UNIVERSITY OF CALGARY

A New Hybrid Estimation Method for the Generalized Pareto Distribution

by

Chunlin Wang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

July, 2011

© Chunlin Wang 2011

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "A New Hybrid Estimation Method for the Generalized Pareto Distribution" submitted by Chunlin Wang in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE.

Dr. Gemai Chen
Supervisor
Department of Mathematics and Statistics

Dr. Xuewen Lu
Department of Mathematics and Statistics

Dr. Tak Shing Fung
External Examiner
Information Technology

07. 19. 2011

Date

Abstract

The generalized Pareto distribution (GPD) is one of the most important distributions in the analysis of extreme values, especially in modeling the exceedances over thresholds in environment, geology, reliability and insurance, etc.

Most of the existing methods for estimating the scale and shape parameters σ and k of the GPD suffer from theoretical and computational problems. Among these methods, the maximum likelihood (ML) estimators may not possess the classical properties because the Cramer's regularity conditions fail to hold for $1/2 < k < 1$, and does not exist for $k > 1$. Even when ML estimators exist in a restricted parameter space, it may still give convergence problem. The maximum goodness-of-fit (MGF) estimators introduced by Luceño (2006) can always be found, but its efficiency is always low, and its bivariate numerical optimization can be complex. To improve the estimation in terms of bias and mean square error (MSE), and to simplify the computation, a new hybrid estimation method for the GPD is proposed in this thesis, which is mainly based on the idea of minimizing a goodness-of-fit measure and incorporating useful maximum likelihood information. Compared with the original ML and MGF methods, we show that this new hybrid method can not only reduce the estimation bias but also improve the MSE in the range $-6 \leq k \leq 2$, and our new hybrid estimators also perform well compared with other estimators suggested in the recent literature.

Key Words: EDF statistics; Estimation bias and MSE; Exceedences over threshold; Extreme values; Maximum goodness-of-fit estimators; Maximum likelihood estimators; Minimum distance estimators; Parameter profiling.

Acknowledgements

Not everyone can be kissed by the fortune. I would like to reserve my deep gratitude to my supervisor Dr. Gemai Chen for the unlimited support and encouragement over the past two years. You motive me to discover and make me taste the happiness in doing research. I was more than lucky to be one of your students.

I wish to express my special thanks to Dr. Hyang Mi Kim for guiding me through the 30 weeks project as research assistant. Under your great direction and patience, I enjoyed and learnt a lot from this precious opportunity.

Many thanks go to all the statistics faculty members at the department for opening a brand new statistics world for me and leading me into many interesting topics. Especially, I benefited a lot from the courses I took by Dr. Xuewen Lu, Dr. Murray Burke, Dr. John Collins and Dr. Gordon H. Fick. All of your profound knowledge and generous sharing impressed me very much.

In addition, I want to address my thanks to my dear statistics friends Bing Xia, Yayuan Zhu, Yifan Zhu, Dongdong Li and etc for the insightful discussions with you.

Moreover, I greatly acknowledged the financial support that I've received from the Department of Mathematics and Statistics at the University of Calgary during my graduate studies, which ensured my life and research in Calgary very enjoyable.

There are still so many people that I am grateful to. My any progress was made possible with your huge efforts that I can never forget.

Dedication

To My Dearest Mom and Dad.

Table of Contents

Abstract ii
Acknowledgements iii
Dedication iv
Table of Contents v
List of Tables vi
List of Figures vii
1 Introduction 1
1.1 The Generalized Pareto Distribution 1
1.2 Application: Peaks Over Thresholds 4
1.3 Framework of This Thesis 5
2 Estimation of the GPD Parameters 7
2.1 A Review of Literature 7
2.2 The Maximum Likelihood Estimation 10
2.2.1 The Estimating Equations 11
2.2.2 Computing the ML Estimates 13
2.3 The Maximum Goodness-of-Fit Estimation 14
2.3.1 EDF Statistics 15
2.3.2 Computing the MGF Estimates 17
2.4 The Empirical Bayesian Method 19
2.4.1 The Prior Information 20
2.4.2 Computing the EBM Estimates 21
2.5 A New Hybrid Estimation Method 22
2.5.1 Motivation 22
2.5.2 Computing the New Hybrid Estimates 24
2.5.3 Inference 25
3 Simulation Study 27
3.1 Comparisons of Finite-Sample Performances 27
3.1.1 Bias Comparisons 29
3.1.2 MSE Comparisons 29
4 An Example 36
5 Final Conclusions and Future Research 41
A Derivative of the Target Function G 44
B Finite-Sample Performances of Different Estimation Methods 45
C An R Function for Computing the New Hybrid Estimates 48
Bibliography 52

List of Tables

2.1 *Number of times a convergence problem occurred when computing the ML estimates based on* $10,000$ *replications, and* $\sigma = 1$ *is fixed.* 13
2.2 *Number of times a convergence problem occurred when computing the MGF estimates using the R function 'gpdmgf' in the package "POT" with the default starting values based on* $10,000$ *replications, and* $\sigma = 1$ *is fixed.* . 17

4.1 *The Bilbao waves data: the zero-crossing hourly mean periods (in seconds) of sea waves measured in the Bilbao bay, Spain, in January, 1997.* 36
4.2 *The estimated GPD parameters for the Bilbao waves exceedences at given thresholds using different estimation methods.* 37

B.1 *Summary of the finite-sample performances for different estimation methods based on 10,000 times simulation with sample size* $n = 50$, *and* $\sigma = 1$ *fixed.* 46
B.2 *Summary of the finite-sample performances for different estimation methods based on 10,000 times simulation with sample size* $n = 200$, *and* $\sigma = 1$ *fixed.* 47

List of Figures

1.1 *The shape of the density functions of the GPD for different values of k.* . 3

2.1 *Bias for estimating the GPD parameters using four MGF estimators and the ML estimators based on 10,000 replications with size* $n = 50$ *and* $\sigma = 1$. 19
2.2 *MSE for estimating the GPD parameters using four MGF estimators and the ML estimators based on 10,000 replications with size* $n = 50$ *and* $\sigma = 1$. 20

3.1 *Bias comparisons for estimating the GPD parameters, based on 10,000 replications with sample size* $n = 20$ *and* $n = 50$, *and* $\sigma = 1$. 30
3.2 *Bias comparisons for estimating the GPD parameters, based on 10,000 replications with sample size* $n = 100$ *and* $n = 200$, *and* $\sigma = 1$. 31
3.3 *MSE comparisons for estimating the GPD parameters, based on 10,000 replications with sample size* $n = 20$ *and* $n = 50$, *and* $\sigma = 1$. 33
3.4 *MSE comparisons for estimating the GPD parameters, based on 10,000 replications with sample size* $n = 100$ *and* $n = 200$, *and* $\sigma = 1$. 34

4.1 *The fitted GPD cdf using different methods versus the EDF for the Bilbao waves data at the threshold* $t = 7.5$. 39
4.2 *The plot of* $G(\theta; x)$ *and its first derivative to check whether the minimum is reached for the Bilbao waves data at the threshold* $t = 7.5$. 40
4.3 *Histograms of 1000 parametric bootstrap samples of* $\hat{\sigma}_{HYB}$ *and* $\hat{k}_{HYB}$ *for the Bilbao waves data at the threshold* $t = 7.5$. 40

Chapter 1

Introduction

The extreme value theory (EVT) is a branch of probability and statistics that studies the behaviors of unusually large or small values, called extremes, in a sequence of random variables. In traditional data analysis, such extremes are often negligible and labelled as outliers. But if the questions asked are related to some rare events that do not occur very frequently, the information contained in such extremes is often important. For example, the analysis of peaks over a high threshold level is of particular interest in many statistical applications. According to Pickands (1975), the distribution of the extreme values exceeding a given high threshold is found to be the generalized Pareto distribution (GPD), which is the model we will concentrate on in this thesis.

1.1 The Generalized Pareto Distribution

The **Generalized Pareto Distribution (GPD)** is a two-parameter family of distributions first introduced by Pickands (1975) [15] with the distribution function (cdf)

$$F(x;\sigma,k) = \begin{cases} 1-(1-kx/\sigma)^{1/k}\,, & \text{if } k \neq 0\,, \\ 1-\exp(-x/\sigma)\,, & \text{if } k = 0\,, \end{cases} \tag{1.1}$$

and the probability density function (pdf)

$$f(x;\sigma,k) = \begin{cases} \sigma^{-1}(1-kx/\sigma)^{1/k-1}\,, & \text{if } k \neq 0\,, \\ \sigma^{-1}\exp(-x/\sigma)\,, & \text{if } k = 0\,, \end{cases} \tag{1.2}$$

where the $\sigma > 0$ and $-\infty < k < \infty$ are the scale and shape parameters, respectively, and the domain of x is $(0,\infty)$ when $k \leq 0$ or $(0,\sigma/k)$ when $k > 0$. We denote the above distribution by GPD(σ,k).

The quantile function of the GPD is given by

$$Q(u;\sigma,k) = F^{-1}(u) = -\sigma \cdot T(1-u,k)\,, \quad 0 < u < 1\,,$$

where $T(\cdot)$ is the **Box-Cox transformation** (or Power transformation) defined by

$$T(z;\lambda) = \begin{cases} \frac{z^\lambda - 1}{\lambda}\,, & \text{if } \lambda \neq 0\,, \\ \ln(z)\,, & \text{if } \lambda = 0\,. \end{cases}$$

The mean of the GPD is $\sigma/(1+k)$ and the variance is $\sigma^2/[(1+k)^2(1+2k)]$, but the mean and variance exist only if $k > -1$ and $k > -1/2$, respectively. In general, the rth central moment of the GPD exists only if $k > -1/r$.

The GPD is important because of its versatility and flexibility. It contains the uniform, the exponential and the standard Pareto distributions as its special cases. Specifically, when $k = 1$, the GPD becomes the uniform distribution in the range $[0,\ \sigma]$; when k tends to 0, the GPD becomes the exponential distribution with mean σ as taken the limit; and when $k < 0$, the GPD reduces to the Pareto distribution (PD). In the situations where the exponential distribution might be used but some robustness is necessary against heavier or lighter tail, the GPD is considered as an appropriate alternative.

Also, some distributions often used in fitting heavy-tailed data may be approximated by a GPD through suitable choice of k and σ. As shown in Choulakian and Stephens (2001), the GPD provides a good approximation to the standard half-Cauchy, Lognormal and Weibull distributions.

We plot the density functions of the GPD for different values of the shape parameter k in Figure 1.1, where the scale parameter σ is fixed at 1. From Figure 1.1, we can see the two special cases of the GPD: the exponential distribution ($k = 0$) and the uniform distribution ($k = 1$). Also it is easily noticed that, when $k > 0$, the range of x is bounded by σ/k. In particular, when $k > 1/2$ the GPD has finite end-points with the pdf $f(x;\sigma,k) > 0$ near each end-point. As k gets bigger, the right tail of the GPD

increases sharply. On the other hand, when $k < 0$, x can take any positive number, and as k gets smaller, the right tail of the GPD becomes heavier. Especially, when $k \leq -1/2$ the GPD has infinite variance, and when $k \leq -1$ the mean of the GPD does not exist.

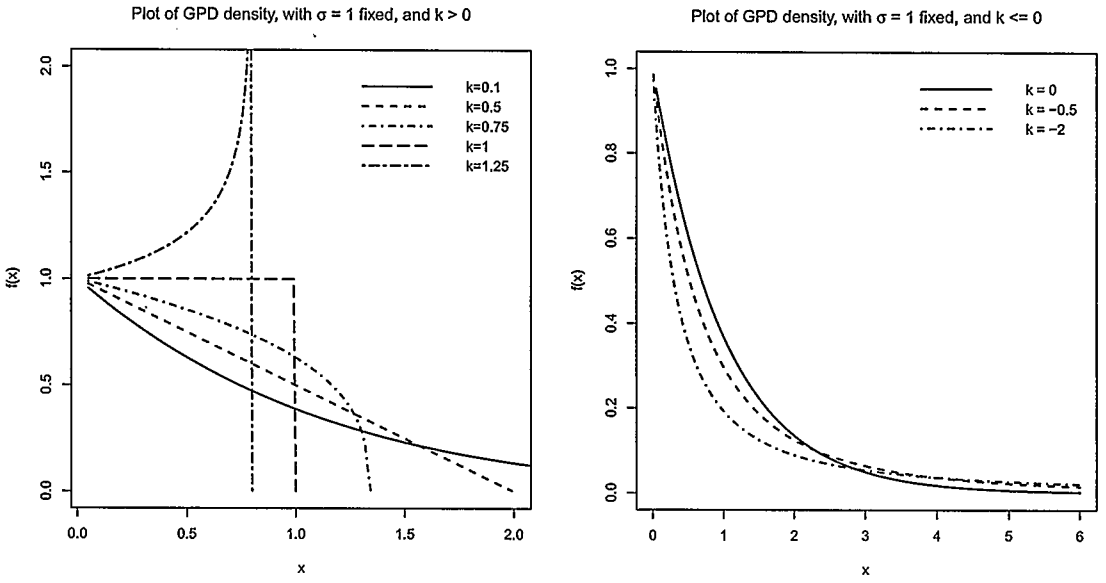

Figure 1.1: *The shape of the density functions of the GPD for different values of k.*

Besides the uniform, the exponential and the standard Pareto distributions, there are some connections between the GPD and other familiar distributions. As shown by Pickands (1975), an interesting such connection is given below.

Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) random variables from the GPD(σ, k) given in (1.1), and suppose that the number N of X_i exceeding a level t follows a Poisson distribution with mean λ. Let $Z_N = \max(X_1, X_2, \ldots, X_N)$. Then Z_N converges to the generalized extreme value distribution (GEVD) as t increases, that is,

$$P(Z_N \leq z) = \exp\left\{-\left(1 - \frac{k(x-t)}{\psi}\right)^{1/k}\right\}, \tag{1.3}$$

where the parameter $\psi = \sigma/\lambda^k$. And the location parameter t in the GEVD is actually called a threshold of interest in the GPD, which will be discussed in the next section.

Furthermore, it is useful to notice from (1.3) that the shape parameters of the GEVD and the GPD are equal.

1.2 Application: Peaks Over Thresholds

In extreme value theory, there are generally two methods for modeling the maximum or minimum observations. Originally introduced by Fisher and Tippett (1928), the classical approach to model the extreme values is based on the limiting distribution of the maxima or minima of a sequence of independent and identically distributed random variables, which turns out to be the GEVD. Because of this, the GEVD is appropriate when the data contain a set of maxima or minima during some fixed periods. Using only the maxima or minima information, this method was criticized by many authors, due to the loss of information contained in other extreme order statistics (see Smith (1990) for a general review of these two most widely used methods). The problem can be addressed by considering the limiting distribution of observations that exceed a given threshold in order to use several largest order statistics rather than the maxima. Hence the GPD was first introduced by Pickands (1975) to model the exceedences over a high threshold, such that the distribution of the exceedences $(X_i - t)$ converges to the GPD, where $\{X_i\}$ are the sample observations and t is a given threshold. For this reason, in extreme value theory the GPD is often referred to as the "Peaks Over Thresholds" (POT) model.

There are many examples of this application. Hosking and Wallis (1987) used the GPD to model the annual maximum flood levels of the River Nidd in England. Smith (1989) discussed an application of the GPD to study the ozone levels in the upper atmosphere. Castillo and Hadi (1997) fit the GPD to the heights of sea waves in the Bay of Biscay, Spain. Choulakian and Stephens (2001) applied the GDP to model the flood levels for 238 Canadian rivers, among others.

An attractive and useful property of the application of GPD in POT is its stability. If X follows a GPD (σ, k) as defined in (1.1), then the conditional distribution of the exceedence $X - t$ given that $X > t$ for any level t, follows the GPD $(\sigma - kt, k)$. This can be shown as follows

$$\begin{aligned} P(X - t \leq x | X > t) &= \frac{P(t < X \leq x + t)}{P(X > t)} \\ &= \frac{F(x + t;\ \sigma, k) - F(t;\ \sigma, k)}{1 - F(t;\ \sigma, k)} \\ &= \frac{(1 - k(x + t)/\sigma)^{1/k} - (1 - kt/\sigma)^{1/k}}{-(1 - kt/\sigma)^{1/k}} \\ &= 1 - \left(1 - \frac{kx}{\sigma - kt}\right)^{1/k} \\ &\sim \text{GPD}(\sigma - kt, k)\ . \end{aligned} \tag{1.4}$$

This property implies that the model is consistent with the data for any given thresholds, that is, the shape parameter k of the GPD stays unchanged with the level of different thresholds. So if a GPD model is appropriate, the shape of the distribution will be stable for any chosen levels of threshold, and $E(X - t \leq x | X > t) = (\sigma - kt)/(1 + k)$.

1.3 Framework of This Thesis

The GPD has received a lot of attention in both practical applications and theoretical research. The existing parameter estimation methods and their drawbacks will be discussed in the next Chapter. We will especially focus on the Maximum Likelihood (ML) method, the Maximum Goodness-of-Fit (MGF) method and the Empirical Bayesian Method (EBM). Then a new hybrid estimation method for the GPD will be developed at the end of Chapter 2. In Chapter 3, the finite-sample performances of the various estimators will be studied, and we will compare our proposed new estimator with other existing estimators in terms of bias and mean squared error (MSE). In Chapter 4, the new hybrid method will be illustrated through analyzing the sea waves data in the

Bay of Biscay, Spain. Finally in Chapter 5, some advantages of our new hybrid estimators will be highlighted, and some challenges encountered in working on this thesis and possible future works will be summarized.

Chapter 2

Estimation of the GPD Parameters

In this Chapter, we will consider the parameter estimation methods for the GPD introduced in Chapter 1. Most of the existing methods have some theoretical or computational issues. In general, the estimation of the GPD parameters σ and k is not an easy task, which still receives great attention in the most recent literature.

2.1 A Review of Literature

For the GPD, the most classical and important method of estimation, the maximum likelihood (ML) method, has been considered by DuMouchel (1983) [9], Davison (1984) [7], Smith (1984, 1985) [18, 19], Hosking and Wallis (1987) [13], Grimshaw (1993) [12], Choulakian and Stephens (2001) [5], and the references therein. In general, the maximum likelihood estimators may not exist in some region of the parameter space, and even when they exist, they may not possess the usual asymptotic properties and may give some computational difficulties. We will present the ML method in more details in Section 2.2.

Hosking and Wallis (1987) [13] and Dupuis and Tsao (1998) [10] studied some alternative estimates to the method of moment (MOM) estimates, and the probability-weighted moment (PWM) estimates of the GPD parameters. The MOM estimates of the GPD parameters are defined as

$$\hat{\sigma}_{\text{MOM}} = \bar{X}(\bar{X}^2/s^2 + 1)/2 \quad \text{and} \quad \hat{k}_{\text{MOM}} = (\bar{X}^2/s^2 - 1)/2\,, \tag{2.1}$$

and the PWM estimates of the GPD parameters are defined as

$$\hat{\sigma}_{\text{PWM}} = 2\bar{X}\alpha/(\bar{X} - 2\alpha) \quad \text{and} \quad \hat{k}_{\text{PWM}} = \bar{X}/(\bar{X} - 2\alpha) - 2\,, \tag{2.2}$$

where $\bar{X}$ and s^2 are the sample mean and sample variance, respectively, and one possible choice of α is $\alpha = n^{-1}\sum_{i=1}^{n} p_i X_{(i)}$, where $p_i = (n-i)/(n-1)$, and $X_{(i)}$ is the i^{th} order statistic of a random sample of size n.

In Hosking and Wallis (1987) [13], the MOM estimators and the PWM estimators were compared with the ML estimators when the range of the shape parameter k is restricted to $-1/2 < k < 1/2$. They concluded that the MOM is unreliable for $k < -0.2$, the PWM method performs well only when $-1/2 < k < 0$, and the ML method needs a sample size as large as $n = 500$ to possess its asymptotic efficiency. Additionally, when $k \leq -1/2$ the MOM estimates do not exist since the GPD has infinite variance, and similarly the PWM estimates do not exist when $k \leq -1$ because the mean of the GPD does not exist. Even when both of the MOM and PWM estimates exist, they may not be acceptable because some of the sample values may fall outside the range $0 < x < \hat{\alpha}/\hat{k}$. The infeasible problem of the PWM estimates has been pointed out by Chen and Balakrishnan (1995) [4].

To remedy this problem of unacceptable estimates from the MOM and the PWM method, Dupuis and Tsao (1998) [10] derived a hybrid method by incorporating a simple constraint on feasibility into the MOM and PWM estimates. They showed that their method can always give valid estimates. However, all of the estimates based on moments have low large-sample efficiencies and can be found only on a very restricted region of the parameter space. Hosking and Wallis (1987) obtained the asymptotic variances for the MOM and PWM estimators given in (2.1) and (2.2) (see formulas (4) and (6) in their paper).

Castillo and Hadi (1997) [3] proposed an elemental percentile method (EPM) to solve the invalid estimation problem for the GPD. This method expanded the estimable parameter space to all possible values of the parameters. The idea was to make full use of the order statistics by initially equating the GPD distribution function to all pairs of

the order statistics, and then use the median as the overall estimates of σ and k. Finally, compared with the MOM and the PWM method in terms of bias and root mean squared error (RMSE), their simulation results indicated that there was no dominant method for the considered range of the shape parameter values $-2 \leq k \leq 2$.

Luceño (2006) [14] brought out a maximum goodness-of-fit (MGF) method based on the family of the empirical distribution function (EDF) statistics. By minimizing any of the EDF statistics measuring the specific distance between the GPD distribution function and the EDF with respect to unknown parameters σ and k, this method can always give estimates for any possible values of the parameters. This technique was shown to be able to deal with the GPD parameters estimation for a given sample, as well as in the context of generalized linear model, even when the ML method and other methods failed. However, the bivariate search for the minimum could be complex and considerably time-consuming. We will carefully investigate the MGF method in Section 2.3, and borrow some of its ideas to develop our new hybrid estimation method.

Zhang (2007) [26] suggested a likelihood moment estimation (LME) method for the GPD to overcome the computational problems faced by the ML method. However, the evaluation of performance of this method is dependent on a careful choice of a constant r ($r < 1$) without any knowledge of the true value of k. Only if the assumption of $r = k$ is true, the LME was proved to be asymptotically efficient for $k < 1/2$. A possible robust choice of r is $r = -1/2$, which was recommended by the author.

Zhang and Stephens (2009) and Zhang (2010) [25, 27] provided a new efficient estimation method based on the likelihood and the empirical Bayesian method (EBM). It is computationally friendly and the data-driven prior can ensure that the parameter estimates are valid. However, it was indicated in their paper that the performance of the EBM estimates are quite sensitive to the choice of the shape parameter of the prior distribution as observed from extensive simulation. In Zhang and Stephens (2009), the

prior distribution was chosen as the GPD itself with the shape parameter $\tilde{k} = -1/2$. Then to improve the poor performance of the EBM estimators in the heavy-tailed situation (such as $k < -1$), a modified EBM (EBM*) was introduced in Zhang (2010). The main conclusion of the paper was that this EBM* generally outperforms the other existing estimation procedures in the extended range $-6 \leq k \leq 1/2$, in terms of estimation efficiency and bias. We will study this method further in Section 2.4.

The formal tests of Goodness-of-Fit for fitting the data to the GPD have been studied by Choulakian and Stephens (2001) [5].

2.2 The Maximum Likelihood Estimation

As the most important and widely used estimation method in statistics, the maximum likelihood estimation of GPD has drawn much attention in the literature. It is preferred because when $k < 1/2$, Smith (1984) showed that under Cramer's regularity conditions the ML estimators possess the classical asymptotic properties, such as consistency, asymptotical normality and asymptotical efficiency. However, problems arise when $k \geq 1/2$, which Smith (1984) identified as the non-regular case since the regularity conditions fail to hold, and also the convergence problems may occur in this case. When $k > 1$, the ML estimators do not exist because the likelihood function near the endpoint tends to infinity as x approaches σ/k. Therefore the ML method is only reliable in a restricted range of $k < 1/2$, and a special examination on the convergence issue is also necessary.

2.2.1 The Estimating Equations

Given a random sample $X = (X_1, X_2, \ldots, X_n)$ from the GPD with the cdf given in (1.1)[1], the log-likelihood function is given by

$$l(\sigma, k; X) = -n \log \sigma - \left(1 - \frac{1}{k}\right) \sum_{i=1}^{n} \log\left(1 - \frac{kX_i}{\sigma}\right) . \tag{2.3}$$

If $k > 1$, it is easy to check that

$$\lim_{\sigma/k \to X_{(n)}} l(\sigma, k; X) = +\infty .$$

Hence, the effort to find the ML estimators should be performed over a constrained parameter space $\mathscr{A} = \{k < 0,\ \sigma > 0\} \cup \{k > 0,\ \sigma/k > X_{(n)}\}$. To find the maximum of the log-likelihood on $\mathscr{A}$, consider the first derivatives of the GPD log-likelihood given in (2.3) with respect to k and σ, and set them to be zero to have the following estimating equations

$$\begin{cases} n(k-1) = \sum_{i=1}^{n} \log\left(1 - \frac{kX_i}{\sigma}\right) + (k-1) \sum_{i=1}^{n} \left(1 - \frac{kX_i}{\sigma}\right)^{-1} , \\ n = -(k-1) \sum_{i=1}^{n} \left(1 - \frac{kX_i}{\sigma}\right)^{-1} . \end{cases}$$

$$\Rightarrow \begin{cases} 1 = \left[1 + n^{-1} \sum_{i=1}^{n} \log\left(1 - \frac{kX_i}{\sigma}\right)\right] \cdot \left[n^{-1} \sum_{i=1}^{n} \left(1 - \frac{kX_i}{\sigma}\right)^{-1}\right] , \\ k = -n^{-1} \sum_{i=1}^{n} \log\left(1 - \frac{kX_i}{\sigma}\right) . \end{cases} \tag{2.4}$$

As pointed out by Davison (1984), the above bivariate maximization over $\mathscr{A}$ can be reduced to a one-dimensional search because in (2.4) the first equation is only dependent on the ratio $\theta = k/\sigma$ ($\theta < 1/X_{(n)}$), and then given a value of θ, a close-form expression for k is available. So it is natural and convenient to reparameterize the (σ, k) to (θ, k), which is a one-to-one mapping defined on the parameter space $\mathscr{A}$. Based on (θ, k) the

[1]For $k \to 0$, the GPD is reduced to the exponential distribution, so the result is already well known. So we will only consider the $k \neq 0$ case in this Section.

log-likelihood function is

$$l^*(\theta, k; X) = -n\log(k/\theta) - \left(1 - \frac{1}{k}\right)\sum_{i=1}^{n}\log\left(1 - \theta X_i\right) . \tag{2.5}$$

Substituting k in (2.5) with

$$k = -n^{-1}\sum_{i=1}^{n}\log\left(1 - \theta X_i\right) , \tag{2.6}$$

we have the profile log-likelihood function of θ given by

$$l(\theta; X) = -n - \sum_{i=1}^{n}\log\left(1 - \theta X_i\right) - n\log\left[-(n\theta)^{-1}\sum_{i=1}^{n}\log\left(1 - \theta X_i\right)\right] . \tag{2.7}$$

Suppose that a local maximum of (2.7) can be found at $\hat{\theta}_{\text{MLE}}$ numerically on the parameter space $\mathscr{B} = \{\theta < 1/X_{(n)}\}$, then the corresponding $\hat{k}_{\text{MLE}}$ and $\hat{\sigma}_{\text{MLE}}$ of (2.3), which are the ML estimators of σ and k, are given by

$$\hat{k}_{\text{MLE}} = -n^{-1}\sum_{i=1}^{n}\log(1 - \hat{\theta}_{\text{MLE}}X_i) \quad \text{and} \quad \hat{\sigma}_{\text{MLE}} = \hat{k}_{\text{MLE}}/\hat{\theta}_{\text{MLE}} . \tag{2.8}$$

It is important to emphasize that the local maximum of the GPD profile log-likelihood of θ over $\mathscr{B}$ corresponds to the local maxima of the GPD log-likelihood over $\mathscr{A}$, since we can easily express σ and k as the one-to-one functions of a single parameter θ, which is actually a ratio containing both k and σ. Then the unique value of $\hat{\theta}_{\text{MLE}}$ maximizing (2.7) gives the estimates of $\hat{\sigma}_{\text{MLE}} = \sigma(\hat{\theta}_{\text{MLE}})$ and $\hat{k}_{\text{MLE}} = k(\hat{\theta}_{\text{MLE}})$.

When $k < 1/2$, Smith (1984) proved that the ML estimators given in (2.8) is asymptotically normally distributed with the asymptotic variances achieving the Cramer-Rao lower bound under some proper regularity conditions. Specifically, we have

$$\begin{bmatrix} \hat{\sigma}_{\text{MLE}} \\ \hat{k}_{\text{MLE}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \sigma \\ k \end{bmatrix}, n^{-1}\begin{bmatrix} 2\sigma^2(1-k) & \sigma(1-k) \\ \sigma(1-k) & (1-k)^2 \end{bmatrix}\right) , \quad k < 1/2. \tag{2.9}$$

2.2.2 Computing the ML Estimates

However, the numerical maximization of (2.7) could still be complex. An algorithm for computing the ML estimates based on (2.7) was designed in Grimshaw (1993). As proved in Grimshaw (1993), there could have more than one root for the first derivative of (2.7) to be zero, and some convergence problem may occur when θ gets closer to its boundary because

$$\lim_{\theta \to 1/X_{(n)}} l(\theta; X) = \infty \, .$$

Also, our simulation results indicate that Grimshaw's algorithm for computing ML estimates is sensitive to the starting values. Therefore, a well chosen initial value $\theta^{(0)}$ is critical. In Grimshaw (1993), the initial value is suggested to be

$$\theta_L = 2(X_{(1)} - \bar{X}) / \left(X_{(1)}^2\right) \quad \text{or} \quad \theta_U = 1/X_{(n)} - \varepsilon \ \text{ with } \varepsilon = 10^{-6}/\bar{X} \, .$$

We use the algorithm provided by Dr. Grimshaw and carry out a simulation study with $10,000$ replications. The scale parameter σ is taken as 1 since the results are invariant for different values of σ. The convergence problem encountered is summarized in the following Table 2.1.

Table 2.1: *Number of times a convergence problem occurred when computing the ML estimates based on* $10,000$ *replications, and* $\sigma = 1$ *is fixed.*

n	$k=-3$	$k=-2$	$k=-1$	$k=-0.2$	$k=0.2$	$k=0.4$	$k=0.9$	$k=1.2$
10	96	315	1310	4501	7680	8949	9889	9973
20	5	0	26	449	1998	4159	9464	9924
50	8	0	0	1	13	107	7619	9914
100	14	1	0	0	0	0	5386	9953
200	29	0	0	0	0	0	2719	9986
500	50	0	0	0	0	0	395	10000

*The algorithm used here is provided by Dr. Grimshaw as proposed in his paper (1993).

The simulation results in Table 2.1 are in agreement with the conclusions of Grimshaw (1993) and Hosking and Wallis (1987). In many real cases where $k > 0$ and $n < 25$, the

ML method of GPD may have convergence problem, and the situation of no maximum can occur when k approaches and beyond 0.5. However, from the above table we can also find that in some very heavy-tailed cases, that is, $k < -2$, some extremely large observations may significantly alter the convergence of the algorithm even though the sample size is large. To overcome this computational problem in such cases, a carefully selected initial point $\theta^{(0)}$ is recommended.

2.3 The Maximum Goodness-of-Fit Estimation

The essential idea to assess the Goodness-of-Fit (GOF) of fitting a continuous probability distribution to data is based on measuring certain "distance" between the empirical distribution function (EDF) and the underlying distribution function. In Luceño (2006), the idea of GOF was borrowed for the parameter estimation purpose for the GPD. The proposed estimator is obtained by minimizing any of the EDF statistics, and is therefore called the maximum goodness-of-fit (MGF) estimation method.

In fact, this method can be dated back to Wolfowitz (1953, 1957) under a more general name of *minimum distance estimation method.* However, to be consistent with the name used in Luceño (2006), and to avoid confusing with other "distance" which is not directly related to EDF statistics[2], we prefer to use the name of MGF method in this thesis.

Most of the materials in Luceño (2006) are under the framework of generalized linear model where the parameters are estimated at the presence of some covariates, which will not be discussed in this thesis.

[2]For example, the Hellinger distance is defined as a measure of the similarity between two probability distributions; and the Mahalanobis distance is defined as a measure of the similarity of a test sample point to a known sample set.

2.3.1 EDF Statistics

Let $F_n(x)$ denote the right-continuous empirical distribution function (EDF) of a given random sample $X = (X_1, X_2, \ldots, X_n)$ from a continuous distribution function $F(x;\theta)$, that is,

$$F_n(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} I_{X_i}(x) \ ,$$

where $I_{X_i}(x) = 1$ if $X_i \leq x$, and $I_{X_i}(x) = 0$ if $X_i > x$.

Then any statistic that measures the discrepancy between $F_n(x)$ and $F(x;\theta)$ is called an EDF statistic. The following **Glivenko-Cantelli theorem** will make sure that as $n \to \infty$, $F_n(x)$ converges uniformly to $F(x,\theta)$.

Lemma 1 *As* $n \to \infty$,

$$\sup_x |F_n(x) - F(x;\theta)| \to 0 \qquad a.s..$$

There are mainly two classes of EDF statistics: the supremum EDF statistics which include the Kolmogorov-Smirnov (KS) statistic, the Kuiper statistic; and the integral EDF statistics which include the Cramér-von Mises (CM) statistic, the Anderson-Darling (AD) statistic, etc. In particular, Luceño (2006) introduced some modified Anderson-Darling statistics, such as the right-sided and left-sided Anderson-Darling (ADR, ADL) statistics and the Anderson-Darling statistics of higher degree. However, due to the non-differentiability of the KS statistic and the poor performances of the second degree AD statistics, we will only discuss the CM statistic, the AD statistic, the modified ADR and ADL statistics in the rest of this section. In terms of the GPD with cdf $F(x;\sigma,k)$, the definition of these four EDF statistics are

$$W^2(\sigma,k;x) = n \int_{-\infty}^{\infty} \{F_n(x) - F(x;\sigma,k)\}^2 \, dF(x;\sigma,k) \ , \tag{2.10}$$

$$A^2(\sigma,k;x) = n\int_{-\infty}^{\infty} \{F_n(x) - F(x;\sigma,k)\}^2 \{F(x;\sigma,k)(1-F(x;\sigma,k))\}^{-1} dF(x;\sigma,k) \ , \tag{2.11}$$

$$R^2(\sigma,k;x) = n\int_{-\infty}^{\infty} \{F_n(x) - F(x;\sigma,k)\}^2 (1-F(x;\sigma,k))^{-1} dF(x;\sigma,k) \ , \tag{2.12}$$

$$L^2(\sigma,k;x) = n\int_{-\infty}^{\infty} \{F_n(x) - F(x;\sigma,k)\}^2 F(x;\sigma,k)^{-1} dF(x;\sigma,k) \ . \tag{2.13}$$

The AD statistic A^2 assigns more weight to the observations in two tails of the distribution than the CM statistic W^2. Similarly, the modified ADR R^2 and ADL L^2 statistics give more weight to observations in the corresponding tail of the distribution function. So these EDF statistics can do different job in detecting the departure of the data from the GPD.

Since the $F_n(x)$ is a step function with jump at each order statistics, the above EDF statistics can be easily expressed in alternative forms for computational purposes. Denoting the i^{th} order statistic by $X_{(i)}$ and applying the probability integral transformation to the ordered sample to get $Z_i = F(X_{(i)};\ \sigma,k)$, $i = 1,\ldots n$. Then we can rewrite the EDF statistics defined in (2.10) to (2.13) as functions of σ and k as follows:

$$W^2(\sigma,k) = \sum_{i=1}^{n}\left(Z_i - \frac{i-1/2}{n}\right)^2 + \frac{1}{12n} \ ,$$

$$A^2(\sigma,k) = -n - \frac{1}{n}\sum_{i=1}^{n}\{(2i-1)\ln Z_i + (2n+1-2i)\ln(1-Z_i)\} \ ,$$

$$R^2(\sigma,k) = \frac{n}{2} - 2\sum_{i=1}^{n} Z_i - \frac{1}{n}\sum_{i=1}^{n}(2n+1-2i)\ln(1-Z_i) \ ,$$

$$L^2(\sigma,k) = -\frac{3n}{2} + 2\sum_{i=1}^{n} Z_i - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\ln Z_i \ .$$

More details on the EDF statistics can be found in Stephens (1986).

2.3.2 Computing the MGF Estimates

Let $X = (X_1, X_2, \ldots, X_n)$ be a random sample from the GPD with cdf $F(\sigma, k; X)$ given in (1.1). The $\hat{\sigma}_{\text{MGF}}$ and $\hat{k}_{\text{MGF}}$ of the GPD parameters will be obtained by minimizing any of the EDF statistics defined in Section (2.3.1) with respect to the unknown parameters σ and k. The minimization should be carefully performed with respect to the parameter space $\mathscr{A} = \{k < 0,\ \sigma > 0\} \cup \{k > 0,\ \sigma/k > X_{(n)}\}$. Again, this two-dimensional numerical optimization may cause convergence problems in some cases, and a well specified starting point $(\sigma^{(0)}, k^{(0)})$ could be very useful. The following Table 2.2 summarizes the convergence problem for computing the MGF estimates using the R function *gpdmgf* in the R package "POT" based on the paper of Luceño(2006). The default starting values used in the algorithm are $(\sigma^{(0)}, k^{(0)}) = (\bar{X}, 0)$. The simulation study is based on $10,000$ replications and the tolerance level is set to be $\varepsilon = 10^{-6}/\bar{X}$. The scale parameter σ is taken as 1 because of its invariant property.

Table 2.2: *Number of times a convergence problem occurred when computing the MGF estimates using the R function 'gpdmgf' in the package "POT" with the default starting values based on* $10,000$ *replications, and* $\sigma = 1$ *is fixed.*

Method	n	$k=-3$	$k=-2$	$k=-1$	$k=-0.2$	$k=0.2$	$k=0.4$	$k=0.9$	$k=1.2$
CM	20	6087	4269	495	0	0	0	0	10
	50	7759	5692	354	0	0	0	0	0
	100	8484	6741	239	0	0	0	0	0
	200	8676	7593	156	0	0	0	0	0
AD	20	3581	1461	40	0	0	0	3	212
	50	7517	4029	718	0	0	0	0	53
	100	9539	7973	3019	3	0	0	0	16
	200	9990	9782	6237	8	0	0	0	5
ADR	20	3978	2455	158	0	0	0	1	322
	50	7820	4730	722	0	0	0	3	128
	100	9727	8099	3019	3	0	0	0	50
	200	9993	9783	6237	8	0	0	0	11
ADL	20	7259	4198	295	0	0	0	0	0
	50	8636	4889	143	0	0	0	0	0
	100	9077	5327	76	0	0	0	0	0
	200	9116	5705	42	0	0	0	0	0

According to the results in Table 2.2, the MGF method generally works well in the cases when k approaches and beyond 0.5, where the ML method has severe problems. But in turns, the MGF method can not successfully deal with the very heavy-tailed cases, e.g., $k < -2$, and it is relatively slow as expected.

In Luceño (2006), the MGF method was also compared with the ML method, MOM, PWM and EPM in terms of bias and root mean squared error (RMSE) in the simulation study. The results showed that the MGF method can successfully handle the estimation of GPD parameters, even for the cases where the other estimation methods may fail. In addition, the results also indicated that there does not exist a unique EDF statistic that can perform uniformly best in the range $-2 \leq k \leq 2$ considered in the paper.

To evaluate the performance of the MGF estimators further, we also conduct a similar simulation study, but the range of the shape parameter k is extended to $-6 \leq k \leq 2$. Without loss of generality, the scale $\sigma = 1$ is fixed. The results of estimation bias and MSE are displayed in Figure 2.1 and Figure 2.2 based on 10,000 simulation samples of size $n = 50$ and the tolerance level is set at $\varepsilon = 10^{-6}/\bar{X}$.

The simulation results in Figure 2.1 and Figure 2.2 verify again that in the wider range of shape parameter $-6 \leq k \leq 2$, there is no MGF estimator that can outperform the others. However, it appears that the MGF estimator based on the CM statistic generally has the worse performance, and the MGF estimator based on the ADL statistic performs better only in the very heavy-tailed case, e.g., $k \leq -2$, while the MGF estimator based on the ADR statistic performs better in the common range $-2 \leq k \leq 2$. So from the practical point of view, the MGF estimator based on the ADR would be more desired. But above all, the MGF estimator based on the AD statistic has the most balanced performance in the whole considered range $-6 \leq k \leq 2$ compared with other estimators, because it gives weights equally to both tails. These results confirm again the conclusions of Luceño (2006), and the MGF method when the AD statistic is used is preferred in the

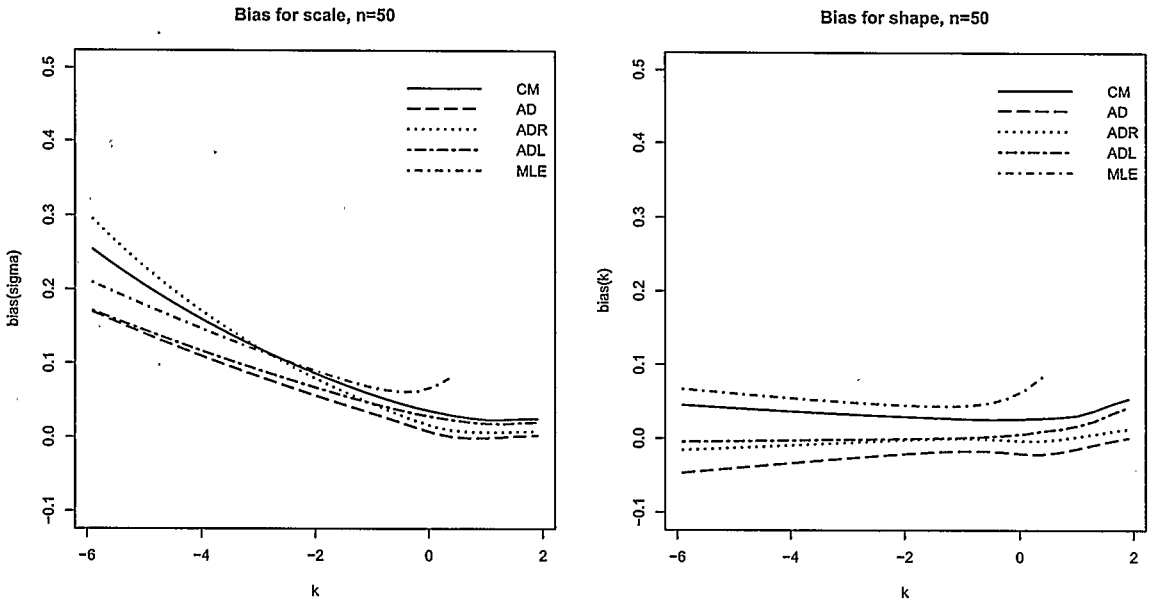

Figure 2.1: *Bias for estimating the GPD parameters using four MGF estimators and the ML estimators based on 10,000 replications with size* $n = 50$ *and* $\sigma = 1$.

rest of this thesis.

In general, the MGF method is based on the EDF statistics to make full use of the information contained in the sample order statistics, so it is always possible to give more robust estimates, even when other methods fail. Moreover, in Pollard (1980) the MGF estimators were proved to be consistent. Nevertheless, the insufficient use of the likelihood information determines that the MGF estimators have low large-sample efficiency, especially when Cramer's regularity conditions hold, in which case, as we have already known that ML estimators is asymptotically optimal.

2.4 The Empirical Bayesian Method

In order to improve the ML method for estimating the GPD parameters and to avoid the computational difficulties, Zhang and Stephens (2009) proposed a new estimation procedure based on the maximum likelihood to gain efficiency, but borrowed the idea

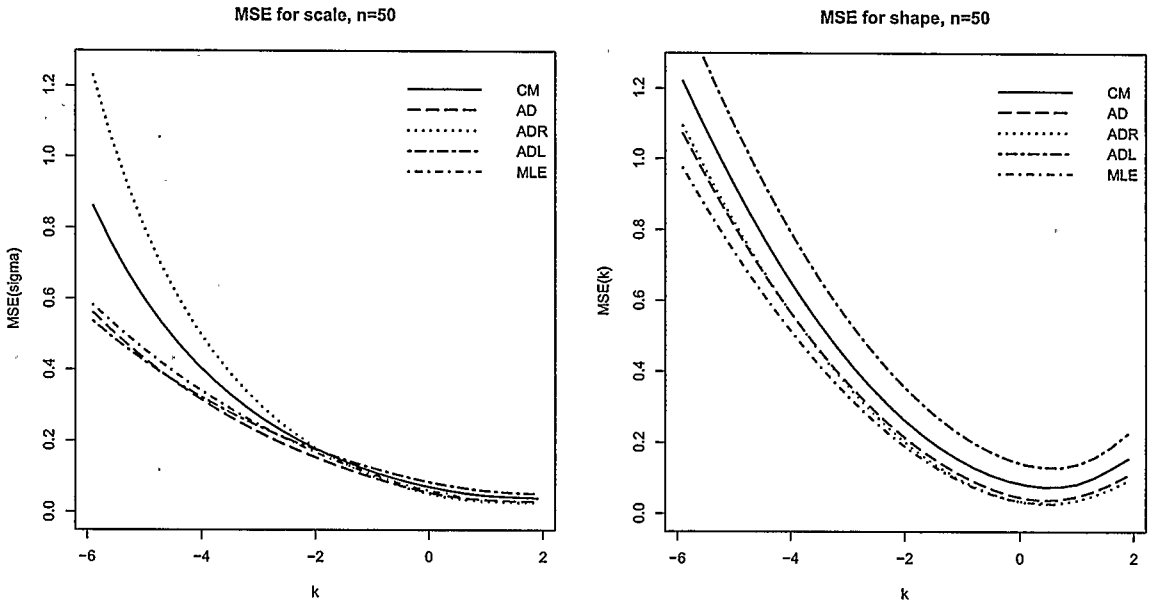

Figure 2.2: *MSE for estimating the GPD parameters using four MGF estimators and the ML estimators based on 10,000 replications with size* $n = 50$ *and* $\sigma = 1$.

from the empirical Bayesian method (EBM) to take the data-driven prior information into account. This proposed method has a simple approximated computational formula, which can result in never producing invalid estimates. Furthermore, in Zhang (2010) the EBM was modified (denoted as EBM*) to be more adaptive for the heavy-tailed cases in a wider range $-6 \leq k \leq 1/2$. As shown by the simulation results in these papers, the EBM and EBM* outperformed the other existing methods in terms of bias and efficiency.

2.4.1 The Prior Information

As mentioned in Section 2.2.1, it is convenient to use the reparameterization of (σ, k) into (θ, k), where $\theta = k/\sigma$. If the ML estimate $\hat{\theta}_{\text{MLE}}$ is known from maximizing the profile likelihood function given in (2.7), the ML estimates $\hat{k}_{\text{MLE}}$ and $\hat{\sigma}_{\text{MLE}}$ can be obtained by using the relation (2.8). Similarly, the EBM estimation procedure is based on the profile likelihood function of θ, and uses the sample-driven prior information through the

empirical Bayesian method. The EBM estimate of θ is defined as the posterior mean

$$\hat{\theta}_{\text{EBM}} = \int \theta \cdot \pi(\theta) L(\theta) d\theta \Big/ \int \pi(\theta) L(\theta) d\theta \,, \tag{2.14}$$

where $\pi(\theta)$ is a data-driven prior distribution of θ, and $l(\theta) = \log(L(\theta))$ is the profile log-likelihood function.

Then a key issue is to choose a reasonable prior distribution for θ to reduce bias and keep efficiency. Zhang and Stephens (2009) pointed out that the EBM estimates are in fact sensitive to the shape of the prior distribution as revealed by extensive simulation studies. To make sure the boundary constraint $\theta < 1/X_{(n)}$ is satisfied, a prior $\pi(\theta) = g(1/X_{(n)} - \theta)$ is considered such that the constraint becomes the requirement that the density function g must have a positive support $(0, \infty)$. Based on the simulation results of Zhang and Stephens (2009), a good choice of the prior distribution turned out to be the GPD density itself, with the data-driven scale $\tilde{\sigma} = 1/6Q_1$ and shape $\tilde{k} = -1/2$, where $Q_1 = X_{(\lfloor n/4+0.5 \rfloor)}$ is the first quartile of the sample data and $\lfloor\ \rfloor$ denotes the floor function.

Zhang (2010) modified this prior distribution by introducing a more reliable and adaptive prior. The updated EBM* was shown to be able to overcome the poor performance of the estimates in the heavy-tailed cases. The modified prior distribution is still the GPD itself, but instead the data-driven prior becomes $\pi^*(\theta) = g^*\left(\frac{n-1}{n+1} X_{(n)}^{-1} - \theta\right)$ with the prior scale and shape parameters replaced by $\tilde{\sigma}^* = 1/2\text{median}(\hat{\sigma}_{0.3}, \hat{\sigma}_{0.4}, \ldots, \hat{\sigma}_{0.9})$ and $\tilde{k}^* = -1$, where

$$\hat{\sigma}_p = \hat{k}_p \hat{x}_{1-p}/(1 - p^{\hat{k}_p}) \,, \quad \hat{k}_p = \log_p(\hat{x}_{1-p^2}/\hat{x}_{1-p} - 1) \,, \quad \text{and} \quad \hat{x}_\alpha = X_{(\lfloor n\alpha+0.5 \rfloor)} \,.$$

2.4.2 Computing the EBM Estimates

For a given prior distribution, the integrals in (2.14) is not easy to compute. To be computationally friendly and without loss of accuracy, an approximated version of (2.14)

is worked out, which is actually a weighted average

$$\hat{\theta}_{\text{EBM}} = \sum_{j=1}^{m} \theta_j \cdot w(\theta_j) \ , \text{ for } j = 1, \ldots, m \ , \tag{2.15}$$

where the weights w_j are given by $w(\theta_j) = L(\theta_j)/\sum_{t=1}^{m} L(\theta_t)$, $m = 20 + \lfloor\sqrt{n}\rfloor$, and the θ_j are the $\frac{j-0.5}{m}$th quantile of the proposed prior distribution, given by $\frac{j-0.5}{m} = 1 - G_{\tilde{\sigma},\tilde{k}}(1/X_{(n)} - \theta_j)$ which leads to the solution

$$\theta_j = \frac{1}{X_{(n)}} + \left(1 - \sqrt{\frac{m}{j-0.5}}\right) \Big/ (3Q_1) \ , \text{ for } j = 1, \ldots, m \ .$$

For the modified prior distribution in Zhang (2010), the θ_j^* are given by

$$\theta_j^* = \frac{n-1}{n+1} X_{(n)}^{-1} - \frac{\tilde{\sigma}^*}{\tilde{k}^*} \left[1 - \left(\frac{j-0.5}{m}\right)^{\tilde{k}^*}\right] \ , \text{ for } j = 1, \ldots, m \ .$$

The above θ_j and θ_j^* can be always within the boundary point $1/X_{(n)}$.

Once the EBM estimators $\hat{\theta}_{\text{EBM}}$ or $\hat{\theta}^*_{\text{EBM}}$ are calculated through (2.15), the EBM estimators $(\hat{\sigma}_{\text{EBM}}, \hat{k}_{\text{EBM}})$ or $(\hat{\sigma}^*_{\text{EBM}}, \hat{k}^*_{\text{EBM}})$ are obtained by using the reparameterization given in (2.8).

2.5 A New Hybrid Estimation Method

To reduce the estimation bias by using the EDF statistics, and to improve the estimation efficiency by incorporating the maximum likelihood information at the same time, a new hybrid estimation method for the GPD is proposed in this section. Compared with the original MGF method, the ML method and other available methods, it is generally less computationally intensive and it performs well in the common situations as well as in the heavy-tailed and non-regular cases.

2.5.1 Motivation

As discussed in Section 2.2 and Section 2.3, the ML estimator are consistent and asymptotically efficient provided that it exists under certain regularity conditions. The MGF

estimators are shown to be consistent as well and can always be found provided a well chosen initial point, with small bias but they have low efficiency compared with the ML estimators. However, in order to find the estimates, the numerical optimization required for these two methods can both be troublesome without an appropriate initial value. For computing the ML estimates, some convergence or existence problems always occur when $k \geq 1/2$ which has been identified as the non-regular case. On the other hand, for computing the MGF estimates, the computational problems mostly happen in the very heavy-tailed cases where $k \leq -2$.

We like the small biases and the availability over the entire parameter space of the maximum goodness-of-fit estimators. We also like the high efficiency of the ML estimators. We, however, do not want the convergence problems associated with both the maximum goodness-of-fit method and the ML method. Motivated by the idea to take advantage of both the MGF and the ML methods, we propose a new hybrid estimation method, which primarily relies on the MGF method to maintain the small bias and then improves the efficiency by incorporating the useful likelihood information. At the same time, the computational effort is also greatly reduced.

Under the reparameterization of $\theta = k/\sigma$ for the GPD, the maximum likelihood estimators of k and θ must satisfy (2.6)

$$k = -n^{-1} \sum_{i=1}^{n} \log(1 - \theta X_i) \ .$$

For any of the four EDF statistics $W^2(\sigma, k; x)$, $A^2(\sigma, k; x)$, $R^2(\sigma, k; x)$ and $L^2(\sigma, k; x)$, we can consider the reparameterized version and substitute the above relationship into it to have a simplified univariate minimization problem. In this thesis, we will only focus on the Anderson-Darling EDF statistic $A^2(\sigma, k; X)$ as recommended in Section 2.3.2, which has the most balanced performance among all the EDF statistics. Specifically, we

consider minimizing the target function G, which is the univariate minimization problem

$$\min_{\theta\in\mathscr{B}} G(\theta;X) = \min_{\sigma,k\in\mathscr{A}} A^2(\sigma,k;X)\ ,$$

where $\theta = k/\sigma$ and k is replaced by the right side of (2.6). Our new hybrid estimator $\hat{\theta}_{\text{HYB}}$ of θ is defined to be the value of θ at which $G(\theta;X)$ is minimized over the parameter space $\mathscr{B} = \{\theta < 1/X_{(n)}\}$.

2.5.2 Computing the New Hybrid Estimates

Given a sample $X = (X_1, X_2, \ldots, X_n)$ is from the GPD with the distribution function defined in (1.1). A target function G^* based on AD statistic can be written in a simple computational form

$$\begin{aligned} G^*(\theta;X) &= -n - \frac{1}{n}\sum_{i=1}^{n}\left\{(2i-1)\log\left[1-(1-\theta X_i)^{-n/\sum_j \log(1-\theta X_j)}\right]\right. \\ &\qquad \left. -n(2n+1-2i)\frac{\log(1-\theta X_i)}{\sum_j \log(1-\theta X_j)}\right\} \\ &= -n - \frac{1}{n}\sum_{i=1}^{n}\left\{(2i-1)\log\left[1-(1-\theta X_i)^{-n/g(\theta)}\right] - n(2n+1-2i)\frac{\log(1-\theta X_i)}{g(\theta)}\right\}, \end{aligned}$$

where $g(\theta) = \sum_{j=1}^{n}\log(1-\theta X_j)$.

In POT applications, special attention is needed for the small sample case. According to Pickands (1975), as the threshold t gets larger, the conditional distribution of $X - t$ given $X > t$ converges to the GPD. But the number of observations exceeding a given high threshold will decrease as the threshold increases, so the sample size of exceedences is usually small in the real applications. Our extensive simulation reveals an small but effective adjustment in the above $G^*(\theta;X)$ can keep the biases even smaller for small n, say $n \leq 50$, which is to replace the first n by $(n-0.5)$ to ensure that as n gets larger,

this adjustment vanishes. Then the adjusted target function G becomes

$$G(\theta;X) = -n - \frac{1}{n}\sum_{i=1}^{n}\left\{(2i-1)\log\left[1-(1-\theta X_i)^{-n/g(\theta)}\right] -(n-0.5)(2n+1-2i)\frac{\log(1-\theta X_i)}{g(\theta)}\right\}. \quad (2.16)$$

Consider the continuous function $G(\theta;X)$ defined in (2.16) and its first derivative given in Appendix A. On the parameter space $\mathscr{B} = \{\theta < 1/X_{(n)}\}$, the numerical search for the value of θ that minimizes $G(\theta;X)$ can be performed using the standard Newton-type algorithm or the bisection search algorithm. Hence by minimizing $G(\theta;X)$ with respect to θ subject to the boundary condition $\theta < 1/X_{(n)}$, the optimal $\hat{\theta}_{\text{HYB}}$ is obtained and it is always feasible. Finally, the new hybrid estimates $\hat{\sigma}_{\text{HYB}}$ and $\hat{k}_{\text{HYB}}$ can be calculated as

$$\hat{k}_{\text{HYB}} = -n^{-1}\sum_{i=1}^{n}\log(1-\hat{\theta}_{\text{HYB}}X_i) \quad \text{and} \quad \hat{\sigma}_{\text{HYB}} = \hat{k}_{\text{HYB}}/\hat{\theta}_{\text{HYB}}\,. \quad (2.17)$$

2.5.3 Inference

After getting the estimators of the GPD parameters, it is often useful to find the variances of the estimators for the purpose of constructing confidence intervals or testing statistical hypotheses. Because the new hybrid method combines both the maximum goodness-of-fit and the maximum likelihood methods, it seems not easy to derive the asymptotic variances of these new estimators. Fortunately, the bootstrap resampling method introduced by Efron (1977) provides us a reasonable, though computationally intensive, alternative to find useful approximations to the distributions of the new hybrid estimators. The bootstrap samples can be drawn directly from the data nonparametrically, or drawn parametrically from the GPD. Then based on the bootstrap samples we can calculate the standard errors of the new estimators.

Actually, the use of bootstrap method to find the standard error for other different estimators for the GPD has already been suggested by many authors, such as Castillo

and Hadi (1997) for their EPM estimates and Zhang and Stephens (2009) for their EBM estimates, etc. A reason for preferring the bootstrap method is that the confidence intervals obtained for the parameters can always make sense by satisfying the endpoint constraints. In Chapter 4, we will employ a real-world data set to illustrate the computational advantages of the new hybrid estimation method, together with the bootstrap standard errors and confidence intervals based on the new hybrid estimators.

In Appendix C, an R function called *gpdhyb* is provided for computing the new hybrid estimators of the GPD parameters. As an option, standard errors and confidence intervals are also provided using the parametric bootstrap method. The other optional arguments include a graphical tool to check the convergence of the minimization of the target function G, and the probability-probability (pp) plot and quantile-quantile (qq) plot as elementary Goodness-of-Fit assessments when fitting the GPD model to data.

Chapter 3

Simulation Study

In this Chapter a finite-sample Monte Carlo simulation study will be conducted to compare the performances of the new hybrid estimators proposed in Section 2.5 with other estimators. As the widely accepted criteria for evaluating the quality of an estimator, the estimation bias and mean squared error (MSE) are calculated for some finite sample sizes.

It is well known that the MOM and the PWM method have extremely poor performance unless $|k| < 1/2$. The EPM was compared with MGF method in Luceño (2006), but the author did not identify preferable estimators. The LME was studied in Zhang and Stephens (2009), and was shown to be no better than the EBM. However, the EBM was then outperformed by the improved EBM* in Zhang (2010), which performed as well as its first version in the common range but had significant improvements in the heavy-tailed cases. Following these facts, in this Chapter we will only consider the classical ML method, the MGF method and the improved EBM* in the finite-sample comparisons. More detailed tables summarizing the performances of all possible estimators are given in Appendix B.

3.1 Comparisons of Finite-Sample Performances

Our finite-sample simulation comparisons are based on $10,000$ random samples where each random sample is generated from the GPD(σ, k) with size $n = 20$, $n = 50$, $n = 100$ or $n = 200$. Without loss of generality, the scale parameter σ is taken to be 1 because the estimates for the GPD are invariant with respect to the values of σ. To be consistent

with the algorithms computing the ML and the MGF estimates, and to guarantee the minimum is reached, the tolerance level is set to be $\epsilon = 10^{-6}/\bar{X}$. The range of the shape parameter k considered in this Chapter is $-6 \leq k \leq 2$, which covers all the ranges used previously in the literature. For example, in Luceño (2006) the range of k used was $-2 \leq k \leq 2$ for the MGF method; in Zhang (2010) the range of k used was $-6 \leq k \leq 1/2$ for the EBM*, and also the commonly used range of k is $-1 < k < 1/2$; the non-regular range of k where the ML method has trouble is $k \geq 1/2$; the range of k where the GPD has infinite variance is $k < -1/2$.

Although there are many arguments to restrict the range of k to the most commonly used range $-1 < k < 1/2$, we have many practical and theoretical reasons to extend the range wider. First of all, as an important distribution encountered in a lot of applications, the uniform distribution is a special case of the GPD with $k = 1$. Furthermore, there are real-world examples supporting the range of $k > 1/2$, such as in Walshaw (1990) where an example was given to illustrate that an estimate of $k > 1/2$ is possible. Also in Castillo and Hadi (1997), two examples were provided and the corresponding estimates were shown to be $k > 1/2$ and $k > 1$. One of these examples, the Bilbao waves data, was adopted again in Luceño (2006) and also in Zhang and Stephens (2009) and the possibility of $k > 1/2$ was verified. Similarly, an estimate of $k \leq -1$ could be observed in many real-life examples, such as in heavy-tailed data and truncated data.

It is already known in Section 2.2.2 that the ML method may have severe convergence problems when $k \geq 1/2$, and has no solution when $k > 1$. In our simulation, these problems actually begin to happen as early as k approaches $1/2$. To deal with such unusual behavior of the ML method in simulation, Luceño (2006) and Zhang and Stephens (2009) gave two different treatments. In Zhang and Stephens (2009), the ML estimates of $(\hat{\sigma}_{\text{MLE}}, \hat{k}_{\text{MLE}})$ was replaced by $(X_{(n)}, 1)$ whenever the algorithm fails to converge or $\hat{k}_{\text{MLE}} > 1$. However, as a result this treatment may cause unnecessary reduction of the

standard errors of the estimates. In Luceño (2006), a quasi-maximum likelihood (QML) method was used which is a combination of the standard ML method and a modified ML method. The idea of this QML method is that as k gets larger, approximately we can assume $\hat{\sigma}_{\text{MLE}} = \hat{k}_{\text{MLE}} X_{(n)}$. Introduce this relation to the log-likelihood function defined in (2.3) based on the remaining sample $\left(X_{(1)}, \ldots, X_{(n-1)}\right)$, and maximizing this quasi log-likelihood with respect to k leads to

$$\tilde{k}_{\text{QML}} = -(n-1)^{-1} \sum_{i=1}^{n-1} \log\left(1 - \frac{X_{(i)}}{X_{(n)}}\right) \quad \text{and} \quad \tilde{\sigma}_{\text{QML}} = \tilde{k}_{\text{QML}} X_{(n)} \,.$$

Actually the treatment of Zhang and Stephens (2009) is just a special case of the QML of Luceño (2006). Therefore, in this section we prefer the QML method and will apply it in the simulation if the ML iterations do not converge or the estimated $\hat{k}_{\text{MLE}} > 1$.

3.1.1 Bias Comparisons

In measuring the accuracy of different estimators, the unbiasness is always desired. The biases for different estimators of σ and k are plotted against k in Figure 3.1 and Figure 3.2 for sample sizes $n = 20$, $n = 50$, $n = 100$ and $n = 200$. From Figure 3.1 and Figure 3.2, we see that the biases of $\hat{\sigma}_{\text{MLE}}$ and $\hat{k}_{\text{MLE}}$ are always positive and relatively larger compared with those of the other estimators. The biases of the EBM* estimators are generally small when k is around 0, and become larger as k gets smaller, and increase dramatically towards the negative side when k is greater than 0.5. The new hybrid method has significantly improved the estimation biases for σ and k, especially when compared with the MGF method and the ML method which supply the original ideas behind it.

3.1.2 MSE Comparisons

In evaluating the overall performance of different estimators, Luceño (2006) used the criterion of root mean squared error (RMSE) for the MGF method, while Zhang and

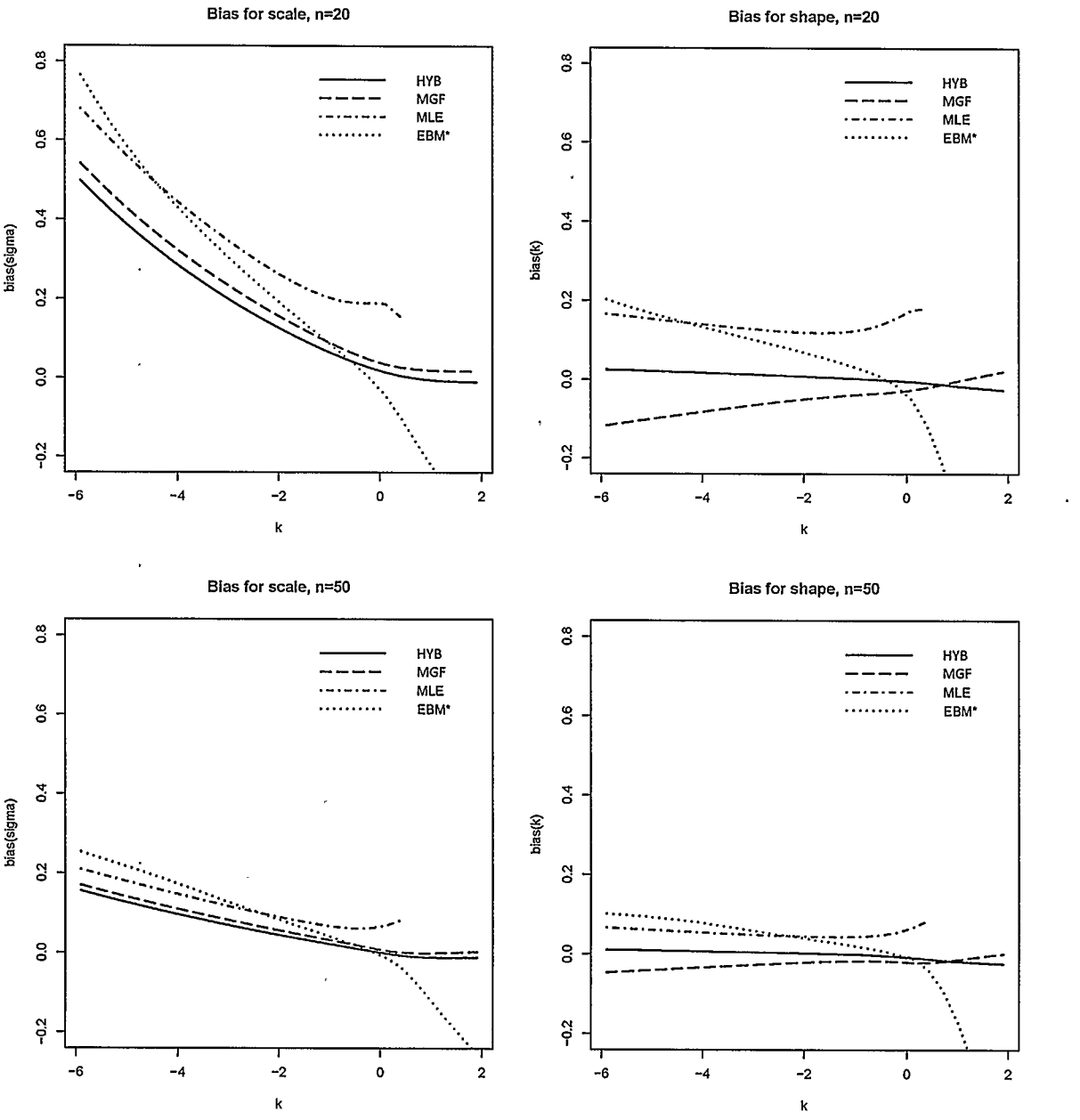

Figure 3.1: *Bias comparisons for estimating the GPD parameters, based on 10,000 replications with sample size* $n = 20$ *and* $n = 50$, *and* $\sigma = 1$.

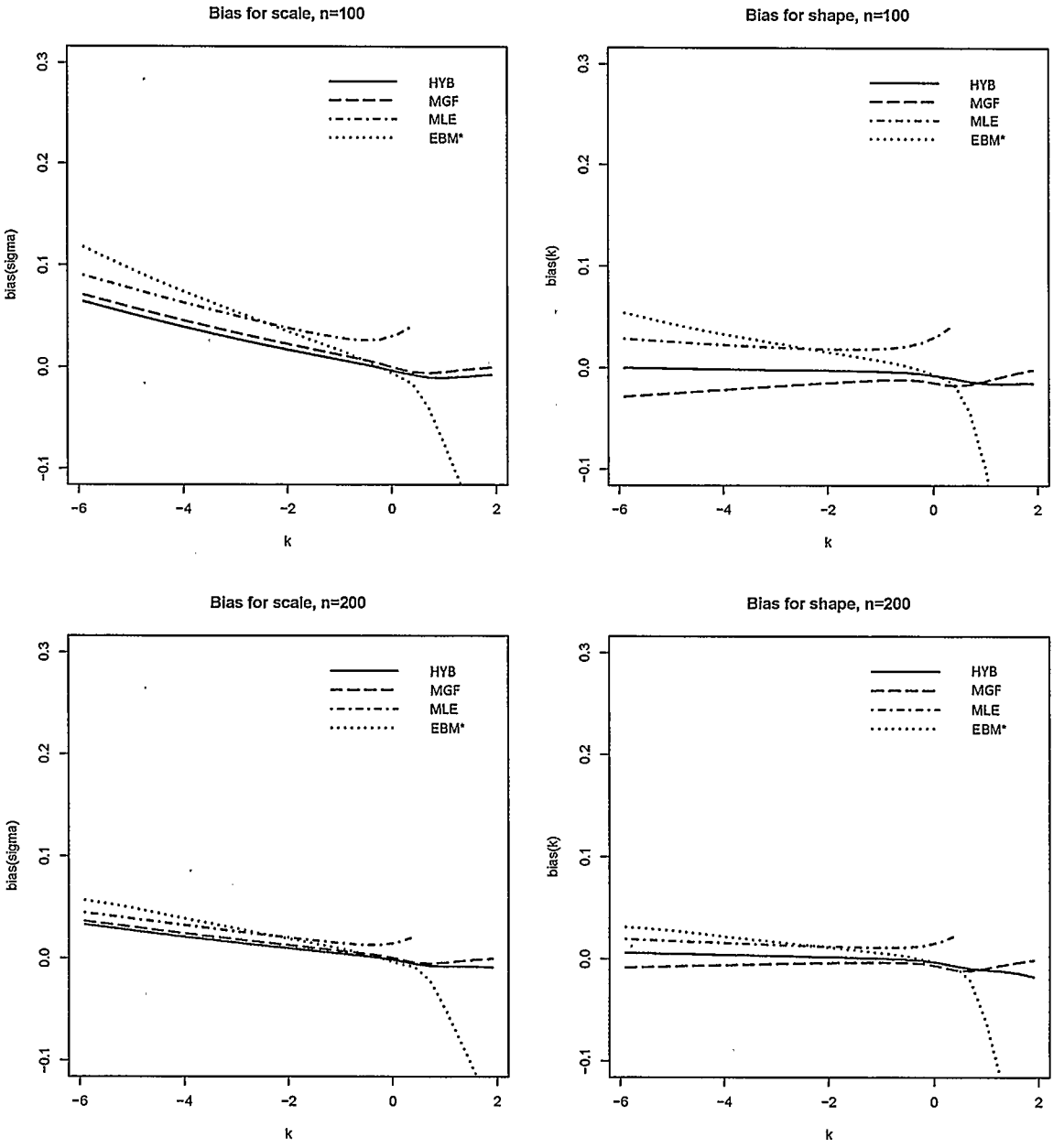

Figure 3.2: *Bias comparisons for estimating the GPD parameters, based on 10,000 replications with sample size* $n = 100$ *and* $n = 200$, *and* $\sigma = 1$.

Stephens (2009) and Zhang (2010) defined the relative efficiency for the EBM and EBM*, which is the MSE divided by the Cramér-Rao lower bound. In our simulation studies, the RMSEs are usually too close to each other to tell the differences graphically, and the Cramér-Rao lower bound does not exist for $k \geq 1/2$. So for our illustration purpose, we use the mean squared error (MSE) for comparing different estimators. The MSE is defined as the average squared difference between the estimates and the true parameter, which is a measure incorporating both the variability and accuracy of an estimator.

The MSEs for different estimators of σ and k are plotted against k in Figure 3.3 and Figure 3.4 for sample sizes $n = 20$, $n = 50$, $n = 100$ and $n = 200$. From Figure 3.3 and Figure 3.4, we see that the MSEs of the EBM* estimators are the smallest only when k falls in a small neighborhood of 0, and become larger than the rest estimators when $k > 1$. The new hybrid estimators always possess comparable MSEs, and improve over the ML method for estimating the scale parameter σ, and over the MGF method for estimating the shape parameter k.

In both the bias comparisons and the MSE comparisons, when the sample size increases from $n = 20$ to $n = 200$, the estimation bias and MSE for all estimators become smaller, implying that all estimators are consistent.

The new hybrid method totally outperforms the traditional ML method by always giving smaller biases and keeping the MSEs as good as those of the ML method. Besides, the new hybrid estimators are free from computational or existence problems in the non-regular case of $k \geq 1/2$.

Compared with the MGF method in Luceño (2006), which supplies the basis idea of our new hybrid method, the new hybrid method still improves over the MGF method in terms of both estimation bias and MSE in the range $-6 \leq k \leq 2$, especially for the estimation of the shape parameter k. Moreover, the computation of the new hybrid

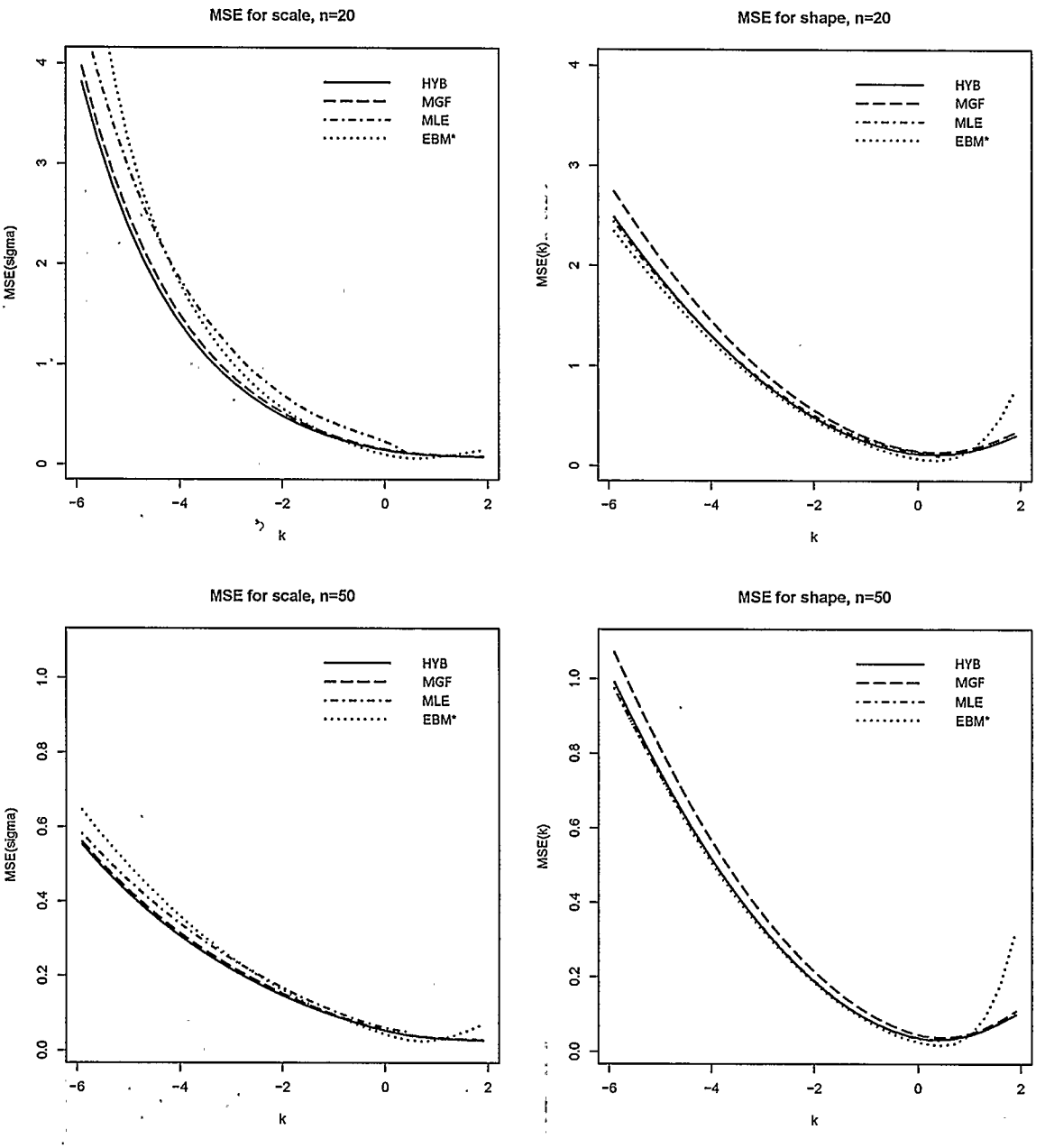

Figure 3.3: *MSE comparisons for estimating the GPD parameters, based on 10,000 replications with sample size* $n = 20$ *and* $n = 50$, *and* $\sigma = 1$.

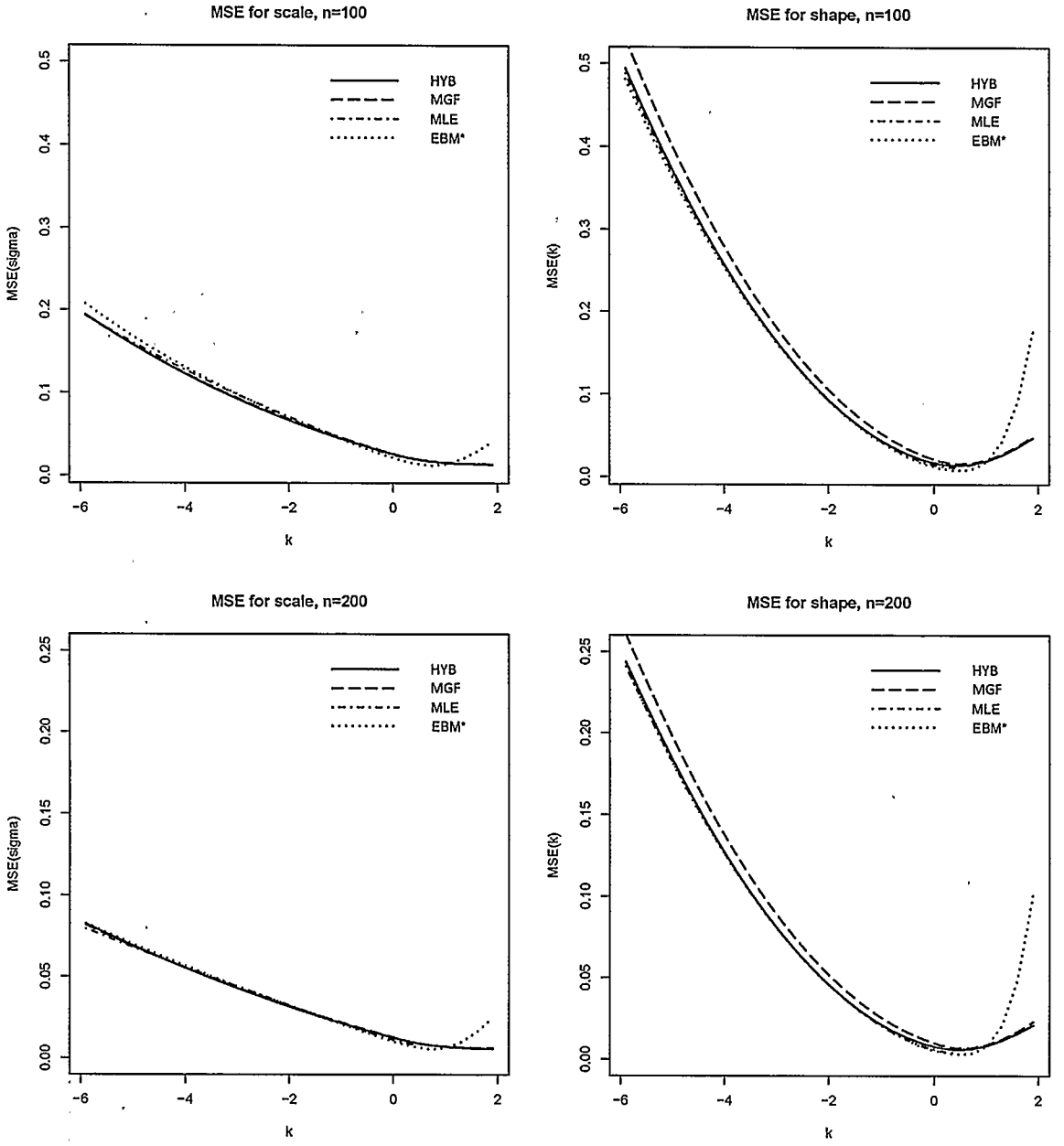

Figure 3.4: *MSE comparisons for estimating the GPD parameters, based on 10,000 replications with sample size* $n = 100$ *and* $n = 200$, *and* $\sigma = 1$.

estimators is much faster and tractable.

Finally, compared with the improved EBM* in Zhang (2010), although the bias and MSE curves cross each other, the new hybrid method generally has a better performance in most of the range $-6 \leq k \leq 2$, especially when n is small.

Chapter 4

An Example

To illustrate the advantages of the new hybrid estimation procedure, a real-world example will be presented in this Chapter. The data considered here is originally analyzed in Castillo and Hadi (1997), which consists of the zero-crossing hourly mean periods (in seconds) of the sea waves measured in a Bilbao buoy, Spain, in January, 1997. Later on, this data set was revisited in Luceño (2006) and in Zhang and Stephens (2009). One purpose of the study was to see the influence of periods on beach morphodynamics and other problems related to the right tail. Only the 197 observations with periods above 7 seconds were taken into consideration, and shown in Table 4.1.

Table 4.1: *The Bilbao waves data: the zero-crossing hourly mean periods (in seconds) of sea waves measured in the Bilbao bay, Spain, in January, 1997.*

7.05	7.12	7.15	7.18	7.19	7.20	7.20	7.20	7.20	7.25	7.26	7.27	7.28	7.30
7.31	7.31	7.32	7.33	7.37	7.40	7.46	7.46	7.47	7.48	7.48	7.52	7.54	7.55
7.55	7.58	7.59	7.59	7.61	7.63	7.65	7.66	7.66	7.67	7.67	7.68	7.69	7.72
7.72	7.72	7.72	7.72	7.77	7.77	7.79	7.79	7.82	7.83	7.83	7.83	7.84	7.85
7.85	7.88	7.88	7.90	7.90	7.91	7.93	7.93	7.93	7.94	7.95	7.95	7.97	7.97
7.97	7.99	8.00	8.03	8.03	8.05	8.06	8.06	8.07	8.10	8.11	8.12	8.15	8.15
8.15	8.18	8.18	8.18	8.19	8.20	8.21	8.23	8.23	8.30	8.30	8.31	8.31	8.32
8.32	8.33	8.40	8.41	8.42	8.43	8.43	8.45	8.48	8.49	8.50	8.50	8.51	8.52
8.53	8.54	8.56	8.58	8.59	8.59	8.60	8.65	8.69	8.71	8.72	8.74	8.74	8.74
8.74	8.79	8.81	8.84	8.85	8.86	8.88	8.88	8.94	8.98	8.98	8.99	9.01	9.03
9.06	9.12	9.16	9.17	9.17	9.18	9.18	9.18	9.21	9.22	9.23	9.24	9.27	9.29
9.30	9.32	9.33	9.36	9.38	9.43	9.46	9.47	9.59	9.59	9.60	9.61	9.62	9.63
9.66	9.74	9.75	9.78	9.79	9.79	9.80	9.84	9.85	9.89	9.90			

*The data is from Castillo and Hadi (1997), and only those observations above 7 seconds are listed.

It is reasonable to model this data set by the GPD because the data points are the exceedences over some fixed thresholds. To demonstrate the use of the new hybrid method, we will fit the data using thresholds at $t = 7.0, 7.5, 8.0, 8.5, 9.0$ and 9.5, following

the above mentioned authors. In Table 4.2, the parameters σ and k are estimated using different estimation methods (see also Table 4 in Castillo and Hadi (1997), Figure 3 in Luceño (2006) and Table 3 in Zhang and Stephens (2009)).

Table 4.2: *The estimated GPD parameters for the Bilbao waves exceedences at given thresholds using different estimation methods.*

		$\hat{\sigma}$						$\hat{k}$					
t	m	MOM	PWM	MLE	EBM*	MGF	HYB	MOM	PWM	MLE	EBM*	MGF	HYB
7.0	179	2.748	2.778	2.501	2.331	2.451	2.445	1.052	1.074	0.861	0.782	0.838	0.837
7.5	154	1.622	1.618	1.860	1.722	1.632	1.626	0.606	0.602	0.768	0.686	0.614	0.620
8.0	106	1.385	1.371	1.647	1.462	1.417	1.410	0.647	0.630	0.864	0.731	0.682	0.688
8.5	69	1.130	1.115	NA	1.146	1.176	1.168	0.722	0.700	NA	0.767	0.789	0.792
9.0	41	0.814	0.809	NA	0.756	0.846	0.837	0.833	0.823	NA	0.760	0.900	0.895
9.5	17	0.626	0.601	NA	0.361	0.521	0.507	1.709	1.601	NA	0.736	1.291	1.257

*m is the number of observations exceeding the given threshold t.

It is easy to see from Table 4.2 that all the estimation methods give estimates of k outside the commonly used range $-1 < k < 1/2$. From Chapter 2, we know that in such a non-regular range of k, the MOM and the PWM estimates are not reliable and may give infeasible results, and the ML estimates are also computationally unstable for the small sample size.

For the threshold at $t = 7.5$, the following Figure 4.1 shows the EDF of the Bilbao waves data, versus the fitted GPD cdf with the parameters estimated as in Table 4.2 using six different methods (see also Figure 2 in Luceño (2006) and Figure 4 in Zhang and Stephens (2009)). From the plots in Figure 4.1, we see that the new hybrid estimators give an overall good fit to the Bilbao data, which improve the fits using the ML estimators and the EBM* estimators.

To check graphically whether the minimum of the target function $G(\theta; X)$ defined in (2.16) is reached at $\hat{\theta}_{\text{HYB}} = \hat{k}_{\text{HYB}}/\hat{\sigma}_{\text{HYB}} = 0.3812$, the $G(\theta)$ and its first derivative are plotted in Figure 4.2 for threshold at $t = 7.5$. The boundary condition for this given data set is $\theta < 1/X_{(n)} = 1/2.4 = 0.4167$. In addition, Figure 4.3 shows the histograms of $B = 1000$ parametric bootstrap samples of $\hat{\sigma}_{\text{HYB}}$ and $\hat{k}_{\text{HYB}}$ for the Bilbao waves data at

the threshold $t = 7.5$. The parametric bootstrap standard errors for the hybrid estimates are calculated to be $se(\hat{\sigma}_{\text{HYB}}) = 0.167$ and $se(\hat{k}_{\text{HYB}}) = 0.090$, and the corresponding 95% bootstrap confidence intervals for σ and k are $(1.288, 1.949)$ based on $\hat{\sigma}_{\text{HYB}}$ and $(0.413, 0.771)$ based on $\hat{k}_{\text{HYB}}$ [1]. The bootstrap confidence interval for the GPD shape parameter k indicates that it is significantly different from 0 and 1, therefore there is no evidence that this data is from exponential or uniform distribution.

[1]This bootstrap standard errors and 95% confidence intervals can be obtained using the R function *gpdhyb* provided in Appendix C.

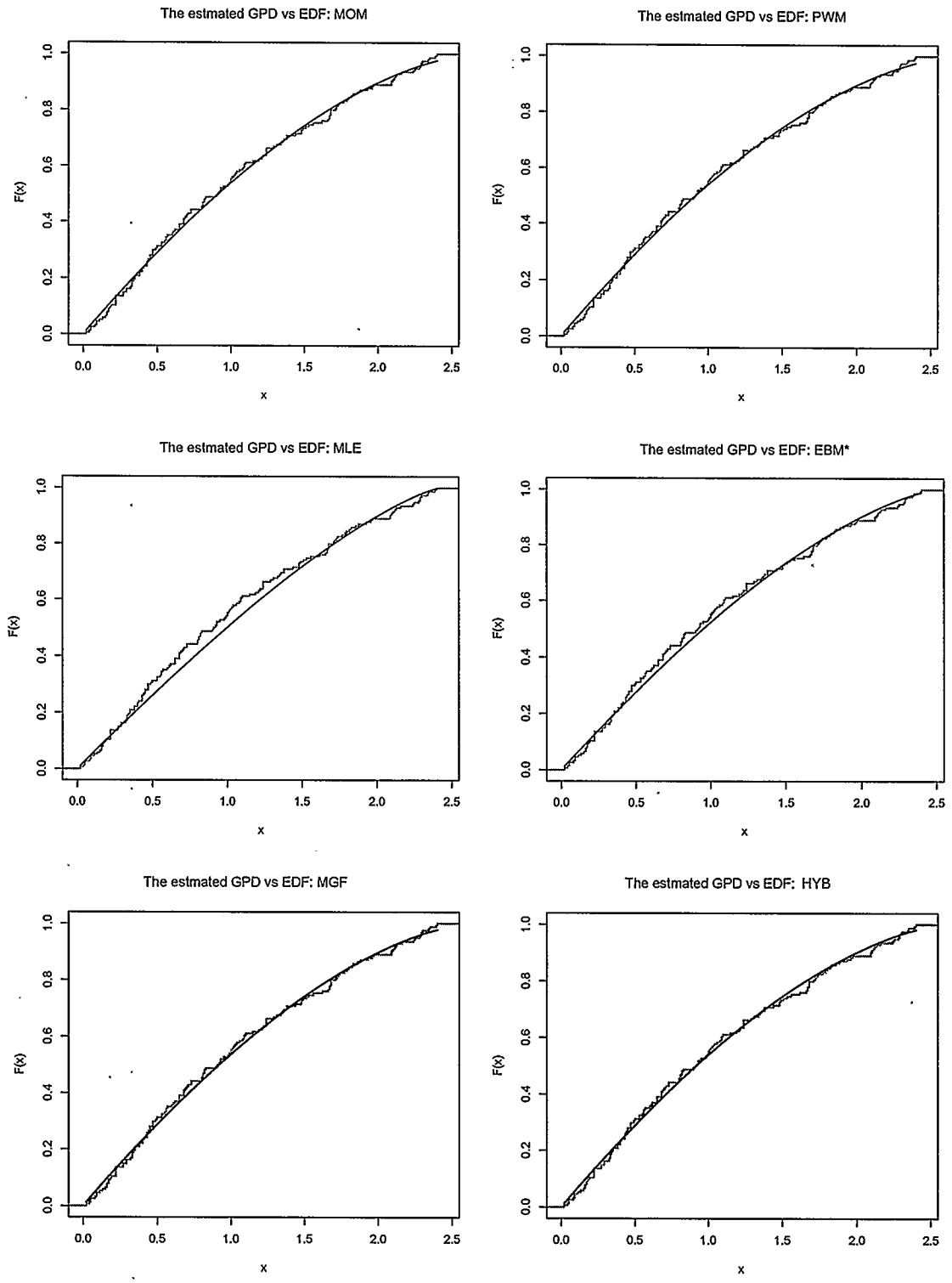

Figure 4.1: *The fitted GPD cdf using different methods versus the EDF for the Bilbao waves data at the threshold* $t = 7.5$.

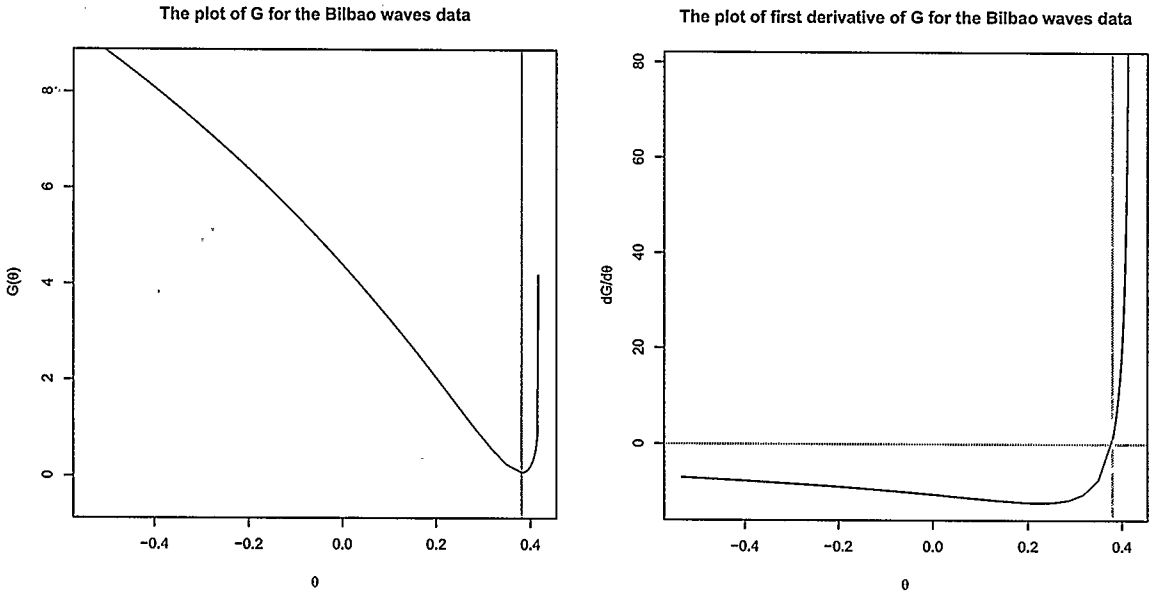

Figure 4.2: *The plot of $G(\theta; x)$ and its first derivative to check whether the minimum is reached for the Bilbao waves data at the threshold $t = 7.5$.*

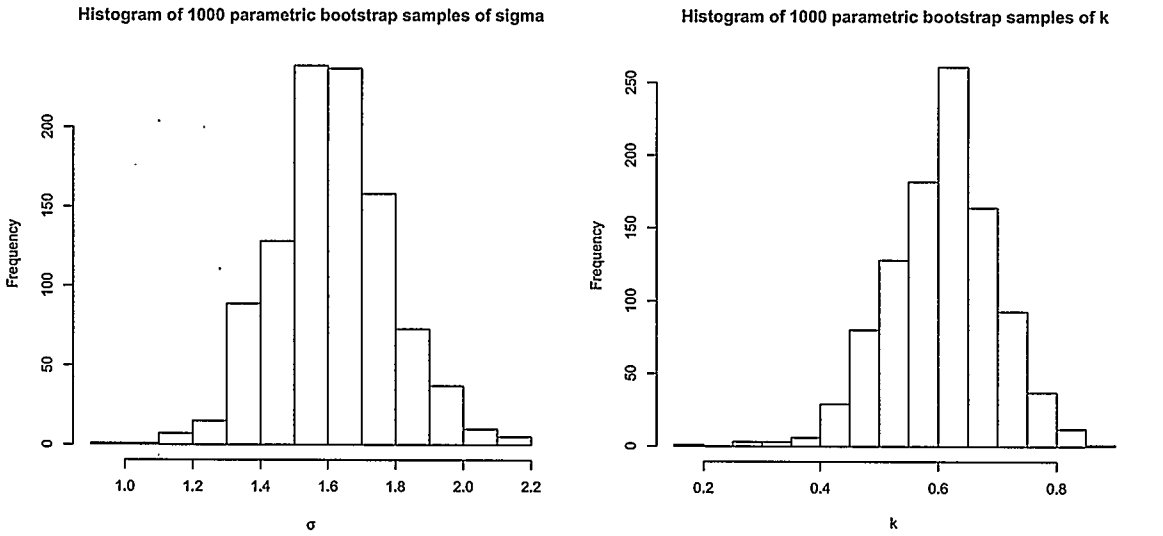

Figure 4.3: *Histograms of 1000 parametric bootstrap samples of $\hat{\sigma}_{HYB}$ and $\hat{k}_{HYB}$ for the Bilbao waves data at the threshold $t = 7.5$.*

Chapter 5

Final Conclusions and Future Research

There are many difficulties in dealing with the parameters estimation of the GPD, especially for the non-regular case of $k \geq 1/2$ where Cramer's regularity conditions are not satisfied. Besides, the difficulties also lie in the numerical optimization of complicated nonlinear functions. A lot of research has been done to overcome these theoretical and computational problems since the GPD was first introduced in 1975.

In this thesis, a new hybrid estimation method has been proposed for estimating the GPD parameters. It is based on the MGF method, but combines the reparameterization revealed by the maximum likelihood equations. This new hybrid estimation method has several advantages. First, the new hybrid method does not suffer from the convergence problem, and is easy and fast to implement by optimizing a single variable function using some standard algorithms, and the existence and feasibility of the hybrid estimates can even be verified graphically. Second, unlike some other existing methods, the new hybrid method can always provide valid estimates for the entire parameter space. Third, coupled with the parametric bootstrap method, our new hybrid method can allow practitioners to conduct reliable and accurate data inference using the generalized Pareto distribution. Although the proposed hybrid method in this thesis mainly focuses on the Anderson-Darling statistic among other EDF statistics, this hybrid idea can be similarly carried out for any other EDF statistic.

To evaluate the finite-sample performance of this new hybrid method and to compare it with other existing methods, we have carried out a simulation study in a wider range of shape parameter values $-6 \leq k \leq 2$. We have found that the proposed hybrid method can greatly improves over the ML method and the MGF method in terms of the

estimation bias and MSE, and that the new hybrid estimators are well compared with the EBM* estimators in most of the cases considered, although their bias and MSE cross each other for a small range of k around 0. In particular, the new hybrid estimators work well in the non-regular case, where the ML estimators have no solution and the EBM* estimators drastically underestimate the GPD parameters.

Finally, this proposed hybrid estimation procedure has been applied to analyze a real-world data set from Castillo and Hadi (1997).The results show that the new hybrid method is easy to use in practice and provides an overall good fit when fitting the GPD to real data.

A useful R program for computing the hybrid estimates and making statistical inference when using the GPD is given in Appendix C.

For future research, the following issues are worth to be worked on:

- Develop a large-sample theory for the proposed hybrid estimators.
- Prove the uniqueness of the new hybrid estimators, that is to prove the target function $G(\theta; X)$ defined in (2.16) and its first derivative given in the Appendix A satisfy the following properties on the parameter space $\mathscr{B} = \{\theta < 1/X_{(n)}\}$

  1.
  $$\lim_{\theta \to -\infty} G(\theta; X) = +\infty \quad \text{and} \quad \lim_{\theta \to 1/X_{(n)}} G(\theta; X) = +\infty ,$$

  2.
  $$\lim_{\theta \to -\infty} \frac{\partial G(\theta; X)}{\partial \theta} = 0^- \quad \text{and} \quad \lim_{\theta \to 1/X_{(n)}} \frac{\partial G(\theta; X)}{\partial \theta} = +\infty ,$$
  $$\frac{\partial G(\theta; X)}{\partial \theta} > 0 \text{ for } \theta \in (\hat{\theta},\ 1/X_{(n)}) \text{ and } \frac{\partial G(\theta; X)}{\partial \theta} < 0 \text{ for } \theta \in (-\infty,\ \hat{\theta}) .$$

- Explore the robustness of the new hybrid method according to the mechanism of contamination faced in real extreme value applications.

- Generalize the hybrid estimation method and apply it to some other distributions suffering from the finite parameter-dependent endpoint problem.
- Develop the idea of this hybrid method for the generalized linear model, where the GPD parameters to be estimated depend on some other covariates, for example, when spatial information is available.
- Improve the performances of the new hybrid estimators when k is around 0.

Appendix A

Derivative of the Target Function G

The first derivative of the target function $G(\theta; X)$ based on AD statistic defined in 2.16 in the section (2.5.2) is given by

$$\begin{aligned}\frac{\partial G(\theta;X)}{\partial \theta} &= -\frac{1}{n}\sum_{i=1}^{n}\Bigg\{(2i-1)\cdot\left[1-(1-\theta X_i)^{-n/g(\theta)}\right]^{-1}\Bigg[\left(\frac{-nX_i}{g(\theta)}\right)(1-\theta X_i)^{-n/g(\theta)-1} \\ &\quad -(1-\theta X_i)^{-n/g(\theta)}\cdot\log(1-\theta X_i)\cdot\frac{n\,g'(\theta)}{g(\theta)^2}\Bigg]-(n-0.5)(2n+1-2i)\cdot \\ &\quad \left[\frac{1}{g(\theta)^2}\cdot\left(\frac{-X_i}{1-\theta X_i}\cdot g(\theta)-\log(1-\theta X_i)\cdot g'(\theta)\right)\right]\Bigg\}\end{aligned}$$

where the

$$g(\theta)=\sum_{j=1}^{n}\log(1-\theta X_j) \qquad \text{and} \qquad g'(\theta)=\sum_{j=1}^{n}\frac{-X_j}{1-\theta X_j}$$

for simplification.

□

Appendix B

Finite-Sample Performances of Different Estimation Methods

The following two tables summarize the performances of several estimation methods for the GPD parameters that have been frequently referred in recent literature, together with the proposed new hybrid estimation method based on the ordinary Anderson-Darling statistic and the right-sided Anderson-Darling statistic. The simulated estimation bias and RMSE are from 10,000 time replications with sample size $n = 50$ and $n = 200$. The random seed used in the simulation is #2011. The scale parameter σ is fixed to be 1 due to its invariance property.

Table B.1: *Summary of the finite-sample performances for different estimation methods based on 10,000 times simulation with sample size $n = 50$, and $\sigma = 1$ fixed.*

							$\hat{\sigma}$					
Methods	MOM	PWM	MLE	LME	EBM	EBM*	CM	AD	ADR	ADL	$\mathbf{HYB}_{\mathrm{AD}}$	$\mathrm{HYB}_{\mathrm{ADR}}$
$k = -5$												
Bias(σ)	Inf	Inf	0.1787	0.4062	0.8552	0.2163	0.2061	0.1398	0.2309	0.1442	0.1258	0.1801
RMSE(σ)	(Inf)	(Inf)	(0.6747)	(1.1650)	(1.4204)	(0.7047)	(0.7735)	(0.6551)	(0.8933)	(0.6502)	(0.6489)	(0.8504)
$k = -3$												
Bias(σ)	Inf	Inf	0.1166	0.2022	0.2309	0.1267	0.1197	0.0815	0.1207	0.0905	0.0688	0.0880
RMSE(σ)	(Inf)	(Inf)	(0.4945)	(0.6418)	(0.5694)	(0.4978)	(0.5190)	(0.4730)	(0.5518)	(0.4901)	(0.4657)	(0.5309)
$k = -1$												
Bias(σ)	7.1646	0.2704	0.0667	0.0770	0.0581	0.0408	0.0570	0.0317	0.0446	0.0452	0.0217	0.0271
RMSE(σ)	(196.93)	(0.4570)	(0.3252)	(0.3282)	(0.3095)	(0.3078)	(0.3364)	(0.3115)	(0.3178)	(0.3505)	(0.3057)	(0.3114)
$k = -0.5$												
Bias(σ)	0.3919	0.0503	0.0607	0.0557	0.0165	0.0184	0.0450	0.0195	0.0294	0.0358	0.0109	0.0154
RMSE(σ)	(0.6334)	(0.2562)	(0.2835)	(0.2720)	(0.2544)	(0.2587)	(0.2985)	(0.2716)	(0.2691)	(0.3189)	(0.2672)	(0.2657)
$k = 0.2$												
Bias(σ)	0.0290	0.0125	0.0717	0.0317	-0.0236	-0.0181	0.0319	0.0029	0.0116	0.0250	-0.0036	0.0018
RMSE(σ)	(0.2097)	(0.2269)	(0.2338)	(0.2101)	(0.1945)	(0.1889)	(0.2523)	(0.2181)	(0.2092)	(0.2779)	(0.2162)	(0.2088)
$k = 1.5$												
Bias(σ)	0.0362	0.0203	NA	0.0077	-0.0756	-0.1969	0.0236	0.0005	0.0066	0.0257	-0.0123	-0.0060
RMSE(σ)	(0.2776)	(0.2515)	(NA)	(0.1645)	(0.1457)	(0.2176)	(0.2033)	(0.1739)	(0.1630)	(0.2379)	(0.1697)	(0.1611)
							$\hat{k}$					
$k = -5$												
Bias(k)	4.5146	4.0097	0.0607	0.1277	0.4896	0.0922	0.0409	-0.0402	-0.0126	-0.0038	0.0090	0.0066
RMSE(k)	(4.5146)	(4.0097)	(0.8589)	(0.9634)	(0.8331)	(0.8570)	(0.9629)	(0.9006)	(0.9064)	(1.0474)	(0.8640)	(0.9063)
$k = -3$												
Bias(k)	2.5202	2.0266	0.0490	0.0895	0.1493	0.0591	0.0326	-0.0268	-0.0056	-0.0017	0.0047	0.0054
RMSE(k)	(2.5203)	(2.0269)	(0.5754)	(0.6222)	(0.5533)	(0.5677)	(0.6565)	(0.6056)	(0.5995)	(0.7392)	(0.5746)	(0.5968)
$k = -1$												
Bias(k)	0.5950	0.2818	0.0440	0.0538	0.0387	0.0205	0.0264	-0.0171	-0.0004	0.0009	-0.0009	0.0042
RMSE(k)	(0.5996)	(0.3234)	(0.2981)	(0.2953)	(0.2778)	(0.2878)	(0.3820)	(0.3260)	(0.3044)	(0.4679)	(0.2977)	(0.2968)
$k = -0.5$												
Bias(k)	0.2243	0.0742	0.0482	0.0453	0.0076	0.0087	0.0255	-0.0174	-0.0010	0.0027	-0.0035	0.0030
RMSE(k)	(0.2542)	(0.1934)	(0.2348)	(0.2237)	(0.2131)	(0.2196)	(0.3282)	(0.2633)	(0.2378)	(0.4157)	(0.2373)	(0.2298)
$k = 0.2$												
Bias(k)	0.0293	0.0090	0.0714	0.0323	-0.0245	-0.0174	0.0266	-0.0215	-0.0036	0.0073	-0.0098	-0.0002
RMSE(k)	(0.1567)	(0.1872)	(0.1750)	(0.1574)	(0.1493)	(0.1409)	(0.2794)	(0.1985)	(0.1702)	(0.3655)	(0.1810)	(0.1650)
$k = 1.5$												
Bias(k)	0.0615	0.0238	NA	0.0121	-0.1247	-0.3565	0.0432	-0.0053	0.0083	0.0525	-0.0219	-0.0093
RMSE(k)	(0.5137)	(0.4517)	(NA)	(0.2516)	(0.2263)	(0.3818)	(0.3426)	(0.2751)	(0.2513)	(0.4455)	(0.2654)	(0.2474)

Table B.2: *Summary of the finite-sample performances for different estimation methods based on 10,000 times simulation with sample size $n = 200$, and $\sigma = 1$ fixed.*

	$\hat{\sigma}$											
Methods	MOM	PWM	MLE	LME	EBM	EBM*	CM	AD	ADR	ADL	$\mathbf{HYB_{AD}}$	$\mathrm{HYB_{ADR}}$
$k = -5$												
Bias(σ)	Inf	Inf	0.0386	0.0865	0.3013	0.0490	0.0437	0.0304	0.0487	0.0312	0.0272	0.0376
RMSE(σ)	(Inf)	(Inf)	(0.2607)	(0.3848)	(0.4635)	(0.2646)	(0.2808)	(0.2619)	(0.3128)	(0.2605)	(0.2616)	(0.3079)
$k = -3$												
Bias(σ)	Inf	Inf	0.0261	0.0458	0.0650	0.0289	0.0267	0.0183	0.0272	0.0200	0.0152	0.0195
RMSE(σ)	(Inf)	(Inf)	(0.2093)	(0.2557)	(0.2296)	(0.2099)	(0.2169)	(0.2076)	(0.2287)	(0.2138)	(0.2068)	(0.2262)
$k = -1$												
Bias(σ)	7.3732	0.1948	0.0148	0.0177	0.0139	0.0094	0.0130	0.0071	0.0103	0.0100	0.0045	0.0059
RMSE(σ)	(111.48)	(0.2616)	(0.1463)	(0.1486)	(0.1448)	(0.1447)	(0.1553)	(0.1482)	(0.1487)	(0.1641)	(0.1465)	(0.1479)
$k = -0.5$												
Bias(σ)	0.2736	0.0178	0.0129	0.0126	0.0026	0.0038	0.0101	0.0042	0.0066	0.0076	0.0019	0.0029
RMSE(σ)	(0.3591)	(0.1260)	(0.1268)	(0.1258)	(0.1236)	(0.1243)	(0.1402)	(0.1317)	(0.1287)	(0.1509)	(0.1299)	(0.1282)
$k = 0.2$												
Bias(σ)	0.0063	0.0022	0.0170	0.0060	-0.0092	-0.0062	0.0066	-0.0020	-0.0002	0.0047	-0.0034	-0.0024
RMSE(σ)	(0.0995)	(0.1090)	(0.0979)	(0.0982)	(0.0932)	(0.0930)	(0.1194)	(0.1053)	(0.0997)	(0.1319)	(0.1048)	(0.1002)
$k = 1.5$												
Bias(σ)	0.0074	0.0039	NA	-0.0033	-0.0673	-0.1037	0.0039	-0.0019	-0.0007	0.0027	-0.0085	-0.0055
RMSE(σ)	(0.1249)	(0.1181)	(NA)	(0.0770)	(0.0902)	(0.1171)	(0.0920)	(0.0812)	(0.0767)	(0.1042)	(0.0786)	(0.0759)
	$\hat{k}$											
$k = -5$												
Bias(k)	4.5036	4.0024	0.0176	0.0343	0.2350	0.0276	0.0122	-0.0073	-0.0001	0.0011	0.0050	0.0045
RMSE(k)	(4.5036)	(4.0024)	(0.4265)	(0.4821)	(0.3930)	(0.4257)	(0.4769)	(0.4441)	(0.4474)	(0.5182)	(0.4284)	(0.4482)
$k = -3$												
Bias(k)	2.5050	2.0067	0.0137	0.0237	0.0497	0.0165	0.0094	-0.0049	0.0006	0.0008	0.0029	0.0032
RMSE(k)	(2.5050)	(2.0068)	(0.2845)	(0.3108)	(0.3004)	(0.2836)	(0.3243)	(0.2983)	(0.2962)	(0.3638)	(0.2844)	(0.2950)
$k = -1$												
Bias(k)	0.5373	0.2002	0.0111	0.0137	0.0104	0.0058	0.0069	-0.0034	0.0007	0.0004	0.0004	0.0016
RMSE(k)	(0.5382)	(0.2234)	(0.1438)	(0.1453)	(0.1412)	(0.1427)	(0.1866)	(0.1598)	(0.1507)	(0.2263)	(0.1458)	(0.1460)
$k = -0.5$												
Bias(k)	0.1476	0.0282	0.0114	0.0112	0.0013	0.0024	0.0063	-0.0037	0.0001	0.0003	-0.0007	0.0007
RMSE(k)	(0.1643)	(0.1093)	(0.1094)	(0.1080)	(0.1066)	(0.1076)	(0.1586)	(0.1286)	(0.1177)	(0.1980)	(0.1148)	(0.1119)
$k = 0.2$												
Bias(k)	0.0068	0.0019	0.0182	0.0065	-0.0083	-0.0051	0.0055	-0.0085	-0.0041	0.0005	-0.0048	-0.0027
RMSE(k)	(0.0730)	(0.0897)	(0.0685)	(0.0703)	(0.0651)	(0.0644)	(0.1301)	(0.0902)	(0.0785)	(0.1671)	(0.0814)	(0.0745)
$k = 1.5$												
Bias(k)	0.0125	0.0041	NA	-0.0055	-0.1104	-0.1752	0.0072	-0.0043	-0.0018	0.0046	-0.0139	-0.0087
RMSE(k)	(0.2318)	(0.2132)	(NA)	(0.1167)	(0.1425)	(0.1929)	(0.1482)	(0.1241)	(0.1162)	(0.1764)	(0.1197)	(0.1148)

Appendix C

An *R* Function for Computing the New Hybrid Estimates

```
gpdhyb <- function(data, threshold, start, std.err=FALSE,
          conv.plot=FALSE, ppqq=FALSE){

 if(missing(threshold)){ threshold = 0 }
  exceed <- (data > threshold)
     dat <- data[exceed]-threshold;  m <- length(dat)

## define the main function G depends on single parameter \theta;
G <- function(dat, theta){
     x <- sort(dat); n <- length(x)
     c <- 1-theta*x; d <- 1-theta*rev(x)

# based on adjusted AD statistic;
dis <- -n-mean((2*(1:n)-1)*(log(1-c^(-n/sum(log(c))))-(n-0.5)*log(d)/sum(log(c))))
 return(dis)
}

if(missing(start)){ start = -0.5 }
## constrained minimization of the target G to obtain the estimate of \theta;
# the 'optimize' function can be alternative which leads to almost same result;
   e <- 1e-6/mean(dat)
```

```
 min <- nlminb(start, G, upper=1/max(dat)-e, control=list(abs.tol=e), dat=dat)

# based on minimum of \theta to get the final estimates of (\sigma,k);
           k <- -mean(log(1-min$par*dat))
         par <- c(scale=k/min$par, shape=k)
 convergence <- min$convergence
        iter <- min$iterations

## graphical assessment of the algorithm convergence;
if(conv.plot==T){
   t <- c(seq(min$par-median(dat), min$par,length=30),
          seq(min$par, 1/max(dat)-(1e-12),length=30) )
 out <- NULL
for(j in 1:length(t)){
 out <- c(out, G(dat, t[j]))
}
X11()
plot(t, out, type="l", ylim=c(min(out), quantile(out,0.9)), xlab="theta",
     ylab="G(theta)", main="The plot of G to assess if the minimum is reached")
 abline(v=min$par, col=8)
}

## the p-p and q-q plot to assess the fit of GPD to data using the hybrid method;
if(ppqq==T){
 n <- length(dat);  x <- sort(dat)
 Sample <- pgpd(x, ,par[1], -par[2])
```

```
 Empirical <- NULL
 for(j in 1:n){ Empirical[j] <- j/(n+1)  }
 X11(); par(mfrow=c(1,2))
  plot(Empirical, Sample); title(main="Probability Plot of the GPD")
  abline(0,1, col="red")

 Sample <- x
 Empirical <- NULL
 for(j in 1:n){ Empirical[j] <- qgpd(j/(n+1),, par[1], -par[2])  }
  plot(Empirical, Sample); title(main="Quantile Plot of the GPD")
  abline(0,1, col="red")
}

## bootstrap standard error, and confidence intervals;
if(std.err==T){
   B <- 1000    # number of independent bootstrap samples;
 b.se <-NULL
 for(j in 1:B){
      b <- rgpd(m, ,par[1], -par[2])
  b.min <- nlminb(min$par, G, upper=1/max(b)-e,
                  control=list(abs.tol=e), dat=b)
    b.k <- -mean(log(1-b.min$par*b))
  b.par <- c(b.k/b.min$par, b.k)
   b.se <- rbind(b.se, b.par)
 }
```

```
se <- apply(b.se, 2, sd);  names(se) <- c("se.scale", "se.shape")
CI <- rbind(quantile(b.se[,1], c(0.025, 0.975)),
            quantile(b.se[,2], c(0.025, 0.975)))
names <- c("95%CI.scale", "95%CI.shape")
dimnames(CI) <- list(names, c("2.5%", "97.5%"))
list(param=par, std.err=se, CIs=CI, convergence=convergence, iterations=iter)
}

else{
list(param=par, convergence=convergence, iterations=iter)
}
}
```

□

Bibliography

[1] Anderson, T. W. and Darling, D. A. (1954). A Test for Goodness-of-Fit. *Journal of the American Statistical Association*, **49**, 300–310.

[2] Bolthausen, E. (1977). Convergence in Distribution of Minimum-Distance Estimators. *Metrika*, **24**, 215–227.

[3] Castillo, E. and Hadi, A. S. (1997). Fitting the Generalized Pareto Distribution to Data. *Journal of the American Statistical Association*, **92**, 1609–1620.

[4] Chen, G. and Balakrishnan, N. (1995). The Infeasibility of Probability Weighted Moments Estimation of Some Generalized Distribution. *Recent Advances in Life-Testing and Reliability*, ed. N. Balakrishnan, London: CRC Press, pp. 565-573.

[5] Choulakian, V. and Stephens, M. A. (2001). Goodness-of-Fit Tests for the Generalized Pareto Distribution. *Technometrics*, **43**, 478–484.

[6] Coles, S. G. (2001). An Introduction to Statistical Modeling of Extreme Values. London: Springer.

[7] Davison, A. C. (1984). Modeling Excesses Over High Thresholds, With an Application. *Statistical Extremes and Applications*, ed. J. Tiago de Oliveira, Dordrecht: Reidel, pp. 461–482.

[8] Davison, A. C., and Smith, R. L. (1990). Models for Exceedances Over High Thresholds (with comments). *Journal of the Royal Statistical Society, Series B*, **52**, 393–442.

[9] DuMouchel, W. (1983). Estimating the Stable Index α in Order to Measure Tail Thickness: A Critique. *The Annals of Statistics*, **11**, 1019–1036.

[10] Dupuis, D. J., and Tsao, M. (1998). A Hybrid Estimator for Generalized Pareto and Extreme-Value Distrubutions. *Communications in Statistics-Theory and Methods*, **27**, 925–941.

[11] Fisher, R. A., and Tippett, L. H. C. (1928). Limiting Forms of the Frequency of the Largest or Smallest Member of a Sample. *Proceedings of the Cambridge Philosophical Society*, **24**, 180–190.

[12] Grimshaw, S. D. (1993). Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution. *Technometrics*, **35**, 185–191.

[13] Hosking, J. R. M and Wallis, J. R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, **29**, 339–349.

[14] Luceño, A. (2006). Fitting the Generalized Pareto Distribution to Data Using Maximum Goodness-of-Fit Estimators. *Computational Statistics & Data Analysis*, **51**, 904–917.

[15] Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, **3**, 119–131.

[16] Pollard, D. (1980). The Minimum Distance Method of Testing. *Metrika*, **27**, 43–70.

[17] R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing.
Available at `http://www.R-project.org`.

[18] Smith, R. L. (1984). Threshold Methods for Sample Extremes. *Statistical Extremes and Applications*, ed. J. Tiago de Oliveira, Dordrecht: Reidel, pp. 621–638.

[19] Smith, R. L. (1985). Maximum Likelihood Estimation in a Class of Non-regular Cases. *Biometrika*, **72**, 67–90.

[20] Smith, R. L. (1989). Extreme Value Analysis of Environmental Time Series: an Application to Trend Detection in Group-Level Ozone. *Statistical Science*, 4, 367–393.

[21] Stephens, M. A. (1986). Tests Based on EDF Statistics. in *Goodness-of-fit Technique*, eds. R. B. D'Agostino and M. A. Stephens, New York: Marcel Dekker, pp. 97–122.

[22] Walshaw, D. (1990). Discussion of "Models for Exceedances Over High Thresholds." by Davison, A. C. and Smith, R. L.. , *Journal of the Royal Statistical Society, Series B*, **52**, 393–442.

[23] Wolfowitz, J. (1953). Estimation by the Minimum Distance Method. *Annals of the Institute of Statistical Mathematics*, **5**, 9–23.

[24] Wolfowitz, J. (1957). The Mimimum Distance Method. *Annals of Mathematical Statistics*, **28**, 75–88.

[25] Zhang, J. and Stephens, M. A. (2009). A New and Efficient Estimation Method for the Generalized Pareto Distribution. *Technometrics*, **51**, 316–325.

[26] Zhang, J. (2007). Likelihood Moment Estimation for the Generalized Pareto Distribution. *Australian and New Zealand Journal of Statistics*, **49**, 69–77.

[27] Zhang, J. (2010). Improving on Estimation for the Generalized Pareto Distribution. *Technometrics*, **52**, 335–339.