributions and their respective domains of attraction

### 2.2.1 Reinsurance: Motor Liability Data

We present in this section a motor third-party liability (MTPL) data set first studied by Albrecher et al. (2017). This data contains 849 insurance European claims between 1995 and 2010 evaluated on 1 January 2011. 59% of these claims were not closed at the time of evaluation. To account for inflation and reflect costs in the calendar year 2011, all amounts have been indexed.

In order to make claim reservations, insurance companies will have experts approximate the ultimate losses still under development alongside aggregate payments and incurred losses. In case a claim is closed during the study period, the ultimate value is then the aggregate payment.

For now we shall work with the ultimate loss projections and assume that they are IID. In order to aid us in identifying the claim size distribution type, we make a



**Figure 2.3:** Scatter plot of ultimates loss data (left) and the corresponding empirical density plot (right)

min	1st quartile	median	mean	3rd quartile	max
90	167,389	286,296	422,435	485,973	4,564,759

**Table 2.2:** Summary of the ultimates data

scatter-plot and histogram of the data in Figure 2.3 and also make a summary table of the data in Table 2.2.

Let us consider a situation in which the insurance company is interested in knowing the probability of getting a claim amount larger than say 1,000,000. Secondly, what is the claim amount that can be exceeded with a probability of 0.05. This situation can be solved by classical statistical techniques.

The CLT suggests that the empirical distribution of the data carries all the information needed for inference, this implies the inability to extrapolate beyond a given population sample. The solution is therefore straight forward since the sample size is large enough and the value 1,000,000 is within the sample. Now denote the random claim by  $Y$ ,

from the 849 claims we find that only 65 claims exceed 1,000,000. The probability can thus be estimated by  $P(Y > 1,000,000) \approx 65/849 = 0.077$ .

To answer the second question we need a 95<sup>th</sup> percentile of the underlying distribution i.e. the value of  $y$  such that  $P(Y > y) = 0.05$ . We sort the data set by values from highest to lowest and discard the highest 5% of the sorted data. The succeeding highest sample is then the 95<sup>th</sup> percentile value for the data set.  $849 \cdot 0.95 = \lceil 787.55 \rceil = 788$ . This is the 788<sup>th</sup> number in the ordered list, which is 1,352,780.

We now consider a situation in which the insurance company is interested in knowing the probability of getting a claim amount larger than the largest claim amount. They might also be interested in knowing, what is the highest possible claim amount that may occur. Classical statistical methods begin to break down. For the first question: any value above the maximum is not observable and therefore the probability would be zero. A conventional statistical solution to the second problem is the maximum observable value contained in the sample. Insurance claims are known to be heavy-tailed and therefore both these answers are not plausible.

We can resort to EVT to answer these questions since it allows us to extrapolate beyond the given range of data. We can naively assert from Figure 2.3 that this data set is heavy-tailed. There are of course more plausible graphical tools for this type of analysis. These are explained in Sections 2.2.2 and 2.2.3.

For the first question, let  $F$  be the distribution function of the ultimates. The problem is that of estimating  $1 - F(x)$  with  $x$  large. Using Equation (2.1.4) it can then be restated as a tail probability for some large threshold  $t$ :

$$\begin{aligned} t(1 - F(a(t)x + U(t))) &= (1 + \gamma x)^{-1/\gamma} \\ 1 - F(a(t)x + U(t)) &= \frac{1}{t} (1 + \gamma x)^{-1/\gamma} \end{aligned}$$



$$\begin{aligned}
1 - F\left(\frac{a(t)x + U(t) - U(t)}{a(t)}\right) &= 1 - F(x) \\
&= \frac{1}{t} \left(1 + \gamma \frac{x - U(t)}{a(t)}\right)^{-1/\gamma}.
\end{aligned}$$

Assuming that the data is indeed heavy-tailed and in the Fréchet domain ( $\gamma > 0$ ), Beirlant et al. (2004) show that a suitable choice of the function  $a(t)$  can be  $a(t) = \gamma t^\gamma \ell_U(t) = \gamma U(t)$  and thus:

$$\begin{aligned}
1 - F(x) &= \frac{1}{t} \left(1 + \gamma \frac{x - U(t)}{\gamma U(t)}\right)^{-1/\gamma} \\
&= \frac{1}{t} \left(\frac{U(t)}{x}\right)^{1/\gamma} > 0.
\end{aligned}$$

To get an estimate of this we can use:

$$\hat{\hat{F}} = k/n \left(\frac{\hat{U}(n/k)}{x}\right)^{1/\hat{\gamma}}, \quad (2.2.12)$$

where  $n$  is the samples size,  $k$  is the number of observations above the threshold,  $\hat{\gamma}$  is an estimate of  $\gamma$  and  $\hat{U}(n/k)$  is an estimate of  $U(t)$ .

For the second question we need to find the right endpoint or tail quantile  $F^\leftarrow(1)=U(\infty)$ , in the  $\gamma$  negative case.

Using (2.1.3), for some positive measurable function  $a$ :

$$\begin{aligned}
\lim_{t \rightarrow \infty} \frac{U(\infty) - U(t)}{a(t)} &= \frac{x^\gamma - 1}{\gamma} \rightarrow \frac{-1}{\gamma} \\
U(\infty) &= U(t) - a(t)/\gamma.
\end{aligned}$$

This is only the case if the data set is short tailed ( $\gamma < 0$ ). If the data is heavy-tailed, the right endpoint does not exist, i.e.  $x_+ = U(\infty) = \infty$ . The insurance company will therefore resort to computing a risk measure known as value at risk (VAR), which in an extreme value context is a quantile value computed at a very small probability.

It is clear by now that the parameter  $\gamma$  plays a very important role in EVT. As will be seen in the next section. A significant amount of research has been dedicated to

the estimation of this parameter. We can make use of graphical methods to assess the goodness of fit of the proposed distribution of the data.

To a great extent, interest in EVT has been statistical fits above high thresholds (the tails). As such quantile-quantile (QQ) plots and mean excess (mean residual life) plots are the most commonly used graphical tools since they are most informative for such purposes.

### 2.2.2 Mean excess

The **Mean excess** function of a random variable  $X$  is defined as:

$$e_X(t) = E(X - t | X > t) = \frac{\int_t^\infty (1 - F)(u) du}{1 - F(t)} \quad (2.2.13)$$

provided that  $E(X) < \infty$ .

Estimates of the mean excesses for sample  $X_1, \dots, X_n$  can be naively estimated by replacing the expectation by the empirical analogue, this yields:

$$\hat{e}_{k,n} = \frac{\int_t^\infty \hat{F}(u) du}{\hat{F}(X_{n-k,n})} = \frac{1}{k} \sum_{j=1}^k X_{n-j+1,n} - X_{n-k,n} \quad (2.2.14)$$

where  $\hat{F}$  is the empirical survival function and  $t = X_{n-k,n}$  the threshold. The empirical function  $\hat{e}_{k,n}$  is plotted against values  $t = X_{n-k,n}$ ,  $k = 1, \dots, n-1$ , the  $(k+1)^{th}$  largest observation.

Using the fact that  $\log X$  is exponentially distributed with mean  $\alpha$  when  $X$  follows a strict Pareto, the mean excess function of the log transformed  $X$  can be written as:

$$e_{\log X}(\log t) = E(\log X - \log t | X > t) = \frac{\int_t^\infty (1 - F)(u) d \log u}{1 - F(t)}. \quad (2.2.15)$$

Replacing again the expectation by the empirical analogue leads to the Hill (1975) statistic:

$$H_{k,n} = \frac{\int_t^\infty \hat{F}(u) d \log u}{\hat{F}(X_{n-k,n})} = \frac{1}{k} \sum_{j=1}^k \log X_{n-j+1,n} - \log X_{n-k,n}. \quad (2.2.16)$$

The Hill estimator is one of the most famous and computationally convenient estimator of  $\gamma > 0$ .

The Mean excess can be very useful when determining a sufficient threshold for a plausible fit of a POT model. As the threshold increases, the plot may become more linear as the empirical distribution of excesses approach the POT model. Thus suggesting a choice of the threshold at the point above the non-linear portion of the plot.

### 2.2.3 QQ-plots

A QQ plot is a graphic tool used to test goodness of fit of a model to some sample population. The idea behind a QQ plot is that we can assess whether a particular model  $F$  provides a good fit to the distribution of the random variable  $X$  without having to estimate the location and scale parameters. If the model is a good fit, a plot of the ordered empirical quantiles  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$  against the quantiles  $Q(1/n), Q(2/n), \dots, Q(n/n + 1)$  associated with the proposed distribution function  $F$  should be linear. This linear relationship can be captured by eye.

The Hill (1975) estimator can be retrieved through estimating the slope  $1/\alpha$  of the regression line on a Pareto QQ plot.

Other useful diagnostic plots that can be paired along with QQ plots are known as derivative plots (see Albrecher et al., 2017, for more details). Table 2.3 <sup>2</sup> shows QQ plot coordinates for some distributions as given in Beirlant et al. (2004).

The corresponding derivative plot coordinates of the distributions in Table 2.3 are given by:

---

<sup>2</sup>Table 2.3 is adapted from Beirlant et al. (2004)

**Table 2.3:** Q-Q plot and derivatives plot coordinates.

Distribution	$F(x)$	QQ Coordinates
<b>Log-normal</b>	$\int_0^x \frac{1}{\sqrt{2\pi\sigma t}} \exp\left(-\frac{(\log t - t)^2}{2\sigma^2}\right) du$ $x > 0; t \in \mathbb{R}, \sigma > 0$	$(\Phi^{-1}(p_{i,n}), \log X_{i,n})$
<b>Exponential</b>	$1 - \exp(-\lambda x)$ $x > 0; \lambda > 0$	$(-\log(1 - p_{i,n}), X_{i,n})$
<b>Pareto</b>	$1 - x^{-\alpha}$ $x > 1; \alpha > 0$	$(-\log(1 - p_{i,n}), \log X_{i,n})$
<b>Weibull</b>	$1 - \exp(-\lambda x^\tau)$ $x > 0; \lambda, \tau > 0$	$(\log(-\log(1 - p_{i,n})), \log X_{i,n})$

**Log-normal**

$$\left( \log X_{n-k,n}, \frac{H_{k,n}}{N_{k,n}} \right)$$

$$N_{k,n} = \frac{n+1}{k+1} \phi \left( \Phi^{-1} \left( 1 - \frac{k+1}{n+1} \right) \right) - \Phi^{-1} \left( 1 - \frac{k+1}{n+1} \right)$$

$\phi$  denotes a standard normal density

**Pareto**

$$(\log X_{n-k,n}, H_{k,n}).$$

## Weibull

$$\left( \log X_{n-k,n}, \frac{H_{k,n}}{W_{k,n}} \right)$$

$$W_{k,n} = \frac{1}{k} \sum_{j=1}^k \log \log \frac{n+1}{j} - \log \log \frac{n+1}{k+1}.$$

The derivative plot of the **Exponential** distribution is the Mean excess plot.

As an initial step in extreme value analysis, it is important that we classify the tail of the distribution function  $F$  as being exponential, heavier than exponential or lighter than exponential. The graphical diagnostics are summarised below.

### Lighter than exponential (LTE):

- The exponential QQ plot will be ultimately convex tipping above the fitted regression line of the coordinates;
- The Mean excess function will be ultimately increasing.

### Exponential tail:

- The exponential QQ plot will be linear;
- The Mean excess plot will be ultimately horizontal;
- The Pareto QQ plot will be ultimately concave and tipping below the fitted regression line.

### Heavier than exponential (HTE):

- The exponential QQ plot will be ultimately concave tipping below the fitted regression line of the coordinates;
- The Mean excess plot will be ultimately decreasing;

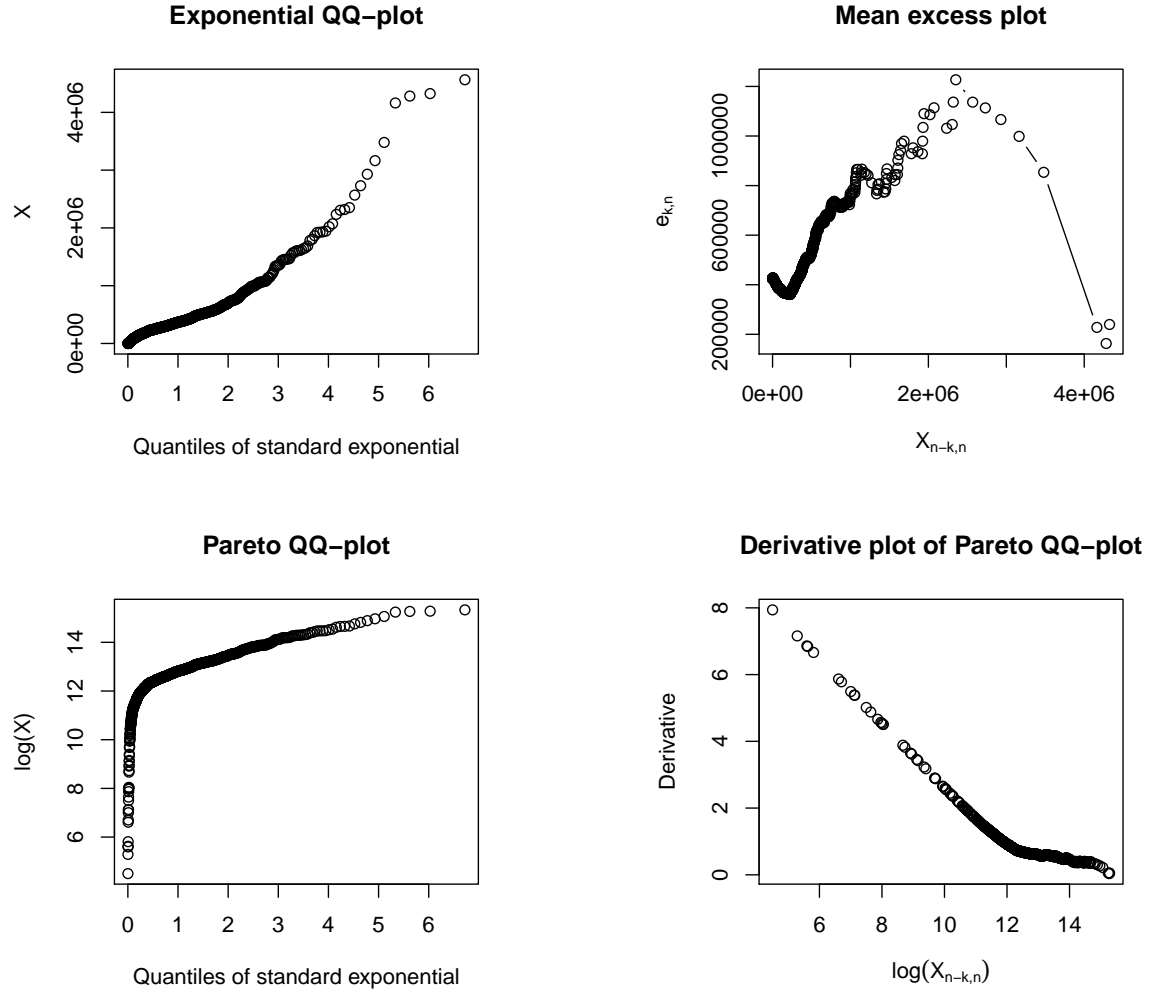
- The Pareto QQ plot will ultimately be linear

If however the practitioner is interested in determining whether the tail of the distribution function is Pareto, heavier than Pareto or lighter than Pareto, a similar approach can be taken replacing the Exponential QQ plot with a Pareto QQ plot and the Mean excess with a Pareto derivative plot. In an analogous manner, distributions in the Weibull domain can be inspected.

Diagnostic plots of the ultimates data set are given in Figure 2.4. These plots confirm that the data does in fact exhibit heavy tails.

It should be noted that data can have different regimes since the natural phenomena governing the process from which data is generated can vary. For example, the way an insurance company deals with small claims will be different from the way they deal with large claims.

It can be seen in Figure 2.4 that the ultimates loss projections appear to exhibit some form of a truncated regime for larger claims.



**Figure 2.4:** Diagnostic plots for the ultimates data

## 2.3 Tail Estimation

The two most common extreme value approaches for estimating the EVI and corresponding tail quantities thereof are the Block maxima method and the POT method. A short description of these methods are given below.

### 2.3.1 Block maxima

The method of block maxima is motivated by the limit behaviour of the normalised maximum as given in (2.1.1), when a non-degenerate distribution exists. The GEV distribution given in (2.1.2), now with the location and scale parameter, can be fitted to the maxima of the sub sample (Gumbel, 1958)

$$G_\gamma(x) = \begin{cases} \exp \left\{ - \left( 1 + \gamma \frac{x-t}{\sigma} \right)^{-1/\gamma} \right\}, & \text{if } \gamma \neq 0 \\ \exp \left\{ - \exp \left( \frac{x-t}{\sigma} \right) \right\}, & \text{if } \gamma = 0 \end{cases} \quad (2.3.17)$$

defined on  $\{x : 1 + \gamma(\frac{x-t}{\sigma}) > 0\}$ , where  $\sigma > 0$ ,  $t \in \mathbb{R}$  and  $\gamma \in \mathbb{R}$ .

If the IID data are divided into say  $m$  blocks then this sub sample consists of the maxima (top observation) extracted from each block. The parameters of the GEV distribution can be estimated either through data analytic methods discussed in Beirlant et al. (2004) or using the Maximum likelihood or Probability weighted moment estimators, to name a few.

Corresponding quantile estimates of the GEV can be obtained by inverting the GEV distribution and replacing the parameters  $(\gamma, \sigma, t)$  with the estimates from an estimation procedure of choice, giving:

$$z_p = \begin{cases} t - \frac{\sigma}{\gamma} \{1 - [-\log(1-p)]^{-\gamma}\}, & \text{if } \gamma \neq 0 \\ t - \sigma \log[-\log(1-p)], & \text{if } \gamma = 0. \end{cases} \quad (2.3.18)$$

The block maxima approach suffers some drawbacks. Useful data (extreme values) is discarded in the process of extracting sub samples. Furthermore if block sizes are too small, this can lead to poor asymptotic approximations. Overly large block sizes lead to fewer observations available for inference thus leading to a high variability in the estimates.



One of the ways to overcome this drawback is to model all the data above a sufficiently high threshold. This method is known as the POT approach.

### 2.3.2 Peaks over threshold approach

Consider the sequence of IID random variables  $X_1, X_2, \dots, X_n$ , the distribution of conditional excesses  $Y = X - t$  above a suitably high threshold  $t$  is given by:

$$F_t(y) = P(X - t \leq y | X > t) = \frac{F(y + t) - F(t)}{1 - F(t)}. \quad (2.3.19)$$

Pickands (1975) showed that for any given function  $F$  under (2.1.1), the generalised Pareto distribution (GPD) arises as the limit distribution of  $F_t(y)$  as  $t \rightarrow \infty$ . The distribution function of the GPD is given by:

$$G_{\gamma, \sigma_t, t} = \begin{cases} 1 - \left\{ 1 + \gamma \left( \frac{x - t}{\sigma_t} \right) \right\}^{-1/\gamma}, & \text{if } \gamma \neq 0 \\ 1 - \exp \left\{ -\frac{x - t}{\sigma_t} \right\}, & \text{if } \gamma = 0 \end{cases} \quad (2.3.20)$$

where  $x > t$ ,  $\sigma_t > 0$ ,  $\left\{ 1 + \gamma \left( \frac{x - t}{\sigma_t} \right) \right\} > 0$ . The scaling parameter  $\sigma_t$  is dependent on  $t$  and the GPD shape parameter is the same as the GEV shape parameter. Retrospectively: the GPD appears as the limiting distribution for scaled excesses over a high threshold while the GEV describes the limit distribution of the normalised maxima.

There exists numerous methods for estimating the parameters of the GPD. The most commonly used method is the maximum likelihood (ML) method (see Smith, 1987). Hosking and Wallis (1987) derived a simple method of moments estimate of  $\gamma$  which only works for  $\gamma < 0.5$ . Castillo and Hadi (1997) further proposed an elemental percentile method which does not impose the  $\gamma < 0.5$  restriction. Coles and Powell (1996) proposed the use of Bayesian methods.

The corresponding quantile estimates of the GPD can be obtained by inverting (2.3.20),

which yields

$$z_p = t + \sigma_t \frac{\left(\frac{N_t}{np}\right)^\gamma - 1}{\gamma} \text{ for } p < \frac{N_t}{n}. \quad (2.3.21)$$

Another approach not baring focus in this thesis is the quantile view which rely on versions of the first order condition in (2.1.1). Examples following this approach include but is not limited to the following:

- **Pickands estimator:**

$$\hat{\gamma}_{P,k} = \frac{1}{\log 2} \log \left( \frac{X_{n-\lceil k/4 \rceil+1,n} - X_{n-\lceil k/2 \rceil+1,n}}{X_{n-\lceil k/2 \rceil+1,n} - X_{n-k+1,n}} \right).$$

- **Moment estimator:**

$$M_{k,n} = H_{k,n} + 1 - \frac{1}{2} \left( 1 - \frac{H_{k,n}^2}{H_{k,n}^{(2)}} \right)^{-1}$$

where

$$H_{k,n}^{(2)} = \frac{1}{k} \sum_{j=1}^k (\log X_{n-j+1,n} - \log X_{n-k,n})^2.$$

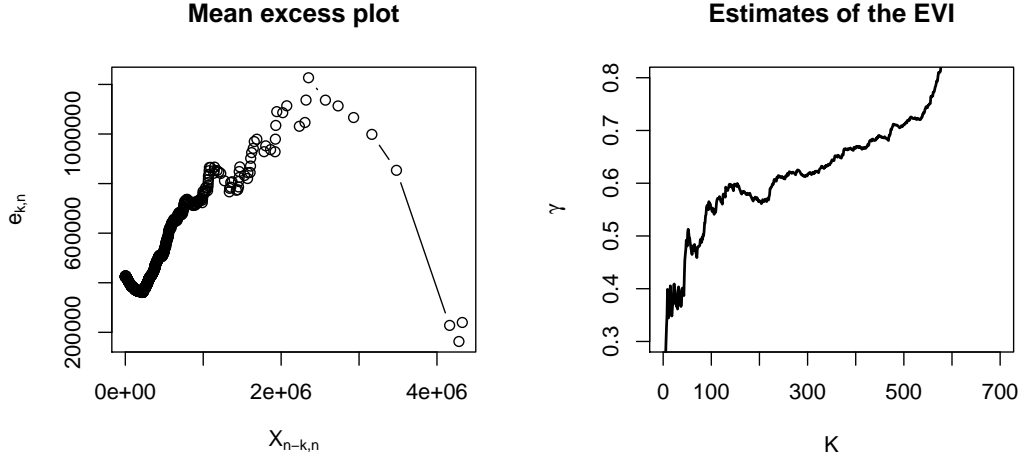
It should be noted that in this thesis, the focus is solely placed on improving POT methods.

### 2.3.3 Fitting Pareto type data

#### Uncensored data

Pareto tail modelling is the most common approach to modelling large insurance claims. This is because large insurance claims generally exhibit heavy tails. We now focus only on a subset of models for which  $\gamma > 0$ . Indeed if  $\gamma > 0$  then the tail quantile function is given by (Beirlant et al., 2004):

$$\bar{F}(x) = x^{-1/\gamma} \ell_F(x) \quad x \uparrow \infty \quad (2.3.22)$$



**Figure 2.5:** Mean Excess (left) and Hill estimates of the EVI (right) of the ultimates

where  $\ell_F(x)$  is a slowly varying function. This statement is equivalent to

$$\frac{\bar{F}(ty)}{\bar{F}(t)} = P\left(\frac{X}{t} > y | X > t\right) \rightarrow y^{-1/\gamma}, \quad t \uparrow \infty \quad (2.3.23)$$

and bears resemblance to (2.3.19) which was stated in terms of the additive (or absolute) excesses  $X - t$ . For Pareto type data it becomes much more convenient to work with multiplicative excesses  $X/t$ .

Let the ultimates be IID random variables  $X_1, \dots, X_n$ . Equation (2.3.23) can be used to retrieve the Hill (1975) estimator, based on the POT approach, by fitting a strict Pareto distribution to the relative excesses using maximum likelihood.

It can be seen from Figure 2.5 (right) that the  $e_{k,n}$  increases with increasing  $X_{n-k,n}$ , indicating that the tail of the ultimates data is heavier than exponential and thus validating the use of an estimator of a positive  $\gamma$  such as the Hill estimator. Figure 2.5 (right) shows a plot of the Hill estimates for the ultimates data as a function of the  $k$  (number of observations above the threshold). The plot has no apparent form of stability. This problem is very common in practice where the Hill estimator is volatile and hard to interpret and hence commonly referred to as the “**Hill horror plot**”. Although there are numerous methods to aid a practitioner in making a suitable

choice of the right value of  $k$ . This choice however is to some degree, subjective. This is further discussed in Section 2.3.4.

### Censored data

One of many difficulties related to tail estimation in EVT is modelling data that is subject to random right censorship. This problem is very common in areas such as reinsurance and survival analysis in clinical trials.

**Clinical trials:** suppose for example a study is conducted to see how long patients can survive when placed on a certain drug. At the end of the study, some of these patients may still be alive. This is an example of right censoring. The quantity in question (time to death) is not fully observable. Gomes and Neves (2011) conducted such a case study to survival data sets.

**Insurance:** here we can revisit the MTPL data studied in Section 2.2. We have discussed the ultimate loss projections made by experts. This data set also contains indexed total incurred values as well as the indexed total paid values.

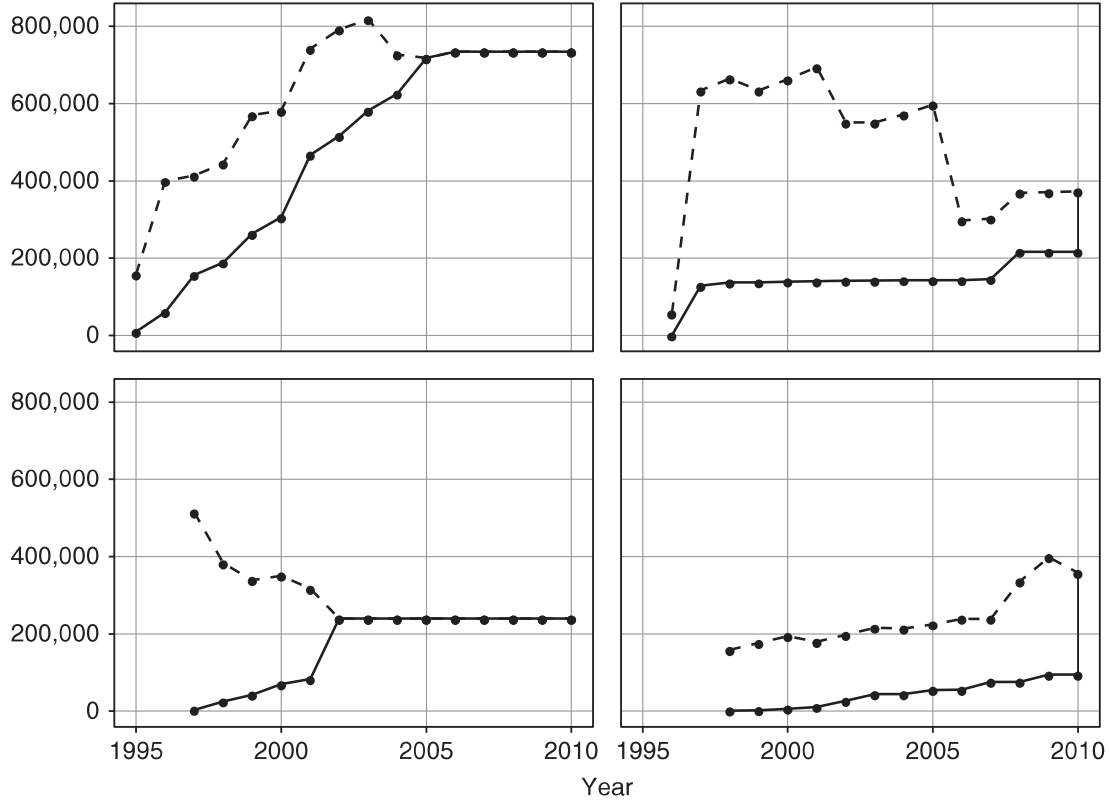
Figure 2.6 <sup>3</sup> illustrates the development of four claims occurring 1995, 1996, 1997 and 1998. The full line indicates the cumulative indexed payment, while the dashed lines indicates the indexed incurred values.

The incurred values at a particular point in time are the sum of the already paid amounts and the amounts reserved as estimated by experts. When the two lines meet i.e. the cumulative payment and the incurred value are equal, and the claim is closed.

This happens in the first and third claim while the second and fourth claim appear to still be in development. These two claims (second and fourth claim) are thus right censored at the time of evaluation (end of 2010).

---

<sup>3</sup>Figure 2.6 is taken from Albrecher et al. (2017) with authorisation from the authors



**Figure 2.6:** Development patterns for four particular claims

There will be an inevitable amount of uncertainty in the approximation of ultimate loss projections. Instead of relying on these ultimate loss projections when modelling tail behaviour we can adopt a right censoring framework and only use the indexed total payments at the end of the evaluation year as lower bounds.

Consider  $X = \text{index total incurred}$  and  $C = \text{indexed total paid}$ . We shall work with observations  $Z = \min(X, C)$ .

- If a claim is closed at portfolio evaluation then  $Z = X$  is observable and the censoring indicator  $\Delta = 1$ ;
- If a claim is still under development at portfolio evaluation then  $Z = C$ , thus not completely observable and the censoring indicator  $\Delta = 0$ ;

Therefore  $Z \leq X$  and we use the ordered sample  $\{Z_{i,n}, \Delta_{i,n}, i = 1, \dots, n\}$ . This case study is conducted with more detail in Section 4.6.

Over the last couple of years, EVT under random right censoring has received considerable attention in literature. A good overview of the subject was made by Beirlant and Guillou (2001) who proposed adapting the Hill estimator under random right censorship.

Under some conditions Delafosse and Guillou (2002) were able to prove almost sure convergence of this adapted Hill estimator under random right censorship. The authors Reiss and Thomas (2007) proposed a POT framework by adapting the likelihood function of the distribution of excesses above a threshold. However, the authors did not study the asymptotic properties of the proposed estimators.

Beirlant et al. (2007) proposed a different approach, generalisation of the POT methodology using the GPD. They also proposed an adaptation of the moment estimator developed by Dekkers et al. (1989).

Einmahl et al. (2008) considered a general adaptation of various estimators of the EVI under random censorship. Under some restrictive conditions, the authors proved the asymptotic normality result of the estimators in detail. Brahimi et al. (2015a) proved the asymptotic normality result of the adapted Hill estimator under more relaxed conditions.

Gomes and Neves (2011) provided a good overview of the EVI estimators under random censoring and proposed an adapted minimum-variance reduced-bias (MVRB) estimator (Caeiro et al., 2005). Using samples from models in the Fréchet domain the authors conducted a simulation study which demonstrated the overall best performance of their adapted MVRB estimator.

Worms and Worms (2014) presented a methodological paper for the estimation of the EVI under random right censorship, based on the ideas of Kaplan-Meier integration and the synthetic data approach of Leurgans (1987). We show in this thesis how to reinterpret their approach using the mean excess function. Under mild censoring, the authors were able to prove consistency of the proposed estimators.

Ameraoui et al. (2016) proposed a Bayesian method to estimate the EVI under random right censoring by developing maximum a posteriori and mean posterior estimators using the Maximal Data Information (MDI), Jeffreys and the conjugate Gamma priors. The authors also establish consistency and asymptotic normality of the proposed estimators.

In this thesis a bias reduced version of the Worms and Worms (2014) estimator is proposed in Chapter 4, and the idea of shrinkage introduced in Chapter 3 is further applied to the proposed estimator. Beirlant, Worms and Worms (2018) study the asymptotic and finite sample behaviour of the large class of estimators first proposed in Worms and Worms (2014), paying special attention to heavy censoring.

### 2.3.4 Threshold selection

An important question in practical applications of statistics of extremes is the threshold choice. There exists a number of methods available for practitioners to aid in choosing a suitable threshold. The threshold selection problem is however still an open research question. The threshold selection problem is a bias-variance trade off problem (Coles, 2001). By counterpoint, if the threshold is too high it will lead to fewer excesses to estimate the parameters and will thus lead to a high variance. However if the threshold

is too low, the asymptotic motivation of the underlying POT model no longer holds and thus a large bias will be present.

The fixed threshold approach is very common in EVT tail estimation. In this approach, the threshold is chosen prior to model fitting. The threshold choice is hereon interchangeably mentioned as; the threshold value from which to consider the data as extremes or the choice of the number ( $k$ ) of upper extremes.

DuMouchel (1983) proposed a simple quantity based rule of thumb to consider the top 10% largest observations as the tail fraction. Loretan and Phillips (1994) proposed a data driven approach to choose  $k = \frac{n^{2/3}}{\log[\log(n)]}$ . Ferreira et al. (2003) proposed to choose the tail fraction as  $k = \sqrt{n}$  through simulation studies.

Hall (1982), Dekkers and de Haan (1993) and de Haan and Peng (1998) proposed choosing a  $k$  that minimizes the asymptotic MSE of any proposed estimator. Beirlant et al. (1999) introduced a heuristically motivated procedure by directly estimating the asymptotic mean squared error. Drees and Kaufmann (1998) proposed a sequential approach to selecting the optimal sample fraction. Their approach requires an arbitrary choice of a tuning parameter. Draisma et al. (1999), Gomes and Oliveira (2001) and Danielsson et al. (2001) used bootstrap methods to choose the optimal sample fraction. Matthys and Beirlant (2000) provide a review of several adaptive threshold selection techniques. See also Section 4.7 in Beirlant et al. (2004).

Dupuis (1999) proposes a method of threshold estimation for the GPD. The procedure assigns to each observation a weight between zero and one. A high weight indicates a good fit and thus the data point should be retained. The author suggests starting with a lower choice of the threshold and increasing it until the weights are all ‘close to one’. Coles (2001) show that diagnostic plots such as the QQ plot, parameter stability plot and the mean excess plot can be used to choose the threshold. Refer to Caeiro



and Gomes (2015) for a recent review on the subject of threshold selection.

A concern with using the fixed threshold approach is that the threshold value consistent with GPD assumptions is regarded as known and extrapolation is then conducted given this known value. Thus threshold uncertainty is not naturally taken into account. Moreover, in comparing the estimation results from different thresholds, it is often found that plausible but different threshold values, lead to different EVI estimates.

### 2.3.5 Bias reduction

In order to address the inadequacies of threshold selection under the POT framework, the concept of bias reduction has been studied extensively. Many of these studies however are focused mainly on the  $\gamma > 0$  case.

Peng (1998), Caeiro and Gomes (2002), Gomes and Martins (2002, 2004) and Gomes et al. (2002, 2005) proposed bias reduction methods that seek to cancel out the bias terms through an indirect approach of using weighted combinations of different estimators.

Beirlant et al. (1999, 2008), Feuerverger et al. (1999) and Gomes et al. (2000, 2007) proposed bias reduction methods based on scaled log-spacings between subsequent high order statistics from a Pareto-type distribution.

Caeiro et al. (2005) proposed minimum variance reduced bias estimators of the EVI, where the bias is reduced without increasing the variance with respect to the Hill estimator. This method is based on adequate external estimation of a pair of parameters of second order slow variation under a third order condition which we do not discuss in this thesis. Following this work, proposals by Gomes and Pestana (2007*a,b*), Gomes

et al. (2008) and Caeiro and Gomes (2008) consider directly subtracting the bias term as given by the asymptotic distribution.

Beirlant et al. (2009) proposed a much more flexible model adequately capturing the deviation between the true excess right tail function and the asymptotic Pareto model. This approach substantially reduced bias by considering the Hall-class of Pareto-type models in (2.2). This approach often leads to much improved results, both in bias and MSE.

### 2.3.6 Mixture modelling in EVT

The main motivation behind use of mixture models in EVT is to improve flexibility of tail models, which generally only focus on the tail part of the data and not the complete data set. This area of research has over the last decade gained considerable interest. Frigessi et al. (2002) proposed an unsupervised tail estimation technique using a dynamic mixtures of two components with a weight function  $\pi = \pi(x)$  smoothly connecting the bulk and the tail of the distribution. Beirlant et al. (2004) suggested the need for a global statistical model describing the entire range of possible claim outcomes using an exponential and a Pareto mixture.

Behrens et al. (2004) considered a Bayesian approach of fitting a parametric mixture model using indirect expert elicited prior distributions. This Bayesian posterior inference does however have a major draw back. Since it treats the prior on the threshold parameter and other parameters as independent, the dependence between the scale and threshold parameters is simply ignored. Tancredi et al. (2006) also accounts for threshold uncertainty by estimating the threshold as one of the parameters in the mixture model.

Carreau and Bengio (2009b) proposed a hybrid Pareto model by stitching a GPD tail with a normal distribution and setting a continuity constraint on the density and its first derivative. Performance of this method is however poor in practice. Improvement

and extension of the performance of this hybrid Pareto model is made by Carreau and Bengio (2009*a*).

MacDonald et al. (2011) considered a mixture approach where the bulk of the distribution is estimated by a standard kernel density estimator. Using penultimate theory Wadsworth and Tawn (2012) account for threshold uncertainty and provide a likelihood ratio testing procedure for the threshold choice.

By applying the inverse CDF of the GPD to more richer Beta distributed draws instead of uniform draws, Papastathopoulos and Tawn (2013) proposed general extensions to the GPD distribution. Following from this work, Naveau et al. (2016) proposed a version of the extended statistical model called the extended generalised Pareto distribution (EGPD) which is in compliance with EVT and allows for a smooth transition between the modal and tail part of the data. Building further from this work, Tencaliec et al. (2018) proposed a flexible semiparametric GP modelling approach by estimating the key component of the EGPD class in Naveau et al. (2016) using Bernstein polynomials.

## 2.4 Data sets and Software

In this section, we give a description of all the data sets used in this thesis for practical illustrations. A description of the software used to conduct simulation studies is also given.

### 1. Secura Belgian Re data

The Secura Belgian Re is studied in Chapter 3 as first studied by Beirlant et al. (2004). The data set contains 371 automobile claims larger than or equal to 1,200,000 Euros for the year 1988 to 2001, collected from various insurance companies in Europe. All amounts have been indexed.

### 2. Motor third-party liability (MTPL) data

Details of this data are given in Section 2.2. The data set is further studied in Chapter 4.

### 3. Mont-Aigoual station rainfall data

This rain fall data is studied in Chapter 5. The data set was first studied by Carreau et al. (2017) and Tencaliec et al. (2018). It contains daily precipitation of all seasons from the year 1976 to 2015 recorded at the Mont-Aigoual station, south of France.

## Software

All simulations in this thesis were conducted using the **R** (R Development Core Team, 2008) software through a Linux based high performance computing unit. The simulation study conducted in Chapter 5 are more extensive and thus resulted in large output data files. To avoid making too many plots, an interactive web application is build using an **R** package called shiny (Chang et al., 2018) and hosted as a standalone application on the following url: <https://phdshinygao.shinyapps.io/ExtendedModels/>. Guidelines regarding usage of the interactive web application are given in Appendix 5.B.

## 2.5 Chapter conclusion

In this chapter, an overview of EVT has been given, along with a review of literature leading up to proposals made in this thesis. A motivating data set was used to give a short illustration of the need for EVT, moreover discuss a growing research interest of an EVT framework for dealing with censored extremes. Finally, descriptions of all the data sets and software used in this thesis were given.

The next chapter marks the beginning of Part I of the thesis which will introduce

the method of shrinkage estimation in Chapter 3 and propose a bias reduced estimator of the EVI in Chapter 4. Chapter 5 in Part II of the thesis will propose a novel bias reduction technique in all domains of attraction and an overall conclusion will be made in Chapter 6.

# Part I

## Second-order Refinements: Heavy Tails ( $\gamma > 0$ )

## Chapter 3

# Using Shrinkage Estimators to Reduce Bias and MSE in Estimation of Heavy Tails

*This chapter is based on Beirlant, J., Maribe, G. and Verster, A., 2019. Using shrinkage estimators to reduce bias and MSE in estimation of heavy tails. REVS-TAT-Statistical Journal, 17(1), pp.91-108.*

### 3.1 INTRODUCTION

In this section we consider the estimation of the EVI  $\gamma$  and tail probabilities  $P(X > x)$  for  $x$  large, on the basis of independent and identically distributed observations  $X_1, X_2, \dots, X_n$  which follow a Pareto-type distribution with right tail function (RTF) given by

$$\bar{F}(x) = 1 - F(x) = P(X > x) = x^{-1/\gamma} \ell(x) \quad (3.1.1)$$

where  $\ell$  is a slowly varying function at infinity, *i.e.*

$$\frac{\ell(ty)}{\ell(t)} \rightarrow 1, \text{ as } t \rightarrow \infty, \text{ for every } y > 1.$$

The most famous estimator of  $\gamma$  was first derived by Hill (1975) as a ML estimator approximating the RTF of the excesses  $\frac{X}{t}|X > t$  over a large threshold  $t$  by a simple Pareto distribution with RTF  $y^{-1/\gamma}$ :

$$\bar{F}(ty)/\bar{F}(t) \approx y^{-1/\gamma}, t \text{ large.} \quad (3.1.2)$$

When setting  $t = X_{n-k,n}$  where  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$  the ML estimator is given by

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^k \log \frac{X_{n-j+1,n}}{X_{n-k,n}}. \quad (3.1.3)$$

A simple estimator of a tail probability  $P(X > x)$  with  $x$  large, introduced in Weissman (1978), is then obtained from (3.1.2) setting  $ty = x$  and estimating  $P(X > t)$  by the empirical proportion  $k/n$ :

$$\hat{p}_{x,k} = \frac{k}{n} \left( \frac{x}{X_{n-k,n}} \right)^{-1/H_{k,n}}. \quad (3.1.4)$$

In practice, a way to verify the validity of model (3.1.1) is to check whether the Hill estimates are stable as a function of  $k$ . However in most cases the stability is not visible, which can be explained by slow convergence in (3.1.2). For this reason bias reduced estimators have been proposed which lead to plots that are much more horizontal in  $k$  which facilitates the analysis of a practical case to a great extent.

Beirlant et al. (2009) proposed a more flexible model capable of capturing the deviation between the true excess RTF  $\bar{F}(ty)/\bar{F}(t)$  and the asymptotic Pareto model. For a heavy tailed distribution (3.1.1), this deviation can be parametrized using a power series expansion (Hall, 1982), or more generally via second-order slow variation (Bingham et al., 1987). More specifically in Beirlant et al. (2009) the subclass  $\mathcal{F}(\gamma, \tau)$  of the Pareto-type tails (3.1.1) was considered satisfying

$$\bar{F}(x) = Cx^{-1/\gamma} (1 + \gamma^{-1}\delta(x)), \quad (3.1.5)$$



with  $\delta(x)$  eventually nonzero and of constant sign such that  $|\delta(x)| = x^\tau \ell_\delta(x)$  with  $\tau < 0$  and  $\ell_\delta$  slowly varying. It was shown that under  $\mathcal{F}(\gamma, \tau)$  as  $t \rightarrow \infty$

$$\sup_{y \geq 1} \left| \frac{\bar{F}(ty)}{\bar{F}(t)} - \bar{G}_{\gamma, \delta, \tau}(y) \right| = o(|\delta(t)|)$$

with  $\bar{G}_{\gamma, \delta, \tau}$  the RTF of the EPD

$$\bar{G}_{\gamma, \delta, \tau}(y) = \{y(1 + \delta - \delta y^\tau)\}^{-1/\gamma}, \quad y > 1, \quad (3.1.6)$$

with  $\tau < 0 < \gamma$  and  $\delta > \max(-1, 1/\tau)$ . This shows that the EPD improves the approximation (3.1.2) with an order of magnitude. Then ML estimation of the parameters  $(\gamma, \delta)$  based on a set of excesses  $(Y_{j,k} := X_{n-j+1,n}/X_{n-k,n}, j = 1, \dots, k)$  was used to obtain a bias reduced estimator  $\hat{\gamma}_{k,n}^{ML}$  of  $\gamma$ . Bias reduction of the Weissman estimator of tail probabilities can analogously be obtained using

$$\hat{p}_{x,k}^{EP} = \frac{k}{n} \bar{G}_{\hat{\gamma}_k, \hat{\delta}_k, \hat{\tau}} \left( \frac{x}{X_{n-k,n}} \right), \quad (3.1.7)$$

where  $(\hat{\gamma}_k, \hat{\delta}_k)$  denote the ML estimators based on the EPD model, and where  $\hat{\tau}$  is a consistent estimator of  $\tau$ , to be specified below, which was shown not to affect the asymptotic distribution of  $(\gamma, \delta)$ .

If  $F$  satisfies  $\mathcal{F}(\gamma, \tau)$ , it is shown in Beirlant et al. (2009) that  $U(x) := Q(1 - x^{-1})$  ( $x > 1$ ), with  $Q(p) = \inf\{x : F(x) \geq p\}$  ( $p \in (0, 1)$ ), satisfies

$$U(x) = C_a \gamma x^\gamma (1 + a(x)) \quad (3.1.8)$$

with  $a(x) = \delta(Q(1 - x^{-1}))(1 + o(1)) = \delta(C_a \gamma x^\gamma)(1 + o(1))$  as  $x \rightarrow \infty$ . In particular  $a$  is eventually nonzero and of constant sign and  $|a(x)| = x^\rho \ell_a(x)$  with  $\ell_a$  slowly varying and  $\rho = \gamma\tau$ . Here we assume  $|\ell_a(x)| = C_a(1 + o(1))$  as  $x \rightarrow \infty$  for some constant  $C_a > 0$ .

The following asymptotic results have been derived for  $H_{k,n}$  and  $\hat{\gamma}_{k,n}^{ML}$  assuming that  $F$  satisfies  $\mathcal{F}(\gamma, \tau)$ , and  $\sqrt{k}a(n/k) \rightarrow \lambda \in \mathbb{R}$  and  $\hat{\rho}_{k,n} = \rho + o_p(1)$  as  $k, n \rightarrow \infty$  and

$k/n \rightarrow 0$ , the following asymptotic results hold for the EPD-ML estimator  $\hat{\gamma}_{k,n}^{ML}$  and  $H_{k,n}$ :

$$\sqrt{k}(H_{k,n} - \gamma) \rightarrow^d \mathcal{N}\left(\lambda \frac{\rho}{1-\rho}, \gamma^2\right), \quad (3.1.9)$$

$$\sqrt{k}(\hat{\gamma}_{k,n}^{ML} - \gamma) \rightarrow^d \mathcal{N}\left(0, \gamma^2 \left(\frac{1-\rho}{\rho}\right)^2\right). \quad (3.1.10)$$

An estimator  $\hat{\rho}_{k,n}$  of  $\rho$  can be taken from Fraga Alves et al. (2003) using  $k = k_1 = \lfloor n^{1-\epsilon} \rfloor$  for some  $\epsilon > 0$  (see, Appendix 3.C). The required consistency for  $\hat{\rho}_{k,n}$  was obtained under (3.1.8).

Asymptotic results of the type (3.1.9) and (3.1.10) are typical for bias reduced estimators when both  $\gamma$  and  $a(n/k)$  or  $\delta$  are jointly estimated at every  $k$  value: for larger values of  $k$  corresponding to  $\sqrt{k}a(n/k) \rightarrow \lambda \neq 0$ , bias reduced estimators still have asymptotic bias 0 in contrast to the Hill estimator, but their variance is increased by a factor  $((1-\rho)/\rho)^2$  compared to  $H_{k,n}$ .

In a pioneering paper, Caeiro et al. (2005) proposed to estimate  $(n/k)^{-\rho}a(n/k)$  at a high level  $k = k_1 = \lfloor n^{1-\epsilon} \rfloor$ , leading to a corrected Hill estimator (denoted below by  $CH_{k,n}$ ) with asymptotic variance  $\gamma^2$  and excellent bias and MSE characteristics. To obtain the normal asymptotic behaviour of such minimum variance reduced bias estimators one needs a third-order slow variation condition which is more restrictive than (3.1.8) or condition  $\mathcal{F}(\gamma, \tau)$ .

Up to now, to the best of our knowledge, the fact that  $\delta(t) \rightarrow 0$  as  $t \rightarrow \infty$ , or  $a(n/k) \rightarrow 0$  as  $n/k \rightarrow \infty$  has not been exploited in the literature. However, this calls for shrinkage estimators. Such shrinkage approach can be implemented by putting a penalty on  $\delta$  in an ML procedure, leading to penalised ML. Alternatively a penalty on  $\delta$  can be naturally introduced in a Bayesian approach putting an appropriate prior on this parameter.

Here we investigate the use of shrinkage estimation when modelling the distribution of the vector of excesses  $\mathbf{Y}_k := (Y_{j,k}, j = 1, \dots, k)$  with an EPD. In Section 3.2 we show that a quadratic penalty, or equivalently a normal prior, on  $\delta$  with zero mean and variance  $\sigma_{k,n}^2$ , depending in an appropriate way on  $k$  and  $n$ , leads to interesting asymptotic MSE results for  $\gamma$ . In Section 3.3 we consider the finite sample behaviour of the penalised likelihood and Bayes approach, and make a comparison with the minimum variance reduced bias estimator, and consider a practical case.

## 3.2 Shrinkage estimators of the EPD parameters

### 3.2.1 penalised likelihood and Bayesian interpretation

ML estimation of the EPD parameters  $(\gamma, \delta)$ , given a value of  $\tau$ , follows by maximising the log-likelihood

$$\begin{aligned} \frac{1}{k} l_{EP}(\gamma, \delta | \mathbf{y}) &= -\log \gamma - \left(\frac{1}{\gamma} + 1\right) \frac{1}{k} \sum_{j=1}^k [\log y_{j,k} + \log(1 + \delta\{1 - y_{j,k}^\tau\})] \\ &\quad + \frac{1}{k} \sum_{j=1}^k \log(1 + \delta\{1 - (1 + \tau)y_{j,k}^\tau\}). \end{aligned} \quad (3.2.11)$$

Shrinkage estimators are then obtained by putting a penalty on  $\delta$ . Below it will be shown that a quadratic penalty is appropriate in view of the asymptotic results for the penalised maximum likelihood (PML) estimators  $(\hat{\gamma}_{k,n}^P, \hat{\delta}_{k,n}^P)$ . These estimators are then obtained by optimising the log-likelihood

$$\frac{1}{k} l_{pen}(\gamma, \delta | \mathbf{y}) = \frac{1}{k} l_{EP}(\gamma, \delta | \mathbf{y}) - \omega \frac{\delta^2}{2k\sigma_{k,n}^2}, \quad (3.2.12)$$

where  $\omega > 0$  serves as a tuning constant regulating the amount of penalty, and  $\sigma_{k,n}^2$  indicating the penalty rate as a function of  $k$ . From the asymptotic analysis below, it follows that  $\sigma_{k,n}^2 = (k/n)^{-2\rho}$  is appropriate.

Alternatively, from a Bayesian perspective, a shrinkage estimator is obtained by considering the posterior mode estimators  $(\hat{\gamma}_k^B, \hat{\delta}_k^B)$  of the log-posterior

$$\frac{1}{k} \log p(\gamma, \delta | \mathbf{y}) = \frac{1}{k} l_{EP}(\gamma, \delta | \mathbf{y}) + \frac{1}{k} \log \pi(\gamma, \delta), \quad (3.2.13)$$

where  $\pi(\gamma, \delta)$  denotes the prior density on  $(\gamma, \delta)$ . In correspondence with the choice for the penalised log-likelihood (3.2.12), we here choose a normal prior on  $\delta$  with mean 0 and variance  $\sigma_{k,n}^2$ . We also truncate it from the left in order to comply with the restriction  $\delta > \max(-1, 1/\tau)$ :

$$\pi(\delta) = \frac{1}{\sqrt{2\pi}\sigma_{k,n}} e^{-\frac{1}{2}\frac{\delta^2}{\sigma_{k,n}^2}} / (1 - \Phi(\max(-1, \tau^{-1})/\sigma_{k,n})). \quad (3.2.14)$$

### 3.2.2 Asymptotic results for the penalised ML estimator $\hat{\gamma}_k^P$

In Appendix 3.A we derive that the first order approximations  $(\hat{\gamma}_k^P, \hat{\delta}_k^P)$  of the penalised ML estimators are given by

$$\begin{aligned} \hat{\gamma}_k^P &= H_{k,n} + \hat{\delta}_k^P (1 - E_{k,n}(\tau)), \\ \hat{\delta}_k^P &= \frac{1 - H_{k,n}\tau}{D_{k,n}^P} \left( E_{k,n}(\tau) - \frac{1}{H_{k,n}\tau} \right) \end{aligned}$$

where

$$E_{k,n}(s) = \frac{1}{k} \sum_{j=1}^k Y_{j,k}^s, \quad s < 0,$$

and

$$\begin{aligned} D_{k,n}^P &= \frac{\omega \hat{\gamma}_k^P}{k \sigma_{k,n}^2} - \left( 1 - 2(1 - \hat{\gamma}_k^P \tau) E_{k,n}(\tau) + (1 - 2\hat{\gamma}_k^P \tau - \hat{\gamma}_k^P \tau^2) E_{k,n}(2\tau) \right. \\ &\quad \left. - \tau(1 - E_{k,n}(\tau)) E_{k,n}(\tau) \right). \end{aligned}$$

These expressions are identical to the asymptotic EPD-ML estimators derived in Beirlant et al. (2009) except for the extra term  $\frac{\omega \hat{\gamma}_k^P}{k \sigma_{k,n}^2}$  in the expression of  $D_{k,n}^P$ . As an external estimator of  $\tau$  we use  $\hat{\tau} = \hat{\rho}_{k,n}/H_{k,n}$  with  $\hat{\rho}_{k,n}$  taken from Fraga Alves et al. (2003). Moreover we set  $\zeta = \gamma^2(1 - 2\rho)(1 - \rho)^2$ . The following result is derived in Appendix 3.A.

**Theorem 3.1.** *Let  $F \in \mathcal{F}(\gamma, \tau)$  with  $|a(x)| = x^\rho C_a(1 + o(1))$  as  $x \rightarrow \infty$ . Assume that  $\sqrt{k}a(n/k) \rightarrow \lambda$  as  $k, n \rightarrow \infty$ ,  $k/n \rightarrow 0$ . Setting  $\sigma_{k,n}^2 = (k/n)^{-2\rho}$ , it follows that  $\gamma_{k,n} := \sqrt{k}(\hat{\gamma}_k^P - \gamma)$  is asymptotically normal with asymptotic mean and variance given by*

$$E_\infty(\gamma_{k,n}) = \frac{\lambda\rho}{1-\rho} \frac{\zeta C_a^2 \omega}{\zeta C_a^2 \omega + \rho^4 \lambda^2}, \quad (3.2.15)$$

$$Var_\infty(\gamma_{k,n}) = \frac{\gamma^2 \rho^8 \lambda^4}{(\rho^4 \lambda^2 + \zeta C_a^2 \omega)^2} \left( \left( \frac{1-\rho}{\rho} \right)^2 + \frac{\zeta^2 C_a^4 \omega^2}{\rho^8 \lambda^4} + 2 \frac{\zeta C_a^2 \omega}{\rho^4 \lambda^2} \right). \quad (3.2.16)$$

▲

Minimising  $MSE_\infty(\gamma_{k,n}) = E_\infty^2(\gamma_{k,n}) + Var_\infty(\gamma_{k,n})$  with respect to  $\omega$ , this leads to the asymptotically optimal value (see Appendix 3.B)

$$\omega_{opt} = C_a^{-2}.$$

One then obtains from (3.2.15) and (3.2.16) that

$$\begin{aligned} E_\infty^{opt}(\gamma_{k,n}) &= \frac{\lambda\rho}{1-\rho} \frac{\zeta}{\zeta + \lambda^2 \rho^4}, \\ Var_\infty^{opt}(\gamma_{k,n}) &= \frac{\gamma^2}{(\lambda^2 \rho^4 + \zeta)^2} \left\{ (1-\rho)^2 \rho^6 \lambda^4 + \zeta^2 + 2\zeta \rho^4 \lambda^2 \right\}, \end{aligned}$$

from which

$$MSE_\infty^{opt}(\gamma_{k,n}) = \gamma^2 + \frac{\lambda^2 \rho^2 \gamma^2 (1-2\rho)}{\gamma^2 (1-2\rho)(1-\rho)^2 + \rho^4 \lambda^2}. \quad (3.2.17)$$

Since the right hand side of (3.2.17) is an increasing function in  $\lambda^2$  it follows that

$$MSE_\infty^{opt}(\gamma_{k,n}) \leq \lim_{\lambda \rightarrow \infty} MSE_\infty^{opt}(\gamma_{k,n}) = MSE_\infty \left( \sqrt{k}(\hat{\gamma}_{k,n}^{ML} - \gamma) \right) = \gamma^2 \left( \frac{1-\rho}{\rho} \right)^2.$$

Also, expanding the right hand side of (3.2.17) for  $\lambda^2 \rightarrow 0$  leads to

$$MSE_\infty^{opt}(\gamma_{k,n}) = \gamma^2 + \lambda^2 \frac{\rho^2}{(1-\rho)^2} (1 + o(1)).$$

We can conclude that the asymptotic MSE of the optimal penalised estimator is uniformly smaller than the MSE of the EPD-ML estimator as given in (3.1.10), while for smaller  $\lambda$  this asymptotic MSE follows the asymptotic MSE of the Hill estimator, given

in (3.1.9), up to terms of order  $\lambda^2$ . Hence with the penalty  $\omega/\sigma_{k,n}^2 = C_a^{-2}(k/n)^{2\rho} = a^{-2}(n/k)$  in (3.2.12), the penalised ML estimator asymptotically follows the better of the two existing estimators as a function of  $\lambda$  or  $k$ .

Replacing  $(\hat{\gamma}_k, \hat{\delta}_k)$  by  $(\hat{\gamma}_k^P, \hat{\delta}_k^P)$  in  $\hat{p}_{x,k}^{EP}$ , it follows from the proof of Theorem 5.2 in Beirlant et al. (2009) that the resulting tail probability estimator  $\hat{p}_{x,k}^P$  satisfies the following asymptotic result under the conditions of the Theorem:

when  $p_n = P(X > x_n)$  satisfies  $np_n/k \rightarrow 0$  and  $\log(np_n)/\sqrt{k} \rightarrow 0$ , then

$$\frac{\sqrt{k}}{\log(k/(np_n))} \gamma\left(\frac{\hat{p}_{x_n,k}^P}{p_n} - 1\right)$$

is asymptotically normal with the same limit distribution as in the Theorem. Hence the asymptotic MSE behaviour for the tail probability estimator has the same characteristics as the tail index estimator.

From the simulations it will follow that the choice  $\omega = 1$  and the use of estimator of  $\rho$  taken from Fraga Alves et al. (2003) yields good results. However, in order to alleviate the problem of choosing the number of top order statistics  $k$  that are used in the estimation procedure, one can choose  $\omega$  adaptively with each sample aiming for a plot of  $\hat{\gamma}_k^P$  as a function of  $k$  which is as horizontal as possible. Setting  $\hat{\gamma}_k^P = \hat{\gamma}_k^P(\omega)$  in order to emphasise the dependence of the penalised ML estimator on  $\omega$ , a possible choice of  $\omega$  is obtained by minimising the variance of the resulting estimators for  $k = 1, \dots, n$ :

$$\omega_{mv} = \operatorname{argmin}_{\omega} s_n^2(\hat{\gamma}^P(\omega)), \quad (3.2.18)$$

with  $s_n^2(\hat{\gamma}^P(\omega)) = \frac{1}{n-1} \sum_{k=1}^n (\hat{\gamma}_k^P(\omega) - \bar{\hat{\gamma}}^P)^2$ .

### 3.3 Simulations and practical case studies

Both the Bayes maximum a posteriori probability estimator and the penalised maximum likelihood estimator are implemented in **R** using the general **optim** function

with default parameters.

We performed a simulation study, taking 1000 repetitions of samples of size  $n = 200, 500, 1000$  studying the finite sample behaviour of  $\hat{\gamma}_{k,n}^P(\omega)$  for different distributions. The bias and root mean squared error (RMSE) are plotted as a function of  $k$ .

The following distributions are used:

- *The extreme value distribution* (EV) with  $F(x) = \exp(-(1+\gamma x)^{-1/\gamma})$  ( $1+\gamma x > 0$ ) taking  $\gamma = 0.25$  in which case  $\rho = -0.25$  and  $C_a = 1$ .
- *The Fréchet distribution* with  $\bar{F}(x) = 1 - \exp(-x^{-1/\gamma})$  taking  $\gamma = 0.5$  in which case  $\rho = -1$  and  $C_a = 0.25$ .
- *The Burr distribution* with  $\bar{F}(x) = (1+x)^{-4/3}$  so that  $\gamma = 0.75$  and  $\rho = -0.75$  and  $C_a = 1$ .
- *The loggamma distribution* with  $\bar{F}(x) \sim \text{constant} \times x^{-2}(\log x)^3$  so that  $\gamma = 0.5$ , which does not belong to the class  $\mathcal{F}(\gamma, \tau)$ .

First in Figures 3.1-3.4 we plotted the bias and the RMSE of the Hill estimator  $H_k$ , the EPD-ML estimator  $\hat{\gamma}_k^{ML}$ , the penalised ML estimator  $\hat{\gamma}_k^P(1)$  with  $\omega = 1$ , the Bayesian estimator  $\hat{\gamma}_k^B(1)$  with  $\omega = 1$ , and the minimum variance reduced bias estimator  $CH_k$  from Caeiro *et al.* (2005) given by

$$CH_k = H_{k,n} \left( 1 - \frac{\hat{\beta}_{k_1}(\hat{\rho}_{k_1})}{1 - \hat{\rho}_{k_1}} \left( \frac{n}{k} \right)^{\hat{\rho}_{k_1}} \right),$$

with

$$\hat{\beta}_k(\rho) = \frac{\left( \frac{k}{n} \right)^\rho \left\{ \left( \frac{1}{k} \sum_{j=1}^k \left( \frac{j}{k} \right)^{-\rho} \right) \left( \frac{1}{k} \sum_{j=1}^k Z_j \right) - \left( \frac{1}{k} \sum_{j=1}^k \left( \frac{j}{k} \right)^{-\rho} Z_j \right) \right\}}{\left( \frac{1}{k} \sum_{j=1}^k \left( \frac{j}{k} \right)^{-\rho} \right) \left( \frac{1}{k} \sum_{j=1}^k \left( \frac{j}{k} \right)^{-\rho} Z_j \right) - \left( \frac{1}{k} \sum_{j=1}^k \left( \frac{j}{k} \right)^{-2\rho} Z_j \right)},$$

where  $Z_j := j(\log X_{n-j+1,n} - \log X_{n-j,n})$  ( $j = 1, 2, \dots$ ), and  $k_1 = \lfloor n^{0.99} \rfloor$ .

In Figure 3.5 we briefly report on the effect of the choice of  $\omega$  using  $\omega = 1$  and  $\omega = \omega_{mv}$  and compare these with the optimal asymptotic RMSE expression from (3.2.17).

We conclude from the simulations that the finite sample behaviour of the proposed estimators follows the characteristics predicted by the asymptotic analysis to a great extent: for small  $k$  the shrinkage estimators  $\gamma_k^P$  and  $\gamma_k^B$  show a similar behaviour as the Hill estimator, while for larger  $k$  the proposed estimators tend to follow the characteristics of the bias reduced EPD-ML estimator. In between these two  $k$ -regions the shrinkage estimators make a transition from the EPD-ML to the Hill RMSE curve. Only in the Fréchet case the Hill estimator shows a smaller RMSE than the shrinkage estimators for small  $k$ , while the shrinkage estimators then still show a much smaller RMSE than the EPD-ML estimator.

The Bayesian implementation shows a smaller RMSE than the penalised ML estimator, except for the Fréchet distribution where both RMSEs are comparable. In the latter case  $\hat{\gamma}_k^B$  shows a negative bias. Also note that the difference between both the Bayesian and penalised likelihood implementation decreases as  $n$  increases.

The results in case of the loggamma distribution are quite good. Hence it appears that the proposed method exhibits some robustness against deviations from the underlying model.

When the plots of the shrinkage estimators are not systematically increasing with increasing  $k$  as in the case of the Fréchet and the Burr distribution, it is useful to use the choice  $\omega = \omega_{mv}$  when using the penalised ML estimator. In the case of the Fréchet distribution with  $\omega_{opt} = 16$ , this adaptive choice of  $\omega$  leads to a clear RMSE improvement in the transition zone (in  $k$ ) between the Hill and EPD-ML RMSE behaviour (see Figure 3.5, top). In the Burr case (see Figure 3.5, bottom) where  $C_a = 1$  and hence  $\omega_{opt} = 1$  the choice  $\omega = 1$  is best, but the adaptive minimum variance choice  $\omega = \omega_{mv}$  is almost as good in RMSE behaviour.

Overall, the proposed shrinkage estimators are competitive with respect to the mini-

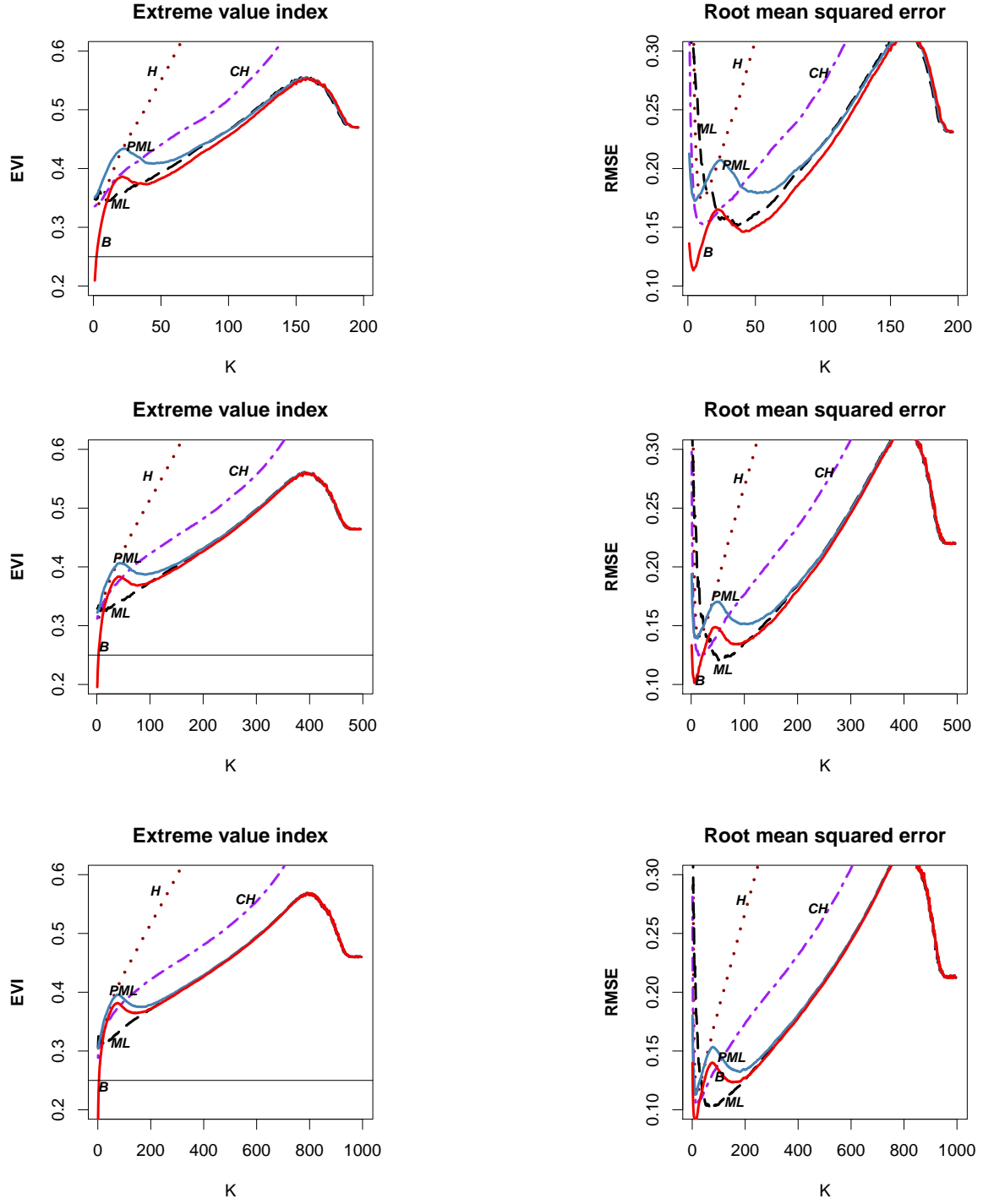


minimum variance reduced bias estimator  $CH_k$ .

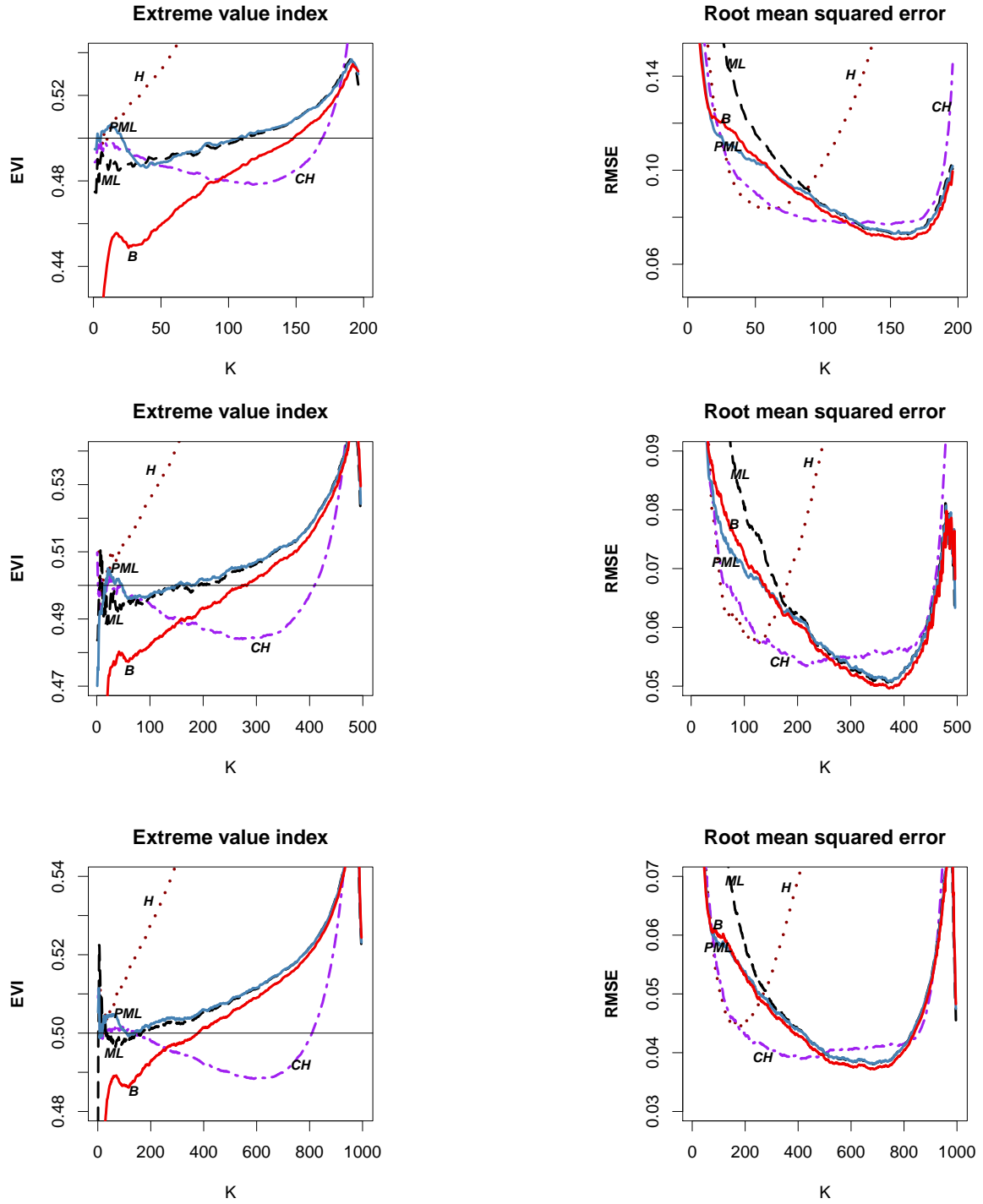
In order to illustrate the use of the proposed method we consider the Secura Belgian Re data introduced in section 6.2 in Beirlant et al. (2004). For  $k \leq 100$  the penalised ML estimator  $\hat{\gamma}_k^P(1)$  is quite constant and follows the Hill estimator quite closely. This is in contrast with the EPD-ML estimates which vary a lot in that region. The Bayesian estimates  $\hat{\gamma}_k^B(1)$  and CH estimates show somewhat lower estimates. Beirlant et al. (2004) concluded that the Hill estimate in this  $k$ -region is an appropriate choice and the adaptive choice  $\hat{k} = 98$  was proposed as one of the largest  $k$ -values in this region. This proposal is also supported by the present analysis, leading to an estimate  $\hat{\gamma}^P(1) = 0.28$ .

### 3.4 Chapter conclusion

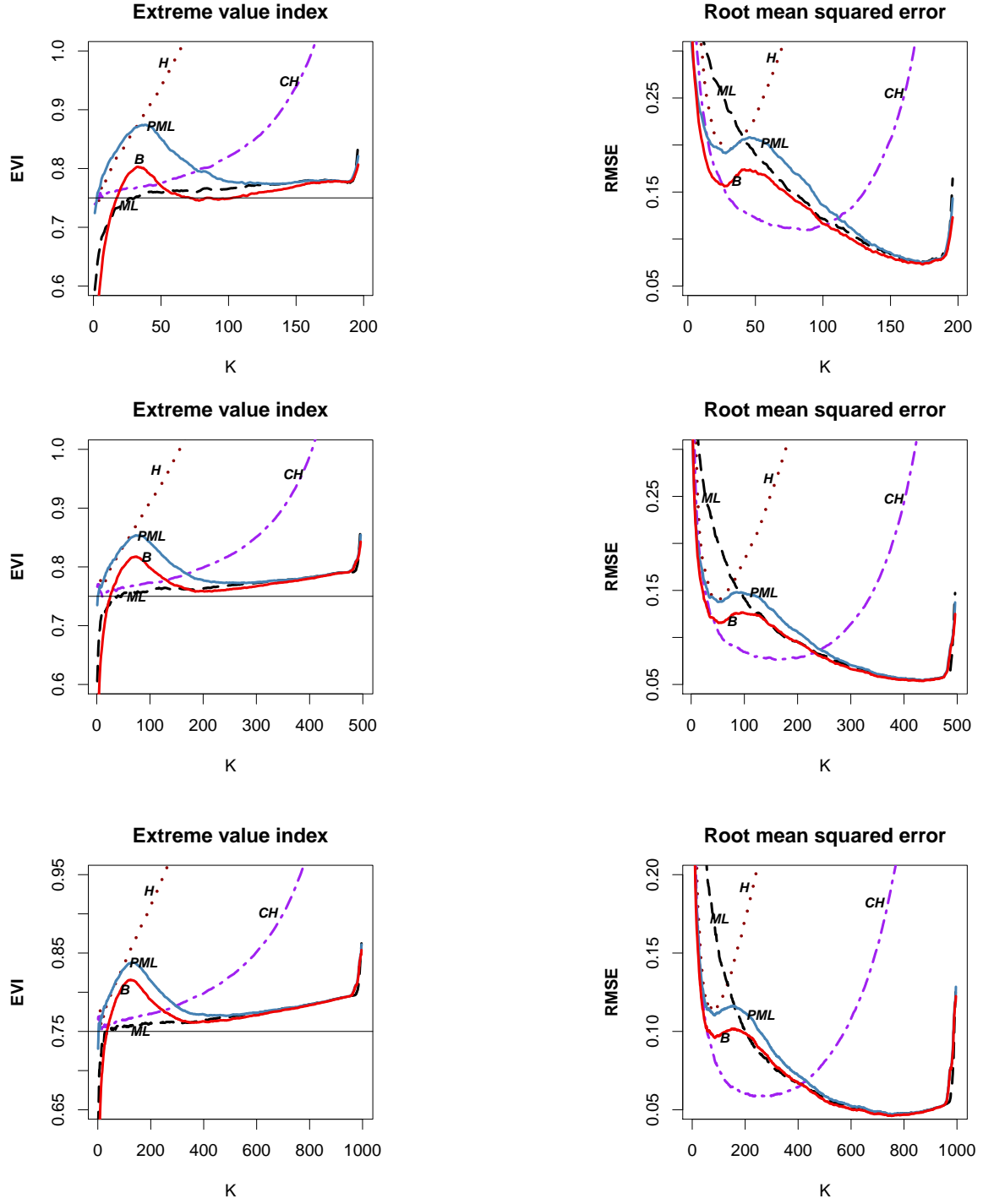
We introduced the use of shrinkage estimators in tail estimation, in order to obtain bias reduction jointly with good MSE behaviour. Shrinkage estimators can be obtained through a penalised ML approach, or through a Bayesian implementation. For larger thresholds the proposed estimators follow the behaviour of the classical Hill estimator with small bias and minimal variance, while the new estimators are never worse than the corresponding bias reduced ML estimators without penalisation. The simulated MSE results are competitive with those of other bias reduced estimators. In contrast to existing minimum variance bias reduced estimators we only use second order slow variation conditions.



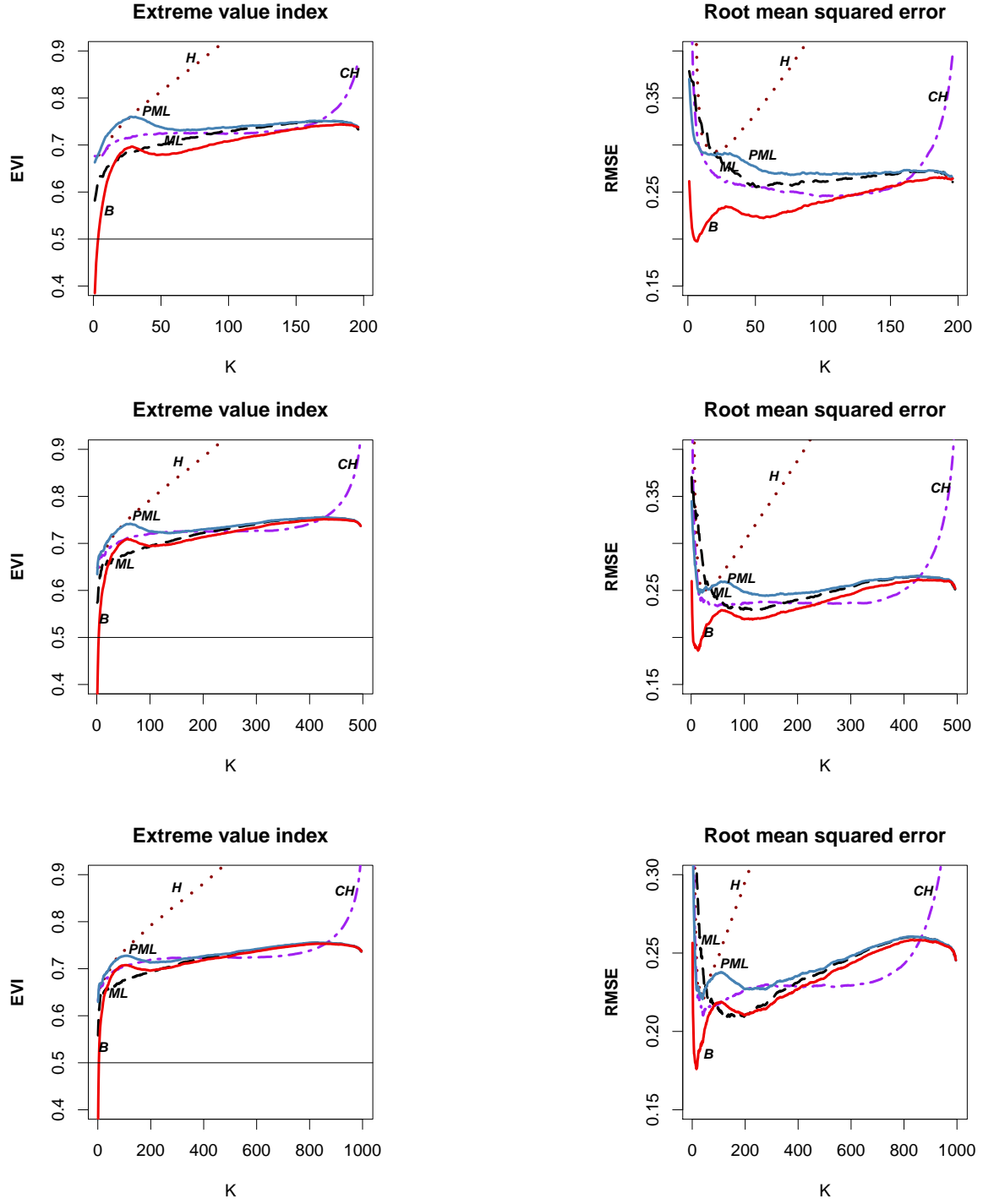
**Figure 3.1:** Bias (left) and root mean squared error (right) in case of the **EV distribution** with  $\gamma = 0.25$  for sample sizes  $n = 200$  (top),  $n = 500$  (middle) and  $n = 1000$  (bottom) for the Hill estimator (H), the EPD-ML estimator  $\hat{\gamma}_k^{ML}$  (ML), the penalised ML estimator  $\hat{\gamma}_k^P(1)$  with  $\omega = 1$  (PML), the Bayesian estimator  $\hat{\gamma}_k^B(1)$  with  $\omega = 1$  (B), and the minimum variance reduced bias estimator  $CH_k$  (CH).



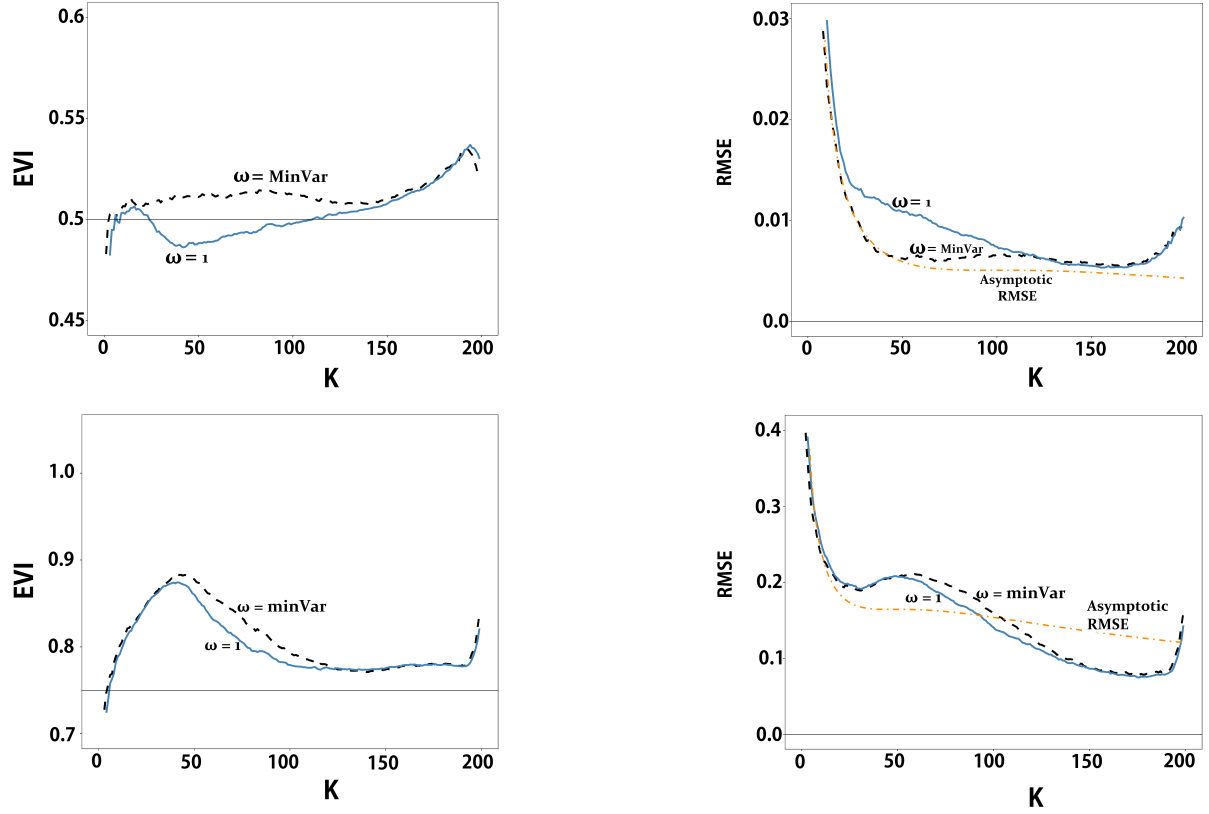
**Figure 3.2:** Bias (left) and root mean squared error (right) in case of the **Fréchet distribution** with  $\gamma = 0.5$  for sample sizes  $n = 200$  (top),  $n = 500$  (middle) and  $n = 1000$  (bottom) for the Hill estimator (H), the EPD-ML estimator  $\hat{\gamma}_k^{ML}$  (ML), the penalised ML estimator  $\hat{\gamma}_k^P(1)$  with  $\omega = 1$  (PML), the Bayesian estimator  $\hat{\gamma}_k^B(1)$  with  $\omega = 1$  (B), and the minimum variance reduced bias estimator  $CH_k$  (CH).



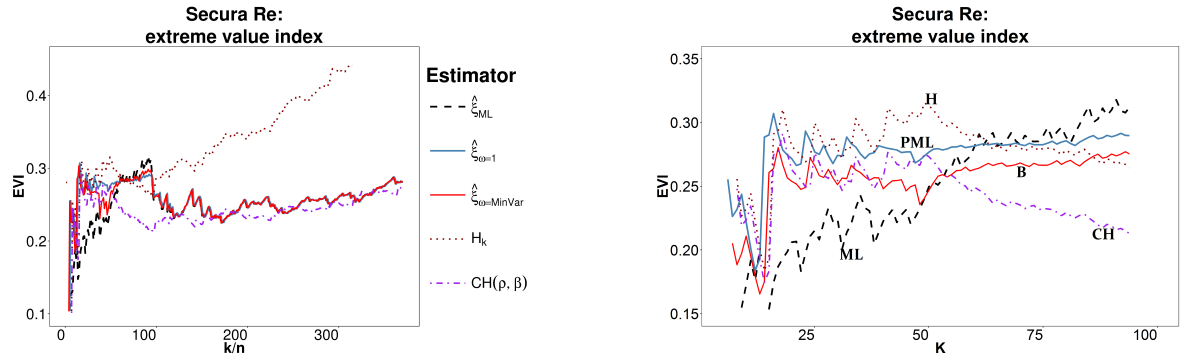
**Figure 3.3:** Bias (left) and root mean squared error (right) in case of the **Burr distribution** with  $\gamma = 0.75$  for sample sizes  $n = 200$  (top),  $n = 500$  (middle) and  $n = 1000$  (bottom) for the Hill estimator (H), the EPD-ML estimator  $\hat{\gamma}_k^{ML}$  (ML), the penalised ML estimator  $\hat{\gamma}_k^P(1)$  with  $\omega = 1$  (PML), the Bayesian estimator  $\hat{\gamma}_k^B(1)$  with  $\omega = 1$  (B), and the minimum variance reduced bias estimator  $CH_k$  (CH).



**Figure 3.4:** Bias (left) and root mean squared error (right) in case of the **loggamma distribution** with  $\gamma = 0.5$  for sample sizes  $n = 200$  (top),  $n = 500$  (middle) and  $n = 1000$  (bottom) for the Hill estimator (H), the EPD-ML estimator  $\hat{\gamma}_k^{ML}$  (ML), the penalised ML estimator  $\hat{\gamma}_k^P(1)$  with  $\omega = 1$  (PML), the Bayesian estimator  $\hat{\gamma}_k^B(1)$  with  $\omega = 1$  (B), and the minimum variance reduced bias estimator  $CH_k$  (CH).



**Figure 3.5:** Bias (left) and root mean squared error (right) in case of the **Fréchet distribution** with  $\gamma = 0.5$  (top) and Burr distribution with  $\gamma = 0.75$  (bottom) for sample size  $n = 200$  comparing the penalised ML estimator  $\hat{\gamma}_k^P(1)$  with  $\omega = 1$ ,  $\omega = \omega_{mv}$  from (3.2.18), and the optimal asymptotic RMSE from (3.2.17) replacing  $\lambda$  by  $C_a \sqrt{k}(k/n)^{-\rho}$ .



**Figure 3.6:** Estimates of  $\gamma$  for Secura Belgian Re data set: results for the Hill estimator (H), the EPD-ML estimator  $\hat{\gamma}_k^{ML}$  (ML), the penalised ML estimator  $\hat{\gamma}_k^P(1)$  with  $\omega = 1$  (PML), the Bayesian estimator  $\hat{\gamma}_k^B(1)$  with  $\omega = 1$  (B), and the minimum variance reduced bias estimator  $CH_k$  (CH) (left), focused plot for  $k = 1, \dots, 100$  (right).

# Appendix

## 3.A Derivation of the expressions of $(\hat{\gamma}_k^P, \hat{\delta}_k^P)$ .

First consider the asymptotic approximations of the penalised ML estimator of  $\gamma$  based on maximisation of (3.2.12). From (3.2.11)-(3.2.12) using expansions in  $\delta \rightarrow 0$  we obtain

$$\begin{aligned} \frac{1}{k} \log l_{pen}(\gamma, \delta | \mathbf{y}) &= - \left(1 + \frac{1}{k}\right) \log \gamma - \frac{1}{k} (1 + \gamma) - \left(\frac{1}{\gamma} + 1\right) \frac{1}{k} \sum_{j=1}^k \log y_{j,k} \\ &\quad - \delta \left(\frac{1}{\gamma} + 1\right) \frac{1}{k} \sum_{j=1}^k (1 - y_{j,k}^\tau) + \delta \frac{1}{k} \sum_{j=1}^k (1 - (1 + \tau) y_{j,k}^\tau) \\ &\quad - \frac{\omega \delta^2}{2k \sigma_{k,n}^2} + \left(\frac{\delta^2}{2}\right) \left(\frac{1}{\gamma} + 1\right) \frac{1}{k} \sum_{j=1}^k (1 - y_{j,k}^\tau)^2 - \frac{\delta^2}{2} \frac{1}{k} \sum_{j=1}^k (1 - (1 + \tau) y_{j,k}^\tau)^2 \\ &\quad + O(\delta^3) + c, \end{aligned}$$

where  $c$  is a constant only depending on  $\sigma_{k,n}^2$  and  $\tau$ . Note that  $\frac{1}{k} \sum_{j=1}^k \log y_{j,k} = H_{k,n}$ .

Then the score functions admit the following expansions in  $\delta \downarrow 0$  for  $j = 1, \dots, k$ :

$$\begin{aligned} \frac{\partial}{\partial \gamma} \log l_{pen}(\gamma, \delta | y_{j,k}) &= -\frac{1}{\gamma} + \frac{1}{\gamma^2} \log y_{j,k} + \frac{\delta}{\gamma^2} (1 - y_{j,k}^\tau) + O(\delta^2), \\ \frac{\partial}{\partial \delta} \log l_{pen}(\gamma, \delta | y_{j,k}) &= -\frac{1}{\gamma} (1 - (1 - \gamma\tau) y_{j,k}^\tau) - \frac{\omega \delta}{k \sigma_{k,n}^2} \\ &\quad + \frac{\delta}{\gamma} (1 - 2(1 - \gamma\tau) y_{j,k}^\tau + (1 - 2\gamma\tau - \gamma\tau^2) y_{j,k}^{2\tau}) + O(\delta^2). \end{aligned}$$

*Derivation of Theorem 3.1.* Note that as  $k, n \rightarrow \infty$ ,  $k/n \rightarrow 0$  and  $\sqrt{ka}(n/k) \rightarrow \lambda$ , we



also have  $k\sigma_{k,n}^2 \rightarrow \lambda^2 C_a^{-2}$ . Also as  $\sqrt{k}a(n/k) \rightarrow \lambda$  we find using  $E_{k,n}(s) \rightarrow 1/(1-\gamma s)$  (see Theorem A.1 in Beirlant et al., 2009) that

$$D_{k,n}^P = -\frac{\gamma C_a^2}{\lambda^2} + \frac{\rho^4}{\gamma(1-2\rho)(1-\rho)^2} + o_p(1).$$

Then, proceeding as in the proof of Theorem 3.1 in Beirlant et al. (2009), we obtain with  $\Gamma_{k,n} = \sqrt{k}(H_{k,n} - \gamma)$ ,  $\mathbb{E}_{k,n}(s) = \sqrt{k}(E_{k,n}(s) - \frac{1}{1-\gamma s})$  ( $s < 0$ ), that

$$\begin{aligned} \sqrt{k}(\hat{\gamma}_k^P - \gamma) &= \sqrt{k}\left(H_{k,n} - \gamma - \hat{\delta}_k^P \frac{\rho}{1-\rho}\right) \\ &= \Gamma_{k,n} - \frac{\rho}{1-\rho} \sqrt{k} \hat{\delta}_k^P \\ &= \Gamma_{k,n} \left(1 + \frac{\rho^2}{\gamma(1-\rho^2)} \frac{1}{\gamma C_a^2/\lambda^2 + \rho^4/\gamma(1-2\rho)(1-\rho)^2}\right) \\ &\quad - \frac{\rho}{\gamma C_a^2/\lambda^2 + \rho^4/\gamma(1-2\rho)(1-\rho)^2} \mathbb{E}_{k,n}(\hat{\tau}) + o_p(1) \\ &= \Gamma_{k,n} \left(1 + \frac{\rho^2(1-2\rho)}{\zeta + \rho^4}\right) + \mathbb{E}_{k,n}(\hat{\tau}) \left(\frac{(-\rho)\gamma(1-2\rho)(1-\rho)^2}{\rho^4 + \zeta}\right) + o_p(1). \end{aligned}$$

Using Theorem A.1 in Beirlant et al. (2009), (3.2.15) and (3.2.16) follow under  $\sqrt{k}a(n/k) \rightarrow \lambda$ .

### 3.B Optimal $\omega$ minimising the $MSE_\infty$

$$MSE_\infty = (E_\infty)^2 + \text{VAR}_\infty$$

$$E_\infty = \frac{\lambda\rho(\zeta C_a^2\omega)}{(1-\rho)(\zeta C_a^2\omega + \lambda^2\rho^4)}, \quad \text{VAR}_\infty = \frac{(\lambda^4\xi^2\rho^8) \left( \frac{C_a^4\zeta^2\omega^2}{\lambda^4\rho^8} + \frac{2\zeta C_a^2\omega}{\lambda^2\rho^4} + \left(\frac{1-\rho}{\rho}\right)^2 \right)}{(\zeta C_a^2\omega + \lambda^2\rho^4)^2}$$

$$MSE_\infty = \frac{(\lambda^4\xi^2\rho^8) \left( \frac{C_a^4\zeta^2\omega^2}{\lambda^4\rho^8} + \frac{2\zeta C_a^2\omega}{\lambda^2\rho^4} + \left(\frac{1-\rho}{\rho}\right)^2 \right)}{(\zeta C_a^2\omega + \lambda^2\rho^4)^2} + \left( \frac{\lambda\rho(\zeta C_a^2\omega)}{(1-\rho)(\zeta C_a^2\omega + \lambda^2\rho^4)} \right)^2$$

Where  $\zeta = \xi^2(1-2\rho)(1-\rho)^2$

$$\begin{aligned}
\partial_\omega(MSE_\infty) &= \frac{2C_a^2\zeta\lambda^2\xi^2\rho^4}{(C_a^2\zeta\omega + \lambda^2\rho^4)^2} - \frac{2C_a^2\zeta\lambda^4\xi^2\rho^8}{(C_a^2\zeta\omega + \lambda^2\rho^4)^3} + \frac{4C_a^2\zeta\lambda^4\xi^2\rho^7}{(C_a^2\zeta\omega + \lambda^2\rho^4)^3} \\
&\quad - \frac{2C_a^2\zeta\lambda^4\xi^2\rho^6}{(C_a^2\zeta\omega + \lambda^2\rho^4)^3} - \frac{2C_a6\zeta^3\xi^2\omega^2}{(C_a^2\zeta\omega + \lambda^2\rho^4)^3} - \frac{2C_a6\zeta^3\lambda^2\rho^2\omega^2}{(1-\rho)^2(C_a^2\zeta\omega + \lambda^2\rho^4)^3} \\
&\quad - \frac{4C_a^4\zeta^2\lambda^2\xi^2\rho^4\omega}{(C_a^2\zeta\omega + \lambda^2\rho^4)^3} + \frac{2C_a^4\zeta^2\xi^2\omega}{(C_a^2\zeta\omega + \lambda^2\rho^4)^2} + \frac{2C_a^4\zeta^2\lambda^2\rho^2\omega}{(1-\rho)^2(C_a^2\zeta\omega + \lambda^2\rho^4)^2} \\
\partial_\omega(MSE_\infty) &= \frac{\lambda^4\xi^2\rho^8 \left( \frac{2C_a^4\zeta^2\omega}{\lambda^4\rho^8} + \frac{2\zeta C_a^2}{\lambda^2\rho^4} \right)}{(\zeta C_a^2\omega + \lambda^2\rho^4)^2} - \frac{2\zeta C_a^2\lambda^4\xi^2\rho^8 \left( \frac{C_a^4\zeta^2\omega^2}{\lambda^4\rho^8} + \frac{2\zeta C_a^2\omega}{\lambda^2\rho^4} + \frac{(1-\rho)^2}{\rho^2} \right)}{(\zeta C_a^2\omega + \lambda^2\rho^4)^3} \\
&\quad - \frac{2\zeta C_a^6\lambda\rho^2\omega^2}{(1-\rho)^2(\zeta C_a^2\omega + \lambda^2\rho^4)^3} + \frac{2\zeta C_a^4\lambda\rho^2\omega}{(1-\rho)^2(\zeta C_a^2\omega + \lambda^2\rho^4)^2}
\end{aligned}$$

Which simplifies to:

$$\begin{aligned}
&= \frac{2C_a^2\zeta\lambda^4\xi^2(2\rho-1)\rho^6}{(C_a^2\zeta\omega + \lambda^2\rho^4)^3} + \frac{2C_a^4\zeta^2\lambda^4\rho^6\omega}{(\rho-1)^2(C_a^2\zeta\omega + \lambda^2\rho^4)^3} \\
&= \frac{2C_a^2\zeta\lambda^4\rho^6(C_a^2\zeta\omega + \xi^2(\rho-1)^2(2\rho-1))}{(\rho-1)^2(C_a^2\zeta\omega + \lambda^2\rho^4)^3}
\end{aligned}$$

Set  $\partial_\omega(MSE_\infty) = 0$  and solve for  $\omega$

$$\omega \rightarrow C_a^{-2} * \left( -\frac{\xi^2(\rho-1)^2(2\rho-1)}{\zeta} \right)$$

After substituting  $\zeta$  back into the equation we get:

$$\omega = C_a^{-2} * \left( -\frac{(\rho-1)^2(2\rho-1)}{(1-2\rho)(1-\rho)^2} \right) = C_a^{-2}$$

$$\partial_\omega^2(MSE_\infty) = \frac{2C_a^4\zeta^2\lambda^4\rho^6}{(\rho-1)^2(C_a^2\zeta\omega + \lambda^2\rho^4)^3} - \frac{6C_a^4\zeta^2\lambda^4\rho^6(C_a^2\zeta\omega + \xi^2(\rho-1)^2(2\rho-1))}{(\rho-1)^2(C_a^2\zeta\omega + \lambda^2\rho^4)^4}$$

Substituting  $\zeta$  and  $\omega = C_a^{-2}$  into the equation:

$$\begin{aligned}
\partial_\omega^2(MSE_\infty) &= -\frac{8C_a^4\lambda^4\xi^4\rho^{10}}{(\xi^2(\rho-1)^2(2\rho-1)-\lambda^2\rho^4)^3} + \frac{24C_a^4\lambda^4\xi^4\rho^9}{(\xi^2(\rho-1)^2(2\rho-1)-\lambda^2\rho^4)^3} \\
&\quad - \frac{26C_a^4\lambda^4\xi^4\rho^8}{(\xi^2(\rho-1)^2(2\rho-1)-\lambda^2\rho^4)^3} + \frac{12C_a^4\lambda^4\xi^4\rho^7}{(\xi^2(\rho-1)^2(2\rho-1)-\lambda^2\rho^4)^3} \\
&\quad - \frac{2C_a^4\lambda^4\xi^4\rho^6}{(\xi^2(\rho-1)^2(2\rho-1)-\lambda^2\rho^4)^3} \\
&= \frac{2C_a^4\lambda^4\xi^4(1-2\rho)^2(1-\rho)^4\rho^6}{(\rho-1)^2(C_a^2\xi^2(1-\rho)^2(1-2\rho)\omega+\lambda^2\rho^4)^3} \\
&\quad - \frac{6C_a^4\lambda^4\xi^4(1-2\rho)^2(1-\rho)^4\rho^6(C_a^2\xi^2(1-2\rho)(1-\rho)^2\omega+\xi^2(\rho-1)^2(2\rho-1))}{(\rho-1)^2(C_a^2\xi^2(1-\rho)^2(1-2\rho)\omega+\lambda^2\rho^4)^4} \\
&= -\frac{2C_a^4\lambda^4\xi^4(1-2\rho)^2(\rho-1)^2\rho^6}{(\xi^2(\rho-1)^2(2\rho-1)-\lambda^2\rho^4)^3}
\end{aligned}$$

This leads to the following series expansion around zero

$$\begin{aligned}
&-\frac{2\lambda^4(C_a^4\rho^6)}{\xi^2(\rho-1)^4(2\rho-1)} + O(\lambda^5) \\
&\approx \frac{2C_a^4\lambda^4\rho^6}{\xi^2(\rho-1)^4(1-2\rho)} > 0
\end{aligned}$$

Using the fact that  $\rho < 0$ , we can conclude that  $\omega = C_a^{-2}$  is the optimal value that minimises the MSE.

### 3.C Algorithm to estimate $\rho$

In this section we provide the algorithm of Fraga Alves et al. (2003) used to estimate  $\rho$ .

1. Given a sample  $(X_1, X_2, \dots, X_n)$ , plot the estimates of

$$\hat{\rho}_{k,r} := \min\left\{0, \frac{3(T_{k,n}^{(\tau)} - 1)}{T_{k,n}^{(\tau)} - 3}\right\}, \quad (3.3.19)$$

for  $\tau = 0$ , where

$$T_{k,n} := \frac{\left(M_{k,n}^{(1)}\right)^\tau - \left(\frac{M_{k,n}^{(2)}}{2}\right)^{\tau/2}}{\left(\frac{M_{k,n}^{(2)}}{2}\right)^{\tau/2} - \left(\frac{M_{k,n}^{(3)}}{6}\right)^{\tau/3}} \quad (3.3.20)$$

and

$$M_{k,n}^{(j)} := \frac{1}{k} \sum_{i \leq k} (\log X_{n-i+1:n} - \log X_{n-k:n})^j, \quad j \geq 1. \quad (3.3.21)$$

since  $\tau = 0$ , the notation  $a^{b\tau} = b \ln a$  is used and therefore Equation 3.3.20 is written as

$$T_{k,n} := \frac{\log(M_{k,n}^{(1)}) - \frac{1}{2} \log\left(\frac{M_{k,n}^{(2)}}{2}\right)}{\frac{1}{2} \log\left(\frac{M_{k,n}^{(2)}}{2}\right) - \frac{1}{3} \log\left(\frac{M_{k,n}^{(3)}}{6}\right)} \quad (3.3.22)$$

2. Consider  $\{\hat{\rho}_{k,n}\}_{k \in \mathcal{K}}$  as given in Equation 3.3.19 for large  $k$ , say  $k \in \mathcal{K} = ([n^{0.990}], [n^{0.999}])$ , compute their median, denoted  $\rho_\tau$  and work with  $\hat{\rho} \equiv \hat{\rho}_{\tau_0} := \hat{\rho}_{k_1}$ , with

$$k_1 = [n^{0.995}] \quad (3.3.23)$$

The choice of the level  $k_1$  is not very important, we can consider any reasonably large value of  $k$  of order  $n^{1-\epsilon}$  from some intermediate sequence of integers  $k = k(n)$  satisfying  $\lim_{n \rightarrow \infty} \sqrt{k} A(n/k) = \infty$ . This estimator uses more upper order statistics than the first order parameter  $\gamma$ .

# Chapter 4

## Penalised Bias Reduction in Extreme Value Estimation for Censored Pareto-type Data

*This chapter is based on Beirlant, J., Maribe, G. and Verster, A., 2018. Penalized bias reduction in extreme value estimation for censored Pareto-type data, and long-tailed insurance applications. Insurance: Mathematics and Economics, 78, pp.114-122.*

### 4.1 Introduction

Extreme value analysis under random right censoring is becoming more popular with applications for example in survival analysis, reliability and insurance. For instance, in certain long-tailed insurance products, such as car liability insurance, long developments of claims are encountered. At evaluation of the portfolio a large proportion of the claims are then not fully developed and hence are censored.

In the setting of random right censoring the variable of interest  $X$  with CDF  $F$  can be

censored by a random variable  $C$  with CDF  $G$ . Moreover observations of  $X$  and  $C$  are assumed to be independent. One then observes  $Z = \min(X, C)$  with CDF  $H$  satisfying  $1 - H = (1 - F)(1 - G)$ , jointly with the indicator  $\Delta = 1_{(X \leq C)}$  which equals 1 if the observation  $Z$  is non-censored. Here we assume that  $X$  and  $C$  both are Pareto-type distributed with EVI  $\gamma_1 > 0$  and  $\gamma_2 > 0$ , i.e.

$$\bar{F}(x) = 1 - F(x) = x^{-1/\gamma_1} \ell_1(x) \text{ and } \bar{G}(y) = 1 - G(y) = y^{-1/\gamma_2} \ell_2(y), \quad x, y > 1,$$

where both  $\ell_1, \ell_2$  are slowly varying at infinity:

$$\ell_j(tx)/\ell_j(t) \rightarrow_{t \rightarrow \infty} 1, \quad \text{for every } x > 1 \quad (j = 1, 2).$$

Note that  $1 - H$  then also belongs to the domain of attraction of an extreme value distribution with positive EVI  $\gamma = \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2}$ . Of course, the smaller  $\gamma_2/\gamma_1$  the heavier the censoring will be. In long-tailed insurance applications as discussed above, the proportion of censored data can well be larger than 50%, so that the situation  $\gamma_2 < \gamma_1$  is then most relevant.

In this section we discuss the estimation of  $\gamma_1$  based on IID observations  $(Z_i, \Delta_i)$  ( $i = 1, \dots, n$ ) with  $Z_i = \min(X_i, C_i)$  and  $\Delta_i = 1_{(X_i \leq C_i)}$ , where  $(X_i, C_i)$  ( $i = 1, \dots, n$ ) are IID. random variables from  $(F, G)$ . In the next section we review the available estimators for  $\gamma_1$  that were published in the literature. In Section 4.3 we propose a new bias reduced estimator which is based on an estimator proposed by Worms and Worms (2014). Moreover we show how shrinkage estimation, as introduced in Beirlant et al. (2019) in the non-censoring case, can also be used in the censoring context. In Section 4.4 a parametric bootstrap algorithm is proposed in order to construct confidence intervals for  $\gamma_1$ . We then report on a simulation study involving all available estimators and the proposed bootstrap algorithm. Finally we make a detailed study of a motor third party liability (MTPL) case study.

## 4.2 A review of estimators of $\gamma_1$

In case there is no censoring (i.e.  $\gamma_2 = \infty$  and  $1 - H = 1 - F$ ), the Hill (1975) estimator is the benchmark estimator for  $\gamma_1 = \gamma$ . Denoting the ordered  $Z$  data by  $Z_{1,n} \leq Z_{2,n} \leq \dots \leq Z_{n,n}$  this estimator is given by

$$\hat{\gamma}_{Z,k}^H = \frac{1}{k} \sum_{j=1}^k \log \frac{Z_{n-j+1,n}}{Z_{n-k,n}}.$$

This estimator follows using maximum likelihood when approximating the distribution of the peaks  $Z/t$  over a threshold  $t$ , given  $Z > t$ , by a simple Pareto distribution with density  $y \mapsto \gamma^{-1} y^{-\gamma^{-1}-1}$ , and taking a top order statistic  $Z_{n-k,n}$  as a threshold  $t$ .

It can also be found back by estimating the functional

$$L_t := \mathbb{E}(\log Z - \log t | Z > t) = \int_1^\infty \frac{\bar{F}(ut)}{\bar{F}(t)} \frac{du}{u}, \quad (4.2.1)$$

which tends to the EVI  $\gamma$  of  $Z$  as  $t \rightarrow \infty$ . In (4.2.1)  $\bar{F}$  is estimated by the empirical survival function  $1 - \hat{F}_n$ , again using  $Z_{n-k,n}$  as a threshold  $t$ . This leads to an alternative writing of  $\hat{\gamma}_{Z,k}^H$  by partial summation:

$$\hat{\gamma}_{Z,k}^{(H)} = \frac{1}{k} \sum_{j=1}^k j (\log Z_{n-j+1,n} - \log Z_{n-j,n}).$$

While both approaches yield the same estimator in the non-censoring case this is no longer the case under random censoring.

- Beirlant et al. (2007) proposed the following estimator of  $\gamma_1$  using the maximum likelihood approach:

$$\hat{\gamma}_{1,k}^{(H)} = \frac{\hat{\gamma}_{Z,k}^{(H)}}{\hat{p}_k}, \quad (4.2.2)$$

with  $\hat{p}_k = \frac{1}{k} \sum_{j=1}^k \Delta_{n-j+1,n}$  the proportion of non-censored observations under the largest  $k$  observations of  $Z$ , where  $\Delta_{n-j+1,n}$  denotes the  $\Delta$  indicator attached to  $Z_{n-j+1,n}$  ( $1 \leq j \leq n$ ). Indeed,  $\hat{\gamma}_{Z,k}^H$  estimates  $\gamma$  while  $\hat{p}_k$  is shown to be a consistent estimator of  $p = \gamma_2/(\gamma_1 + \gamma_2)$ . Einmahl et al. (2008) enhanced the

asymptotic analysis of this estimator and generalised this approach by considering any classical EVI estimator  $\hat{\gamma}_{Z,k}^{(\cdot)}$  of  $\gamma$ , proposing the estimators  $\hat{\gamma}_{1,k}^{(\cdot)} = \frac{\hat{\gamma}_{Z,k}^{(\cdot)}}{\hat{p}_k}$ . See also Gomes and Oliveira (2003), Gomes and Neves (2011), and Brahim et al. (2015a) for other papers in this spirit.

- Worms and Worms (2014) essentially used the second approach estimating (4.2.1) by substituting  $1 - F$  with the Kaplan-Meier estimator

$$1 - \hat{F}_n^{KM}(x) = \Pi_{Z_{i,n} \leq x} \left( 1 - \frac{1}{n - i + 1} \right)^{\Delta_{i,n}}$$

, setting  $1 - \hat{F}_n^{KM}(Z_{n,n}) = 0$ :

$$\hat{\gamma}_{1,k}^{(W)} = \sum_{j=1}^k \frac{1 - \hat{F}_n^{KM}(Z_{n-j+1,n})}{1 - \hat{F}_n^{KM}(Z_{n-k,n})} (\log Z_{n-j+1,n} - \log Z_{n-j,n}). \quad (4.2.3)$$

Worms and Worms (2014) also introduced

$$\hat{\gamma}_{1,k}^{(KM)} = \sum_{j=1}^k \frac{1 - \hat{F}_n^{KM}(Z_{n-j+1,n})}{1 - \hat{F}_n^{KM}(Z_{n-k,n})} \frac{\Delta_{n-j+1,n}}{j} (\log Z_{n-j+1,n} - \log Z_{n-k,n}). \quad (4.2.4)$$

Through simulations the estimator  $\hat{\gamma}_{1,k}^{(W)}$  was found to have the best RMSE behaviour for smaller values of  $k$ . Below we will then concentrate on  $\hat{\gamma}_{1,k}^{(W)}$ . Unfortunately, the asymptotic distribution of  $\hat{\gamma}_{1,k}^{(W)}$  is not known up to now. The asymptotic normality of a fixed threshold version of  $\hat{\gamma}_{1,k}^{(KM)}$  can be derived from Worms and Worms (2018) in case  $\gamma_1 < \gamma_2$ . Brahim et al. (2015b) consider closely related estimators but also considered asymptotic results in case  $\gamma_1 < \gamma_2$  or  $p > 1/2$ .

- In an objective Bayesian approach (see Zellner, 1971), Amaraoui et al. (2016) recently proposed several other estimators. They considered the maximum posterior (m) and the mean posterior (e) estimators of the posterior density of  $\gamma_1$ . The maximal data information (M) prior and a conjugate gamma prior with parameters  $(a, b)$  were considered. It was also shown that Jeffreys prior lead to special cases of the conjugate prior estimators setting  $a = b = 0$ . This then leads



to the following estimators:

$$\hat{\gamma}_{1,k}^{(m,M)} = \frac{2k\hat{\gamma}_{Z,k}^{(H)}}{1 + k\hat{p}_k + \sqrt{(1 + k\hat{p}_k)^2 + 4k\hat{\gamma}_{Z,k}^{(H)}}}, \quad (4.2.5)$$

$$\hat{\gamma}_{1,k}^{(e,a,b)} = \frac{k\hat{\gamma}_{Z,k}^{(H)} + b}{k\hat{p}_k + a}, \quad (4.2.6)$$

$$\hat{\gamma}_{1,k}^{(m,a,b)} = \frac{k\hat{\gamma}_{Z,k}^{(H)} + b}{k\hat{p}_k + a - 1}. \quad (4.2.7)$$

It is well-known that extreme value estimators often suffer from severe bias. In the random censoring case Einmahl et al. (2008) first derived the asymptotic bias of  $\hat{\gamma}_{1,k}^{(H)}$ , which was further detailed in Beirlant et al. (2016) under more specific assumptions on the slowly varying functions  $\ell_1$  and  $\ell_2$ , which are commonly proposed in EVT:

$$\begin{aligned} 1 - F(x) &= C_1 x^{-1/\gamma_1} (1 + D_1 x^{-\beta_1} (1 + o(1))), \quad x \rightarrow \infty, \\ 1 - G(y) &= C_2 y^{-1/\gamma_2} (1 + D_2 x^{-\beta_2} (1 + o(1))), \quad y \rightarrow \infty, \end{aligned}$$

where  $\beta_1, \beta_2, C_1, C_2$  are positive constants and  $D_1, D_2$  are real constants. Taking the bias of the Hill estimator  $\hat{\gamma}_{Z,k}^{(H)}$  as a reference, it was observed that especially when  $\beta_1 \leq \beta_2$  the bias of  $\hat{\gamma}_{1,k}^{(H)}$  increases with decreasing value of  $p$ , i.e. for smaller  $\gamma_2/\gamma_1$ . Within the maximum likelihood approach, a bias reduced estimator was then proposed for the censoring case following the technique from Beirlant et al. (2009) where the distribution of the excesses  $X/t|X > t$  is approximated by the EPD with CDF  $\mathbb{P}(Y \leq y) = 1 - (y\{1 + \delta_t(1 - y^{-\beta_1})\})^{-1/\gamma_1}$  where  $\delta_t = \gamma_1 D_1 t^{-\beta_1} (1 + o(1))$  as  $t \rightarrow \infty$ . The resulting estimator is given by

$$\hat{\gamma}_{1,k}^{(EP)} = \hat{\gamma}_{1,k}^{(H)} + C_{\hat{\gamma}_{1,k}^{(H)}, \beta_*} \frac{H_{Z,k}^{(-\beta_*)}}{\hat{p}_k} \left\{ H_{Z,k}^{(-\beta_*)} - \hat{\gamma}_{1,k}^{(H)} E_{Z,k}^{(c)}(-\beta_*) \right\} \quad (4.2.8)$$

where,  $\beta_* = \min(\beta_1, \beta_2)$  and

$$\begin{aligned} H_{Z,k}^{(-\beta_*)} &= \frac{1}{\beta_*} \left( 1 - \frac{1}{k} \sum_{j=1}^k \left( \frac{Z_{n-j+1,n}}{Z_{n-k,n}} \right)^{-\beta_*} \right), \\ E_{Z,k}^{(c)}(-\beta_*) &= \frac{1}{k} \sum_{j=1}^k \Delta_{n-j+1,n} \left( \frac{Z_{n-j+1,n}}{Z_{n-k,n}} \right)^{-\beta_*}, \\ C_{\gamma,\beta_*} &= -\frac{(1 + \gamma\beta_*)^3(1 + 2\gamma\beta_*)}{\gamma^4\beta_*^3}. \end{aligned}$$

In this estimation procedure  $\beta_*$  is assumed to be known. In fact, since in the definition of the EP distribution the term  $y^{-\beta_*}$  is multiplied by the  $\delta_t$ -factor, the asymptotic distribution of tail estimators based on the EP distribution will not depend on the asymptotic distribution of an estimator of  $\beta_*$ . One can also impute estimators of the parameter  $\beta_*$  of the distribution  $H$  of  $Z$  without increasing the bias in estimating  $\gamma_1$ . Estimators of  $\rho_* = -\gamma\beta_*$  were discussed in Fraga Alves et al. (2003). An estimator for  $\beta_*$  is then given by  $-\rho_*/\hat{\gamma}_{Z,k}^{(H)}$ . In the simulations the sensitivity of the choice of  $\rho_*$  was examined. In (4.2.8) one can reparametrize  $\beta_*\hat{\gamma}_{Z,k}^{(H)}$  by  $-\rho_*$  with  $\rho_* < 0$ , leading to

$$\hat{\gamma}_{1,k}^{(EP)}(\rho_*) = \hat{\gamma}_{1,k}^{(H)} - \frac{\hat{\gamma}_{1,k}^{(H)}(1 - \rho_*)^2(1 - 2\rho_*)}{\rho_*^3} \left\{ H_{Z,k}^{(\rho_*/\hat{\gamma}_{Z,k}^{(H)})} - \hat{\gamma}_{1,k}^{(H)} E_{Z,k}^{(c)}(\rho_*/\hat{\gamma}_{Z,k}^{(H)}) \right\}. \quad (4.2.9)$$

It was shown that when using the correct value of  $\beta_*$  or  $\rho_*$ , the asymptotic bias of  $\hat{\gamma}_{1,k}^{(EP)}$  is 0 as long as  $\sqrt{k}(k/n)^{\beta_*} = O(1)$ , whereas the asymptotic bias of the original estimator  $\hat{\gamma}_{1,k}^{(H)}$  is only 0 when  $\sqrt{k}(k/n)^{\beta_*} \rightarrow 0$  as  $k, n \rightarrow \infty$ . Hence the bias is reduced for a longer set of values of  $k \geq 1$  when choosing  $t$  as  $X_{n-k,n}$ . At the other hand the variance of this bias reduced estimator was shown to be increased by the factor  $\left(\frac{1+\gamma\beta_*}{\gamma\beta_*}\right)^2$  in comparison with the estimator  $\hat{\gamma}_{1,k}^{(H)}$ .

Before comparing the different estimators through simulations, we next derive a bias reduced estimator starting from the Worms & Worms estimator  $\hat{\gamma}_{1,k}^{(W)}$  from (4.2.3). Following Beirlant et al. (2019), we then also apply shrinkage estimation on  $\delta_t$  forcing this parameter to decrease to 0 as  $t \rightarrow \infty$  or  $k \downarrow 1$  as it is the case in the mathematical definition of  $\delta_t$ .

### 4.3 Bias reduction of the Worms & Worms estimator and penalised estimation of bias

Using the EP approximation to the survival function  $\frac{\bar{F}(ut)}{\bar{F}(t)}$  of the excesses  $X/t|X > t$ , leads to the following approximation of the integral expression of  $L_t$  in (4.2.1): as  $t \rightarrow \infty$

$$L_t = \int_1^\infty \frac{\bar{F}(ut)}{\bar{F}(t)} \frac{du}{u} = \gamma_1 - \delta_t \frac{\beta_1 \gamma_1}{1 + \beta_1 \gamma_1} (1 + o(1)). \quad (4.3.10)$$

Similarly, considering

$$E_t(-\beta_1) := \mathbb{E} \left( \left( \frac{X}{t} \right)^{-\beta_1} | X > t \right) = 1 + \int_1^\infty \frac{\bar{F}(ut)}{\bar{F}(t)} du^{-\beta_1}$$

leads to

$$(1 + \gamma_1 \beta_1) E_t(-\beta_1) = 1 + \frac{\delta_t}{\gamma_1} \frac{(\beta_1 \gamma_1)^2}{1 + 2\beta_1 \gamma_1} (1 + o(1)), \quad t \rightarrow \infty. \quad (4.3.11)$$

Substituting  $\gamma_1$  in the left hand side of (4.3.11) by the expression  $L_t + \delta_t \frac{\beta_1 \gamma_1}{1 + \beta_1 \gamma_1}$  which follows from (4.3.10), one obtains for  $t \rightarrow \infty$  that

$$\begin{aligned} & (1 + \beta_1 L_t) E_t(-\beta_1) \\ &= 1 + \left\{ \frac{\delta_t}{\gamma_1} \frac{(\beta_1 \gamma_1)^2}{1 + 2\beta_1 \gamma_1} - \delta_t \frac{\beta_1^2 \gamma_1}{1 + \beta_1 \gamma_1} E_t(-\beta_1) \right\} (1 + o(1)) \\ &= 1 + \frac{\delta_t}{\gamma_1} (\beta_1 \gamma_1)^2 \left[ \frac{1}{1 + 2\beta_1 \gamma_1} - \frac{1}{(1 + \beta_1 \gamma_1)^2} \right] (1 + o(1)) \\ &= 1 + \frac{\delta_t}{\gamma_1} \frac{(\beta_1 \gamma_1)^4}{(1 + \beta_1 \gamma_1)^2 (1 + 2\beta_1 \gamma_1)} (1 + o(1)), \end{aligned}$$

where in the second step we approximated  $E_t(-\beta_1)$  by  $(1 + \gamma_1 \beta_1)^{-1}$ . We now conclude that

$$\delta_t = \frac{L_t (1 + \beta_1 L_t)^3 (1 + 2\beta_1 L_t)}{(\beta_1 L_t)^4} \left\{ E_t(-\beta_1) - \frac{1}{1 + \beta_1 L_t} \right\} (1 + o(1)). \quad (4.3.12)$$

Estimating  $L_t$  at a random threshold  $Z_{n-k,n}$  by  $\hat{\gamma}_{1,k}^{(W)}$  and similarly  $E_t(-\beta_1)$  by

$$\hat{E}_k(-\beta_1) = 1 + \sum_{j=1}^k \frac{1 - \hat{F}_n^{KM}(Z_{n-j+1,n})}{1 - \hat{F}_n^{KM}(Z_{n-k,n})} \left( Z_{n-j+1,n}^{-\beta_1} - Z_{n-j,n}^{-\beta_1} \right) / Z_{n-k,n}^{-\beta_1},$$

we obtain the following bias reduced estimator for  $\gamma_1$  combining (4.3.10) and (4.3.12):

$$\hat{\gamma}_{1,k}^{(BR,W)} = \hat{\gamma}_{1,k}^{(W)} + \frac{\hat{\gamma}_{1,k}^{(W)}(1 + \beta_1 \hat{\gamma}_{1,k}^{(W)})^2(1 + 2\beta_1 \hat{\gamma}_{1,k}^{(W)})}{(\beta_1 \hat{\gamma}_{1,k}^{(W)})^3} \left\{ \hat{E}_k(-\beta_1) - \frac{1}{1 + \beta_1 \hat{\gamma}_{1,k}^{(W)}} \right\}. \quad (4.3.13)$$

In (4.3.13) one can reparametrize  $\beta_1 \hat{\gamma}_{1,k}^{(W)}$  by  $-\rho_1$  with  $\rho_1 < 0$ , leading to

$$\hat{\gamma}_{1,k}^{(BR,W)}(\rho_1) = \hat{\gamma}_{1,k}^{(W)} - \frac{\hat{\gamma}_{1,k}^{(W)}(1 - \rho_1)^2(1 - 2\rho_1)}{\rho_1^3} \left\{ \hat{E}_k(\rho_1/\hat{\gamma}_{1,k}^{(W)}) - \frac{1}{1 - \rho_1} \right\}, \quad (4.3.14)$$

and we will study the sensitivity of the estimator with respect to the choice of  $\rho_1$ . In fact our objective will be to look for an appropriate choice of  $\rho_1$  such that the plot of the estimates as a function of  $k$  is most constant in order to assist practitioners.

Also here the variance of the bias reduced estimator can be expected to be inflated compared with the corresponding estimator (here  $\hat{\gamma}_{1,k}^{(W)}$ ). This will be confirmed by the simulations in the next section. However, Beirlant et al. (2019) showed that this problem can be alleviated forcing the bias estimator  $\frac{\hat{\gamma}_{1,k}^{(W)}(1-\rho_1)^2(1-2\rho_1)}{(-\rho_1)^3} \left\{ \hat{E}_k(\rho_1/\hat{\gamma}_{1,k}^{(W)}) - \frac{1}{1-\rho_1} \right\}$  to decrease to 0 as  $t \rightarrow \infty$  or  $k \downarrow 1$ . Formally applying the shrinkage procedure introduced in Chapter 3, leads to a penalised version of (4.3.12):

$$\delta_t^s = \frac{1 + \beta_1 L_t}{\frac{\omega L_t}{k\sigma_{1,k,n}^2} + \frac{(\beta_1 L_t)^4}{L_t(1 + \beta_1 L_t)^2(1 + 2\beta_1 L_t)}} \left\{ E_t(-\beta_1) - \frac{1}{1 + \beta_1 L_t} \right\},$$

where  $\sigma_{1,k,n}^2 = (k/n)^{-2\rho_1}$  and  $\omega$  is a weight factor that allows to control the penalisation. The term  $(\omega L_t)/(k\sigma_{1,k,n}^2)$  makes the bias correction shrink for smaller values of  $k$ , i.e. when the original estimator  $\hat{\gamma}_{1,k}^{(W)}$  is asymptotically unbiased, namely  $k\sigma_{1,k,n}^2 \rightarrow 0$ . This then leads to the penalised estimator

$$\hat{\gamma}_{1,k}^{(s,W)}(\rho_1) = \hat{\gamma}_{1,k}^{(W)} - \frac{\rho_1}{\frac{\omega \hat{\gamma}_{1,k}^{(W)}}{k\sigma_{1,k,n}^2} + \frac{\rho_1^4}{\hat{\gamma}_{1,k}^{(W)}(1 - \rho_1)^2(1 - 2\rho_1)}} \left\{ \hat{E}_k(\rho_1/\hat{\gamma}_{1,k}^{(W)}) - \frac{1}{1 - \rho_1} \right\}. \quad (4.3.15)$$

In a similar way the bias component in  $\hat{\gamma}_{1,k}^{(EP)}(\rho_*)$  can be penalised for smaller values

of  $k$ :

$$\hat{\gamma}_{1,k}^{(s,EP)}(\rho_*) = \hat{\gamma}_{1,k}^{(H)} - \frac{\rho_*}{\frac{\omega \hat{\gamma}_{1,k}^{(H)}}{k\sigma_{*,k,n}^2} + \frac{\rho_*^4}{\hat{\gamma}_{1,k}^{(H)}(1-\rho_*)^2(1-2\rho_*)}} \left\{ H_{Z,k}^{(\rho_*/\hat{\gamma}_{Z,k}^{(H)})} - \hat{\gamma}_{1,k}^{(H)} E_{Z,k}^{(c)}(\rho_*/\hat{\gamma}_{Z,k}^{(H)}) \right\} \quad (4.3.16)$$

with  $\sigma_{*,k,n}^2 = (k/n)^{-2\rho_*}$ .

## 4.4 Bootstrap confidence intervals for $\gamma_1$

Given the lack of any distribution theory for the Worms and Worms estimator  $\hat{\gamma}_{1,k}^{(W)}$  we here present a parametric bootstrap algorithm in order to construct confidence intervals for  $\gamma_1$ .

The main idea behind this bootstrap procedure is that for a value of  $k$  where the bias of an estimator of  $\gamma_1$  is 0, one can as well simulate from simple Pareto distributions rather than from the true Pareto-type distribution  $F$  and  $G$  in order to construct samples of estimators. Also note that

$$\hat{\gamma}_{2,k}^{(W)} = \sum_{j=1}^k \frac{1 - \hat{G}_n^{KM}(Z_{n-j+1,n})}{1 - \hat{G}_n^{KM}(Z_{n-k,n})} (\log Z_{n-j+1,n} - \log Z_{n-j,n}),$$

where  $1 - \hat{G}_n^{KM}(y) = \Pi_{Z_{i,n} \leq y} \left(1 - \frac{1}{n-i+1}\right)^{1-\Delta_{i,n}}$  denotes the Kaplan-Meier estimator of  $G$ , jointly with its bias reduced versions constructed in a similar way as in the preceding section (replacing  $\Delta_{n-j+1,n}$  by  $1 - \Delta_{n-j+1,n}$ ), lead to estimates of  $\gamma_2$ . The procedure then runs as follows:

- Given a value of  $\hat{k}_1$ , respectively  $\hat{k}_2$ , where the bias of the estimator  $\hat{\gamma}_{1,k}^{(W)}$ , respectively  $\hat{\gamma}_{2,k}^{(W)}$ , is judged to be negligible, one can perform a parametric bootstrap using samples of size  $n$  from  $\left(\min(\hat{X}_i, \hat{C}_i), 1_{(\hat{X}_i \leq \hat{C}_i)}\right)$  ( $i = 1, \dots, n$ ) where  $\hat{X}_i$ , respectively  $\hat{C}_i$ , are simulated from a standard Pareto distribution with survival function  $x^{-1/\hat{\gamma}_{1,\hat{k}_1}^{(W)}}$ ,  $x > 1$ , respectively  $y^{-1/\hat{\gamma}_{2,\hat{k}_2}^{(W)}}$ ,  $y > 1$ .

- The values  $\hat{k}_j$ ,  $j = 1, 2$ , are chosen from

$$\hat{k}_j = \max\{k : |\hat{\gamma}_{j,k}^{(W)} - \hat{\gamma}_{j,k}^{(s,W)}(\rho_j)| \leq \epsilon\}$$

for a small value of  $\epsilon$ . From each bootstrap sample one then retains a bootstrap estimate  $\hat{\gamma}_{1,\hat{k}_1}^{(*,s,W)}$  of  $\gamma_1$ .

- Finally, repeating this bootstrap sampling step  $N$  times, we consider the empirical distribution of the values  $\hat{\gamma}_{1,\hat{k}_1}^{(*,s,W)}(j)$  ( $j = 1, \dots, N$ ), and more specifically the  $\lfloor N\alpha/2 \rfloor$ ,  $\lfloor N(1-\alpha)/2 \rfloor$  empirical quantiles, in order to construct a  $100(1-\alpha)\%$  confidence interval for  $\gamma_1$ .

In order to test this bootstrap procedure in the next section we will apply this procedure to several simulated censored samples under different values of the proportion of non-censoring.

## 4.5 Finite sample simulations

As the asymptotic distribution of  $\hat{\gamma}_{1,k}^{(W)}$  and hence also of  $\hat{\gamma}_{1,k}^{(BR,W)}(\rho_1)$  and  $\hat{\gamma}_{1,k}^{(s,W)}(\rho_1)$  is not known, we here consider a comparison using finite sample simulations. We report the simulation results for sample size  $n = 500$  from

- the Burr  $(\eta, \tau, \lambda)$  distribution with right tail function

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

with  $\eta, \tau, \lambda > 0$ , and  $\gamma = 1/(\tau\lambda)$ ,  $\beta = \tau$ ,  $D = -\lambda\eta$ ;

- the Fréchet  $(\alpha)$  distribution with right tail function

$$1 - F(x) = 1 - \exp(-x^{-\alpha}), \quad x > 0,$$

with  $\alpha > 0$ , and  $\gamma = 1/\alpha$ ,  $\beta = \alpha$ ,  $D = -1/2$ .

Here we present results concerning the bias and the root mean squared error of the different estimators of  $\gamma_1$  discussed above, and of the bootstrap algorithm, in case

- Burr (10, 2, 2) censored by Burr (10, 5, 2) with  $\gamma_1 = 0.25$  and  $\gamma_2 = 0.10$ , leading to heavy censoring with the proportion of non-censoring  $p = 0.286$ ; see Figure 4.1;
- Burr (10, 2, 1) censored by Burr (10, 2, 1) with  $\gamma_1 = \gamma_2 = 0.5$  so that  $p = 0.5$ ; see Figure 4.2;
- Burr (10, 5, 2) censored by Burr (10, 2, 2) with  $\gamma_1 = 0.10$  and  $\gamma_2 = 0.25$ , with light censoring  $p = 0.714$ ; see Figure 4.3;
- Fréchet (2) censored by Fréchet (1) with  $\gamma_1 = 0.5$  and  $\gamma_2 = 1$ , so that  $p = 2/3$ ; see Figure 4.4.

In each of these four cases we consider the results for

- $\hat{\gamma}_{1,k}^{(H)}$  from (4.2.2),  $\hat{\gamma}_{1,k}^{(EP)}(\rho_*)$  from (4.2.9), and  $\hat{\gamma}_{1,k}^{(s,EP)}(\rho_*)$  from (4.3.16) with  $\omega = 1$  and for different values of  $\rho_*$  (left in Figures 4.1-4.4),
- $\hat{\gamma}_{1,k}^{(W)}$  from (4.2.3),  $\hat{\gamma}_{1,k}^{(KM)}$  from (4.2.4),  $\hat{\gamma}_{1,k}^{(BR,W)}(\rho_1)$  from (4.3.14), and  $\hat{\gamma}_{1,k}^{(s,W)}(\rho_1)$  from (4.3.15) with  $\omega = 1$  and for different values of  $\rho_1$  (see middle of Figures 4.1-4.4),
- the Bayesian estimators  $\hat{\gamma}_{1,k}^{(m,M)}$ ,  $\hat{\gamma}_{1,k}^{(e,1,2)}$  and  $\hat{\gamma}_{1,k}^{(m,1,2)}$  from (4.2.5), (4.2.6) and (4.2.7) (right in Figures 4.1-4.4).

One observes that in case  $p < 0.5$  the likelihood based estimator  $\hat{\gamma}_{1,k}^{(H)}$  and its bias reduced versions, and the Bayesian estimators have larger bias than the estimators derived from estimating the functional form  $L_t$  in (4.2.1). Bias reduction of  $\hat{\gamma}_{1,k}^{(H)}$  helps only partially in such cases. In these cases the bias reduced and penalised estimators  $\hat{\gamma}_{1,k}^{(BR,W)}$  and  $\hat{\gamma}_{1,k}^{(s,W)}(\rho_1)$  perform good with low bias and low RMSE for a long interval of values of  $k$  which is quite helpful in choosing an appropriate value of  $k$ . This is

in contrast with the bias of  $\hat{\gamma}_{1,k}^{(W)}$  and  $\hat{\gamma}_{1,k}^{(KM)}$  which is systematically decreasing with decreasing value of  $k$ . In order to evaluate the effect of the penalisation in  $\hat{\gamma}_{1,k}^{(s,W)}(\rho_1)$ , we focused the scale for the bias and RMSE plots in the Burr cases, see Figure 4.5. Especially in case  $p \leq 0.5$  and for smaller values of  $k$  the mean of the shrinkage estimator is much more stable and ultimately for small  $k$  the behaviour of this estimator follows that of  $\hat{\gamma}_{1,k}^{(W)}$ . The resulting RMSE is then also a lower envelope of the RMSE curves of  $\hat{\gamma}_{1,k}^{(W)}$  and  $\hat{\gamma}_{1,k}^{(BR,W)}$ .

On the other hand when  $p > 0.5$ , and especially in the Fréchet case, the basic estimators  $\hat{\gamma}_{1,k}^{(H)}$ ,  $\hat{\gamma}_{1,k}^{(W)}$  and  $\hat{\gamma}_{1,k}^{(m,M)}$  work almost equally well, and this also holds for the different approaches to bias reduction. Within the group of Bayes estimators clearly  $\hat{\gamma}_{1,k}^{(m,M)}$  works best.

We also tested the proposed bootstrap procedure in the same cases as considered in Figures 4.1-4.4. We applied the algorithm with  $\alpha = 0.05$  and  $N = 1000$  to 1000 samples of size  $n = 500$ . In Figure 4.4 the 1000 confidence intervals are given when choosing  $\hat{k}_1$  and  $\hat{k}_2$  adaptively using  $\epsilon = 0.01$ , and when keeping  $\hat{k}_1 = \hat{k}_2$  fixed to  $25 = 0.05 \times n$  throughout (this value of  $k$  appears appropriate on the basis of Figures 4.1 -4.4). Further simulations showed that for  $n = 1000$  keeping  $\hat{k}_1 = \hat{k}_2$  fixed to  $0.04 \times n$  leads to confidence intervals that attain the required 95% level closely. The confidence intervals missing the correct value of  $\gamma_1$  are put in dark grey. Of course when the  $k$  values are chosen adaptively, it is more difficult to attain the confidence level  $1 - \alpha = 0.95$ .

A deeper understanding of the distribution of  $\hat{\gamma}_{1,k}^{(W)}$  and  $\hat{E}_k(-\beta_1)$  appears necessary to enhance the adaptive choice of  $k_1, k_2$  and the performance of the bootstrap algorithm.



## 4.6 A case study from car insurance

Finally, in order to illustrate the merits of the newly proposed method, we consider a data set with indexed total payments from a motor third party liability insurance company operating in the European union (EU), with records from 1995 till 2010 with  $n = 849$  claims of which only 340 were completely developed at the end of 2010. For every claim the indexed cumulative payments are given at the end of every year until development. In Figure 4.6 we plotted the proportions of non-censored data  $\hat{p}_k$  which are situated in the top  $100k/n\%$  of cumulative payments at the end of 2010 as a function of  $k/n$ . In practice most companies substitute the censored observations by ultimate predictions obtained through reserving techniques. Here we show how the extreme value methods for censored data can also be used directly without ultimates in order to obtain relevant extreme value predictions.

Due to the long-tail nature of such portfolios, only the claims with arrival year between 1995 and 1999 can be considered to satisfy the condition of weak censoring  $p > 0.5$  or  $\gamma_1 < \gamma_2$  when using the information up to 2010. For this group of early claims 29% is censored at the end of 2010, while the percentage of censoring is 60% when considering all claims. Note also that when considering all claims the largest 20 % are all censored, whereas this amount increases to 40% for the claims arriving after 2003. Needless to say that extreme value methodology is quite challenging in such a case.

In order to illustrate the stability of the proposed bias reduction technique based on the Worms and Worms estimator over different percentages of censoring, we also split the full data set in groups along the arrival times in 1995-1999, 1998-2002, 2001-2005, 2004-2008, 2007-2010. In Figure 4.7 the original estimates  $\hat{\gamma}_{1,k}^{(W)}$  and  $\hat{\gamma}_{1,k}^{(BR,W)}(-3)$  are given as a function of  $k/n$  for each of these subgroups and for the complete data set. The value  $\rho_1 = -3$  was chosen as this value yields the most constant plots as a function of  $k$ . The plots of  $\hat{\gamma}_{1,k}^{(W)}$  are steepest for the claims which are most recent in 2010. The

stability of the bias reduced estimates over these subgroups is quite convincing leading to estimates of  $\gamma_1$  between 0.6 and 0.7.

As another validity check, in Figure 4.8 we consider only the claims from 1995-1999 with their cumulative payments as of 2000 till 2010 in steps of 2 years. Note that in 2000 only 9% of those claims were fully developed, while at 2010 this percentage rose to 71%. Again the bias reduced estimates  $\hat{\gamma}_{1,k}^{(BR,W)}(-3)$  are remarkably stable over  $k$ .

We also applied the bootstrap algorithm to this case study. Using  $\epsilon = 0.01$  leads to  $\hat{k}_1 = 73$ ,  $\hat{k}_2 = 50$ ,  $\hat{\gamma}_{1,73}^{(W)} = 0.725$ , and  $\hat{\gamma}_{2,50}^{(W)} = 0.652$ , see Figure 4.10 (left). The confidence intervals for the different values of  $k$  are given in Figure 4.10 (right) with special attention for the case  $k = 73$  which leads to the interval 95% confidence interval (0.48; 0.91). Choosing  $\hat{k}_1 = \hat{k}_2$  fixed at 4 to 5% of the sample size  $n = 849$ , as suggested in the simulation section, leads to lower bounds that are somewhat lower than 0.48, as can be seen from Figure 4.10 (right).

## 4.7 Chapter conclusion

The estimator  $\hat{\gamma}_{1,k}^{(W)}$  from Worms and Worms (2014) has the best RMSE behaviour between all available first order estimators of  $\gamma_1$ . In order to enhance the practical use of this estimator we proposed bias reduction and penalisation techniques which lead to improved bias and RMSE behaviour. Moreover a bootstrap procedure is proposed in order to construct confidence intervals. This is especially useful with long-tailed insurance products. In order to enhance the adaptive choice of the number of extreme data  $k$  asymptotic representations of the estimators involved are needed for all cases, but especially in case of heavy censoring. This will be the subject of future work.

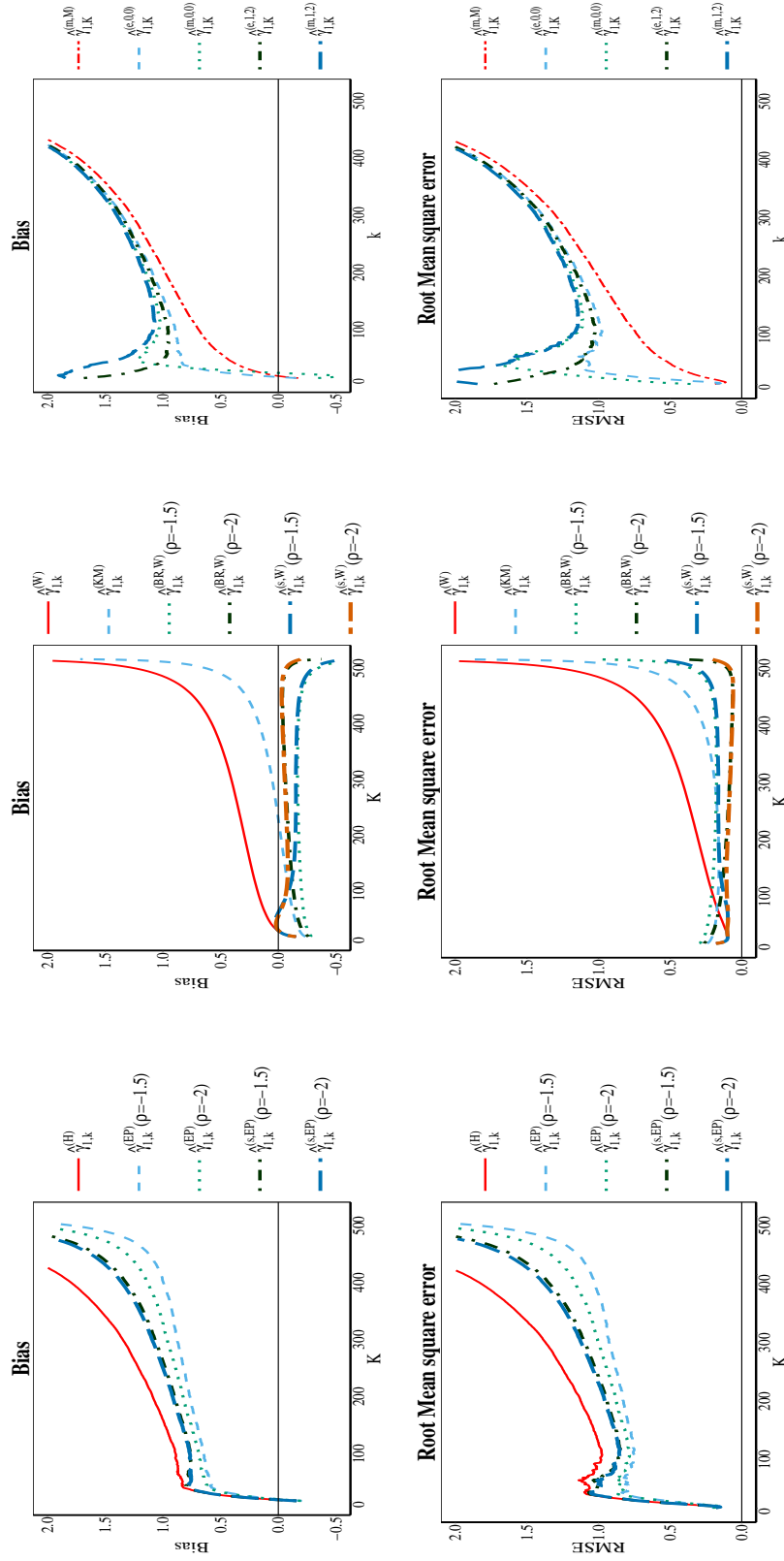


Figure 4.1: Bias and RMSE for Burr(10,2,2) censored by Burr(10,5,2)

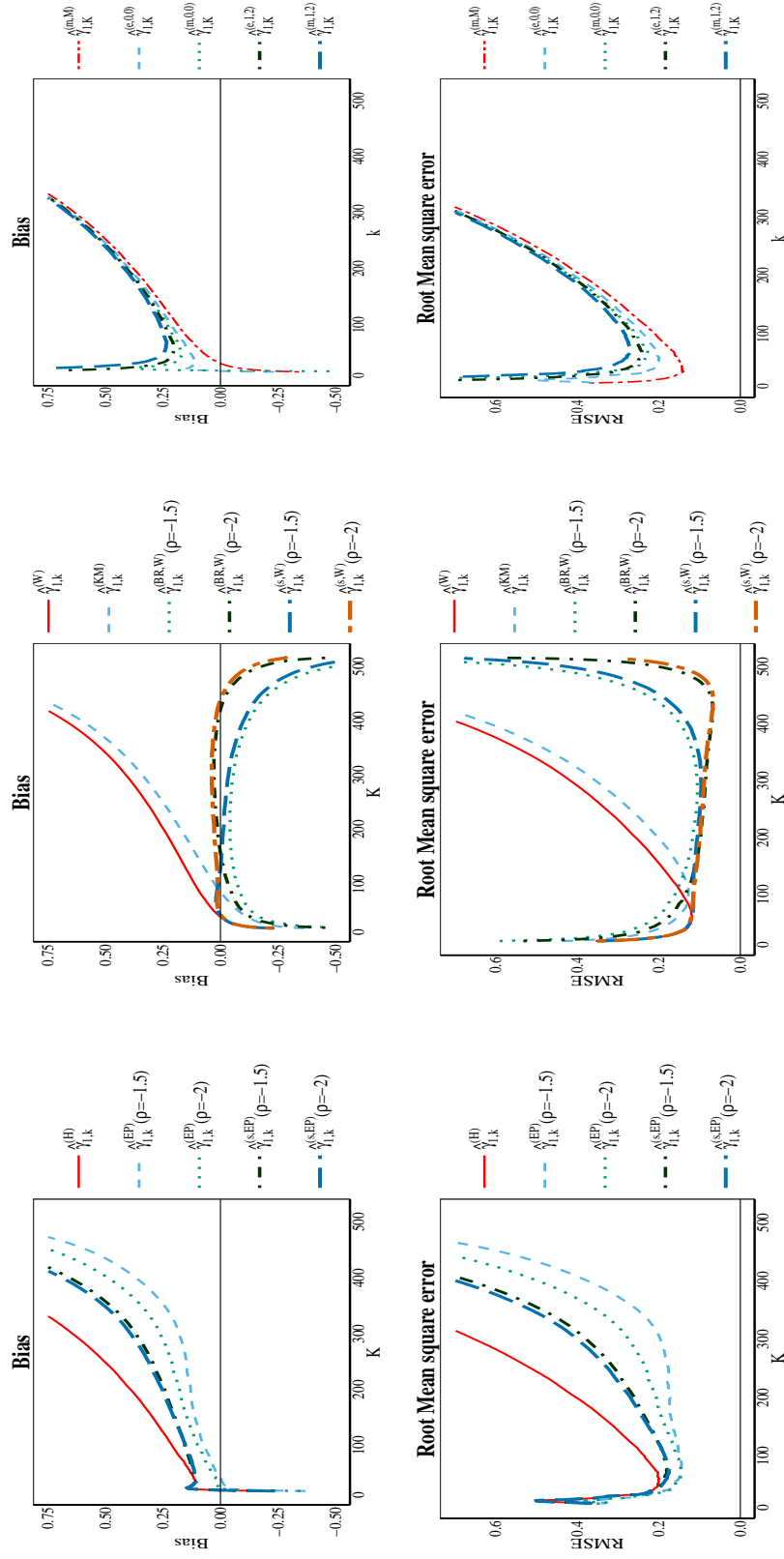


Figure 4.2: Bias and RMSE for Burr(10,2,1) censored by Burr(10,2,1)

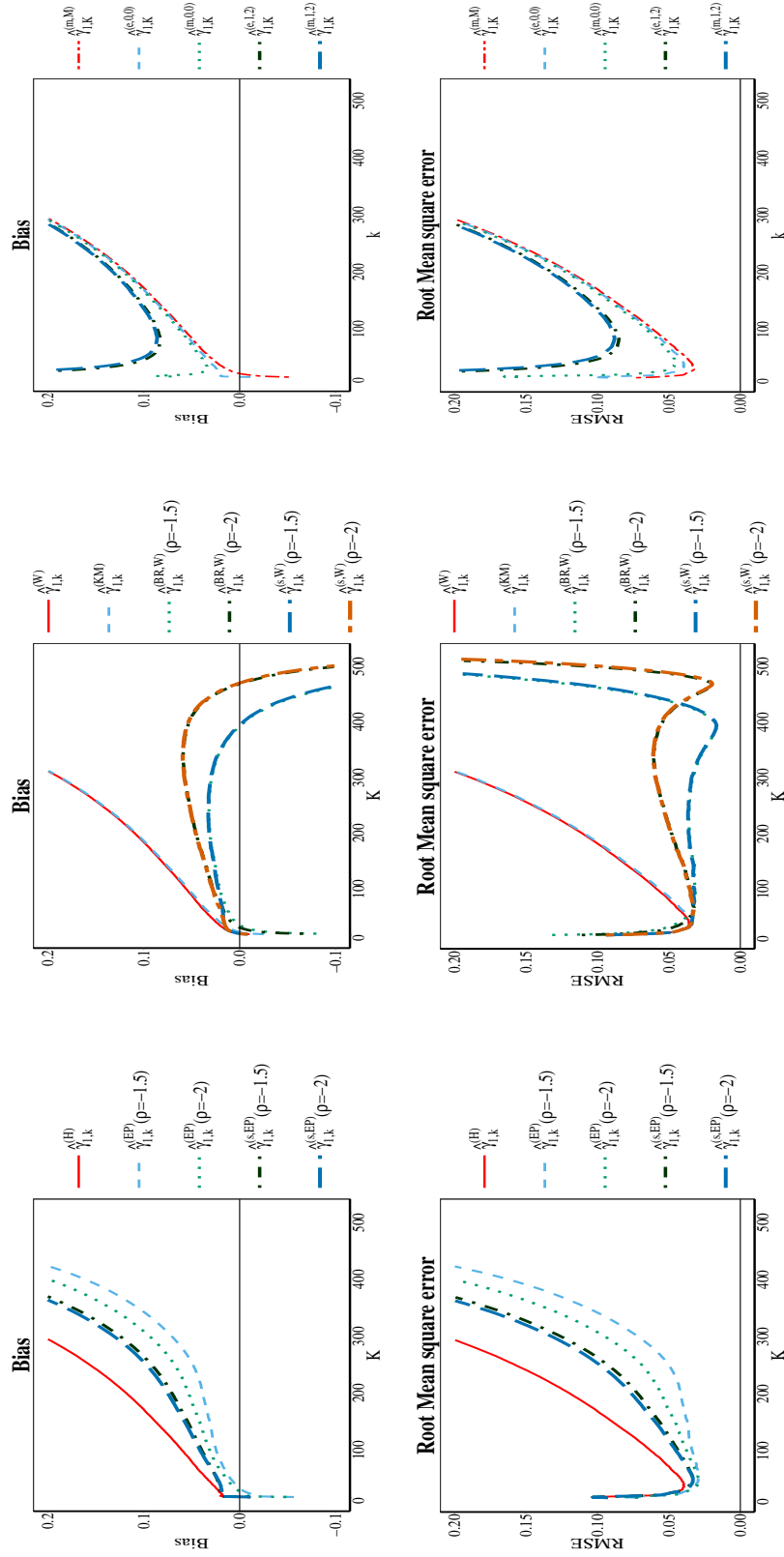


Figure 4.3: Bias and RMSE for Burr(10,5,2) censored by Burr(10,2,2)

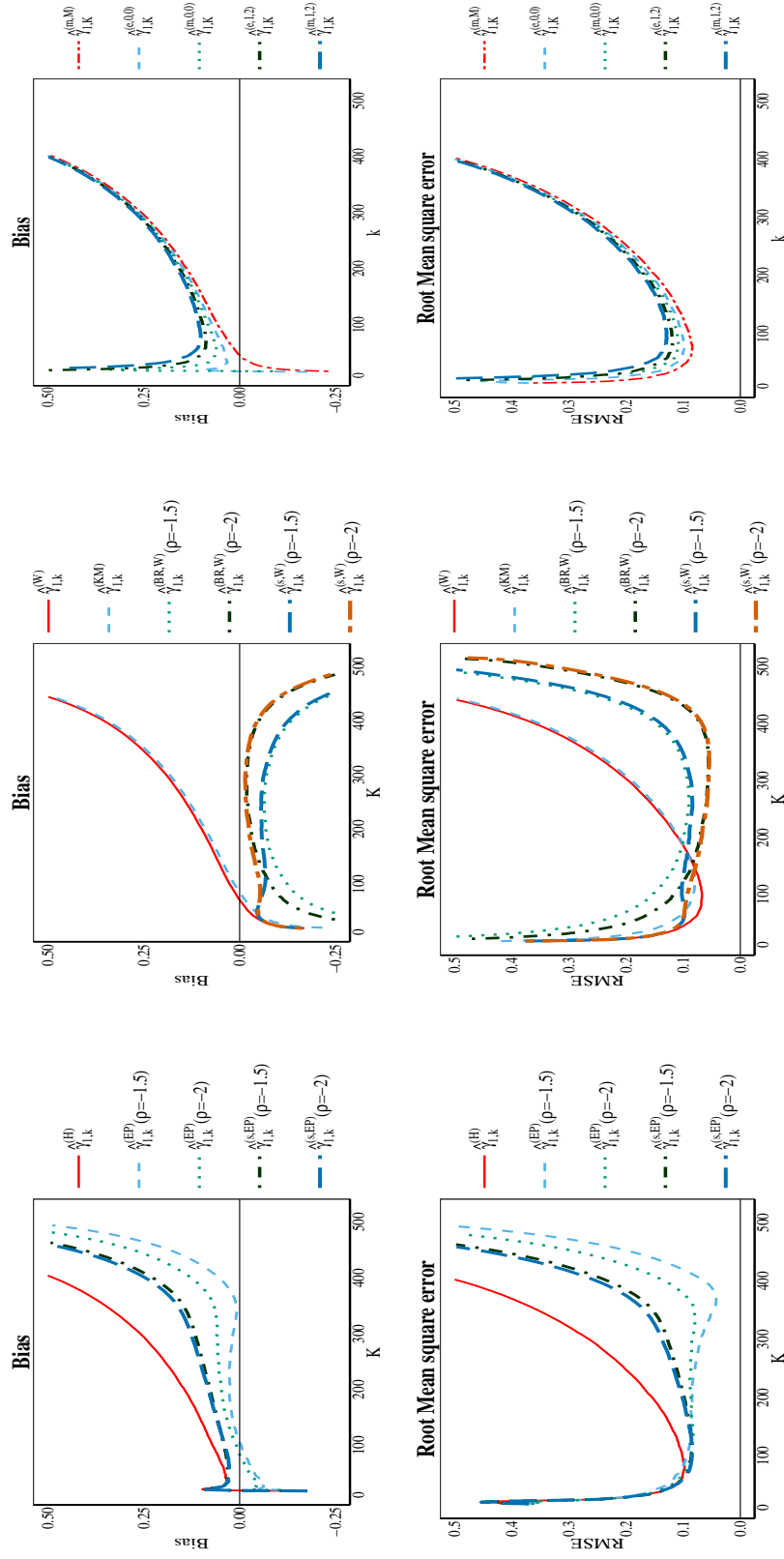
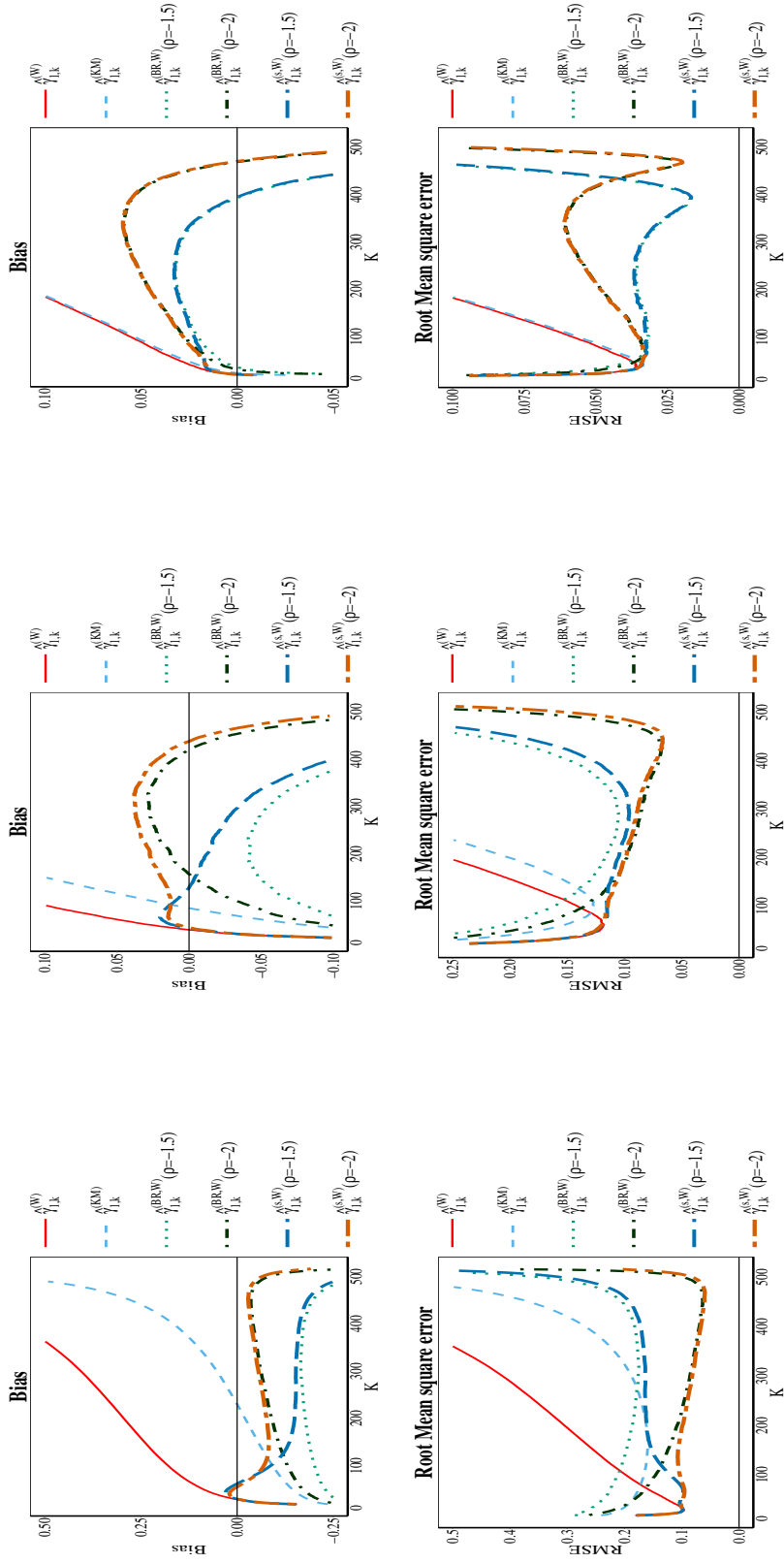
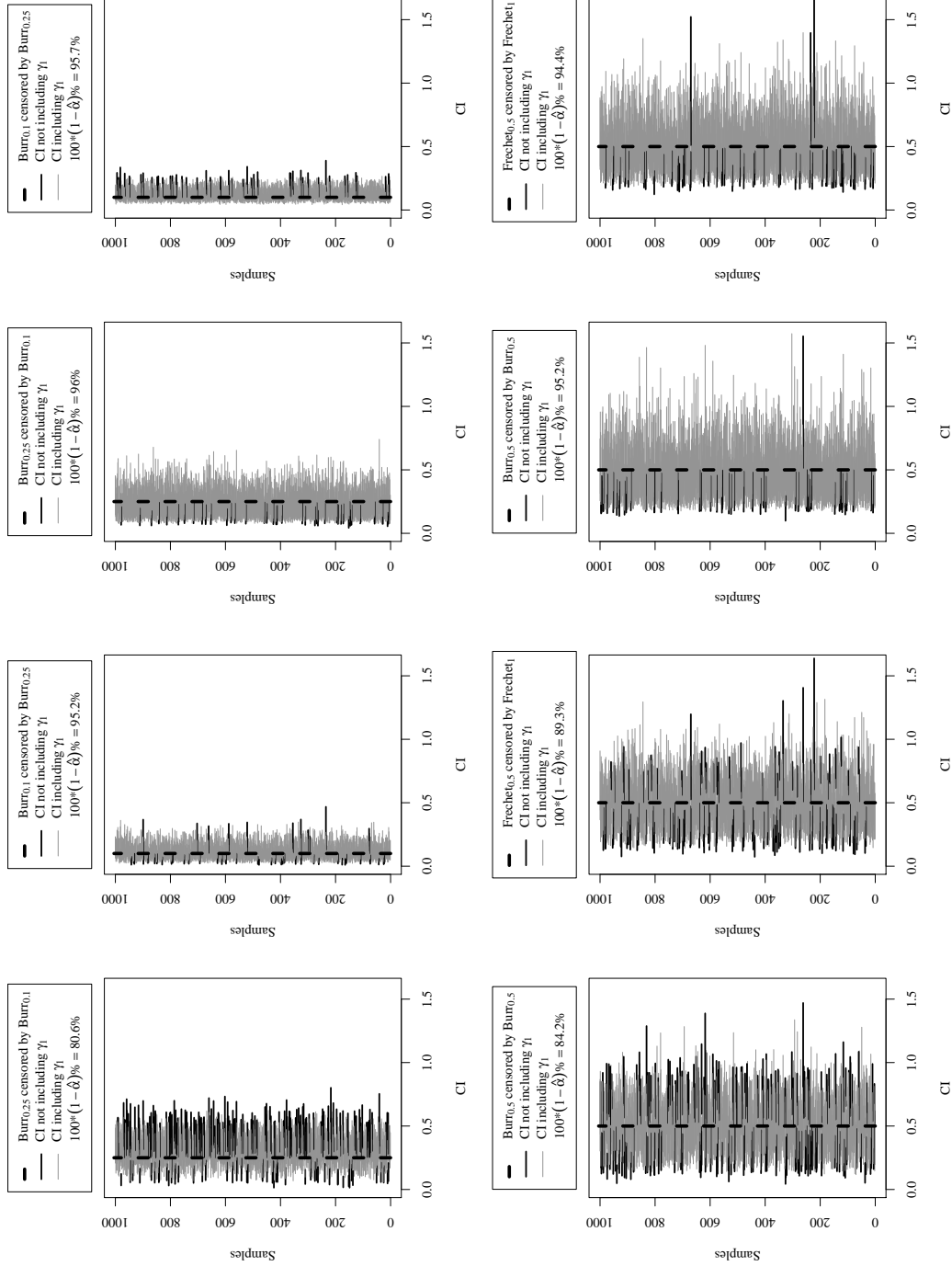


Figure 4.4: Bias and RMSE for Fréchet(2) censored by Fréchet(1)

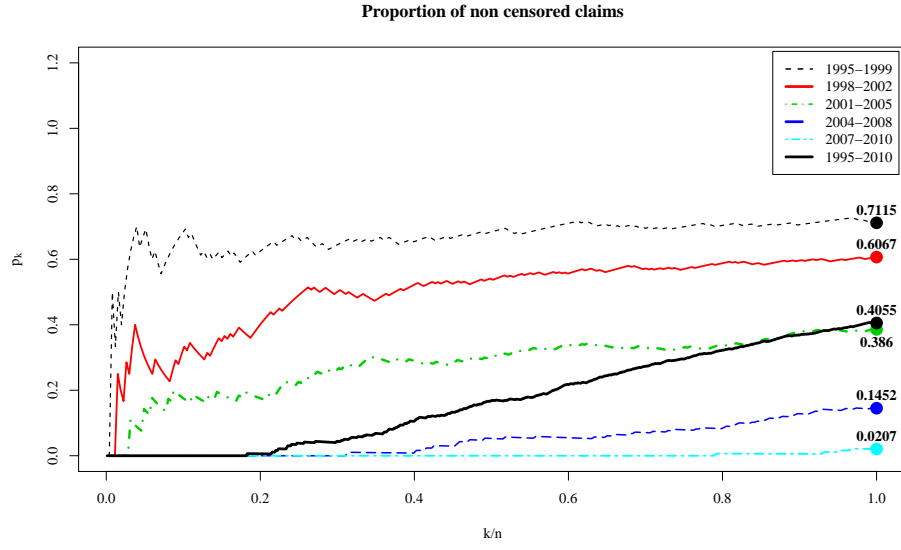


**Figure 4.5:** Bias (top) and RMSE (bottom) for  $\hat{\gamma}_{1,k}^{(W)}$ ,  $\hat{\gamma}_{1,k}^{(BR,W)}(-1.5)$ ,  $\hat{\gamma}_{1,k}^{(BR,W)}(-2)$ ,  $\hat{\gamma}_{1,k}^{(s,W)}(-1.5)$ ,  $\hat{\gamma}_{1,k}^{(s,W)}(-2)$  in case of Burr(10,2,2) censored by Burr(10,2,1) (left); Burr(10,2,1) censored by Burr(10,2,1) (middle); and Burr(10,5,2) censored by Burr(10,2,2) (right)

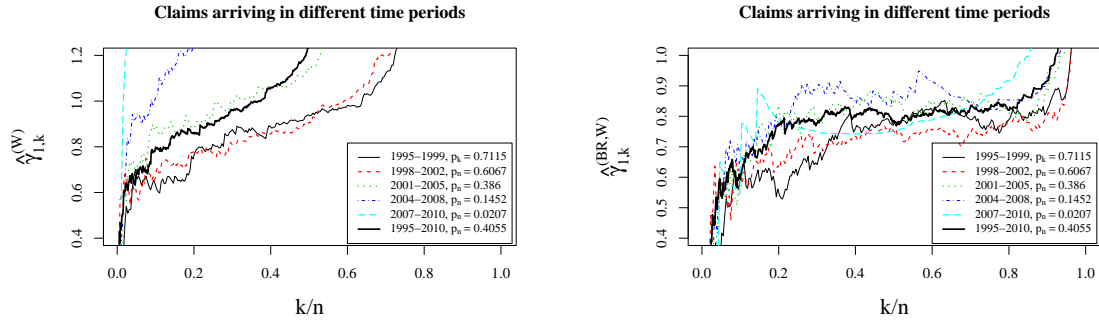


**Figure 4.6:** Simulated bootstrap 95% confidence intervals with adaptive choice of  $k_1$  and  $k_2$  using  $\epsilon = 0.01$  (4 left frames) and using  $k_1 = k_2 = 25 = 5\%n$  for every sample (4 right frames)

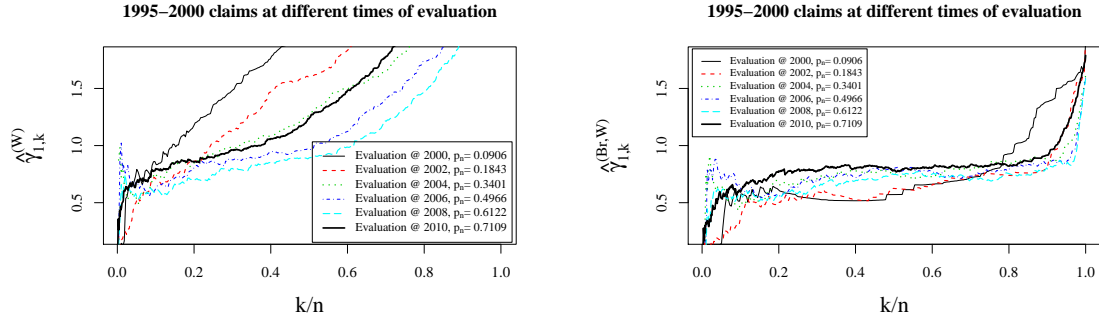




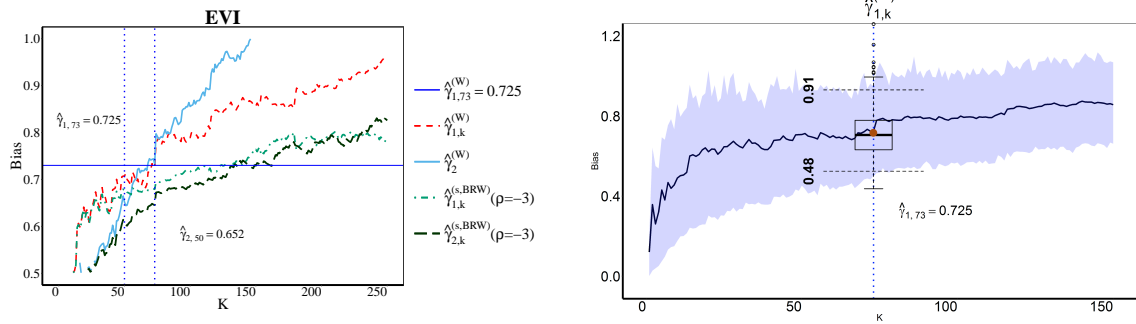
**Figure 4.7:** Car liability data: plot of  $\hat{p}_k$  as a function of  $k/n$  for all claims jointly (thick full line) and for claims that arrived in different time intervals.



**Figure 4.8:** Car liability data:  $\hat{\gamma}_{1,k}^{(W)}$  (left) and  $\hat{\gamma}_{1,k}^{(BR,W)}(-3)$  (right) based on indexed cumulative payments at end of 2010, as a function of  $k/n$  for all claims together (thick full line) and for claims arriving in different time periods



**Figure 4.9:** Car liability data:  $\hat{\gamma}_{1,k}^{(W)}$  (left) and  $\hat{\gamma}_{1,k}^{(BR,W)}(-3)$  (right) for all claims arriving in 1995–1999 and based on indexed cumulative payments at end of 2000(02)2010, as a function of  $k/n$ .



**Figure 4.10:** Car liability data, using all claims:  $\hat{\gamma}_{1,k}^{(W)}$ ,  $\hat{\gamma}_{1,k}^{(s,W)}(-3)$ ,  $\hat{\gamma}_{2,k}^{(W)}$ ,  $\hat{\gamma}_{2,k}^{(s,W)}(-3)$  (left); 95% bootstrap confidence intervals (right).

## Part II

**Semi-parametric Refinements:**

**General Tails ( $\gamma \in \mathbb{R}$ )**

# Chapter 5

## Bias Reduced Peaks over Threshold Tail Estimation

*This chapter is based on Beirlant, J., Maribe, G., Ph. Naveau and Verster, A., 2018. Bias Reduced Peaks over Threshold Tail Estimation, available on: arXiv:1810.01296*

### 5.1 Introduction

Extreme value methodology starts from the assumption that the distribution of the available sample  $X_1, X_2, \dots, X_n$  belongs to the domain of attraction of a GEV distribution, i.e. there exists sequences  $(b_n)_n$  and  $(a_n > 0)_n$  such that as  $n \rightarrow \infty$

$$\frac{\max(X_1, X_2, \dots, X_n) - b_n}{a_n} \rightarrow_d Y_\gamma, \quad (5.1.1)$$

where  $\mathbb{P}(Y_\gamma > y) = \exp(-(1 + \gamma y)^{-1/\gamma})$ , for some  $\gamma \in \mathbb{R}$  with  $1 + \gamma y > 0$ . The parameter  $\gamma$  is termed the EVI. It is well-known (see e.g. Beirlant et al., 2004; de Haan and Ferreira, 2006) that (5.1.1) is equivalent to the existence of a positive function  $t \mapsto \sigma_t$ , such that

$$\mathbb{P}\left(\frac{X - t}{\sigma_t} > y | X > t\right) = \frac{\bar{F}(t + y\sigma_t)}{\bar{F}(t)} \rightarrow_{t \rightarrow x_+} \bar{H}_\gamma^{GP}(y) = (1 + \gamma y)^{-1/\gamma}, \quad (5.1.2)$$

where  $\bar{F}(x) = \mathbb{P}(X > x)$  and  $x_+$  denotes the endpoint of the distribution of  $X$ . The conditional distribution of  $X - t$  given  $X > t$  is called the POT distribution, while  $\bar{H}_\gamma^{GP}$  is the survival function of the GPD.

In case  $\gamma > 0$ , the limit in (5.1.1) holds if and only if  $F$  is of Pareto-type, i.e.

$$\bar{F}(x) = x^{-1/\gamma} \ell(x), \quad (5.1.3)$$

for some slowly varying function  $\ell$ , i.e. satisfying  $\frac{\ell(yt)}{\ell(t)} \rightarrow 1$  as  $t \rightarrow \infty$ , for every  $y > 1$ .

Pareto-type distributions satisfy a simpler POT limit result: as  $t \rightarrow \infty$

$$\mathbb{P}\left(\frac{X}{t} > y | X > t\right) \rightarrow \bar{H}_\gamma^P(y) := y^{-1/\gamma}, y > 1. \quad (5.1.4)$$

Estimation of  $\gamma$  and tail quantities such as return periods is then based on fitting a GPD to the observed excesses  $X - t$  given  $X > t$ , respectively a simple Pareto distribution with survival function  $y^{-1/\gamma}$  to  $X/t$  given  $X > t$  in case  $\gamma > 0$ . The main difficulty in such an EV application is the choice of the threshold  $t$ . Most often, the threshold  $t$  is chosen as one of the top data points  $X_{n-k,n}$  for some  $k \in \{1, 2, \dots, n\}$  where  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$  denotes the ordered sample.

The limit results in (5.1.2) and (5.1.4) require  $t$  to be chosen as large as possible (or, equivalently,  $k$  as small as possible) for the bias in the estimation of  $\gamma$  and other tail parameters to be limited. However, in order to limit the estimation variance,  $t$  should be as small as possible, i.e. the number of data points  $k$  used in the estimation should be as large as possible. Several adaptive procedures for choosing  $t$  or  $k$  have been proposed, but mainly in the Pareto-type case with  $\gamma > 0$  under further second-order specifications of (5.1.3) or (5.1.4), see for instance Chapter 3 in Beirlant et al. (2004), or Matthys and Beirlant (2000).

In case of a real-valued EVI, the selection of an appropriate threshold is even more difficult and only a few methods are available. Beirlant et al. (2009) proposed a penultimate limit distributions in (5.1.2) and (5.1.4). In case  $\gamma > 0$ , under the mathematical

theory of second-order slow variation, i.e. assuming that

$$\frac{\ell(yt)}{\ell(t)} - 1 = \delta_t (y^{-\beta} - 1), \quad (5.1.5)$$

where  $\delta_t = \delta(t) = t^{-\beta} \tilde{\ell}(t)$ , with  $\beta > 0$  and  $\tilde{\ell}$  slowly varying at infinity (see section 2.3 in de Haan and Ferreira, 2006), the left hand side of (5.1.4) equals

$$\frac{\bar{F}(yt)}{\bar{F}(t)} = y^{-1/\gamma} \frac{\ell(yt)}{\ell(t)} = y^{-1/\gamma} (1 + \delta_t (y^{-\beta} - 1)), \quad y > 1.$$

This then leads to the extension of the Pareto distribution to approximate the distribution of  $X/t$  given  $X > t$  as  $t \rightarrow \infty$ :

$$\bar{H}_{\gamma, \delta}^{EP}(y) := y^{-1/\gamma} (1 + \delta_t ((y^{-1/\gamma})^{\beta\gamma} - 1)), \quad y > 1, \quad (5.1.6)$$

with  $\delta_t$  satisfying  $\delta_t \downarrow 0$  as  $t \rightarrow \infty$ . In cases where the second order model (5.1.5) holds, such a mixture model  $\bar{H}_{\gamma, \delta}^{EP}$  will improve the approximation of  $(\frac{X}{t} > u | X > t)$  for values of  $t$  which are smaller than the appropriate  $t$ -values when modelling the POTs using  $\bar{H}_{\gamma}^P$ . So the extension can work when modelling large and moderate extremes. As a byproduct however, at instances, it may even work for the full sample.

In Beirlant et al. (2009), using an external estimator of  $\rho = -\beta\gamma$ , the parameters  $(\gamma, \delta)$  are estimated fitting the EPD (slightly adapted, with survival function  $\left\{y(1 + \tilde{\delta}_t - \tilde{\delta}_t y^{-\beta})\right\}^{-1/\gamma}$  and  $\tilde{\delta}_t = \delta_t \gamma$ ) by maximum likelihood on excesses over a random threshold  $X_{n-k,n}$ ,  $k = 1, 2, \dots, n$ . The result of this procedure is two-fold:

- First, the estimates  $\hat{\gamma}_k^{EP}$  of  $\gamma$  are more stable as a function of  $k$  compared to the original ML estimator derived by Hill (1975)

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^k \log \frac{X_{n-j+1,n}}{X_{n-k,n}}$$

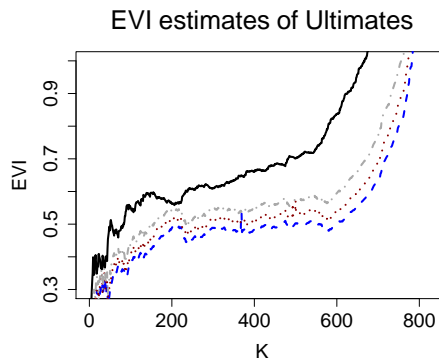
which is obtained by fitting the Pareto distribution  $\bar{H}_{\gamma}^P$  to the excesses  $\left\{\frac{X_{n-j+1,n}}{X_{n-k,n}}, j = 1, \dots, k\right\}$  following (5.1.4). Indeed, the bias in the simple POT model (5.1.4) is estimated when fitting  $\bar{H}_{\gamma, \delta}^{EP}$  and it is shown that, under the assumption that the EP model for the excesses  $X/t$  is correct and that  $\beta$  is estimated consistently,

the asymptotic bias of  $\hat{\gamma}_k^{EP}$  is 0 as long as  $k(k/n)^{2\beta\gamma} \rightarrow \lambda \geq 0$  as  $k, n \rightarrow \infty$ , while the asymptotic bias of  $H_{k,n}$  is only 0 when  $k(k/n)^{2\beta\gamma} \rightarrow 0$ .

- On the other hand, the asymptotic variance of  $\hat{\gamma}_k^{EP}$  equals  $\left(\frac{1-\rho}{\rho}\right)^2 \frac{\gamma^2}{k}$ , where  $\frac{\gamma^2}{k}$  is the asymptotic variance of  $H_{k,n}$ .

As an example Figure 5.1 shows both the Hill estimates  $H_{k,n}$  and the bias reduced estimates  $\hat{\gamma}_k^{EP}$ , obtained from maximum likelihood fitting of (5.1.6) using  $\rho = -\gamma\beta = -0.25, -0.5$  and  $-1$ , as a function of  $k$  for a data set of Belgian ultimate car insurance claims from 1995 and 2010 discussed in more detail in Albrecher et al. (2017). Note that the bias reduced estimates helps to interpret the original Hill “horror” plot.

Here from the bias reduced estimator a  $\gamma$  level around 0.5 becomes apparent for  $k \geq 200$  and a lower value between 0.3 and 0.4 for smaller values of  $k$ . In fact in insurance claim data mixtures in the ultimate tail do appear quite often. Moreover the EPD fit appears to extend quite well down to the lower threshold value, i.e. with  $k$  up to 600 (but not when using almost all data,  $k > 600$ ). In this sense, classical first order extreme value modelling can in some cases be extended using mixture modelling in order to capture the characteristics of the bulk of the data.



**Figure 5.1:** Ultimates of Belgian car insurance claims: bias reduction of Hill estimator (full line) using  $\bar{H}_{\gamma,\delta}^{EP}$  with  $\rho = -0.25$  (dashed line),  $\rho = -0.5$  (dotted line) and  $\rho = -1$  (dash-dotted line).

In this section we concentrate on bias reduction when using the GPD approximation to the distribution of POTs  $X - t|X > t$ , on which the literature is quite limited. This allows to extend bias reduction to the general case  $\gamma > -1/2$ . We apply the flexible semiparametric GP modelling introduced in Tencaliec et al. (2018) to the POT distributions. We also extend the second-order refined POT approach using  $\bar{H}_{\gamma,\delta}^{EP}$  from (5.1.6) to all max-domains of attraction. Here the corresponding basic second order regular variation theory can be found in Theorem 2.3.8 in de Haan and Ferreira (2006) stating that

$$\lim_{t \rightarrow x_+} \frac{\mathbb{P}(X - t > y\sigma_t|X > t) - (1 + \gamma y)^{-1/\gamma}}{\delta(t)} = (1 + \gamma y)^{-1-1/\gamma} \Psi_{\gamma,\tilde{\rho}}((1 + \gamma y)^{1/\gamma}), \quad (5.1.7)$$

with  $\delta(t) \rightarrow 0$  as  $t \rightarrow x_+$  and  $\Psi_{\gamma,\tilde{\rho}}(x) = \frac{1}{\tilde{\rho}} \left( \frac{x^{\gamma+\tilde{\rho}}-1}{\gamma+\tilde{\rho}} - \frac{x^\gamma-1}{\gamma} \right)$  which for the cases  $\gamma = 0$  and  $\tilde{\rho} = 0$  is understood to be equal to the limit as  $\gamma \rightarrow 0$  and  $\tilde{\rho} \rightarrow 0$ . We further allow more flexible second-order models than the ones arising from second-order regular variation theory such as in (5.1.7) using non-parametric modelling of the second-order component. These new methods are also applied to the specific case of Pareto-type distributions.

In the next section we propose our transformed and extended GPD models, and detail the estimation methods. Some basic asymptotic results are provided in Section 5.3. In the final section we discuss simulation results of the proposed methods and some practical case studies. We then also discuss the evaluation of the overall goodness-of-fit behaviour of the fitted models.

## 5.2 Transformed and extended GPD models

Recently, Naveau et al. (2016), generalising Papastathopoulos and Tawn (2013), proposed to use full models for rainfall intensity data that are able to capture low, moderate and heavy rainfall intensities without a threshold selection procedure. These authors,



considering only applications with a positive EVI however, propose to model all data jointly using transformation models with survival function

$$\bar{F}(x) = 1 - \bar{G}_0 \left( H_\gamma^{GP} \left( \frac{x}{\sigma} \right) \right) =: G_0 \left( \bar{H}_\gamma^{GP} \left( \frac{x}{\sigma} \right) \right), \quad (5.2.8)$$

with  $\bar{G}_0$  and  $G_0$  distribution functions on  $[0, 1]$  linked by  $G_0(u) = 1 - \bar{G}_0(1 - u)$  ( $0 < u < 1$ ), and satisfying constraints to preserve the classical tail GPD fit and a power behaviour for small rainfall intensities:

- $\lim_{u \downarrow 0} \frac{G_0(u)}{u} = a$ , for some  $a > 0$ ,
- $\lim_{u \downarrow 0} \frac{\bar{G}_0(u)}{u^\kappa} = c$ , for some  $c > 0$  and  $\kappa > 0$ .

In Naveau et al. (2016) the authors propose parametric examples for  $G_0$ , such as

$$G_0(u) = \frac{1+D}{D} u \left( 1 - \frac{u^D}{1+D} \right), u \in (0, 1) \text{ with } D > 0.$$

In Tencaliec et al. (2018) a non-parametric approach is taken using Bernstein polynomials of degree  $m$  to approximate  $G_0$ , i.e. using  $G_0^{(m)}(u) = \sum_{j=0}^m G_0(\frac{j}{m}) b_{j,m}(u)$  with beta densities

$$b_{j,m}(u) = \binom{m}{j} u^j (1-u)^{m-j}, u \in (0, 1).$$

In Naveau et al. (2016) and Tencaliec et al. (2018) the primary goal is the search for a model fitting the whole outcome set, while the fit of the proposed model to POT values  $X - t | X > t$  for extrapolation purposes in order to estimate extreme quantiles and tail probabilities is imposed using the condition  $\lim_{u \downarrow 0} \frac{G_0(u)}{u} = a$ . However the bias and MSE properties of the estimators of  $\gamma$  and  $\sigma$  are still to be analyzed.

To encompass the above mentioned methods from Beirlant et al. (2009), Naveau et al. (2016) and Tencaliec et al. (2018) we propose to approximate  $\mathbb{P}(X - t > y | X > t)$  with a transformation model with right tail function of the type

$$(\mathcal{T}) : \quad \bar{F}_t^T(y) = G_t \left( \bar{H}_\gamma^{GP} \left( \frac{y}{\sigma} \right) \right)$$

where  $G_t(u)/u \rightarrow 1$  for all  $u \in (0, 1)$  as  $t \rightarrow x_+$ . Note here that for  $u \in (0, 1)$  and  $Y =_d X - t | X > t$ ,

$$G_t(u) = \mathbb{P} \left( \bar{H}_\gamma^{GP} \left( \frac{Y}{\sigma} \right) \leq u \right). \quad (5.2.9)$$

We also consider a sub-model of  $(\mathcal{T})$ , approximating the POT distribution with an extended GPD model

$$(\mathcal{E}) : \quad \bar{F}_t^E(y) = \bar{H}_\gamma^{GP} \left( \frac{y}{\sigma} \right) \left\{ 1 + \delta_t B_\eta \left( \bar{H}_\gamma^{GP} \left( \frac{y}{\sigma} \right) \right) \right\},$$

where

- $\delta_t = \delta(t) \rightarrow 0$  as  $t \rightarrow x_+$ ,
- $B_\eta(1) = 0$  and  $\lim_{u \rightarrow 0} u^{1-\epsilon} B_\eta(u) = 0$  for every  $1 > \epsilon > 0$ ,
- $B_\eta$  is twice continuously differentiable.

Here the parameter  $\eta$  represents a second order (nuisance) parameter. For negative  $\delta$ -values one needs  $\delta_t > \{\min_u (1 - \frac{d}{du} (u B_\eta(u)))\}^{-1}$  to obtain a valid distribution. At  $t = 0$  the function  $u \mapsto u(1 + \delta_0 B_\eta(u))$  then corresponds to  $u \mapsto G_0(u)$  in (5.2.8), while  $\frac{G_t(u)}{u} \rightarrow 1$  as  $t \rightarrow \infty$  leads to the GPD survival function  $\bar{H}_\gamma^{GP}(x/\sigma)$  at large thresholds. Note that model  $(\mathcal{E})$  is a direct generalisation of the EPD model (5.1.6) replacing the Pareto distribution  $y^{-1/\gamma}$  by the GPD  $\bar{H}_\gamma^{GP}$  and considering a general function  $B_\eta(u)$  rather than  $u^{\beta\gamma} - 1 = u^{-\rho} - 1$ .

Now several possibilities for bias reduction appear:

- (1) **Estimation under the transformed model  $(\mathcal{T})$ .** Modelling the distribution of  $Y = X - t | X > t$  with model  $(\mathcal{T})$  and estimating  $G_t$  and  $(\gamma, \sigma)$  for every  $t$ , we propose to use the algorithm from Tencaliec et al. (2018) for every  $t$  or  $k = 1, \dots, n$ . This approach is further denoted with  $(T\bar{p})$ .

Here we apply Bernstein approximation and estimation of  $G_t$  which is the distribution function of  $\bar{H}_\gamma^{GP}(Y/\sigma)$ . The Bernstein approximation of order  $m$  of a

continuous distribution function  $G$  on  $[0, 1]$  is given by

$$G^{(m)}(u) = \sum_{j=0}^m G\left(\frac{j}{m}\right) \binom{m}{j} u^j (1-u)^{m-j}, \quad u \in [0, 1].$$

As in Babu et al. (2002) one then replaces the unknown distribution function  $G$  itself with the empirical distribution function  $\hat{G}_n$  of the available data in order to obtain a smooth estimator of  $G$ :

$$\hat{G}_n^{(m)}(u) = \sum_{j=0}^m \hat{G}_n\left(\frac{j}{m}\right) \binom{m}{j} u^j (1-u)^{m-j}.$$

In the present application, data from  $G_t$  are only available after imputing a value for  $(\gamma, \sigma)$ . This then leads to the iterative algorithm from Tencaliec et al. (2018), which is applied to every threshold  $t$ , or every number of top  $k$  data. We here detail the algorithm for excesses  $Y_{j,k} = X_{n-j+1,n} - X_{n-k,n}$  ( $j = 1, \dots, k$ ), using the reparametrization  $(\gamma, \tau)$  with  $\tau = \gamma/\sigma$ :

*Algorithm ( $A_{\mathcal{T}}$ )*

- (i) Set starting values  $(\hat{\gamma}_k^{(0)}, \hat{\tau}_k^{(0)})$ . Here one can use  $(\hat{\gamma}_k^{ML}, \hat{\tau}_k^{ML})$  from using  $G_t(u) = u$ .
- (ii) Iterate for  $r = 0, 1, \dots$  until the difference in loglikelihood taken in  $(\hat{\gamma}_k^{(r)}, \hat{\tau}_k^{(r)})$  and  $(\hat{\gamma}_k^{(r+1)}, \hat{\tau}_k^{(r+1)})$  is smaller than a prescribed value
  - i. Given  $(\hat{\gamma}_k^{(r)}, \hat{\tau}_k^{(r)})$  construct rv's  $\hat{Z}_{j,k} = \left(1 + \hat{\tau}_k^{(r)} Y_{j,k}\right)^{-1/\hat{\gamma}_k^{(r)}}$
  - ii. Construct Bernstein approximation based on  $\hat{Z}_{j,k}$  ( $1 \leq j \leq k$ )

$$\hat{G}_k^{(m)}(u) = \sum_{j=0}^m \hat{G}_k\left(\frac{j}{m}\right) \binom{m}{j} u^j (1-u)^{m-j}$$

with  $\hat{G}_k$  the empirical distribution function of  $\hat{Z}_{j,k}$

iii. Obtain new estimates  $(\hat{\gamma}_k^{(r+1)}, \hat{\tau}_k^{(r+1)})$  with ML:

$$(\hat{\gamma}_k^{(r+1)}, \hat{\tau}_k^{(r+1)}) = \operatorname{argmax} \left\{ \sum_{j=1}^k \log \{ \hat{g}_k^{(m)} ((1 + \tau \hat{Z}_{j,k})^{-1/\gamma}) \} \right. \\ \left. + \sum_{j=1}^k \log \left\{ \frac{\tau}{\gamma} (1 + \tau \hat{Z}_{j,k})^{-1-1/\gamma} \right\} \right\}$$

with  $\hat{g}_k^{(m)}$  denoting the derivative of  $\hat{G}_k^{(m)}$ .

The final estimates of  $(\gamma, \tau)$  and  $G_t$  are denoted here by  $(\hat{\gamma}_k^T, \hat{\tau}_k^T)$  and  $\hat{G}_k^T$ . As noted in Tencaliec et al. (2018) the theoretical study of these estimates is difficult. In the simulation study the finite sample characteristics of these estimators  $\hat{\gamma}_k^T$  are given using  $m = k^a$  with a fixed value of  $a$  using  $\hat{a} = \operatorname{argmin} \sum_{k=2}^n (\hat{\gamma}_k^T - \bar{\gamma}_n^T)^2$  in order to stabilise the plots of the estimates of  $\gamma$  as much as possible. Note that this estimation procedure is computationally demanding.

Estimates of small tail probabilities  $\mathbb{P}(X > c)$  can be obtained through

$$\hat{\mathbb{P}}_k^T(X > c) = \frac{k}{n} \hat{G}_k^T \left( \bar{H}_{\hat{\gamma}_k^T} \left( \frac{\hat{\tau}_k^T}{\hat{\gamma}_k^T} (c - X_{n-k,n}) \right) \right).$$

Finally, bias reduced estimators of extreme  $1-p$  quantiles for small  $p$  are obtained by setting the above formulas equal to  $p$  and solving for  $c$ .

(2) **Estimation under the extended model ( $\mathcal{E}$ ).** Modelling the distribution of the exceedances  $Y$  with model ( $\mathcal{E}$ ) leads to maximum likelihood estimators based on the excesses  $Y_{j,k} = X_{n-j+1,n} - X_{n-k,n}$  ( $j = 1, \dots, k$ ):

$$(\hat{\gamma}_k^E, \hat{\tau}_k^E, \hat{\delta}_k) = \operatorname{argmax} \left\{ \sum_{j=1}^k \log (1 + \delta_k b_\eta ((1 + \tau Y_{j,k})^{-1/\gamma})) \right. \\ \left. + \sum_{j=1}^k \log \left\{ \frac{\tau}{\gamma} (1 + \tau Y_{j,k})^{-1-1/\gamma} \right\} \right\} \quad (5.2.10)$$

with  $b_\eta(u) = \frac{d}{du}(uB_\eta(u))$  for a given choice of  $B_\eta$ .

Estimates of small tail probabilities  $\mathbb{P}(X > c)$  are then obtained through

$$\hat{\mathbb{P}}_k^E(X > c) = \frac{k}{n} \bar{H}_{\hat{\gamma}_k^E}^{GP} \left( \frac{\hat{\tau}_k^E}{\hat{\gamma}_k^E} (c - X_{n-k,n}) \right) \left( 1 + \hat{\delta}_k^E \hat{B}_\eta \left( \bar{H}_{\hat{\gamma}_k^E}^{GP} \left( \frac{\hat{\tau}_k^E}{\hat{\gamma}_k^E} (c - X_{n-k,n}) \right) \right) \right).$$

As in Naveau et al. (2016), respectively Tencaliec et al. (2018), two approaches can be taken towards the bias function  $B_\eta$ : a parametric approach, respectively a non-parametric approach.

- (a) *In the parametric approach*, denoted  $(Ep)$ , the second-order result (5.1.7) leads to the parametric choice  $B_{\gamma, \tilde{\rho}}(u) = \frac{u^\gamma}{\tilde{\rho}} \left( \frac{u^{-\gamma-\tilde{\rho}}-1}{\gamma+\tilde{\rho}} - \frac{u^{-\gamma}-1}{\gamma} \right)$  in case  $\gamma + \tilde{\rho} \neq 0$  and  $\gamma \neq 0$ .

Model  $(\mathcal{E})$  allows for bias reduced estimation of  $(\gamma, \tau)$  under the assumption that the corresponding second-order model (5.1.7) is correct for the POTs  $X - t | X > t$ . Note that  $(Ep)$  generalises the approach taken in Beirlant et al. (2009) to all max-domains of attraction. When model  $(\mathcal{E})$  is used as a model for all observations, i.e. taking  $t = 0$ , this model directly encompasses the models from Frigessi et al. (2002) and Naveau et al. (2016).

Here

$$b_\eta(u) = u^{-\tilde{\rho}} \left( \frac{1 - \tilde{\rho}}{\tilde{\rho}(\gamma + \tilde{\rho})} \right) + u^\gamma \left( \frac{1 - \gamma}{\gamma(\gamma + \tilde{\rho})} \right) - \frac{1}{\gamma\tilde{\rho}},$$

in which the classical estimator of  $\gamma$  (with  $\delta_k = 0$ ), or an appropriate value  $\gamma_0$ , is used to substitute  $\gamma$ , next to an appropriate value of  $\tilde{\rho}$ . One can also choose a value of  $(\gamma_0, \tilde{\rho})$  minimising the variance in the plot of the resulting estimates of  $\gamma$  as a function of  $k$ .

- (b) Alternatively, a *non-parametric approach* (denoted  $E\bar{p}$ ) can be performed using the Bernstein polynomial algorithm from Tencaliec et al. (2018). In fact in practice a particular distribution probably follows laws of nature, environment or business and does not have to follow the second-order regular variation assumptions as in (5.1.7). Moreover in the case of a real-valued EVI, the function  $B_\eta$  can take different mathematical forms depending on  $(\gamma, \tilde{\rho})$  and  $\gamma + \tilde{\rho}$  being close to 0 or not.

Here a Bernstein type approximation is obtained for  $u \mapsto uB_\eta(u)$  from

$\hat{G}_{k_*}^{(m)}(u) - u$  obtained through algorithm  $(A_{\mathcal{T}})$ , and reparametrizing  $\delta_k$  by  $\delta_k/\delta_{k_*}$  with  $k_*$  an appropriate value of the number of top data used. The function  $b_{\eta}(u)$  is then substituted by  $-1 + \frac{d}{du}\hat{G}_{k_*}^{(m)}(u)$ .

The methods described above of course can be rewritten for the specific case of Pareto-type distributions where the distribution of POTs  $Y = \frac{X}{t}|X > t$  are approximated by transformed Pareto distributions. The models are then rephrased as

$$(\mathcal{T}^+) : \quad \bar{F}_t^E(y) = G_t(\bar{H}_{\gamma}^P(y)),$$

where for  $u \in (0, 1)$

$$G_t(u) = \mathbb{P}(\bar{H}_{\gamma}^P(Y) \leq u),$$

and

$$(\mathcal{E}^+) : \quad \bar{F}_t^E(y) = \bar{H}_{\gamma}^P(y) \{1 + \delta_t B_{\eta}(\bar{H}_{\gamma}^P(y))\}.$$

The above algorithms, now based on the exceedances  $Y_{j,k} = X_{n-j+1,n}/X_{n-k,n}$  ( $j = 1, \dots, k$ ), are then adapted as follows:

- In algorithm  $(A_{\mathcal{T}})$  step (ii.c) is replaced by

$$\hat{\gamma}_k^{(r+1)} = \operatorname{argmax} \left\{ \sum_{j=1}^k \log\{\hat{g}_k^{(m)}((\hat{Z}_{j,k})^{-1/\gamma})\} + \sum_{j=1}^k \log\left\{\frac{1}{\gamma}(\hat{Z}_{j,k})^{-1-1/\gamma}\right\} \right\},$$

with  $\hat{Z}_{j,k} = Y_{j,k}^{-1/\hat{\gamma}_k^{(r)}}$ . The resulting estimates are denoted with  $\hat{\gamma}_k^{T+}$  and  $\hat{G}_k^{T+}$ .

- In approach  $(\mathcal{E})$  the likelihood solutions are given by

$$(\hat{\gamma}_k^{E+}, \hat{\delta}_k^{E+}) = \operatorname{argmax} \left\{ \sum_{j=1}^k \log\left(1 + \delta_k b_{\eta}(Y_{j,k}^{-1/\gamma})\right) + \sum_{j=1}^k \log\left\{\frac{1}{\gamma}(Y_{j,k})^{-1-1/\gamma}\right\} \right\}. \quad (5.2.11)$$

Note that the  $(Ep^+)$  approach using the parametric version  $B_{\eta}(u) = u^{-\rho} - 1$  for a particular fixed  $\rho < 0$  equals the EPD method from Beirlant et al. (2009), while  $(E\bar{p}^+)$  is new.

Estimators of tail probabilities are then given by

$$\hat{\mathbb{P}}_k^{T+}(X > c) = \frac{k}{n} \hat{G}_k^{T+} \left( \bar{H}_{\hat{\theta}_k^{T+}}(c/X_{n-k,n}) \right),$$

respectively

$$\hat{\mathbb{P}}_k^{E+}(X > c) = \frac{k}{n} \bar{H}_{\hat{\gamma}_k^{E+}}^P(c/X_{n-k,n}) \left( 1 + \hat{\delta}_k^{E+} \hat{B}_\eta \left( \bar{H}_{\hat{\gamma}_k^{E+}}^P(c/X_{n-k,n}) \right) \right).$$

### 5.3 Basic asymptotics under model $(\mathcal{E})$

We discuss here in detail the asymptotic properties of the maximum likelihood estimators solving (5.2.10) and (5.2.11). To this end, as in Beirlant et al. (2009) we develop the likelihood equations up to linear terms in  $\delta_k$  since  $\delta_k \rightarrow 0$  with decreasing value of  $k$ . Below we set  $\bar{H}_\theta(y) = (1 + \tau y)^{-1/\gamma}$  when using extended GPD modelling, while  $\bar{H}_\theta(y) = y^{-1/\gamma}$  when using EP modelling under  $\gamma > 0$ .

*Extended Pareto POT modelling.* The likelihood problem (5.2.11) was already considered in Beirlant et al. (2009) in case of parametric modelling for  $B_\eta$ . We here propose a more general treatment. The limit statements in the derivation can be obtained using the methods from Beirlant et al. (2009). The likelihood equations following from (5.2.11) are given by

$$\begin{cases} \frac{\partial}{\partial \gamma} \ell &= -\frac{k}{\gamma} + \frac{1}{\gamma^2} \sum_{j=1}^k \log Y_{j,k} + \frac{\delta_k}{\gamma^2} \sum_{j=1}^k \frac{b'_\eta(\bar{H}_\theta(Y_{j,k})) \bar{H}_\theta(Y_{j,k}) \log Y_{j,k}}{1 + \delta_k b_\eta(\bar{H}_\theta(Y_{j,k}))} \\ \frac{\partial}{\partial \delta_k} \ell &= \sum_{j=1}^k b_\eta(\bar{H}_\theta(Y_{j,k})) - \delta_k \sum_{j=1}^k b_\eta^2(\bar{H}_\theta(Y_{j,k})). \end{cases} \quad (5.3.12)$$

*Extended Generalised Pareto POT modelling.* The likelihood equations following from (5.2.10) up to linear terms in  $\delta_k$  are now given by

$$\begin{cases} \frac{\partial}{\partial \gamma} \ell &= -\frac{k}{\gamma} + \frac{1}{\gamma^2} \sum_{j=1}^k \log(1 + \tau Y_{j,k}) + \frac{\delta_k}{\gamma^2} \sum_{j=1}^k b'_\eta(\bar{H}_\theta(Y_{j,k})) \bar{H}_\theta(Y_{j,k}) \log(1 + \tau Y_{j,k}) \\ \frac{\partial}{\partial \tau} \ell &= \frac{k}{\gamma \tau} \left\{ -1 + (1 + \gamma) \frac{1}{k} \sum_{j=1}^k \frac{1}{1 + \tau Y_{j,k}} \right. \\ &\quad \left. - \frac{\delta_k}{k} \sum_{j=1}^k b'_\eta(\bar{H}_\theta(Y_{j,k})) (\tau Y_{j,k}) (1 + \tau Y_{j,k})^{-1-1/\gamma} \right\} \\ \frac{\partial}{\partial \delta_k} \ell &= \sum_{j=1}^k b_\eta(\bar{H}_\theta(Y_{j,k})) - \delta_k \sum_{j=1}^k b_\eta^2(\bar{H}_\theta(Y_{j,k})), \end{cases}$$

from which

$$\left\{ \begin{array}{l} \hat{\delta}_k = \frac{\sum_{j=1}^k b_\eta(\bar{H}_{\hat{\theta}_k}(Y_{j,k}))}{\sum_{j=1}^k b_\eta^2(\bar{H}_{\hat{\theta}_k}(Y_{j,k}))}, \\ \frac{1}{k} \sum_{j=1}^k \log(1 + \hat{\tau}_k Y_{j,k}) = \hat{\gamma}_k - \frac{\hat{\delta}_k}{k} \sum_{j=1}^k b'_\eta(\bar{H}_{\hat{\theta}_k}(Y_{j,k})) \bar{H}_{\hat{\theta}_k}(Y_{j,k}) \log(1 + \hat{\tau}_k Y_{j,k}), \\ \frac{1}{k} \sum_{j=1}^k \frac{1}{1 + \hat{\tau}_k Y_{j,k}} = \frac{1}{1 + \hat{\gamma}_k} + \frac{\hat{\delta}_k}{1 + \hat{\gamma}_k} \left\{ \frac{1}{k} \sum_{j=1}^k b'_\eta(\bar{H}_{\hat{\theta}_k}(Y_{j,k})) \bar{H}_{\hat{\theta}_k}(Y_{j,k}) \right. \\ \left. - \frac{1}{k} \sum_{j=1}^k b'_\eta(\bar{H}_{\hat{\theta}_k}(Y_{j,k})) \bar{H}_{\hat{\theta}_k}(Y_{j,k}) \frac{1}{1 + \hat{\tau}_k Y_{j,k}} \right\}. \end{array} \right. \quad (5.3.13)$$

Under the extended model we now state the asymptotic distribution of the estimators  $\hat{\gamma}_k^{E+}$  and  $\hat{\gamma}_k^E$ . To this end let  $Q$  denote the quantile function of  $F$ , and let  $U(x) = Q(1 - x^{-1})$  denote the corresponding tail quantile function. Model assumption  $(\mathcal{E})$  can be rephrased in terms of  $U$ :

$$(\tilde{\mathcal{E}}) : \frac{\frac{U(vx) - U(v)}{\sigma_{U(v)}} - h_\gamma(x)}{\delta(U(v))} \rightarrow_{v \rightarrow \infty} x^\gamma B_\eta(1/x),$$

where  $h_\gamma(x) = (x^\gamma - 1)/\gamma$  and  $\delta(U)$  regularly varying with index  $\tilde{\rho} < 0$ . Moreover in the mathematical derivations one needs the extra condition that for every  $\epsilon, \nu > 0$ , and  $v, vx$  sufficiently large

$$(\tilde{\mathcal{E}}_2) : \left| \frac{\frac{U(vx) - U(v)}{\sigma_{U(v)}} - h_\gamma(x)}{\delta(U(v))} - x^\gamma B_\eta(1/x) \right| \leq \epsilon x^\gamma |B_\eta(1/x)| \max\{x^\nu, x^{-\nu}\}.$$

Similarly,  $(\mathcal{E}^+)$  is rewritten as

$$(\tilde{\mathcal{E}}^+) : \frac{\frac{U(vx)}{U(v)} - x^\gamma}{\gamma \delta(U(v))} \rightarrow_{v \rightarrow \infty} x^\gamma B_\eta(1/x).$$

The analogue of  $(\tilde{\mathcal{E}}_2)$  in this specific case is given by

$$(\tilde{\mathcal{E}}_2^+) : \left| \frac{\frac{U(vx)}{U(v)} - x^\gamma}{\gamma \delta(U(v))} - x^\gamma B_\eta(1/x) \right| \leq \epsilon x^\gamma |B_\eta(1/x)| \max\{x^\nu, x^{-\nu}\},$$



with  $\delta(U)$  regularly varying with index  $\rho < 0$ .

Finally, in the expression of the asymptotic variances we use

$$Eb_\eta^2 = \int_0^1 b_\eta^2(u)du, \quad EB_\eta = \int_0^1 B_\eta(u)du, \quad EC_\eta = \int_0^1 u^\gamma B_\eta(u)du.$$

The proof of the next theorem is outlined in Appendix 5.A. It allows to construct confidence intervals for the estimators of  $\gamma$  obtained under the extended models.

**Theorem 5.1.** *Let  $k = k_n$  be a sequence such that  $k, n \rightarrow \infty$  and  $k/n \rightarrow 0$  such that  $\sqrt{k}\delta(U(n/k)) \rightarrow \lambda \in \mathbb{R}$ . Moreover assume that in (5.2.10) and (5.2.11),  $B_\eta$  is substituted by a consistent estimator as  $n \rightarrow \infty$ . Then*

i. *when  $\gamma > 0$  with  $(\tilde{\mathcal{E}}_2^+)$*

$$\sqrt{k}(\hat{\gamma}_k^{E+} - \gamma) \rightarrow^d \mathcal{N}\left(0, \gamma^2 \frac{Eb_\eta^2}{Eb_\eta^2 - (EB_\eta)^2}\right),$$

ii. *when  $\gamma > -1/2$  with  $(\tilde{\mathcal{E}}_2)$*

$$\begin{aligned} & \left( \sqrt{k}(\hat{\gamma}_k^E - \gamma), \sqrt{k}\left(\frac{\hat{\tau}_k^E}{\tau} - 1\right) \right) \rightarrow^d \mathcal{N}_2(\mathbf{0}, \Sigma) \\ & \Sigma = \frac{\gamma^2}{D} \begin{pmatrix} \frac{1}{(1+\gamma)^2(1+2\gamma)} - \frac{(EC_\eta)^2}{Eb_\eta^2} & \frac{1}{\gamma(1+\gamma)^3} - \frac{EB_\eta EC_\eta}{\gamma(1+\gamma)Eb_\eta^2} \\ \frac{1}{\gamma(1+\gamma)^3} - \frac{EB_\eta EC_\eta}{\gamma(1+\gamma)Eb_\eta^2} & \frac{1}{\gamma^2(1+\gamma)^2} \left(1 - \frac{(EB_\eta)^2}{Eb_\eta^2}\right) \end{pmatrix} \end{aligned}$$

where

$$D = \left( \frac{1}{(1+\gamma)^2(1+2\gamma)} - \frac{(EC_\eta)^2}{Eb_\eta^2} \right) \left( 1 - \frac{(EB_\eta)^2}{Eb_\eta^2} \right) - \left( \frac{1}{(1+\gamma)^2} - \frac{EB_\eta EC_\eta}{Eb_\eta^2} \right)^2,$$

▲

**Remark:** The asymptotic variance of  $\hat{\gamma}_k^{E+}$  is larger than the asymptotic variance  $\gamma^2$

of the Hill estimator  $H_{k,n}$ . Indeed,

$$\begin{aligned} (EB_\eta)^2 &= \left( \int_0^1 \log(1/u) b_\eta(u) du \right)^2 \\ &= \left( \int_0^1 (\log(1/u) - 1) b_\eta(u) du \right)^2 \\ &\leq \left( \int_0^1 (\log(1/u) - 1)^2 du \right) \left( \int_0^1 b_\eta^2(u) du \right) \\ &= (Eb_\eta^2), \end{aligned}$$

where the above inequality follows using the Cauchy-Schwarz inequality.

Similarly, one can show that

$$(EC_\eta)^2 = \gamma^{-2} \left( \int_0^1 \left( u^\gamma - \frac{1}{1+\gamma} \right) b_\eta du \right)^2 \leq \frac{1}{(1+2\gamma)(1+\gamma)^2} (Eb_\eta^2).$$

The asymptotic variance of  $\hat{\gamma}_k^E$  equals

$$\frac{(1+\gamma)^2}{k} \frac{1 - (1+\gamma)^2(1+2\gamma)(EC_\eta)^2/(Eb_\eta^2)}{1 - \frac{(1+\gamma)^4(1+2\gamma)}{\gamma^2} (Eb_\eta^2)^{-1} [(EC_\eta)^2 - 2\frac{(EC_\eta)(EB_\eta)}{(1+\gamma)^2} + \frac{(EB_\eta)^2}{(1+\gamma)^2(1+2\gamma)}]}$$

which can be shown to be larger than the asymptotic variance  $(1+\gamma)^2/k$  of the classical GPD maximum likelihood estimator. In the parametric case with  $B_\eta(u) = \frac{u^\gamma}{\tilde{\rho}} \left( \frac{u^{-\gamma-\tilde{\rho}}-1}{\gamma+\tilde{\rho}} - \frac{u^{-\gamma}-1}{\gamma} \right)$ , one obtains  $EB_\eta = (1+\gamma)^{-1}(1-\tilde{\rho})^{-1}$ ,  $EC_\eta = (1+\gamma)^{-1}(1+2\gamma)^{-1}(\gamma-\tilde{\rho}+1)^{-1}$  and  $Eb_\eta^2 = 2(1+2\gamma)^{-1}(1-2\tilde{\rho})^{-1}(\gamma-\tilde{\rho}+1)^{-1}$ . It then follows that the asymptotic variance of  $\hat{\gamma}_k^E$  equals  $\frac{(1+\gamma)^2}{k} \left( \frac{1-\tilde{\rho}}{\tilde{\rho}} \right)^2$ .

In case  $\gamma > 0$  with  $B_\eta(u) = u^{-\rho} - 1$ , the asymptotic variance of  $\hat{\gamma}_k^{E+}$  is given by  $\frac{\gamma^2}{k} \left( \frac{1-\rho}{\rho} \right)^2$  as already found in Beirlant et al. (2009).

Since in model  $(\mathcal{E})$  the  $B_\eta$  factor is multiplied by  $\delta_t$ , the asymptotic distribution of tail estimators based on  $(\mathcal{E})$  will not depend on the asymptotic distribution of the estimator of  $B_\eta$ . As in Beirlant et al. (2009) when using the EPD model in the Pareto-type setting, one can rely in the parametric approach on consistent estimators of the nuisance parameter  $\eta$  using a larger proportion  $k_*$  of the data. Alternatively, one can also consider different values of  $\eta$  in the parametric approach, and of  $(k_*, m)$  in the non-parametric setting, and search for values of this nuisance parameter which stabilises

the plots of the EVI estimates as a function of  $k$  using the minimum variance principle for the estimates as a function of  $k$ . Clearly one loses the asymptotic unbiasedness in Theorem 5.1 if  $B_\eta$  is not consistently estimated. However as becomes clear from the simulation results in many instances the EVI estimators are not very sensitive to such a misspecification, especially in the non-parametric approaches  $T\bar{p}$ ,  $T\bar{p}^+$ ,  $E\bar{p}$  and  $E\bar{p}^+$ , and the proposed estimators can still outperform the classical maximum likelihood estimators based on the first order approximations of the POT distributions.

## 5.4 Simulations and case studies

Simulation results and practical cases are proposed on

<https://phdshinygao.shinyapps.io/ExtendedModels/>

Under *Simulations* one finds simulation results with sample sizes  $n = 200$  for different distributions from each max-domain of attraction. The bias and MSE for the different estimators are plotted as a function of the number of exceedances  $k$ . Using the notation from the preceding sections one has a choice to apply the technique with  $\bar{H}_\theta$  equal to the GPD, respectively the simple Pareto distribution (only when  $\gamma > 0$ ).

In case of the transformed models ( $\mathcal{T}$ ) one finds a slider to adapt the degree  $m$  of the Bernstein polynoms along  $m = k^a$  for different values of  $a \in (0, 1)$ . One can also choose  $a$  adaptively per sample so as to minimise the variance of  $\hat{\gamma}_k$  over  $k = 2, \dots, n$  (in order to have stable plots over  $k$ ).

In case of the extended models ( $\mathcal{E}$ ) one finds sliders for the following parameters:

- in case of Pareto modelling:  $\rho$  in  $Ep^+$ , and  $(k_*, m)$  in  $E\bar{p}^+$  estimation;
- in case of GPD modelling:  $\tilde{\rho}$  in  $Ep$ , and  $(k_*, m)$  in  $E\bar{p}$  estimation.

Again one can indicate to choose these parameters so as to minimise the variance of  $\hat{\gamma}_k$  over  $k = 2, \dots, n$ . The value of  $\gamma$  in the parametric function  $B_{\gamma, \tilde{\rho}}$  in  $Ep$  is imputed with the classical GPD-ML estimator at the given value of  $k$ .

Also bias and RMSE plots of the corresponding tail probability estimates of  $p = \mathbb{P}(X > c)$  are given, where  $c$  is chosen so that these probabilities equal  $p = 0.005$  or  $p = 0.003$ . Here the bias, respectively RMSE, are expressed as the average, respectively the average of squared values, of  $\log(p/\hat{p})$ .

One can also change the vertical scale of the plots, smooth the figures by taking moving averages of a certain number of estimates. Finally one can download the figures as a ‘.pdf’ file type.

While on the above link, several other distributions are used and sliders are provided for the different parameters  $a$ ,  $m$ ,  $\rho$ ,  $\tilde{\rho}$ , and  $k_*$ , we collect here the resulting figures for estimation of  $\gamma$  and estimating 0.003 tail probabilities, when using the minimum variance principle for all parameters, in case of the following subset of models:

- *The Burr( $\tau, \lambda$ ) distribution* with  $\bar{F}(x) = (1 + x^\tau)^{-\lambda}$  for  $x > 0$  with  $\tau = 1$  and  $\lambda = 2$ , so that  $\gamma = \frac{1}{\tau\lambda} = \frac{1}{2}$  and  $\rho = -\frac{1}{\lambda} = -\frac{1}{2}$ .
- *The Fréchet(2) distribution* with  $\bar{F}(x) = 1 - \exp(-x^{-2})$  for  $x > 0$ , so that  $\gamma = \frac{1}{2}$  and  $\rho = \tilde{\rho} = -1$ .
- *The standard normal distribution* with  $\gamma = 0$  and  $\tilde{\rho} = 0$ .
- *The Exponential distribution* with  $\bar{F}(x) = e^{-\lambda x}$  for  $x > 0$ , so that  $\gamma = 0$  and  $\tilde{\rho} = 0$ .
- *The Reversed Burr distribution* with  $\bar{F}(x) = (1 + (1 - x)^{-\tau})^{-\lambda}$  for  $x < 1$  with  $\tau = 5$  and  $\lambda = 1$ , so that  $\gamma = -1/(\tau\lambda) = -\frac{1}{5}$  with  $\tilde{\rho} = -1/\lambda = -1$ .

- *The extreme value Weibull distribution* with  $F(x) = e^{-(1-x)^\alpha}$  for  $x < 1$  with  $\alpha = 4$ , so that  $\gamma = -\frac{1}{4}$  with  $\tilde{\rho} = -1$ .

In general the minimum variance principle works well, though in some cases some improved results can be obtained by choosing specific values of the parameters  $a$ ,  $\rho$ ,  $m$  and  $k_*$ . This is mainly the case for the Pareto-type models when using  $T\bar{p}$  and  $E\bar{p}$ , such as for the Fréchet distribution. Also, in case of tail probability estimation using  $Ep$  for cases with  $\gamma < 0$  particular choices of the corresponding parameters lead to improvements over the minimum variance principle.

When using *GPD* modelling of the exceedances, overall the  $Ep$  approach yields the best results, both in estimation of  $\gamma$  and tail probabilities. The improvement over the classical GPD maximum likelihood approach is smaller for  $E\bar{p}$ , and in case of situations where the second order parameter  $\tilde{\rho}$  equals 0 then  $E\bar{p}$  basically equals the ML estimators.

In case of simple Pareto modelling for  $\gamma > 0$  cases (see Figures 5.2 and 5.4) the  $Ep^+$  and  $E\bar{p}^+$  approaches yield serious improvements over the Hill estimator, with small bias for  $Ep^+$  and  $E\bar{p}^+$ , while parametric approach  $Ep^+$  naturally exhibits the best RMSE. Note that when  $\tilde{\rho} = 0$  the conditions of the main theorem are not met, in which case the GPD and the bias reductions are known to exhibit a large bias. This is typically the case when  $\gamma = 0$ . This is also known to be the case using simple Pareto modelling when  $\rho = 0$ .

Under *Applications* the app also offers the analysis of some case studies, some of which are discussed here in more detail. We use the ultimates of the Belgian ultimate car insurance claims used in Figure 5.1, in order to illustrate  $T\bar{p}^+$ ,  $E\bar{p}^+$  and  $Ep^+$ , and the winter rainfall data at Mont-Aigoual station (1976-2015) already used in Tencaliec et al. (2018) to illustrate  $T\bar{p}$ ,  $E\bar{p}$  and  $Ep$ . We then present estimates of  $\gamma$ ,  $\sigma$  and tail probabilities  $\mathbb{P}(X > x_{n,n})$  with  $x_{n,n}$  denoting the largest observation, so that the esti-

mated probability is supposed to be close to  $1/n$ . An option is provided to construct confidence intervals for  $\gamma$  on the basis of Theorem 5.1.

In case  $k = n$  the exceedances correspond to the reversely ordered data, i.e.  $Y_{j,n} = X_{n-j+1,n}$ . The goodness-of-fit for the *complete data set* can be analyzed choosing a specific value  $\theta_0 = (\gamma_0, \sigma_0)$  using the estimates of  $(\gamma, \sigma)$  which were obtained as a function of  $k$ , and by estimating the transformation  $G$  using one (ii.b) step from the transformation algorithm  $(A_7)$  with starting value  $(\gamma_0, \sigma_0)$  and with a chosen value  $m = n^a$  with  $a \in (0, 1)$  (slider). We then construct transformed P-P plots

$$\begin{aligned} & \left( -\log(1 - \hat{F}_n(X_{n-j+1,n})); -\log \hat{G}(\bar{H}_{\theta_0}(X_{n-j+1,n})) \right) \\ &= \left( \log \frac{n+1}{j}; -\log \hat{G}_n^{(m)}(\bar{H}_{\theta_0}(X_{n-j+1,n})) \right), \quad j = 1, \dots, n, \end{aligned} \quad (5.4.14)$$

where  $\hat{F}_n$  denotes the empirical distribution function based on  $X_{n-j+1,n}$  ( $j = 1, \dots, n$ ). The closer the plot lies to the diagonal, the better the fit of the model defined by the survival function  $\hat{G}^{(m)}(\bar{H}_{\theta_0})$ .

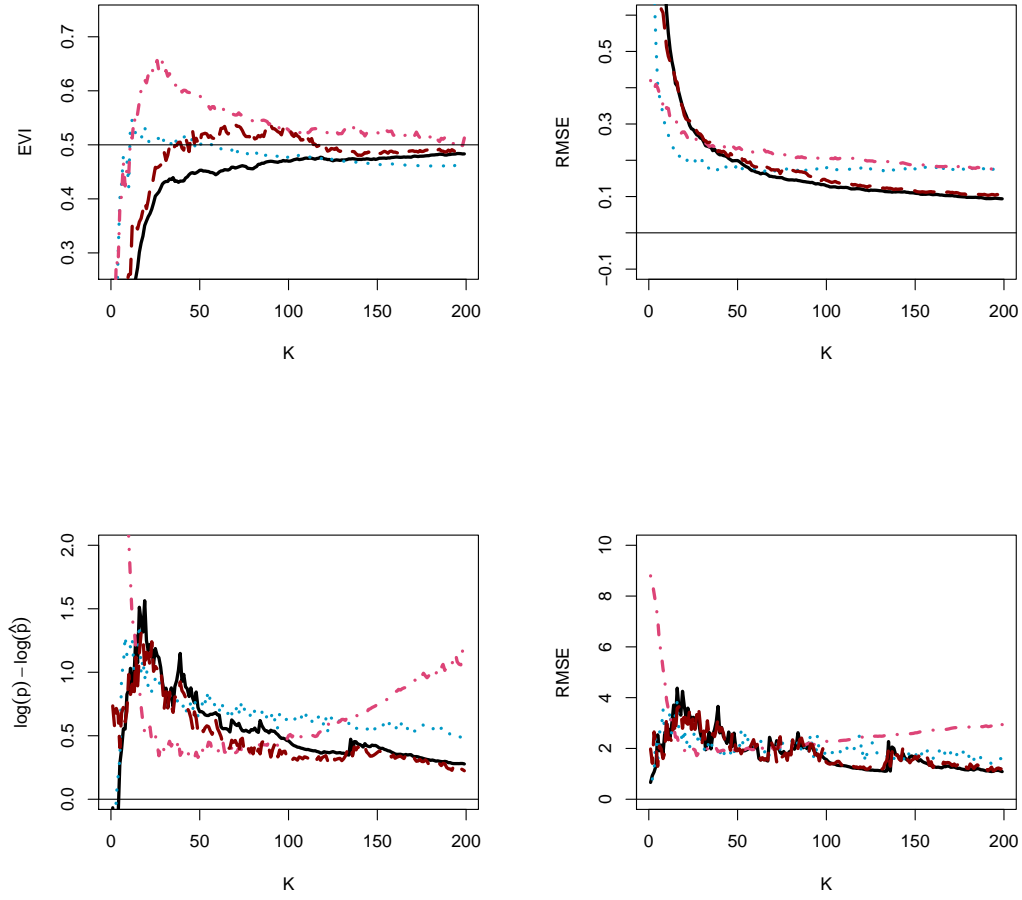
In Figure 5.9, the estimates of  $\gamma$  and  $\mathbb{P}(X > 4\,564\,759)$  using the minimum variance principle are given, next the goodness-of-fit plot as defined in (5.4.14) with  $\gamma_0 = 0.28$  and  $m = n^{0.99}$ . The estimates of  $\gamma$  obtained from  $Ep^+$  and  $E\bar{p}^+$  are clearly most stable as a function of  $k$  indicating the EVI value 0.4. The tail probability above the largest ultimate observation is also most stable for  $Ep^+$  and  $E\bar{p}^+$  indicating a value close to  $1/n$  (indicated by the horizontal line). While the goodness-of-fit plot shows some deviations from the 45 degree line for the fitted transformation model, the presented overall global fit is useful (correlation equals 0.97).

In Figure 5.10, for the winter rainfall data the results for  $E\bar{p}$  with  $m = 57$ ,  $k_* = 43$  indicate two levels for  $\gamma$  (0.4 and ultimately at small  $k$  close to 0),  $\sigma$  (10 and 40 for small  $k$ ) and tail probability  $\mathbb{P}(X > 162.05)$  (0.005 and 0.002 for small  $k$ , to be compared

with  $1/n = 0.0018$ ), which indicates a change of statistical tail behaviour near the top. Method  $Ep$  with  $\rho = -1$  yields almost the same result as the classical GPD-ML method, except for the tail probability where it is quite unstable. Finally the transformation approach  $T\bar{p}$  yields very stable plots at compromise values 0.2 for  $\gamma$ , 10 for  $\sigma$  and 0.003 for the tail probability. While the goodness-of-fit plot with  $m = n^{0.99}$ ,  $\sigma_0 = 10$  and  $\gamma_0 = 0.21$  has a correlation 0.997, the transformation approach seems to be unable to catch the deviating tail component near the top data with an EVI value close to 0.

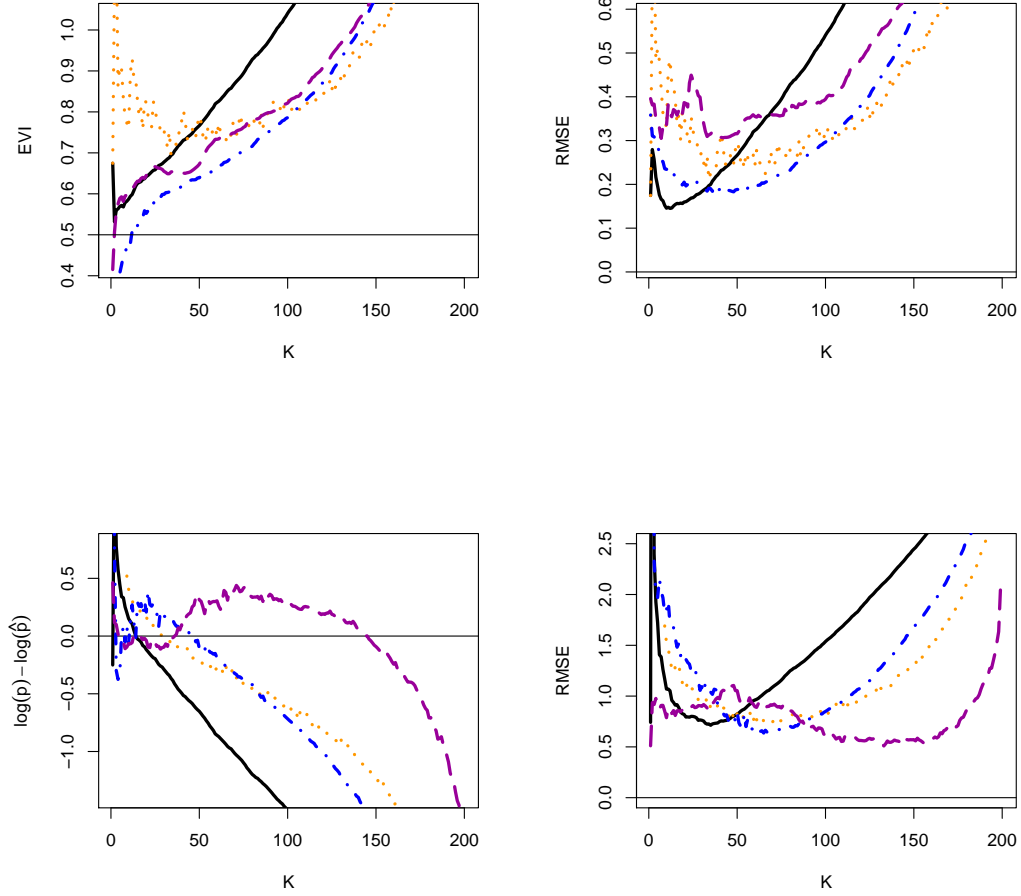
## 5.5 Chapter conclusion

In this contribution we have constructed bias reduced estimators of tail parameters extending the classical POT method using the GPD. The bias can be modelled parametrically (for instance based on second order regular variation theory), or non-parametrically using Bernstein polynomial approximations. A basic asymptotic limit theorem is provided for the estimators of the extreme value parameters which allows to compute asymptotic confidence intervals. A shinyapp has been constructed with which the characteristics and the effectiveness of the proposed methods are illustrated through simulations and practical case studies. From this it follows that within the proposed methods it is always possible to improve upon the classical POT method both in bias and RMSE.

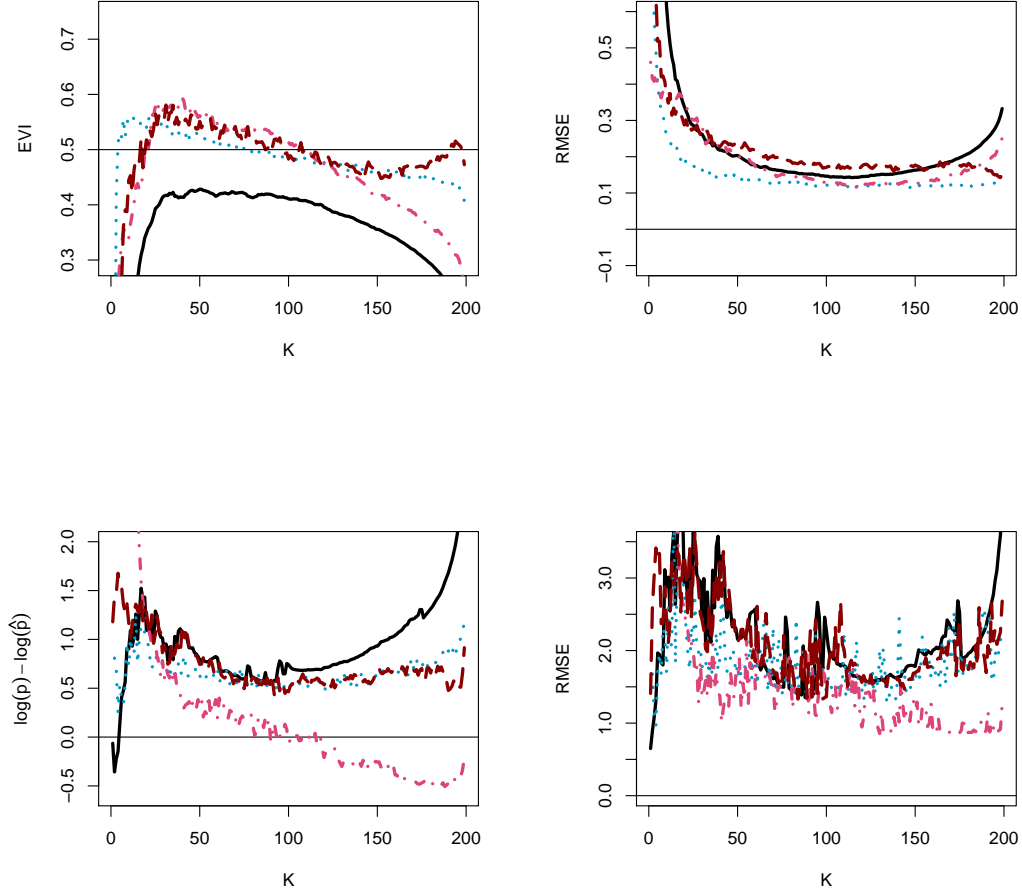


**Figure 5.1:** Burr distribution with  $\gamma = 0.5$  and  $\rho = -0.5$ . Estimation of  $\gamma$  (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line),  $T\bar{p}$  (dotted),  $Ep$  (dash-dotted) and  $E\bar{p}$  (dashed).

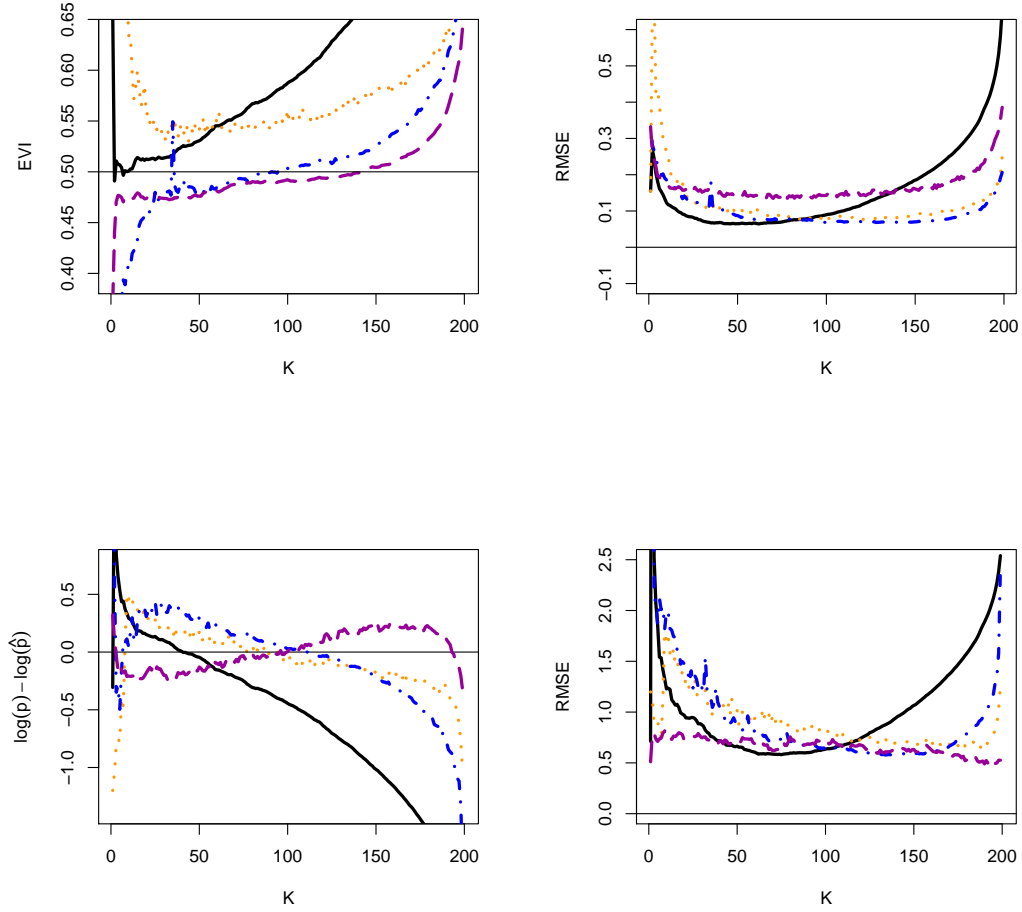




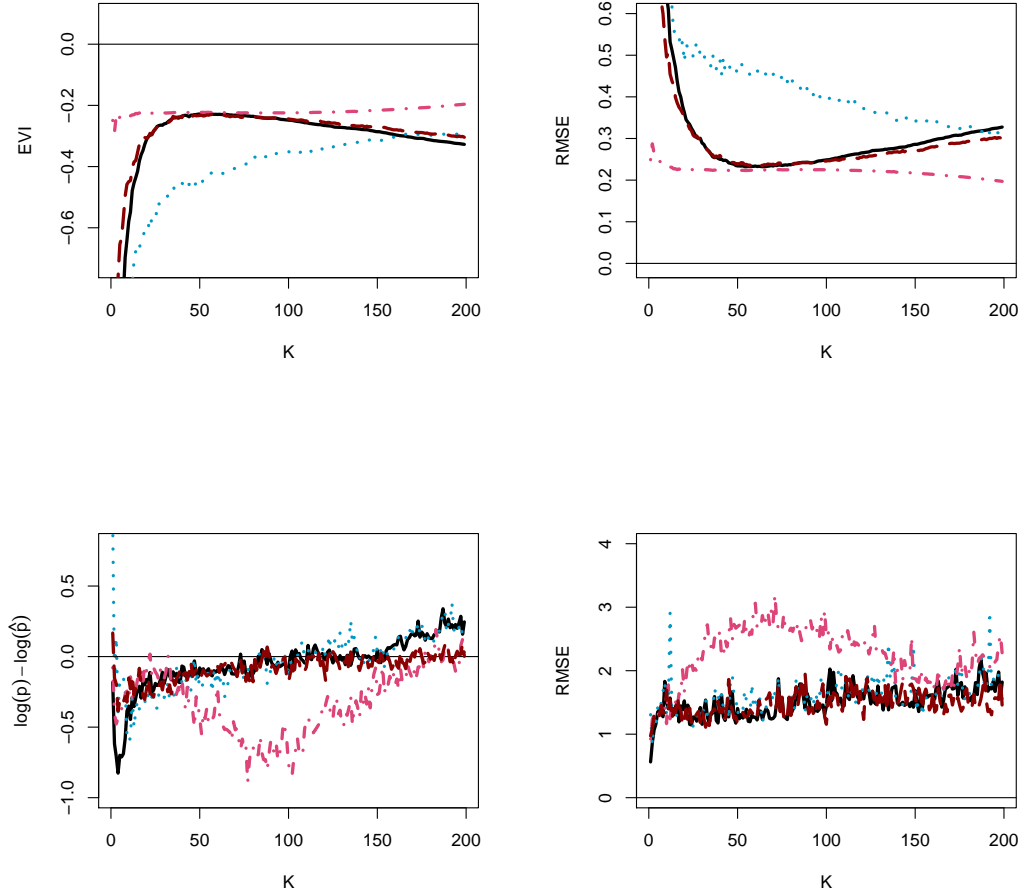
**Figure 5.2:** Burr distribution with  $\gamma = 0.5$  and  $\rho = -0.5$ . Estimation of  $\gamma$  (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): Pareto-ML (full line),  $T\bar{p}^+$  (dotted),  $Ep^+$  (dash-dotted) and  $E\bar{p}^+$  (dashed).



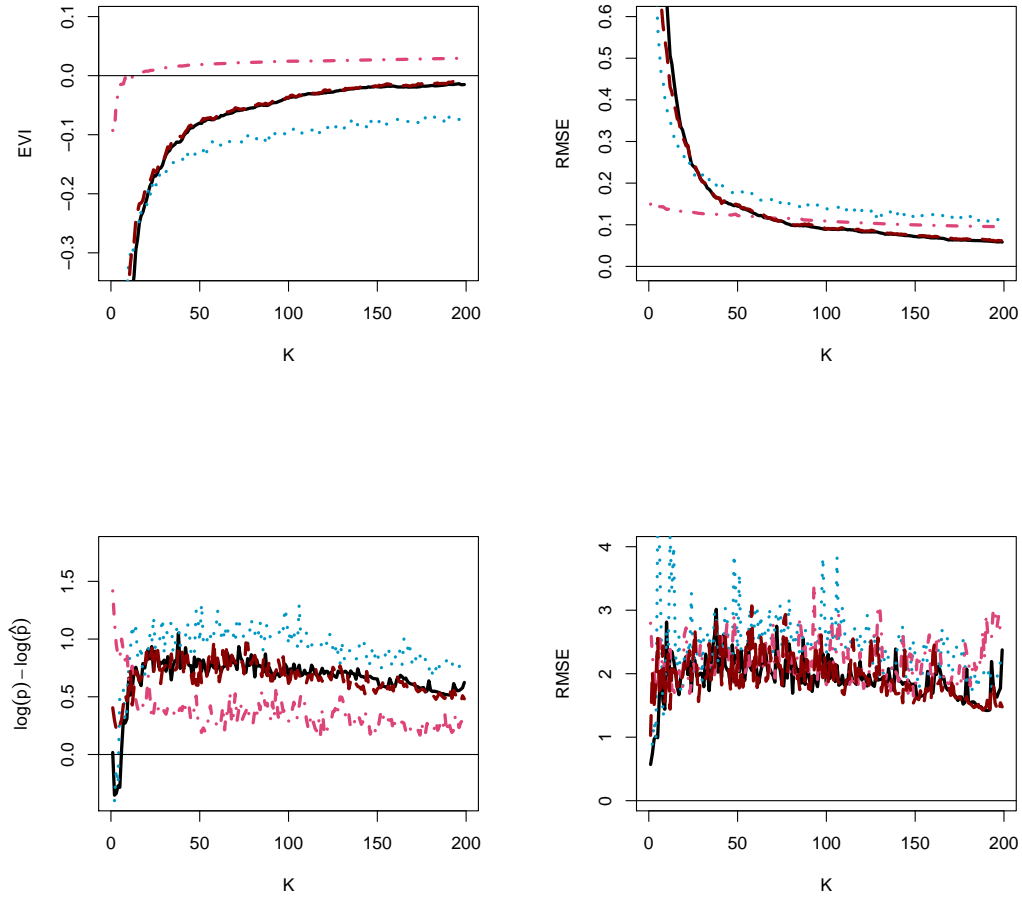
**Figure 5.3:** Fréchet distribution with  $\gamma = 0.5$ . Estimation of  $\gamma$  (top) and tail probability (bottom), bias (left), RMSE (right): GPD-ML (full line),  $T\bar{p}$  (dotted, using minimum variance principle),  $Ep$  with  $\rho = -2$  (dash-dotted) and  $E\bar{p}$  with  $(k_*, m) = (190, 150)$  (dashed).



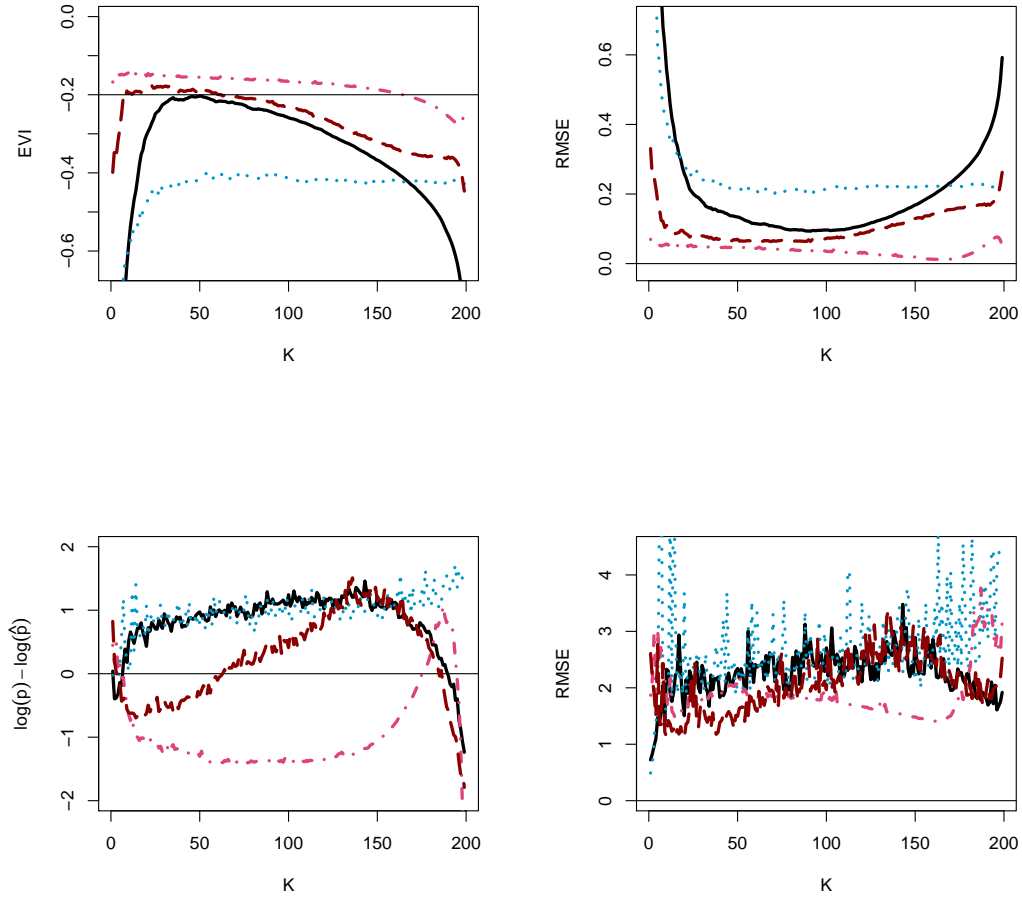
**Figure 5.4:** Fréchet distribution with  $\gamma = 0.5$ . Estimation of  $\gamma$  (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): Pareto-ML (full line),  $T\bar{p}^+$  (dotted),  $Ep^+$  (dash-dotted) and  $E\bar{p}^+$  (dashed).



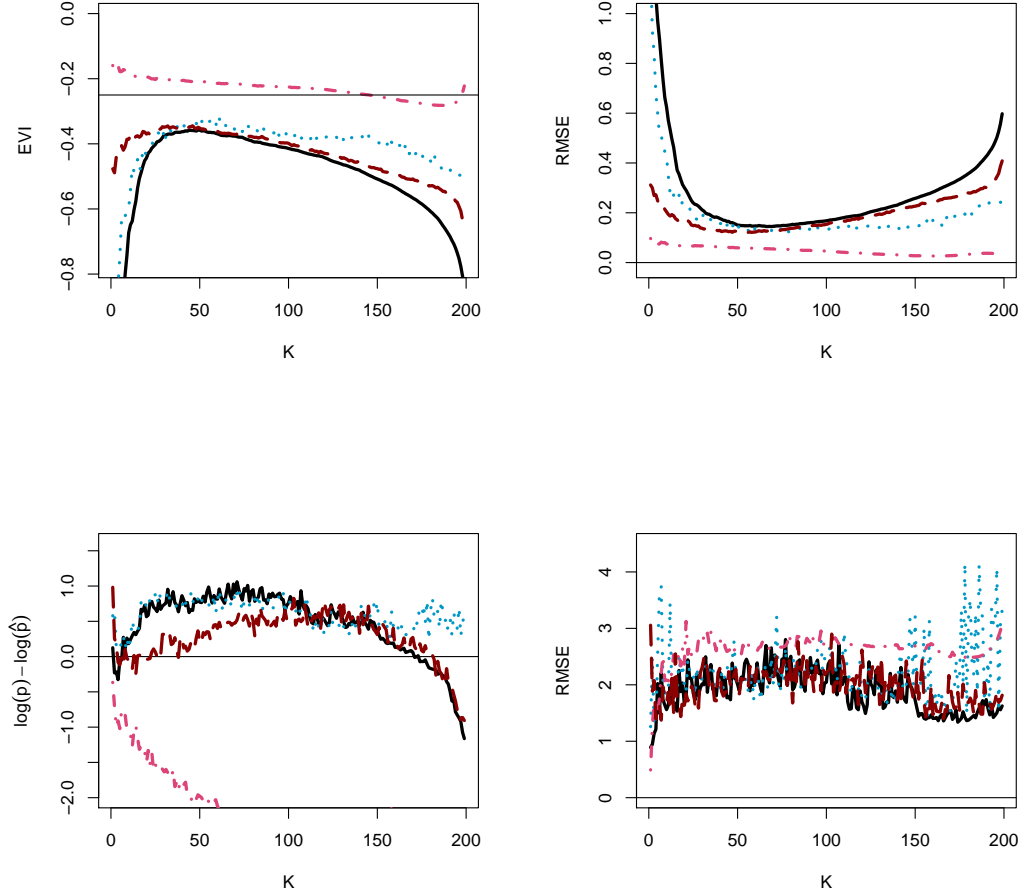
**Figure 5.5:** Standard normal distribution ( $\gamma = 0$  and  $\tilde{\rho} = 0$ ). Estimation of  $\gamma$  (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line),  $T\bar{p}$  (dotted),  $E\bar{p}$  (dash-dotted) and  $E\bar{p}$  (dashed).



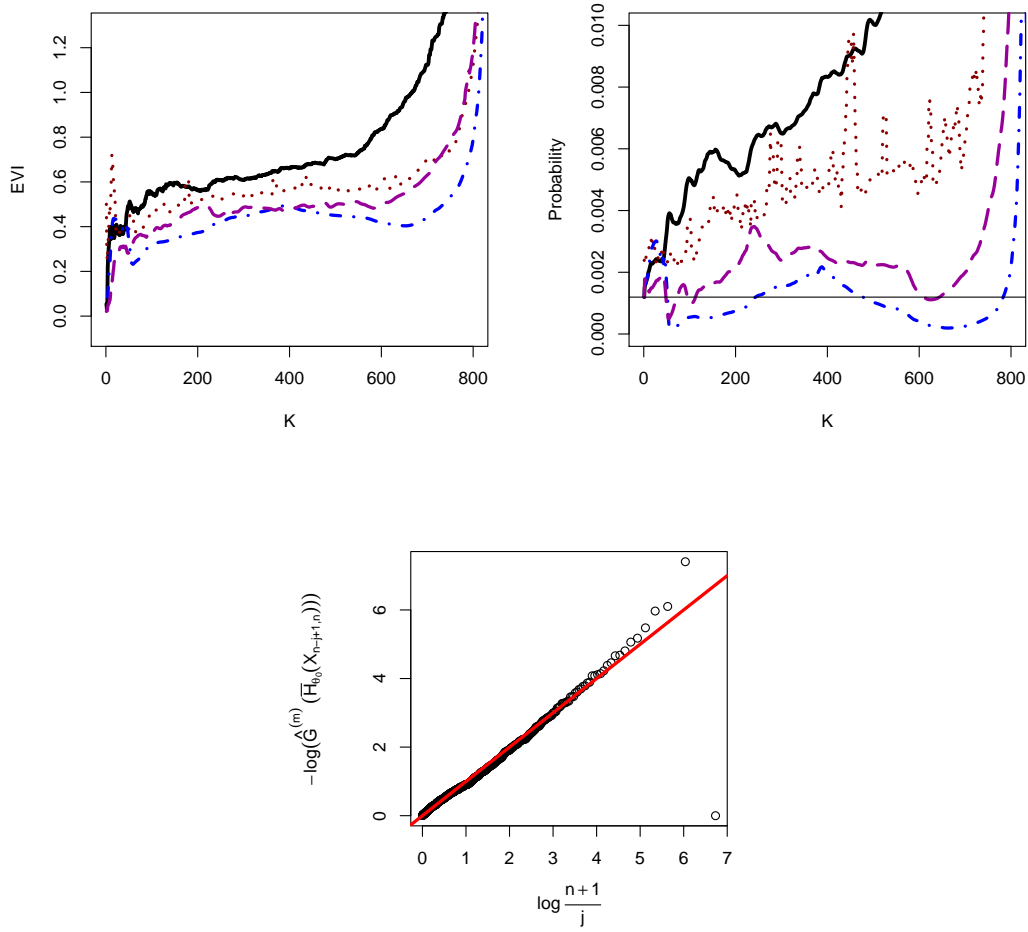
**Figure 5.6:** The exponential distribution ( $\gamma = 0$  and  $\tilde{\rho} = 0$ ). Estimation of  $\gamma$  (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line),  $T\bar{p}$  (dotted),  $Ep$  (dash-dotted) and  $E\bar{p}$  (dashed).



**Figure 5.7:** Reversed Burr distribution ( $\gamma = -0.2$  and  $\tilde{\rho} = -1$ ). Estimation of  $\gamma$  (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line),  $T\bar{p}$  (dotted),  $E\bar{p}$  (dash-dotted) and  $E\bar{p}$  (dashed).

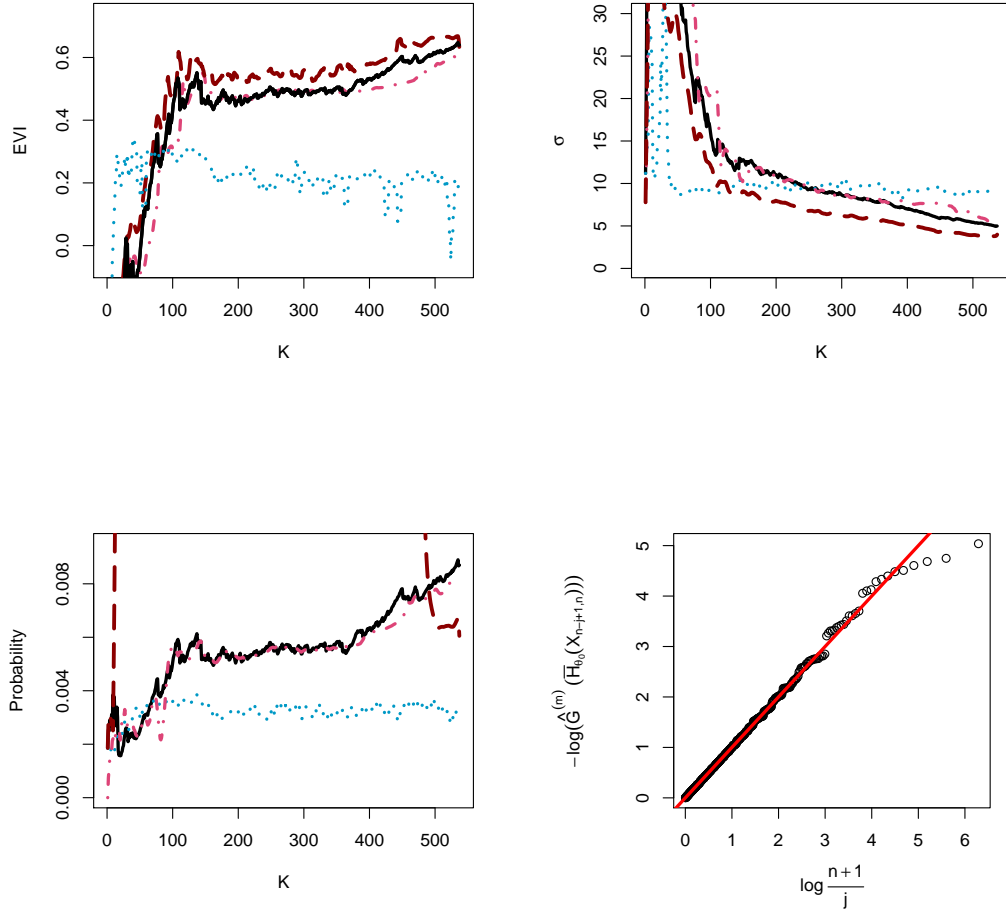


**Figure 5.8:** Extreme value Weibull distribution ( $\gamma = -0.25$  and  $\tilde{\rho} = -1$ ). Estimation of  $\gamma$  (top) and tail probability (bottom) using minimum variance principle, bias (left), RMSE (right): GPD-ML (full line),  $T\bar{p}$  (dotted),  $Ep$  (dash-dotted) and  $E\bar{p}$  (dashed).



**Figure 5.9:** Ultimates of Belgian car insurance claims: estimation of  $\gamma$  (top left), tail probability at maximum observation (top right): Pareto-ML (full line),  $T\bar{p}^+$  (dotted),  $E\bar{p}^+$  (dashed) and  $E\bar{p}^+$  (dash-dotted). Goodness-of-fit plot (bottom).





**Figure 5.10:** Winter rain data at Mont-Aigoual: estimation of  $\gamma$  and  $\sigma$  (top) and tail probability (bottom left) using minimum variance principle: GPD-ML (full line),  $T\bar{p}$  (dotted),  $E\bar{p}$  (dashed) and  $E\bar{p}$  (dash-dotted). Goodness-of-fit plot (bottom right).

# Appendix

## 5.A Proof of Theorem 5.1

In this section we provide details concerning the proof of Theorem 5.1.

*Asymptotic distribution of  $\hat{\gamma}_k^{E+}$ .*

From (5.3.12) we obtain up to linear terms in  $\delta_k$  that (denoting  $\hat{\gamma}_k$  for  $\hat{\gamma}_k^{E+}$ )

$$\begin{cases} \hat{\delta}_k &= \frac{\sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\hat{\gamma}_k})}{\sum_{j=1}^k b_\eta^2(Y_{j,k}^{-1/\hat{\gamma}_k})} \\ \hat{\gamma}_k &= H_{k,n} + \hat{\delta}_k B_k^{(1)}, \end{cases}$$

with  $B_k^{(1)} = \frac{1}{k} \sum_{j=1}^k b'_\eta(Y_{j,k}^{-1/\hat{\gamma}_k}) Y_{j,k}^{-1/\hat{\gamma}_k} \log Y_{j,k}$ . As  $k, n \rightarrow \infty$  and  $k/n \rightarrow 0$  we have  $B_k^{(1)} \rightarrow_p -\gamma \int_0^1 b'_\eta(u) u \log u du = -\gamma E B_\eta$ .

Using a Taylor expansion on the numerator of the right hand side of the first equation leads to

$$\frac{1}{k} \sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\hat{\gamma}_k}) = \frac{1}{k} \sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\gamma}) - (\hat{\gamma}_k - \gamma) \gamma^{-1} (E B_\eta) (1 + o_p(1)),$$

so that, with  $\frac{1}{k} \sum_{j=1}^k b_\eta^2(Y_{j,k}^{-1/\hat{\gamma}_k}) \rightarrow_p E b_\eta^2$ , up to lower order terms

$$\hat{\delta}_k = \frac{1}{E b_\eta^2} \frac{1}{k} \sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\gamma}) - (\hat{\gamma}_k - \gamma) \gamma^{-1} \frac{E B_\eta}{E b_\eta^2} (1 + o_p(1)).$$

Hence, inserting this expansion into  $\hat{\gamma}_k = H_{k,n} + \hat{\delta}_k B_k^{(1)}$ , finally leads to

$$\begin{aligned}
& \sqrt{k}(\hat{\gamma}_k - \gamma)(1 + o_p(1)) \\
&= \frac{Eb_\eta^2}{Eb_\eta^2 - (EB_\eta)^2} \sqrt{k} (H_{k,n} - \gamma) - \frac{\gamma EB_\eta}{Eb_\eta^2 - (EB_\eta)^2} \sqrt{k} \left( \frac{1}{k} \sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\gamma}) \right) \\
&= \frac{Eb_\eta^2}{Eb_\eta^2 - (EB_\eta)^2} \sqrt{k} (H_{k,n} - \gamma - \gamma \delta_k EB_\eta) \\
&\quad - \frac{\gamma EB_\eta}{Eb_\eta^2 - (EB_\eta)^2} \sqrt{k} \left( \frac{1}{k} \sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\gamma}) - \delta_k Eb_\eta^2 \right),
\end{aligned}$$

with  $\delta_k = \delta(U(n/k))$ . We now show that this final expression is a linear combination of two zero centered statistics (up to the required accuracy) which is asymptotically normal with the stated asymptotic variance. To this end let  $Z_{n-k,n} \leq Z_{n-k+1,n} \leq \dots \leq Z_{n,n}$  denote the top  $k+1$  order statistics of a sample of size  $n$  from the standard Pareto distribution with distribution function  $z \mapsto z^{-1}$ ,  $z > 1$ . Then from  $(\tilde{\mathcal{E}}_2^+)$

$$\begin{aligned}
H_{k,n} &= \frac{1}{k} \sum_{j=1}^k (\log U(Z_{n-j+1,n}) - \log U(Z_{n-k,n})) \\
&= \frac{1}{k} \sum_{j=1}^k \log \left\{ \left( \frac{Z_{n-j+1,n}}{Z_{n-k,n}} \right)^\gamma \left[ 1 + \gamma \delta(U(Z_{n-k,n})) B_\eta \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) \right. \right. \\
&\quad \left. \left. + o_p(1) |\delta(U(Z_{n-k,n}))| B_\eta \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) \left| \left( \frac{Z_{n-j+1,n}}{Z_{n-k,n}} \right)^\epsilon \right| \right] \right\} \\
&= \gamma \frac{1}{k} \sum_{j=1}^k \log \frac{Z_{n-j+1,n}}{Z_{n-k,n}} + \gamma \delta(U(Z_{n-k,n})) B_\eta \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) \\
&\quad + o_p(1) |\delta(U(Z_{n-k,n}))| B_\eta \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) \left| \left( \frac{Z_{n-j+1,n}}{Z_{n-k,n}} \right)^\epsilon \right|.
\end{aligned}$$

Now  $\log Z_{n-j+1,n} - \log Z_{n-k,n} =_d E_{k-j+1,k}$ , the  $(k-j+1)$ th smallest value from a standard exponential sample  $E_1, \dots, E_k$  of size  $k$ , so that  $\frac{1}{k} \sum_{j=1}^k \log \frac{Z_{n-j+1,n}}{Z_{n-k,n}} =_d \frac{1}{k} \sum_{j=1}^k E_j$  and  $\frac{1}{k} \sum_{j=1}^k B_\eta \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) =_d \frac{1}{k} \sum_{j=1}^k B_\eta(e^{-E_j}) =_d \frac{1}{k} \sum_{j=1}^k B_\eta(U_j)$  where  $U_1, \dots, U_k$  is a uniform  $(0,1)$  sample. Hence, since  $\delta(U(Z_{n-k,n}))/\delta(U(n/k)) \rightarrow_p 1$  and  $\frac{1}{k} \sum_{j=1}^k B_\eta(U_j) \rightarrow_p EB_\eta$ , we have that  $H_{k,n} - \gamma - \gamma \delta_k EB_\eta$  is asymptotically equivalent to  $\frac{1}{k} \sum_{j=1}^k \gamma(E_j - 1)$  as  $\sqrt{k} \delta_k \rightarrow \lambda$ .

Similarly

$$\begin{aligned}
& \frac{1}{k} \sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\gamma}) \\
&= \frac{1}{k} \sum_{j=1}^k b_\eta \left( \left[ \frac{U \left( \frac{Z_{n-j+1,n}}{Z_{n-k,n}} Z_{n-k,n} \right)}{U(Z_{n-k,n})} \right]^{-1/\gamma} \right) \\
&= \frac{1}{k} \sum_{j=1}^k b_\eta \left( \left( \frac{Z_{n-j+1,n}}{Z_{n-k,n}} \right)^{-1} \left[ 1 + \gamma \delta(U(Z_{n-k,n})) B_\eta \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) \right. \right. \\
&\quad \left. \left. + o_p(1) |\delta(U(Z_{n-k,n}))| B_\eta \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) \left| \left( \frac{Z_{n-j+1,n}}{Z_{n-k,n}} \right)^\epsilon \right| \right]^{-1/\gamma} \right) \\
&= \frac{1}{k} \sum_{j=1}^k b_\eta \left( \left( \frac{Z_{n-j+1,n}}{Z_{n-k,n}} \right)^{-1} \left[ 1 - \delta(U(Z_{n-k,n})) B_\eta \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) \right. \right. \\
&\quad \left. \left. + o_p(1) |\delta(U(Z_{n-k,n}))| B_\eta \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) \left| \left( \frac{Z_{n-j+1,n}}{Z_{n-k,n}} \right)^\epsilon \right| \right] \right) \\
&= \frac{1}{k} \sum_{j=1}^k b_\eta(e^{-E_j}) \\
&\quad - \delta(U(Z_{n-k,n})) \frac{1}{k} \sum_{j=1}^k b'_\eta \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) B_\eta \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) (1 + o_p(1)).
\end{aligned}$$

Since  $\delta(U(Z_{n-k,n}))/\delta_k \rightarrow_p 1$  and  $\frac{1}{k} \sum_{j=1}^k b'_\eta \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) B_\eta \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) \left( \frac{Z_{n-k,n}}{Z_{n-j+1,n}} \right) \rightarrow_p -Eb_\eta^2$  it follows that  $\frac{1}{k} \sum_{j=1}^k b_\eta(Y_{j,k}^{-1/\gamma}) - \delta_k Eb_\eta^2$  is asymptotically equivalent to

$$\frac{1}{k} \sum_{j=1}^k b_\eta(e^{-E_j}) =_d \frac{1}{k} \sum_{j=1}^k b_\eta(U_j)$$

as  $\sqrt{k}\delta_k \rightarrow \lambda$ , which is centered at 0 since  $E(b_\eta(U)) = 0$ .

*Asymptotic distribution of  $\hat{\gamma}_k^E$ .*

This derivation follows similar lines starting from (5.3.13):

$$\begin{cases} \frac{1}{k} \sum_{j=1}^k b'_\eta(\bar{H}_{\hat{\theta}_k}(Y_{j,k})) \bar{H}_{\hat{\theta}_k}(Y_{j,k}) \log(1 + \hat{\tau}_k Y_{j,k}) \rightarrow_p -\gamma EB_\eta, \\ \frac{1}{k} \sum_{j=1}^k b_\eta^2(\bar{H}_{\hat{\theta}_k}(Y_{j,k})) \rightarrow_p Eb_\eta^2, \\ \frac{1}{k} \sum_{j=1}^k b'_\eta(\bar{H}_{\hat{\theta}_k}(Y_{j,k})) \bar{H}_{\hat{\theta}_k}(Y_{j,k}) \rightarrow_p b_\eta(1), \\ \frac{1}{k} \sum_{j=1}^k b'_\eta(\bar{H}_{\hat{\theta}_k}(Y_{j,k})) \bar{H}_{\hat{\theta}_k}(Y_{j,k}) \frac{1}{1 + \hat{\tau}_k Y_{j,k}} \rightarrow_p \gamma(1 + \gamma) EC_\eta + b_\eta(1), \end{cases}$$

as  $k, n \rightarrow \infty$  and  $k/n \rightarrow \infty$ , so that the system of equations is asymptotically equivalent to

$$\begin{cases} \hat{\delta}_k = \frac{\frac{1}{k} \sum_{j=1}^k b_\eta(\bar{H}_{\hat{\theta}_k}(Y_{j,k}))}{Eb_\eta^2}, \\ \frac{1}{k} \sum_{j=1}^k \log(1 + \hat{\tau}_k Y_{j,k}) = \hat{\gamma}_k + \hat{\gamma}_k \hat{\delta}_k EB_\eta \\ \frac{1}{k} \sum_{j=1}^k \frac{1}{1 + \hat{\tau}_k Y_{j,k}} = \frac{1}{1 + \hat{\gamma}_k} - \hat{\gamma}_k \hat{\delta}_k EC_\eta. \end{cases}$$

Using a Taylor expansion on the numerator of the right hand side of the first equation leads to

$$\hat{\delta}_k Eb_\eta^2 = \frac{1}{k} \sum_{j=1}^k b_\eta(\bar{H}_\theta(Y_{j,k})) - \frac{EB_\eta}{\gamma} (\hat{\gamma}_k - \gamma) + (1 + \gamma) EC_\eta \left( \frac{\hat{\tau}_k}{\tau} - 1 \right).$$

Imputing this in the second and third equation in  $\gamma$  and  $\tau$ , and expanding these equations linearly around the correct values  $(\gamma, \tau)$ , while using, as  $k, n \rightarrow \infty$  and  $k/n \rightarrow 0$

$$\frac{1}{k} \sum_{j=1}^k \frac{\tau Y_{j,k}}{1 + \tau Y_{j,k}} \rightarrow_p \frac{\gamma}{1 + \gamma} \text{ and } \frac{1}{k} \sum_{j=1}^k \frac{\tau Y_{j,k}}{(1 + \tau Y_{j,k})^2} \rightarrow_p \frac{\gamma}{(1 + \gamma)(1 + 2\gamma)},$$

leads to the linearized equations

$$\begin{cases} (\hat{\gamma}_k - \gamma) \left( -1 + \frac{(EB_\eta)^2}{Eb_\eta^2} \right) + \left( \frac{\hat{\tau}_k}{\tau} - 1 \right) \left( \frac{\gamma}{1 + \gamma} - \gamma(1 + \gamma) \frac{EB_\eta EC_\eta}{Eb_\eta^2} \right) \\ \quad = - \left( \frac{1}{k} \sum_{j=1}^k \log(1 + \tau Y_{j,k}) - \gamma \right) + \frac{\gamma EB_\eta}{Eb_\eta^2} \frac{1}{k} \sum_{j=1}^k b_\eta(\bar{H}_\theta(Y_{j,k})), \\ (\hat{\gamma}_k - \gamma) \left( \frac{1}{(1 + \gamma)^2} - \frac{EB_\eta EC_\eta}{Eb_\eta^2} \right) + \left( \frac{\hat{\tau}_k}{\tau} - 1 \right) \left( -\frac{\gamma}{(1 + \gamma)(1 + 2\gamma)} + \gamma(1 + \gamma) \frac{(EC_\eta)^2}{Eb_\eta^2} \right) \\ \quad = - \left( \frac{1}{k} \sum_{j=1}^k \frac{1}{1 + \tau Y_{j,k}} - \frac{1}{1 + \gamma} \right) - \frac{\gamma EC_\eta}{Eb_\eta^2} \frac{1}{k} \sum_{j=1}^k b_\eta(\bar{H}_\theta(Y_{j,k})). \end{cases} \quad (5.1.15)$$

It follows that the right hand sides in (5.1.15) can be rewritten as linear combination of two zero centered statistics from which the asymptotic normality of  $\left( \sqrt{k}(\hat{\gamma}_k^E - \gamma), \sqrt{k}(\frac{\hat{\tau}_k^E}{\tau} - 1) \right)$

can be obtained, as stated in Theorem 5.1:

$$\left\{ \begin{array}{l} (\hat{\gamma}_k - \gamma) \left( -1 + \frac{(EB_\eta)^2}{Eb_\eta^2} \right) + \left( \frac{\hat{\tau}_k}{\tau} - 1 \right) \left( \frac{\gamma}{1+\gamma} - \gamma(1+\gamma) \frac{EB_\eta EC_\eta}{Eb_\eta^2} \right) \\ \quad = - \left( \frac{1}{k} \sum_{j=1}^k \log(1 + \tau Y_{j,k}) - \gamma - \gamma \delta_k EB_\eta \right) + \frac{\gamma EB_\eta}{Eb_\eta^2} \left( \frac{1}{k} \sum_{j=1}^k b_\eta(\bar{H}_\theta(Y_{j,k})) - \delta_k Eb_\eta^2 \right), \\ \\ (\hat{\gamma}_k - \gamma) \left( \frac{1}{(1+\gamma)^2} - \frac{EB_\eta EC_\eta}{Eb_\eta^2} \right) + \left( \frac{\hat{\tau}_k}{\tau} - 1 \right) \left( -\frac{\gamma}{(1+\gamma)(1+2\gamma)} + \gamma(1+\gamma) \frac{(EC_\eta)^2}{Eb_\eta^2} \right) \\ \quad = - \left( \frac{1}{k} \sum_{j=1}^k \frac{1}{1+\tau Y_{j,k}} - \frac{1}{1+\gamma} + \gamma \delta_k EC_\eta \right) - \frac{\gamma EC_\eta}{Eb_\eta^2} \left( \frac{1}{k} \sum_{j=1}^k b_\eta(\bar{H}_\theta(Y_{j,k})) - \delta_k Eb_\eta^2 \right). \end{array} \right.$$

This is done using similar derivations as in the case  $\hat{\gamma}_k^{E+}$ .

## 5.B Shiny application

The main aim of this shiny application is to make the simulation study more interactive, so as to study the behaviour of the estimators in detail. This application is hosted on: <https://phdshinygao.shinyapps.io/ExtendedModels/>. For each of the windows given in Figures 5.B.1 to 5.B.3, each graphical control element (or widget) is marked by a numerical label next to the elliptic or rectangular shape with the same color as the shape outline. These elements are described below:

### Window 1

1. This tab is the page that contains the finite simulation studies conducted;
2. This tab is the page that contains all the case studies conducted using real data sets;
3. This is a drop down menu, it contains all distributions from which the simulated data is generated. The distributions change as and when you change the domain of attraction (item 14);
4. This radio button is used to select the model we wish to transform or extend;

5. Under the transformed models this is the choice of  $m = k^a$  for all subsequent  $k$  integers for some choice of  $a \in (0, 1)$ . Note if the slider is moved all the way to the right, this choice of  $a$  is the averaged estimates over the minimum variance estimates across all  $k$ , i.e.  $\hat{a} = \sum_{k=2}^n (\hat{\gamma}_k^T - \bar{\gamma}_n^T)^2$ ;
6. Under the extended models this slider is used to choose the  $\rho$  when using the parametric model: in the case of the Pareto extension this function is  $B(u) = u^{-\rho} - 1$ ;
7. This slider is used to select the  $k$  value at which we fix the function  $B(u)$  when using the non-parametric extension;
8. This slider is used to select a value of  $m$  (number of Bernstein polynomials) to estimate the function  $B(u)$  under the extended models;
9. If this check box is not checked, the plots will be a result of the choice of values selected using sliders under the extended models. If the box is checked the simulations will be averaged over the minimum variance estimates across all  $k$ , i.e.  $(\hat{m}, \hat{K}^*, \hat{\rho}) = \operatorname{argmin} \sum_{k=2}^n (\hat{\gamma}_k^E - \bar{\gamma}_n^E)^2$ ;
10. This block of check boxes can be used to modify the plot aesthetics, the buttons below are used to download plots;
11. These two range sliders can be used to adjust the plotting window of the figures (item 19);
12. If used, the moving average slider is used to smooth out the noise of estimators as a function of  $k$  by averaging over 2 or more estimates as chosen by the user;
13. These are plots that result from selected choices of inputs made by the user;
14. These tabs differentiate the three domains of attractions from which the data generating distributions belong;
15. This tab provides a choice of the EVI or tail probability plots;

## Window 2

16. This block contains a radio button to choose between two application data sets: the Mont-Aigoual station rainfall data and the wind speed data studied in Beirlant et al. (2004) (This data is not studied in this thesis). This block also contains a drop down menu with the different seasons of the rainfall data (or different states of the wind speed data);
17. These two tabs are used to select between case studies with a real  $\gamma \in \mathbb{R}$  and positive  $\gamma > 0$ ;
18. These four tabs can be used to view either the EVI, sigma, tail probability or goodness of fit plot;
19. If item 22 is checked this purple fill depicts the confidence intervals of the EVI estimates;

## Window 3

20. When this check box is ticked, the gpd confidence intervals are included;
21. This is the  $\xi_0$  choice used in computing the confidence intervals;
22. This is the check box used to display the confidence intervals as shown in item 19; furthermore if the confidence interval check box is checked, the choice of the model type will have to be made, either show the confidence bands for the parametric, or non parametric model (both cannot be shown at the same time);
23. This is a group of sliders which can be used to select parameter values  $(\gamma_0, \sigma_0)$  in order to analyze the P-P plot for the complete data set using the non-parametric model. The parameter values should be chosen as  $\hat{\gamma}, \hat{\sigma}$  obtained as a function of  $k$  and the function  $B(u)$  estimated using  $m = n^a$  with  $a \in (0, 1)$ ;
24. This text label indicates the correlation coefficient of the P-P plot;



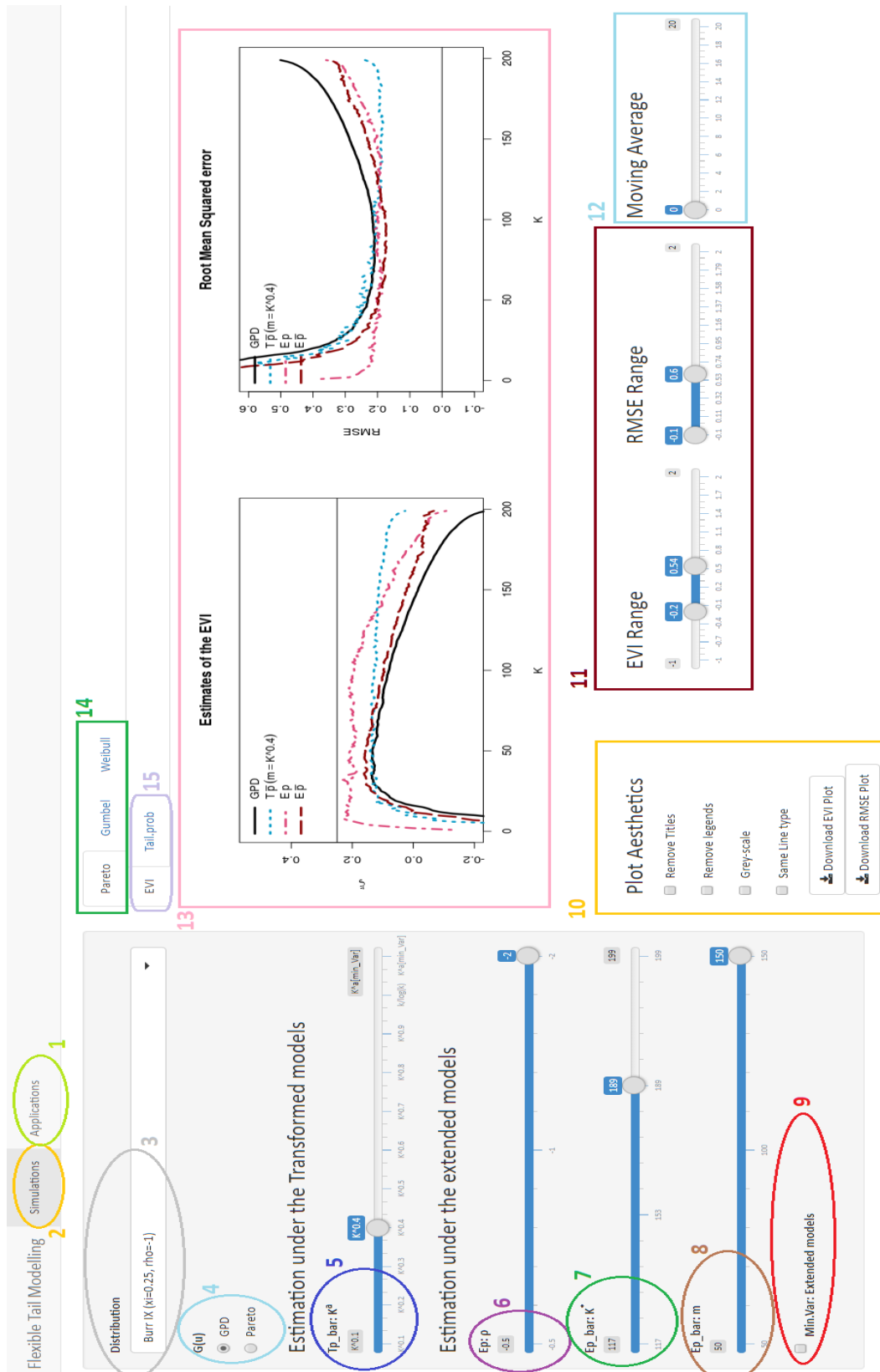


Figure 5.B.1: Window 1

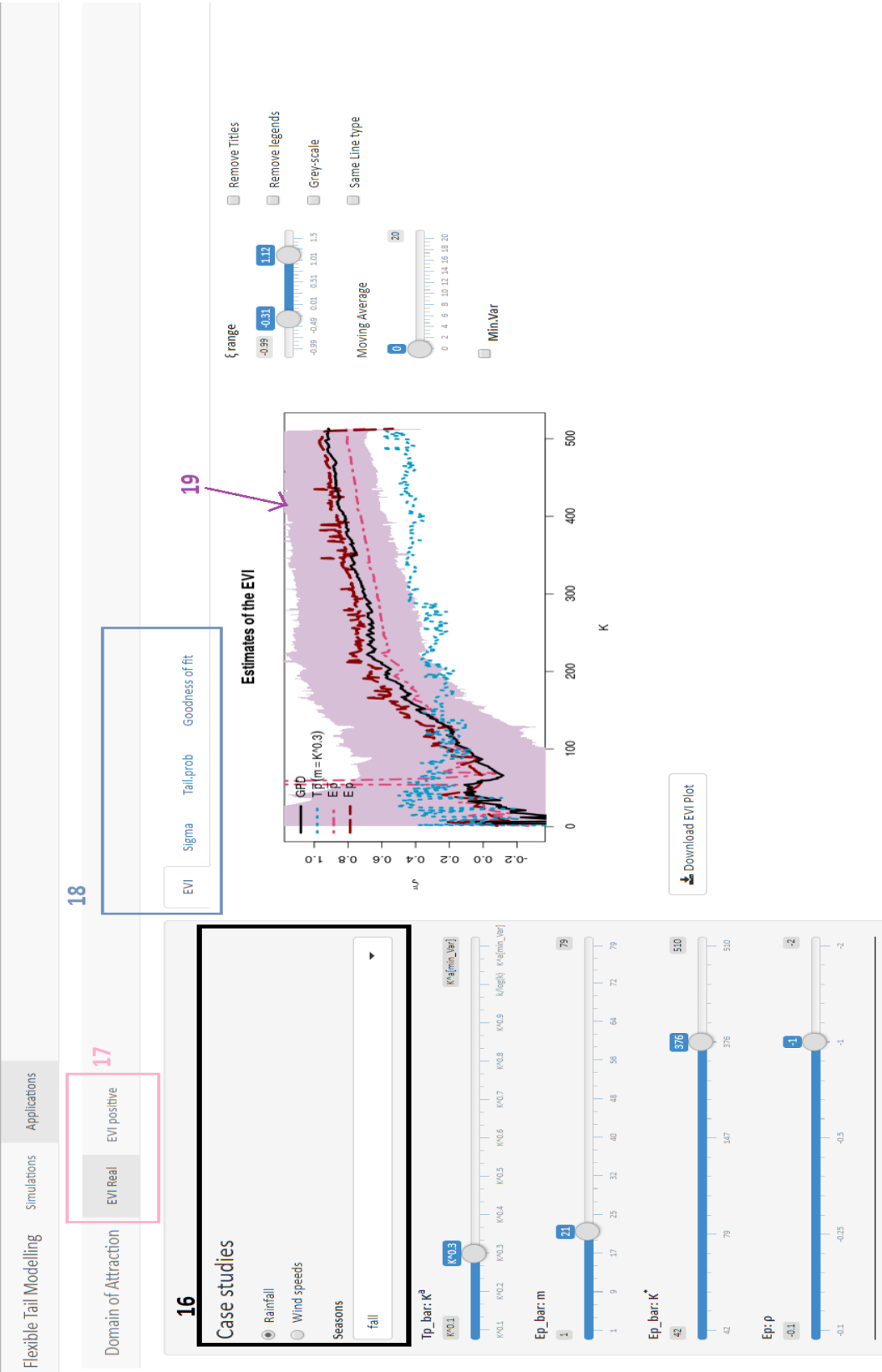


Figure 5.B.2: Window 2

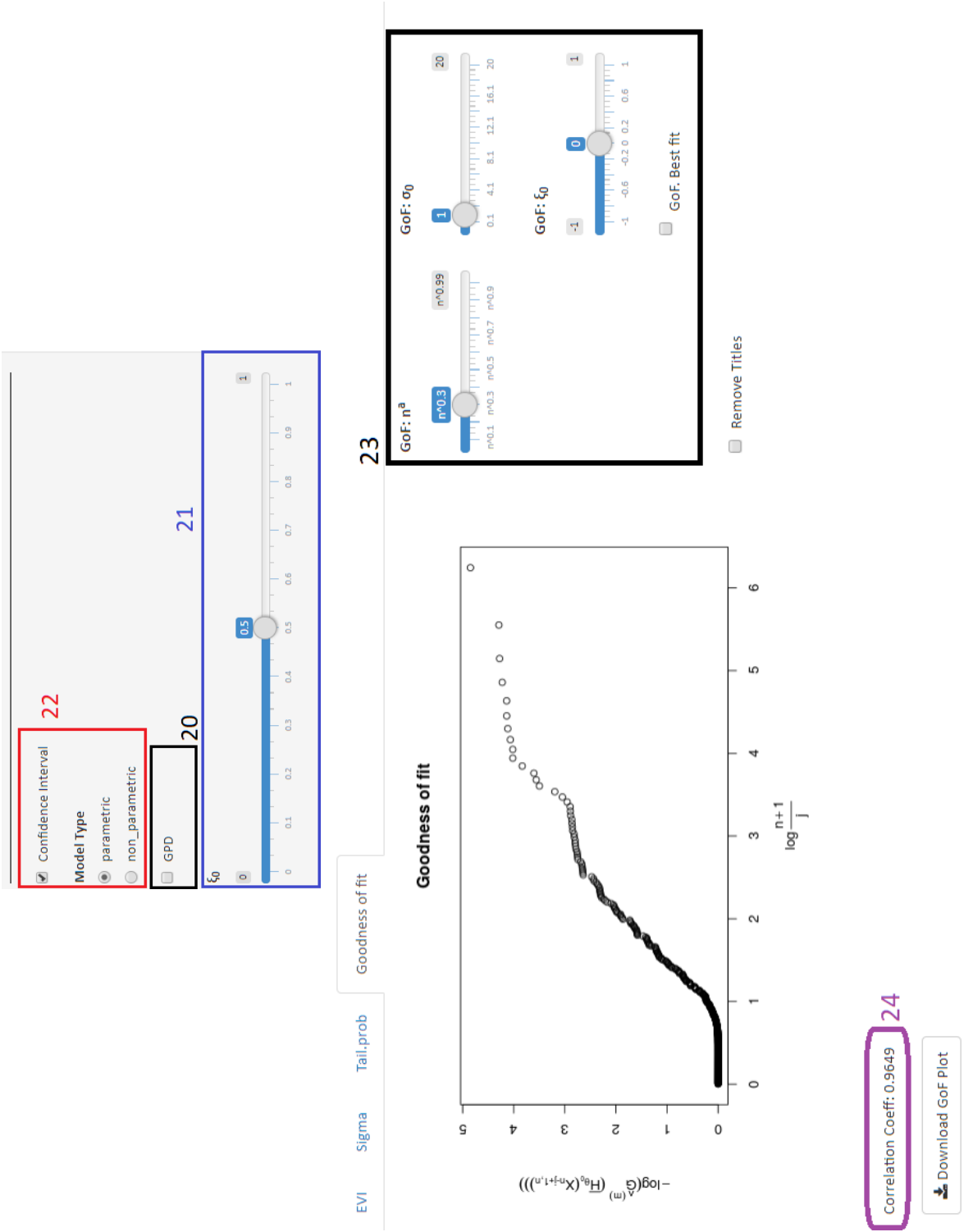


Figure 5.B.3: Window 3

# Chapter 6

## Thesis conclusion

In Chapter 3 we used the mathematical fact that for high thresholds, the distribution of the right tail function is strictly Pareto. We proposed both a penalised ML approach and a Bayesian a posteriori approach of applying some penalty on the parameter  $\delta$ . We showed that the asymptotic MSE of the optimal penalised estimator is uniformly smaller than the MSE of the EPD-ML estimator. We conducted a simulation study to assert these findings. A short case study on the Secura Belgian Re data was also conducted.

In Chapter 4, we proposed a bias reduced version of the Worms and Worms (2014). We also applied the shrinkage estimation procedure proposed in Chapter 3 and proposed in part a penalised version which led to much improved results. Due to the lack of any distribution theory of the Worms and Worms (2014) we conducted finite sample simulations for the proposed estimators and presented a bootstrap algorithm in order to construct confidence intervals for the adapted EVI. Effectiveness of the proposed method were assessed with a long-tailed insurance case-study. The proposed estimators were shown to have an overall better performance.

In Chapter 5, we generalise the bias reduction approach of Beirlant et al. (2009) to all

domains of attraction using a GP model. We use the flexible GP modelling introduced in Tencaliec et al. (2018) to model the second order component. We also proposed a method to assess the goodness of fit of the proposed model. To the best of our knowledge bias reduction of the GPD has not been considered in literature.

## 6.1 Outlook and future work

Careful considerations are needed when deciding exactly which portion of the data can be classified as the tail i.e. choosing a threshold. Choosing a threshold is ultimately subject to a practitioner's judgement often based on graphical diagnostics. This approach can be very tricky when multiple data sources are considered. This is one of the reasons research focus on building models that can capture the bulk and tail component of a distribution is gaining significance.

This thesis considered methods based on second-order modelling. In most cases within the EVT framework, applying second order expansions allows us to conveniently derive explicit expressions which do not suffer from complicated computational aspects faced by mixture models. The proposed methods in Chapter 5 were computationally intensive as the algorithms were implemented in the **R** programming language. Using a lower level language such as **C++** can greatly improve methods in Chapter 5.

The bias-reduction methods proposed in Chapter 5 can be extended to cases where data is censored but not heavy tailed. One can also consider methods proposed by Naveau et al. (2016) and Tencaliec et al. (2018) to model the entire range of censored data.

Lastly, focus in this thesis was primarily placed on the estimation of the EVI and occasionally on small tail probabilities. Other tail quantities can be further derived from the methods proposed.

# References

- Albrecher, H., Teugels, J. L. and Beirlant, J. (2017), *Reinsurance: Actuarial and Statistical Aspects*, John Wiley & Sons.
- Ameraoui, A., Boukhetala, K. and Dupuy, J.-F. (2016), ‘Bayesian estimation of the tail index of a heavy tailed distribution under random censoring’, *Computational Statistics & Data Analysis* **104**, 148–168.
- Balkema, A. A. and de Haan, L. (1974), ‘Residual life time at great age’, *The Annals of probability* pp. 792–804.
- Behrens, C. N., Lopes, H. F. and Gamerman, D. (2004), ‘Bayesian analysis of extreme events with threshold estimation’, *Statistical Modelling* **4**(3), 227–244.
- Beirlant, J., Bardoutsos, A., de Wet, T. and Gijbels, I. (2016), ‘Bias reduced tail estimation for censored pareto type distributions’, *Statistics & Probability Letters* **109**, 78–88.
- Beirlant, J., Dierckx, G., Goegebeur, Y. and Matthys, G. (1999), ‘Tail index estimation and an exponential regression model’, *Extremes* **2**(2), 177–200.
- Beirlant, J., Figueiredo, F., Gomes, M. I. and Vandewalle, B. (2008), ‘Improved reduced-bias tail index and quantile estimators’, *Journal of Statistical Planning and Inference* **138**(6), 1851–1870.
- Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004), *Statistics of Extremes:*

- Theory and Applications*, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd.
- Beirlant, J. and Guillou, A. (2001), ‘Pareto index estimation under moderate right censoring’, *Scandinavian Actuarial Journal* **2001**(2), 111–125.
- Beirlant, J., Guillou, A., Dierckx, G. and Fils-Villetard, A. (2007), ‘Estimation of the extreme value index and extreme quantiles under random censoring’, *Extremes* **10**(3), 151–174.
- Beirlant, J., Joossens, E. and Segers, J. (2009), ‘Second-order refined peaks-over-threshold modelling for heavy-tailed distributions’, *Journal of Statistical Planning and Inference* **139**(8), 2800–2815.
- Beirlant, J., Maribe, G., Naveau, P. and Verster, A. (2018), ‘Bias reduced peaks over threshold tail estimation’, *arXiv preprint arXiv:1810.01296* .
- Beirlant, J., Maribe, G. and Verster, A. (2018), ‘Penalized bias reduction in extreme value estimation for censored pareto-type data, and long-tailed insurance applications’, *Insurance: Mathematics and Economics* **78**, 114–122.
- Beirlant, J., Maribe, G. and Verster, A. (2019), ‘Using shrinkage estimators to reduce bias and mse in estimation of heavy tails’, *REVSTAT–Statistical Journal* **17**(1), 91–108.
- Beirlant, J., Worms, J. and Worms, R. (2018), ‘Estimation of the extreme value index in a censorship framework: asymptotic and finite sample behaviour’, *arXiv preprint arXiv:1804.06583* .
- Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987), ‘Regular variation (encyclopedia of mathematics and its applications)’.



- Brahimi, B., Meraghni, D. and Necir, A. (2015*a*), ‘Gaussian approximation to the extreme value index estimator of a heavy-tailed distribution under random censoring’, *Mathematical Methods of Statistics* **24**(4), 266–279.
- Brahimi, B., Meraghni, D. and Necir, A. (2015*b*), ‘Nelson-aalen tail product-limit process and extreme value index estimation under random censorship’, *arXiv preprint arXiv:1502.03955*.
- Caeiro, F. and Gomes, M. I. (2002), ‘A class of asymptotically unbiased semi-parametric estimators of the tail index’, *Test* **11**(2), 345–364.
- Caeiro, F. and Gomes, M. I. (2008), ‘Minimum-variance reduced-bias tail index and high quantile estimation’, *Revstat* **6**(1), 1–20.
- Caeiro, F. and Gomes, M. I. (2015), ‘Threshold selection in extreme value analysis’, *Extreme Value Modeling and Risk Analysis: Methods and Applications* pp. 69–82.
- Caeiro, F., Gomes, M. I. and Pestana, D. (2005), ‘Direct reduction of bias of the classical hill estimator’, *Revstat* **3**(2), 113–136.
- Carreau, J. and Bengio, Y. (2009*a*), ‘A hybrid pareto mixture for conditional asymmetric fat-tailed distributions’, *IEEE transactions on neural networks* **20**(7), 1087–1101.
- Carreau, J. and Bengio, Y. (2009*b*), ‘A hybrid pareto model for asymmetric fat-tailed data: the univariate case’, *Extremes* **12**(1), 53–76.
- Carreau, J., Naveau, P. and Neppel, L. (2017), ‘Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation’, *Water Resources Research* **53**(5), 4407–4426.
- Castillo, E. and Hadi, A. S. (1997), ‘Fitting the generalized pareto distribution to data’, *Journal of the American Statistical Association* **92**(440), 1609–1620.

- Chang, W., Cheng, J., Allaire, J., Xie, Y. and McPherson, J. (2018), *shiny: Web Application Framework for R*. R package version 1.1.0.  
**URL:** <https://CRAN.R-project.org/package=shiny>
- Coles, S. G. (2001), *An introduction to statistical modelling of extreme values*, Vol. 208, Springer.
- Coles, S. G. and Powell, E. A. (1996), ‘Bayesian methods in extreme value modelling: a review and new developments’, *International Statistical Review/Revue Internationale de Statistique* pp. 119–136.
- Danielsson, J., de Haan, L., Peng, L. and de Vries, C. G. (2001), ‘Using a bootstrap method to choose the sample fraction in tail index estimation’, *Journal of Multivariate analysis* **76**(2), 226–248.
- de Haan, L. (1970), On regular variation and its application to the weak convergence of sample extremes, PhD thesis, Mathematisch Centrum, Amsterdam.
- de Haan, L. (1984), Slow variation and characterization of domains of attraction, in ‘Statistical extremes and applications’, Springer, pp. 31–48.
- de Haan, L. and Ferreira, A. (2006), *Extreme value theory: an introduction*, Springer Science & Business Media.
- de Haan, L. and Peng, L. (1998), ‘Comparison of tail index estimators’, *Statistica Neerlandica* **52**(1), 60–70.
- de Haan, L. and Stadtmüller, U. (1996), ‘Generalized regular variation of second order’, *Journal of the Australian Mathematical Society* **61**(3), 381–395.
- Dekkers, A. L. and de Haan, L. (1993), ‘Optimal choice of sample fraction in extreme-value estimation’, *Journal of Multivariate Analysis* **47**(2), 173–195.
- Dekkers, A. L., Einmahl, J. H. and de Haan, L. (1989), ‘A moment estimator for the index of an extreme-value distribution’, *The Annals of Statistics* pp. 1833–1855.

- Delafosse, E. and Guillou, A. (2002), ‘Almost sure convergence of a tail index estimator in the presence of censoring’, *Comptes Rendus Mathematique* **335**(4), 375–380.
- Dodd, E. L. (1923), ‘The greatest and the least variate under general laws of error’, *Transactions of the American Mathematical Society* **25**(4), 525–539.
- Draisma, G., de Haan, L., Peng, L. and Pereira, T. T. (1999), ‘A bootstrap-based method to achieve optimality in estimating the extreme-value index’, *Extremes* **2**(4), 367–404.
- Drees, H. and Kaufmann, E. (1998), ‘Selecting the optimal sample fraction in univariate extreme value estimation’, *Stochastic Processes and their Applications* **75**(2), 149–172.
- DuMouchel, W. H. (1983), ‘Estimating the stable index  $\alpha$  in order to measure tail thickness: a critique’, *the Annals of Statistics* pp. 1019–1031.
- Dupuis, D. J. (1999), ‘Exceedances over high thresholds: A guide to threshold selection’, *Extremes* **1**(3), 251–261.
- Einmahl, J. H., Fils-Villetard, A., Guillou, A. et al. (2008), ‘Statistics of extremes under random censoring’, *Bernoulli* **14**(1), 207–227.
- Ferreira, A., de Haan, L. and Peng, L. (2003), ‘On optimising the estimation of high quantiles of a probability distribution’, *Statistics* **37**(5), 401–434.
- Feuerverger, A., Hall, P. et al. (1999), ‘Estimating a tail exponent by modelling departure from a pareto distribution’, *The Annals of Statistics* **27**(2), 760–781.
- Fisher, R. A. and Tippett, L. H. C. (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample, in ‘Mathematical Proceedings of the Cambridge Philosophical Society’, Vol. 24, Cambridge University Press, pp. 180–190.
- Fraga Alves, M., Gomes, M. I. and de Haan, L. (2003), ‘A new class of semi-parametric estimators of the second order parameter’, *Portugaliae Mathematica* **60**(2), 193–214.

- Fraga Alves, M., Gomes, M. I., de Haan, L. and Neves, C. (2007), ‘A note on second order conditions in extreme value theory: linking general and heavy tail conditions’, *REVSTAT Statistical Journal* **5**(3), 285–304.
- Frigessi, A., Haug, O. and Rue, H. (2002), ‘A dynamic mixture model for unsupervised tail estimation without threshold selection’, *Extremes* **5**(3), 219–235.
- Fréchet, M. (1927), ‘On the law of probability of the maximum gaps’, *Annales de la societe Polonaise de mathematique(Cracow)* **6**, 93–117.
- Gnedenko, B. (1943), ‘On the limit distribution of the maximum term of a random series’, *Annals of mathematics* pp. 423–453.
- Gomes, M. I., de Haan, L. and Peng, L. (2002), ‘Semi-parametric estimation of the second order parameter in statistics of extremes’, *Extremes* **5**(4), 387–414.
- Gomes, M. I., de Haan, L. and Rodrigues, L. H. (2008), ‘Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(1), 31–52.
- Gomes, M. I. and Martins, M. J. (2002), ‘“asymptotically unbiased” estimators of the tail index based on external estimation of the second order parameter’, *Extremes* **5**(1), 5–31.
- Gomes, M. I. and Martins, M. J. (2004), ‘Bias reduction and explicit semi-parametric estimation of the tail index’, *Journal of statistical planning and inference* **124**(2), 361–378.
- Gomes, M. I., Martins, M. J. and Neves, M. (2000), ‘Alternatives to a semi-parametric estimator of parameters of rare events—the jackknife methodology’, *Extremes* **3**(3), 207–229.
- Gomes, M. I., Martins, M. J. and Neves, M. (2007), ‘Improving second order reduced bias extreme value index estimation’, *Revstat* **5**(2), 177–207.

- Gomes, M. I. and Neves, M. M. (2011), ‘Estimation of the extreme value index for randomly censored data’, *Biometrical Letters* **48**(1), 1–22.
- Gomes, M. I. and Oliveira, O. (2001), ‘The bootstrap methodology in statistics of extremes—choice of the optimal sample fraction’, *Extremes* **4**(4), 331–358.
- Gomes, M. I. and Oliveira, O. (2003), ‘Censoring estimators of a positive tail index’, *Statistics & probability letters* **65**(3), 147–159.
- Gomes, M. I., Pereira, H. and Miranda, M. C. (2005), ‘Revisiting the role of the jackknife methodology in the estimation of a positive tail index’, *Communications in Statistics-Theory and Methods* **34**(2), 319–335.
- Gomes, M. I. and Pestana, D. (2007a), ‘A simple second-order reduced bias’ tail index estimator’, *Journal of Statistical Computation and Simulation* **77**(6), 487–502.
- Gomes, M. I. and Pestana, D. (2007b), ‘A sturdy reduced-bias extreme quantile (var) estimator’, *Journal of the American Statistical Association* **102**(477), 280–292.
- Gumbel, E. J. (1958), *Statistics of Extremes*, Columbia University Press.
- Hall, P. (1982), ‘On some simple estimates of an exponent of regular variation’, *J.R. Statist. Soc. B* **44**(1), 37–42.
- Hall, P. and Welsh, A. (1985), ‘Adaptative estimates of parameters of regular variation’, *Ann. Statist.* **13**, 331–341.
- Hill, B. M. (1975), ‘A simple general approach to inference about the tail of a distribution’, *The annals of statistics* pp. 1163–1174.
- Hosking, J. R. and Wallis, J. R. (1987), ‘Parameter and quantile estimation for the generalized pareto distribution’, *Technometrics* **29**(3), 339–349.
- Leadbetter, M. R., Lindgren, G. and Rootzén, H. (2012), *Extremes and related properties of random sequences and processes*, Springer Science & Business Media.

- Leurgans, S. (1987), ‘Linear models, random censoring and synthetic data’, *Biometrika* **74**(2), 301–309.
- Loretan, M. and Phillips, P. C. (1994), ‘Testing the covariance stationarity of heavy-tailed time series’, *Journal of empirical finance* **1**(2), 211–248.
- MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M. and Russell, G. (2011), ‘A flexible extreme value mixture model’, *Computational Statistics & Data Analysis* **55**(6), 2137–2157.
- Matthys, G. and Beirlant, J. (2000), ‘Adaptive threshold selection in tail index estimation’, *Extremes and Integrated Risk Management* pp. 37–49.
- Naveau, P., Huser, R., Ribereau, P. and Hannart, A. (2016), ‘Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection’, *Water Resources Research* **52**(4), 2753–2769.
- Papastathopoulos, I. and Tawn, J. A. (2013), ‘Extended generalised pareto models for tail estimation’, *Journal of Statistical Planning and Inference* **143**(1), 131–143.
- Peng, L. (1998), ‘Asymptotically unbiased estimators for the extreme-value index’, *Statistics & Probability Letters* **38**(2), 107–115.
- Pickands, J. (1975), ‘Statistical inference using extreme order statistics’, *Annals of Statistics* **3**(1), 119–131.
- R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.  
**URL:** <http://www.R-project.org>
- Reiss, R.-D. and Thomas, M. (2007), *Statistical analysis of extreme values*, Vol. 2, Springer.

- Smith, R. L. (1987), ‘Estimating tails of probability distributions’, *The annals of Statistics* pp. 1174–1207.
- Tancredi, A., Anderson, C. and O’Hagan, A. (2006), ‘Accounting for threshold uncertainty in extreme value estimation’, *Extremes* **9**(2), 87.
- Tencaliec, P., Favre, A.-C., Naveau, P. and Prieur, C. (2018), ‘Flexible semiparametric generalized pareto modeling of the entire range of rainfall amount’, *Environmetrics* .
- Wadsworth, J. and Tawn, J. (2012), ‘Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74**(3), 543–567.
- Worms, J. and Worms, R. (2014), ‘New estimators of the extreme value index under random right censoring, for heavy-tailed distributions’, *Extremes* **17**(2), 337–358.
- Worms, J. and Worms, R. (2018), ‘Extreme value statistics for censored data with heavy tails under competing risks’, *Metrika* **81**(7), 849–889.
- Zellner, A. (1971), *An introduction to Bayesian inference in econometrics*, Wiley.