People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

Faculty of Exact Sciences, Natural Sciences and Life Sciences

Laboratory of Applied Mathematics

Department of Mathematics

A Third Cycle Doctoral Thesis

Presented for the Degree of

DOCTOR in Applied Mathematics

In the field of STATISTICS

By

KHEMISSI Zahia

Title :

On the Estimation of the Distribution Tail Index

Examination Committee :

Mr.	**YAHIA Djabrane**	Professor	University of Biskra	President
Mr.	**BRAHIMI Brahim**	Professor	University of Biskra	Supervisor
Mr.	**BENATIA Fatah**	Professor	University of Biskra	Co.Supervisor
Mrs.	**ABDELLI Jihane**	M.C.(A)	University of Biskra	Examiner
Mr.	**AISSAOUI Adel**	Professor	University of El-Oued	Examiner

March 2023

Dedication

I DEDICATE THIS EFFORT TO,

My Parents, for their love and encouragement throughout my studies May God bless them all with good health and a full life. . .

My husband, who has been my loyal support throughout all my endeavours.

My sons TAMIME & MOEZ

My whole family, my sisters and brothers, friends and to all. who helped me in any way, no matter how small or large their contribution is. . .

I SAY ,THANK YOU

Acknowledgments

I address my thanks to my *GOD* who gave me the patience and courage that allowed me to complete my thesis.

First and foremost, I would like to express my deepest sense of gratitude to my supervisor, Pr. *BRAHIMI BRAHIM* and Pr. *BENATIA FATAH* for the instructions and guidance he gave me and for accompanying me throughout the academic course as well as the moral supports. I do not have enough words to express my deep and sincere appreciation. Thank you and the best of luck in your future endeavors.

I would like to thank the examination committee members for having devoted some of their precious time to very careful readings of this thesis and for their comments, criticisms and suggestions

Scientific Contributions

I) Paper :

Khemissi, Z., Brahimi.B., Benatia, F. (2022). *Heavy tail index estimator through Weighted Least Squares rank regression*. Journal of Siberian Federal University Mathematics and Physics , $15(6), 797 - 805$.

II) Conferences paper :

1. Khemissi, Z., Brahimi.B., Benatia, F. (October $28 - 29, 2019$). *Comparison of three estimation methods for Frechet distribution*. International Conference on Financial Mathematics: Tools and Applications (MFOA'2019), University of Bedjaia, Algeria.

2. Khemissi, Z. (December $15-19, 2019$). *Estimate the weighted least squares using simple weights*. International Conference on Advances in Applied Mathematics (ICAAM'2019), University of Sfax, Tunis.

3. Khemissi, Z., Brahimi.B., Benatia, F. (April $10, 2019$). *Index estimator through weighted least-squares rank regression*. National Conference of Applied Mathematics (JMA'19), University of Ziane Achour, Djelfa, Algeria.

4. Khemissi, Z., Brahimi.B., Benatia, F. (April $28, 2019$). *The Estimation of Frechet Distribution by a Weighted Least Square Method*. National Conference of Applied Mathematics (JMA'19), University of Abdelhafid Boussouf, Mila, Algeria.

5. Khemissi, Z., Brahimi.B. (March $03, 2018$). *Index estimator through weighted least-squares rank regression*. Journée de Mathématiques Appliquées, University of Mohamed Khider, Biskra, Algeria.

Abstract

This thesis is devoted to the study of a regression estimator for estimating the tail index of the heavy-tailed distribution. In particular, it is shown that the considered estimator is in general based on the method of weighted least squares.

The main objective of the thesis is extend the work of Zyl and schall, 2012, *for estimating the shape parameter of the Frechet distribution. By deriving the large sample variances and using the inverse of the approximate variance to calculate the weights for this estimator.*

Simulation study using R statistical software is carried out to evaluate performance of a new estimator wich has been shown to perform better than other considered methods estimator based on order statistics for small and large sample size, and in case of real data.

Keywords: Extreme value Theory, Extreme value index, Heavy-tails, Least squares estimator, Weighted least squares, Rank regression, Frechet distribution.

Abbreviations and Notations

$(\Omega, \mathcal{A}, \mathcal{P})$	probability space
rv	random variable
X	rv dened on $(\Omega,\mathcal{A},\mathcal{P})$, population
$\mathbb{E}[X]$	expectation of (or mean of X)
$Var[X]$	variance of X
pdf	probability density function
df	distribution function
F_n	empirical df
$F^{\leftarrow}$	generalized inverse of F,quantile function
F	df of X
f	pdf of X
$\mathcal{Q}$	quantile function,generalized inverse of X
$\mathcal{Q}_n$	empirical quantile function
$X_{1,n} \leq ... \leq X_{n,n}$	order statistics pertaining to the sample $(X_1, ..., X_n)$
k	numbers of top statistics (upper observations)
x_F	upper endpoint
EVI	extreme value index
EVT	extreme value theory
$GEVD$	generalized extreme value distribution
GPD	generalized Pareto distribution

$\mathcal{D}(.)$	domain of attraction
$\mathcal{RV}_\alpha$	regular variation at ∞ with index α
$\mathcal{RV}_0$	regular variation at 0 with index α
$\xrightarrow{a.s}$	almost sure convergence
$\xrightarrow{P}$	convergence in probability
$\xrightarrow{d}$	convergence in distribution
iid	independent identically distributed
$i.e.$	in other words
$\mathcal{N}(\mu, \delta^2)$	normal or Gaussian distribution
α	tail index
γ	extreme value index
MLE	maximum likelihood estimator
$\exp$ or e	exponential
$\log$	logarithm
LSE	least squares estimator
WLS	weighted least squares
$MPSE$	maximum product of spacings estimation
$RMSE$	root mean squared error

Contents

List of Figures

List of Tables

Introduction

In statistical studies, the excessive behavior of any phenomenon is the only element of interest because we can encounter many cases of extreme situations. Most of the extreme events, some of which were catastrophic, caused widespread damage, for example, *fires*, *earthquakes* and *extreme weather conditions* (*extremely low or high*), along with the virus that swept the world and was an extreme case that caused significant deaths around the world. To counter these extreme situations mentioned above, *Extreme Value Theory* could give a great help.

In particular, extreme value theory, also known as tail distributions, plays an increasingly important role in the treatment of rare event modeling, as it focuses on the tail of the distribution that generates the studied various extreme phenomenon. EVT, is mainly depends on the limit distributions of the extremes and their domain of attractions, and there are two models : Generalized Extreme Value Distributions ($GEVD$) and Generalized Pareto Distributions (GPD).

Thus, it all began with the development of this theory for the first time in the context of independent observations, the preliminary work of Frechet, Fisher and Tippett [24], 1928 shows that, under certain conditions, only limited distributions to the extremes are the distributions of Frechet, Gumbel and Weibull. This allows us to classify most distribution into three domains of attraction where each domain is identified by the descriptions on the distribution functions (see Embrechts et al., [22], 1997; De Haan and Ferreira [16], 2006). A parametrization of the three limiting behaviors into a single distribution, namely the

$GEVD$ is due to Von Mises [56], 1954. One of the important parameters is the index of extreme values (tail index) which describes the heaviness of the tail of the distribution.

In extreme value statistics, estimating the tail index parameter is one of the primal objectives in EVT. For heavy-tailed distributions the Hill estimator [34], 1975 is the most popular way to estimate the tail index parameter, and the Hill estimator has also been improved by recent work in various ways, for example the double bootstrap method by Danielsson et al., [12] 2001, or a model based on Kolmogorov-Smirnov distance by Danielsson et al., [13] 2016. In this memory, we are interested in the study of estimation in the field of regression. The term "*linear regression*" dates back to Francis Galton in 1886 [26] in a paper referring to the phenomenon of "average regression" of son height as a function of father height. In addition, the term is also used for certain curve fitting methods. It consists of techniques for modeling the relationship between a dependent variable and one or more independent variables.

In regression, the dependent variable is modeled as a function of independent variables, corresponding regression parameters (coefficients), and a random error term, the parameters of the regression models can be estimated using different methods : the least squares method LS and the weighted least squares WLS method. It consists of solving a linear system by minimizing the quadratic error between the data and the estimates. The main objective of this memory is to study linear regression, which we present in a new empirical method, which can estimate tail index parameters well and may also be useful in all sample sizes.

The outline of this thesis is a blend of tow parts, we start as preliminary Chapters : 1, 2 and 3. Then second part discuse main results. This is organized as follows :

Chapter 1 :

We adopted to study some essential theoretical elements of the theory of extreme values, we start with foundations definitions like the empirical distribution function, the survival function... etc. Also, we make the main results concerning the limit distributions of the largest observations of a sample as well as the domains of attraction.

Chapter 2 :

The second chapter contains some mathematical preliminaries to the design of regression and their properties. We begins with a few reminders on basic concepts such as regression function , regression model and simple linear regression model. Also, we give the most important result about parameter estimation methods, in order that we can leverage them in the next.

Chapter 3 :

In the last chapter in preliminary part we discuses the important part of our research is devoted to the extension of the *Weigthed Least Squares* estimation method. It worth to mention that we present the different work in expressions of the weigth, in order to estimating the parameters of different distributions, such as Weibull, Gumbel and Pareto. The performance of the proposed estimators is proved by making use of expressions the weigth illustrated through some simulations.

Chapter 4 :

The chapter deals with the regression estimator for the tail index. The main objective of this chapter is to propose a method WLS from introduce a new estimator of the cumulative distribution function for heavy tailed of the Frechet distribution. A simulation study and application to real data were performed to the efficiency of this new estimator proposed .

Part I

Preliminary Theory

Chapter 1

Extreme value theory

The theory of extreme values EVT was developed for the estimation of probabilities of occurrences of rare events. It makes it possible to extrapolate the behavior of the tail of the distribution from the largest data observed (the extreme data of the sample).

For a detailed description of this study, see the excellent books in the works of Embrechts, Klüuppelberg and Mikosch [22] 1997, Coles [9], Reiss and Thomas [51] 2001; Beirlant et al., [4] 2004, De Haan and Ferreira [16] 2006, David [15] 1970, and Balakrishnan [3] 1991.

This chapter we review some of the basic notions of elementary probability and statistics. Then, we introduce various asymptotic models available in the classical EVT.

1.1 Foundations definition

Definition 1.1.1 ***(Distribution and survival functions)*** *If X is a rv defined on aprobability space (Ω, F, P) then, its df and survival function (also called hazard function) are respectively defined on $\mathbb{R}$ by*

$$F(x) := P(X \leq x) \text{ and } \overline{F}(x) := 1 - F(x)$$

Definition 1.1.2 *(**The empirical distribution function**) Let the sample $X_1, ..., X_n$ of a positive r.v X, for $n \geq 1$ size, with the df F. The empirical distribution function F_n is defined by:*

$$F_n := \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq x\}} \quad \forall x \geq 0$$

where $I_{\{A\}}$ is the indicator function of the set A. So we can conclude that F_n is the proportion of the n variables which are less than or equal to x.

Definition 1.1.3 *(**The empirical survival function**) Let the sample $X_1, ..., X_n$ of a positive r.v X and of $n \geq 1$ size, where S its a survival function. The empirical survival function noted by S_n, is given by:*

$$S_n := 1 - F_n = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i > x\}} \quad \forall x \geq 0$$

So S_n is the proportion of observations that exceeds x.

Definition 1.1.4 *(**Quantile function**) The quantile function of F is generalized inverse function of F defined by : for all $0 < s < 1$,*

$$Q(s) = F^{\leftarrow}(s) := \inf\{x : F(x) \geq s\}$$

with the convention that $\inf(\varnothing) = \infty$.

Definition 1.1.5 *(**Empirical quantile function**) The empirical quantile function of the sample $(X_1, ..., X_n)$ is defined by : for all $0 < s < 1$:*

$$Q_n(s) := \inf\{x : F_n(x) \geq s\} := \inf\left\{x : \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq x\}} \geq s\right\}$$

where F_n is the empirical distribution function.

1.1.1 Order statistics

Definition 1.1.6 *(**Order statistics**) Let the sample* $X_1, ..., X_n$ *of an independent and identically distributed r.v of the same df* F*. The order statistics of* $X_1, ..., X_n$ *is the increasing rearrangement of the previous sample, noted*

$$X_{1,n} \leq ... \leq X_{n,n},$$

and the rv $X_{k,n}$ *is the* k^{th} *order statistics for* $1 \leq k \leq n$*.*

Definition 1.1.7 *(**Extreme order statistics**) Two order statistics are particular-interesting for the study of extreme events. noted by* $X_{(1,n)}$ *and* $X_{(n,n)}$ *are defined respectively by:*

$$X_{(1,n)} := \min X_{(i)} \quad and \quad X_{(n,n)} := \max X_{(i)}$$

Definition 1.1.8 *(**Extreme order statistics distributions**) The distributions* $F_{X_{1,n}}$ *and* $F_{X_{n,n}}$ *of the extreme order statistics* $X_{1,n}$ *and* $X_{n,n}$ *are respectively defined by:*

$$F_{X_{1,n}}(x) := 1 - [1 - F(x)]^n$$
$$F_{X_{n,n}}(x) := [F_X(x)]^n.$$

Pdf of $X_{1,n}$ *and* $X_{n,n}$ *are respectively defined by:*

$$f_{X_{1,n}}(x) := nf(x)[1 - F(x)]^{n-1}$$
$$f_{X_{n,n}}(x) := n[F(x)]^{n-1}f(x).$$

Definition 1.1.9 *(**Distribution function of the** K **th upper order statistic**) For* $k = 1, ..., n$ *let* $F_{X_{k,n}}$ *denote the df of* $X_{k,n}$ *, then*

$$F_{k,n}(x) := \sum_{r=0}^{k-1} \begin{pmatrix} n \\ r \end{pmatrix} \overline{F}^r(x) F^{n-r}(x).$$

If F is continuous, then

$$F_{k,n}(x) := \int_{-\infty}^{x} f_{k,n}(z) dF(z),$$

where

$$f_{k,n}(x) := \frac{n!}{(k-1)!(n-i)!}[F(x)]^{k-1}[1-F(x)]^{n-k} f(x)$$

i.e. $f_{k,n}$ is a density of $F_{k,n}$ with respect to F.

Definition 1.1.10 *(**Empirical df**) The empirical df of the sample $(X_1, ..., X_n)$ is evaluated using order statistics as follows:*

$$F_n(x) := \begin{cases} 0 & si \quad x < X_{1,n} \\ \frac{i-1}{n} & si \quad X_{i-1,n} \leq x < X_{i,n} \quad 2 \leq i \leq n \\ 1 & si \quad x \geq X_{n,n} \end{cases}$$

Definition 1.1.11 *(**Upper end point**) We denote by x_F (resp x_F^*) the upper extreme point (resp. Lower) of the distribution F (i.e. the greatest possible value for $X_{k,n}$ which can take the value $+\infty$ (resp $-\infty$)) in the sense that:*

$$x_F := sup\{x : F(x) < 1\} \leq \infty$$

and

$$x_F^* := \inf\{x : F(x) > 0\}$$

1.2 Distribution of extreme values

This section is concerned with classical EVT. The result is the Fisher Tippett theorem [24] 1928, which specifies the form of the limit distribution for centred and normalised maxima.

Analogously to the central limit theorem, the theory of extreme values shows that there are sequences $\{a_n\}$ and $\{b_n\}$, $n \in \mathbb{N}^*$, with $a_n > 0$ and $b_n \in \mathbb{R}$, as

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

where $\mathcal{H}$ is a non-degenerate df. Since extreme value df's are continuous on $\mathbb{R}$, assumption 1.1 is equivalent to the following weak convergence assumption

$$\frac{X_{n,n} - a_n}{b_n} \xrightarrow{d} \mathcal{H} \quad \text{as } n \to \infty$$

Remark 1.2.1 *The sequences* $\{a_n\}$ *and* $\{b_n\}$, $n \geq 1$ *are called sequences of normalization, the constants* $a_n \in \mathbb{R}_+^*$ *and* $b_n \in \mathbb{R}$ *are called constants of normalization and the random variable* $\frac{1}{a_n}(X_{n,n} - b_n)$ *is called the normalized maximum.*

1.2.1 Limit distributions

We shall find all distribution functions $\mathcal{H}$ that can occur as this limit. These distributions are called extreme value distributions. The class of distributions F satisfying 1.1 is called the maximum domain of attraction or simply domain of attraction of $\mathcal{H}$. we have a similar notion in identify all extreme value distributions and their domains of attraction.

The following theorem gives a necessary and sufficient condition for the existence of a non-degenerate limit distribtion for the maximum.

Theorem 1.2.1 ***(Fisher & Tippett)*** *Let* $(X_n)_{n\geq 1}$*be a sequence of random variables* $(i.i.d)$*random variables with distribution function* F*. If there exists two real normalizing sequences* $(a_n)_{n\geq 1} > 0$ *and* $(b_n)_{n\geq 1} \in \mathbb{R}$ *, and a non-degenerate law of distribution* $\mathcal{H}$ *such*

$$\lim_{n\to\infty} P\left[\frac{X_{n,n} - a_n}{b_n} \leq x\right] = \mathcal{H}_\alpha(x),$$

where $\mathcal{H}$ *is the distribution of extreme values. The distribution function of the limit is of* *the type of the following three classes :*

$$
\begin{aligned}
&Gumbel: \quad \mathcal{H}_0(x) = \Lambda(x) = \exp[-\exp(-x)] \qquad x \in \mathbb{R} \\
&Frechet: \quad \mathcal{H}_\alpha(x) = \Phi_\alpha(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-1/\alpha}) & x > 0 \end{cases} \quad \alpha > 0. \\
&Weibull: \quad \mathcal{H}_\alpha(x) = \Psi_\alpha(x) = \begin{cases} 1 & x \geq 0 \\ \exp(-(-x)^{-\alpha}) & x < 0 \end{cases} \quad \alpha < 0.
\end{aligned}
$$

Proposition 1.2.1 ***(Density function of extreme values)*** *The density functions of the distribution of standard extreme values and the different types of extreme distribution, are as follows:*

$$
\begin{aligned}
&\textit{Gumble}: \lambda(x) = \exp[-\left\{x + e^{-x}\right\}] \qquad x \in \mathbb{R} \\
&\textit{Frechet}: \phi(x) = \alpha x^{-\alpha-1}\exp(-x^{-1/\alpha}) \quad x > 0 \\
&\textit{Weibull}: \psi(x) = \gamma(-x)^{-\alpha-1}\exp(-(-x)^{-\alpha}) \quad x < 0
\end{aligned}
$$

Figure **??** illustrates the density functions of $\lambda(x), \phi(x)$ and $\psi(x)$, we chose $\alpha = 1$ for the Frechet and the Weibull distributions.

The three previous formulas can be combined in theorem 1.2.1 into a single type of distribution ($Weibull$, $Gumbel$ and $Frechet$), called a the generalized extreme values distribution ($GEVD$). A better analysis is offered thanks to the work of von Mises [56] 1954, and Jenkinson [38] 1955.

Definition 1.2.1 ***(Generalized extreme values distribution)*** *Let $\gamma \in \mathbb{R}$, we call the GEVD any df $\mathcal{H}_\gamma$ or any probability law which has $\mathcal{H}_\gamma$ as a function of distribution, for all $x \in \mathbb{R}$ such that $1 + \gamma x > 0$, as follows:*

$$
\mathcal{H}_\gamma(x) = \begin{cases} \exp\left\{-\left[1 + \gamma x\right]^{-1/\gamma}\right\} & \textit{if} \ \ \gamma \neq 0 \\ \exp\left(-\exp(-x)\right) & \textit{if} \ \ \gamma = 0 \end{cases} \tag{1.2}
$$

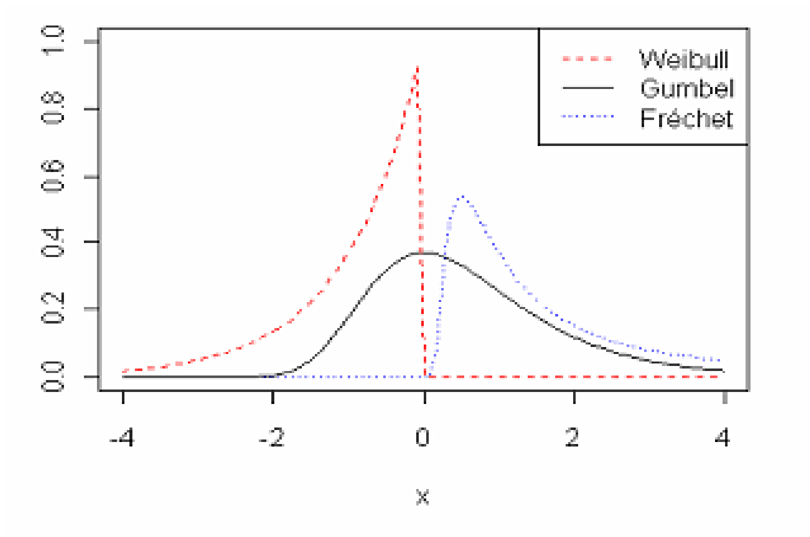

Figure 1.1: Densities of the standard extreme value distributions.

where the parameter γ is called the index of extreme values (EVI).

Remark 1.2.2 *The GEVD H_γ can be written in a more general form so for $\left\{1+\frac{\gamma}{\sigma}(x-\mu)>0\right\}$:*

$$\mathcal{H}_{\gamma,\mu,\sigma}(x) = \begin{cases} \exp\left\{-\left[1+\gamma\left(\dfrac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\} & \gamma \neq 0 \\ \exp\left(-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right) & \gamma = 0 \end{cases} \quad x \in \mathbb{R}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are respectively the location and scale parameters.

Gnedenko [29],1943 accomplished an important result on this issue.

In the applications of the theory of extreme values, which make it possible to classify the three types of extreme distributions $Frechet, Weibull$ and *Gumbel* in a single type which is the type of generalized extreme value distribution. This proposition gives us a very important result. Indeed, we have the following proposition:

Proposition 1.2.2 ***(Ferreira, 2006)*** *Let $\mathcal{H}_\gamma(\gamma \in \mathbb{R})$ be the generalized extreme value distribution and Λ, Φ_α and $\Psi_\alpha$the distribution of standard extreme values with $\alpha > 0$ we have :*

$$\mathcal{H}_\gamma(x) := \begin{cases} \Phi_{1/\gamma}\left(1+\gamma x\right) & \text{if } \gamma > 0 \\ \Psi_{-1/\gamma}\left\{-\left(1+\gamma x\right)\right\} & \text{if } \gamma < 0 \\ \Lambda(x) & \text{if } \gamma = 0 \end{cases}$$

$\forall x \in \mathbb{R}$ such that $1 + \gamma x > 0$.

In other words, Hence the three extreme value distributions can be characterized by the sign of the tail index γ : Frechet type to $\gamma > 0$, Weibull type to $\gamma < 0$ and Gumbel type corresponds to $\gamma = 0$.

Definition 1.2.2 *(**Generalized Pareto Dstribution** GPD) The generalized Pareto distribution any function distribution $\mathcal{G}_\gamma$, for all $\gamma \in \mathbb{R}$, such that $1 + \gamma x > 0$, as follows:*

$$\mathcal{G}_\gamma(x) = \begin{cases} 1-\left(1+\gamma x\right)^{-1/\gamma} & \text{if } \gamma \neq 0 \\ 1-\exp\left(-x\right) & \text{if } \ \gamma = 0 \end{cases} \quad \forall\ x \geq 0$$

Remark 1.2.3 *The GPD $\mathcal{G}_\gamma$ can be written in a more general form that we denote by $G_{\alpha,\mu,\sigma}$ a parameter is shown of localization $\mu \in \mathbb{R}$ and a scale parameter $\sigma > 0$ for $1 + \alpha\left(\frac{x-\mu}{\sigma}\right) > 0$ and $\forall x \geq \mu$, as follows:*

$$G_{\alpha,\mu,\sigma}(x) = \begin{cases} 1-\left(1+\alpha\left(\dfrac{x-\mu}{\sigma}\right)\right)^{-1/\alpha} & \text{if } \alpha \neq 0 \\ 1-\exp\left(-\dfrac{x-\mu}{\sigma}\right) & \text{if } \ \alpha = 0 \end{cases}.$$

The parameter $\alpha \in \mathbb{R}$ is called the "tail index" shape parameter from which we can thus see the generalized Pareto distribution.

1.3 Domain of attraction

In this section, we recall the necessary and sufficient conditions on the distribution function F so that it belongs to one of the domains of attraction of one of the three limit laws of extreme values. These conditions, basically due to von Mises [56], 1936 called *von Mises condition.*

Definition 1.3.1 *(**Domain of attraction**) We say that a distribution function F belongs to the domain of attraction of $\mathcal{H}_\gamma$, if F verifies theorem 1.2.1, denoted by $F \in \mathcal{D}(\mathcal{H}_\gamma)$.*

Now in the following theorem, we shall establish necessary and sufficient conditions for a distribution function F to belong to the domain of attraction of $\mathcal{H}_\gamma$.

Theorem 1.3.1 *According to the sign of γ, let x_F its right endpoint. The distribution function F is in the domain of attraction of the extreme value distribution $\mathcal{D}(H_\gamma)$ if and only if*

1. For $\gamma > 0$, x_F is infinite and

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

for all $x > 0$. This means that the function $1 - F$ is regularly varying at infinity with index $-1/\gamma$.

2. For $\gamma < 0 : x_F < \infty$ and for all $x > 0$

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

3. For $\gamma = 0 :$ here the right endpoint x_F may be finite or infinite and

$$\lim_{t \uparrow x_F} \frac{1 - F(t - f(t + xf(t)))}{1 - F(x_F - t)} = e^x$$

for all real x, where f is a positive suitable function.

1.3.1 Characterizations of domain attraction

Before characterizing the domain of attraction, we define the functions with variations. For more details, refer to Bingham et al., [7] 1987 where many results on regularly varying functions are given.

Definition 1.3.2 ***(Regularly varying and slowly varying functions)*** *Let a measurable function $G : \mathbb{R}^+ \to \mathbb{R}^+$ is regularly varying at ∞ with the index ρ ($G \in \mathcal{RV}_\rho$), if*

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

A measurable function $l :]a, +\infty[\to \mathbb{R}^+$with $(t > 0)$ is said slowly varying at infinity, if:

$$\lim_{x\to\infty} \frac{l(tx)}{l(x)} = 1$$

Theorem 1.3.2 ***(Kramata representation)*** *Every slowly varying function l (i.e $l \in \mathcal{RV}_0$) if and only if can be represented as : for all $x > 0$,*

$$l(x) = c(x) \exp\left\{ \int_1^x t^{-1}\varepsilon(t)dt \right\},$$

where c and ε are two measurable functions,

$$\lim_{x\to\infty} c(x) = c \in]0, +\infty[\quad \textit{and} \quad \lim_{t\to\infty} \varepsilon(t) = 0$$

If the function c is constant, we say that l normalized.

Proof. See Resnick [52],1987; Corollary 2.1 ■

Different characterizations of three domain of attraction of Frechet, Weibull and Gumbel, according to the sign of γ, we can distinguish three domain of attraction :

Characterization of $\mathcal{D}(\Phi_\gamma)$:

If $\gamma > 0$, we say that $F \in \mathcal{D}\left(\Phi_\gamma\right)$, and F has an infinite right end point ($x_F = +\infty$), this domain of attraction of heavy-tailed distributions, that is, which have a polynomial decay survival function. The result below stated by Gnedenko [29],1943 and a simple proof of which can be found in Resnick's book [$Proposition$ 1.11].

Theorem 1.3.3 *$F \in \mathcal{D}\left(\Phi_\gamma\right)$ with parameter $\gamma > 0$, $x_F = +\infty$ if and only if :*

$$1 - F(x) = x^{-1/\gamma} l(x)$$

where l is a slowly varying function. In this case, a possible choice for the sequences a_n and b_n are $a_n = F^{-1}(1 - \frac{1}{n})$ and $b_n = 0$.

Characterization of $\mathcal{D}\left(\Psi_\gamma\right)$:

If $\gamma < 0$, we say that $F \in \mathcal{D}\left(\Psi_\gamma\right)$, and F has a finite right end point ($x_F < +\infty$). This domain of attraction of survival functions whose support is bounded above. The following result (see Gnedenko [29],1943 ;Resnick [52],1987[45, Proposition 1.13]) shows that we pass from the domain of attraction of Frechet to that of Weibull by a simple change of variable in the distribution function.

Theorem 1.3.4 *$F \in \mathcal{D}\left(\Psi_\gamma\right)$ with $\gamma < 0$ iff $x_F = +\infty$ and $1 - F^*$ is a function with regular variations of index α*

$$1 - F(x) = x_F - x^{-1} = x^{-1/\gamma} l(x)$$

where the function l slowly varying of index $1/\gamma$. In this case, a possible choice for the sequences a_n and b_n is

$$a_n = x_F - F^{-1}(1 - \frac{1}{n}) \quad and \quad b_n = x_F$$

this domain of attraction has been considered by Falk [23],1995; Gardes [27], 2010 to give an endpoint estimator of a distribution.

Characterization of $\mathcal{D}(\Lambda)$:

If $\gamma = 0$ we say that $F \in \mathcal{D}(\Lambda)$ the end point x_F can then be finite or not.This domain of attraction of distributions with light tails, that is to say which have an exponentially decaying survival function. The result below is proved notably in Resnick [52],1987 [Proposition 1.4].

Theorem 1.3.5 *A distribution function F belongs to the Gumbel domain of attraction if and only if there exists $z < x_F < \infty$ such that*

$$\overline{F}(x) = c(x) \exp\left\{ -\int_z^{\infty} \frac{1}{a(t)} dt \right\}, \quad z < x < x_F$$

where $c(x) \to c > 0$ when $x \to x_F$ and $a(.)$ is a positive and differentiable function with derivative $\acute{a}(.)$ such that $\lim_{x \to x_F} \acute{a}(.) \to 0$.

The tables 1.1, 1.2 and 1.3 give different examples of standard distributions in these three domains of attraction.

Distributions	$\overline{F}(x)$ or density f	γ
Burr(β, τ, λ) $\beta > 0, \tau > 0, \lambda > 0$	$\left(\frac{\beta}{\beta - x^{\tau}}\right)^{\lambda}$	$\frac{1}{\lambda^{\tau}}$
Frechet $\left(\frac{1}{\alpha}\right)$, $\alpha > 0$	$1 - \exp(-x^{-\alpha})$	$\frac{1}{\alpha}$
Loggamma $\lambda > 0, m \succ 0$	$\frac{\lambda^m}{\Gamma(m)} \int_x^{\infty} (\log(u))^{m-1} u^{-(\lambda+1)} du$	$\frac{1}{\lambda}$
Log-logistic $\beta > 0, \alpha > 0$	$\frac{1}{1+\beta^{\alpha}}$	$\frac{1}{\alpha}$
Pareto $\alpha > 0$	$x^{-\alpha}$, $x > 0$	$\frac{1}{\alpha}$

Table 1.1: Some distributions associated with a positive index

Distribution	$\overline{F}(x)$	γ
Uniforme $[0, 1]$	$1 - x$	-1
Inverse Burr $(\beta, \tau, \lambda, x_\tau), \beta, \tau, \lambda > 0$	$\left(\frac{\beta}{\beta + (x_\tau + x)^{-\tau}}\right)^\lambda$	$-\frac{1}{\lambda}$

Table 1.2: Some distributions associated with a negative index.

Distributions	The $\overline{F}(x)$ or density f	γ
Gamma (m, λ), $m \in \mathbb{N}, \lambda > 0$	$f(x) = \frac{\lambda^m}{\Gamma(m)} \int\limits_x^\infty u^{m-1} \exp(-\lambda u) du$	0
Gumbel (μ, β), $\mu \in \mathbb{R}, \beta > 0$	$f(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\beta}\right)\right)$	0
Logistic	$\overline{F}(x) = \frac{2}{1+\exp(x)}$	0
Log nomale (μ, σ), $\mu \in \mathbb{R}, \sigma > 0$	$f(x) = \frac{1}{2\pi} \int\limits_x^\infty \frac{1}{\mu} \exp(-\frac{1}{2\sigma^2}(\log u - u)^2) du$	0
Weibull (λ, τ), $\lambda > 0, \tau > 0$	$\overline{F}(x) = \exp(-\lambda x^\tau)$	0

Table 1.3: Some distributions associated with a null index.

1.4 Tail Index Estimators

The estimate of the tails index, plays an important role in limiting an extreme law, when it exists, is indexed by a parameter called *extreme value index*, there are two methods for estimating the extreme value index : *parametric methods*, meaning that the data follow an exact GEV distribution, and *semi-parametric methods*, where the parameter has both a finite-dimensional and an infinite-dimensional and are therefore based on partial properties of the underlying distribution, such as the Pickands [50], Hill [34] and Moment [19] estimators.

In the following, we briefly review the estimators that have been proposed for tail index estimation.

1.4.1 Semi-parametric estimators

We present here different estimators constructed under the domain of attraction conditions. That is, the data $(X_1, ..., X_n)$ are assumed to be drawn from a population X with df F. This semi-parametric statistical procedures don't assume the knowledge of the whole distribution but only focus on the distribution tails. The case $\gamma > 0$ has got more interest because data sets in most real-life applications, exhibit heavy tails. The two most common estimators in the literature are the estimators from Hill and Pickands. We give a study in this section of estimators with some of their statistical properties.

Let $X_{1,n} \leq ... \leq X_{n,n}$ be the order statistics based on the sample $X_1, ..., X_n$, with distribution F and $X_{k,n}$ is the k^{th} upper order statistic. The intermediate order statistics $X_{n-k,n} \to \infty$ and $k = k_n$ be a sequence of positive numbers satisfying the conditions

$$1 \leq k_n \leq n, \ k_n \to \infty \text{ and } \frac{k_n}{n} \to 0 \quad \text{as} \quad n \to \infty$$

By the way, it is necessary to calculate this estimator on the tails of the distribution. A sub-sample that is too small does not allow the estimators to reach their level of stability on the contrary choosing k too high generates the risk of taking non-outliers into account, thus we will note that a non-parametric approach is only possible if one has a large number of observations : if the samples small, then we'll turn on to *the parametric approach.*

Pickand's estimator

James Pickands proposed his estimator in 1975, [50] for any $\gamma \in \mathbb{R}$ and $k = k_n$ series of integers with $1 < k < n$. Let $X_{1,n} \leq ... \leq X_{n,n}$ the order statistics of $X_1, ..., X_n$ from F such that $F \in \mathcal{D}\ (\Phi_{\frac{1}{\gamma}})$, the Pickand estimator is defined by:

$$\widehat{\gamma}^{(P)} = \widehat{\gamma}_k^{(P)} := (\log 2)^{-1} \log \left(\frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}} \right)$$

A full analysis on $\widehat{\gamma}^{(P)}$ is to be found in Dekkers and de Haan [19], 1989 where improvements of this estimator were introduced in particular by Drees [21],1995 from which the following result is taken.

Theorem 1.4.1 ***(Asymptotic Properties of*** $\widehat{\gamma}^{(P)}$ *) Assume that* $F \in \mathcal{D}(\mathcal{H})$, $\gamma \in \mathbb{R}$, $k \to \infty$ *and* $\dfrac{k}{n} \to 0$ *when* $n \to \infty$.

1. Weak Consistency :

$$\widehat{\gamma}^{(P)} \xrightarrow{P} \gamma \quad when \quad n \to \infty$$

2. Strong consistency : if $k/\log\log n \to \infty$ when $n \to \infty$, then

$$\widehat{\gamma}^{(P)} \xrightarrow{a.s} \gamma \quad when \quad n \to \infty$$

3. Asymptotic normality: under further conditions on k and F ,

$$\sqrt{k}(\widehat{\gamma}_k^{(P)} - \gamma) \xrightarrow{d} \mathcal{N}(0, \eta^2) \quad when \quad n \to \infty,$$

where

$$\eta^2 := \frac{\gamma^2(2^{2\gamma+1}+1)}{(2(2^\gamma - 1)\log 2)^2}$$

A generalization of the Pickands estimator was introduced by Yun [58],2002 as follows

$$\widehat{\gamma}_{n,k;u,v}^{(Y)} := (\log v)^{-1} \log\left(\frac{X_{n-k+1,n} - X_{n-[uk]+1,n}}{X_{n-[vk]+1,n} - X_{n-[uvk]+1,n}}\right),$$

where, u, v are positive real numbers different from 1 such that $[uk], [vk]$ and $[uvk]$ do not exceed n. For $u = v = 2$, we have $\widehat{\gamma}_k^{(P)}$.

Hill's Estimator

Hill's estimator, is one of the most common estimators for the tail index of heavy tailed distributions, where research has mainly focused on when the EVI is positive $(\gamma = \frac{1}{\alpha} > 0)$ because data sets in most real applications , which corresponds to the distributions belonging to the domain of attraction of Fréchet $F \in \mathcal{D}(\Phi_{\frac{1}{\gamma}})$, that is, when the distribution tail has a Pareto shape, identified by Hill [34],1975 :

$$\widehat{\gamma}^{H}=\widehat{\gamma}_{k}^{(H)}:=\frac{1}{k}\sum_{j=1}^{k}\log X_{n-j+1,n}-\log X_{n-k,n} \tag{1.3}$$

The construction of this estimator is given in the books by De Haan et al.,[**16**],2006 and Beirlant et al.,[4], 2016. Other estimators have been proposed in particular by Beirlant et al., [4],2016 who use an exponential regression model base to the Hill estimator and by Csörgö et al.,[11], 1985 who use a kernel in the Hill estimator.

The asymptotic properties of Hill's estimator are summarized in the following theorem.

Theorem 1.4.2 *(**Asymptotic Properties of** $\widehat{\gamma}^{(H)}$) Assume that* $F\in\mathcal{D}\left(\Phi_{\frac{1}{\gamma}}\right)$, $\gamma>0$, $k\rightarrow\infty$ *and* $\frac{k}{n}\rightarrow 0$ *when* $n\rightarrow\infty$.

1. Mason [**46**],1982 has proven weak consistency :

$$\widehat{\gamma}^{(H)}\overset{P}{\rightarrow}\gamma \quad when\ n\rightarrow\infty$$

2. Strong consistency was established by Deheuvels et all. [**18**],1985 under the condition that : $k/\log\log n\rightarrow\infty$, then

$$\widehat{\gamma}^{(H)}\overset{a.s}{\rightarrow}\gamma \quad when \quad n\rightarrow\infty,$$

and more recently by Necir [**47**], 2006.

3. Asymptotic normality was established under a suitable extra assumption, known as the second-order regular variation condition (see De haan and Stadtmüller [**17**],1996 and De haan and Ferreira [**16**],2006), with mean γ and variance γ^2/k :

$$\sqrt{k}\left(\frac{\widehat{\gamma}^{(H)}-\gamma}{\gamma}\right)\overset{d}{\rightarrow}\mathcal{N}\left(0,1\right)$$

Figure 1.2, show that the Hill estimator against k performs well with both the Frechet distribution, the sample size is $n = 1000$:

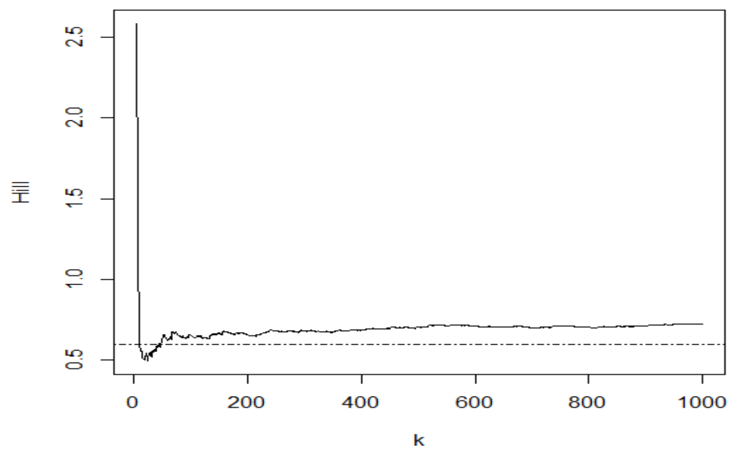

Figure 1.2: Hill estimator for samples of a Frechet distribution, with parameter $\gamma = 0.6$

Moment estimator

Another estimator which can be considered as an adaptation of Hill's estimator, to obtain the consistency for all $\gamma \in \mathbb{R}$, has been proposed by Dekkers et al., [19],1989. This is the moment estimator, given by

$$\hat{\gamma}^{(M)} = \hat{\gamma}_k^{(M)} := M_1 + 1 - \frac{1}{2}\left(1 - \frac{(M_{(k)}^{(1)})^2}{M_{(k)}^{(2)}}\right)^{-1},$$

$$M_k^{(r)} := \frac{1}{k}\sum_{i=0}^{k}\left(\log X_{n-i+1,n} - \log X_{n-k;n}\right)^r, \quad r = 1, 2.$$

Theorem 1.4.3 ***(Asymptotic properties of*** $\hat{\gamma}^{(M)}$***)*** *Suppose that* $F \in \mathcal{D}(\mathcal{H})$, $\gamma \in R$, $k \to \infty$ *and* $k/n \to 0$ *when* $n \to \infty$ *:*

1. Weak consistency :

$$\hat{\gamma}^{(M)} \xrightarrow{P} \gamma \quad \text{when} \quad n \to \infty.$$

2. Strong consistency : if $k/(\log n)^{\delta} \to \infty$ when $n \to \infty$ for certain $\delta > 0$, so

$$\hat{\gamma}^{(M)} \overset{a.s}{\to} \gamma \quad when \;\; n \to \infty.$$

3. Asymptotic normality: (see Theorem 3.1 and Corollary 3.2 of [19])

$$\sqrt{k}(\hat{\gamma}^{(M)} - \gamma) \overset{d}{\to} \mathcal{N}(0, \eta^2) \quad when \;\; n \to \infty,$$

or

$$\eta^2 := \begin{cases} 1 + \gamma^2 & \gamma \geq 0, \\ (1-)(1-2\gamma)\left(4 - 8\dfrac{1-2\gamma}{1-3\gamma} + \dfrac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\right), & \gamma < 0. \end{cases}$$

The normality of this estimator was established by Dekkers et al., [19] under suitable regularity conditions.

1.4.2 Parametric estimators

Maximum likelihood estimator

The maximum likelihood estimator is built from the observations of the maxima, it involves estimating the index of extreme values as well as the two normalizing sequences a_n et b_n:

Let $X_1, ..., X_n$ be a sample of n maxima, in the case $\gamma \neq 0$ the log-likelihood function obtained from the definition is written:

$$\mathcal{L}((\gamma, b_n, a_n); X) = -n \log a_n - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{n} \log\left(1 + \gamma \frac{X_i - b_n}{a_n}\right) - \sum_{i=1}^{n} \left(1 + \gamma \frac{X_i - b_n}{a_n}\right)^{-1/\gamma}$$

In the case where $\gamma = 0$,

$$\mathcal{L}((0, b_n, a_n); X) = -n \log a_n - \sum_{i=1}^{n} \exp - \left(\frac{X_i - b_n}{a_n}\right) - \sum_{i=1}^{n} \frac{X_i - b_n}{a_n}$$

Smith [54],1985: demonstrated the consistency properties and the asymptotic normality of this estimator when $\gamma > 1/2$ and $m \to \infty$:

$$\sqrt{m}\left(\left(\widehat{\gamma}, \widehat{a}_n, \widehat{b}_n\right) - (\gamma, a_n, b_n)\right) \to \mathcal{N}\left(0, I^{-1}\right)$$

where I is the Fisher information matrix estimated by its empirical version

$$I(\theta) = -E\left(\frac{\partial^2 \mathcal{L}(X;\theta)}{\partial \theta^2}\right)$$

$\mathcal{L}(X;\theta)$ is the log-likelihood function associated with the law of the random variable X, θ parameterized by a set of parameters θ.
For $\gamma > -1$, Zhou [61],2009 and Dombry [20],2013 proved that the maximum likelihood estimator exists and is consistent. Then Zhou [62],2010 ; also obtained the asymptotic normality for $-1 < \gamma < -1/2$.

Weighted moment estimator

This method, which dates back to Hosking et al., [36], 1985 is based on the following quantity, called the weighted moment of order r :

$$\omega_r := E(X\ \mathcal{H}\ _{\gamma,\mu,\sigma}^{r}(x)), \quad r \in \mathbb{N}.$$

This quantity exists for $\gamma < 1$ and given by:

$$\omega_r := \frac{1}{r+1}\left\{\mu - \frac{\sigma}{\gamma}\left(1 - \Gamma(1-\gamma)(r+1)^{\gamma}\right)\right\},$$

where $\mathbf{\Gamma}$ is Euler's gamma function. In this case, three weighted moments are enough to calculate μ, σ and γ.

$$\begin{cases} \hat{\omega}_0(\theta) = \mu - \frac{\sigma}{\gamma}\left(1 - \Gamma(1-\gamma)\right), \\ 2\hat{\omega}_1(\theta) - \hat{\omega}_0(\theta) = \frac{\sigma}{\gamma}\Gamma(1-\gamma)(2^\gamma - 1), \\ \dfrac{3\hat{\omega}_2(\theta) - \hat{\omega}_0(\theta)}{2\hat{\omega}_1(\theta) - \hat{\omega}_0(\theta)} = \dfrac{3^\gamma - 1}{2^\gamma - 1}. \end{cases}$$

Thus by replacing respectively ω_r, $r \in \{0, 1, 2\}$ by its empirical estimator

$$\hat{\omega}_{r,n} := \frac{1}{n}\sum_{i=1}^{n} X_{i,n}\left(\frac{i-1}{n}\right)^r.$$

The weighted moment estimator (WME) is obtained by solving the system of three equations

$$\omega_r = \hat{\omega}_{r,n} \quad , \quad r = 0, 1, 2.$$

The solution to this equation is the WM estimator $\hat{\gamma}$ of γ. The other parameters σ and μ are estimated respectively by:

$$\hat{\sigma} = \frac{(2\hat{\omega}_1 - \hat{\omega}_0)\hat{\gamma}}{\Gamma(1-\gamma)(2^\gamma - 1)},$$

and

$$\hat{\mu} = \hat{\omega}_0 + \frac{\hat{\sigma}}{\hat{\gamma}}\left(1 - \Gamma(1-\hat{\gamma})\right).$$

Regression estimator

The parameters of the distribution of extreme values can be estimated by the regression method, Gumbel [28] 1958 and Kinnison [42] 1985, presented this method consists of four steps, as follows:

• Choose the maximum profitability from a set of daily profitability. At each date, we observe a realization of the variable X. After n time units, we therefore have n observations denoted $X_1, X_2, ..., X_n$, from which we extract the greatest value denoted $Y_{1,n}$. Of the following n observations, we extract again the maximum term called $Y_{2,n}$. If we have

$N^{obs} = n.N$ observations, then N observations of maxima $Y_{1,n}, Y_{2,n}, , ..., Y_{N,n}$.

• Row the sequence$Y_{1,n}, Y_{2,n}, , ..., Y_{N,n}$ in ascending order to obtain ordered statistics $\acute{Y}_{1,n}, \acute{Y}_{2,n}, , ..., \acute{Y}_{N,n}$ is verify: $\acute{Y}_{1,n} \leq \acute{Y}_{2,n}, \leq ..., \acute{Y}_{N,n}$.

• Use random frequencies $F_Y(\acute{Y}_{1,n}), F_Y(\acute{Y}_{2,n}), , ..., F_Y(\acute{Y}_{N,n})$ and assumes that the extreme observations are exactly taken from the extreme value distribution $\mathcal{H}_\gamma$ given by the theorem 1.2.1. For each value of i, $F_Y(\acute{Y}_{i,n})$, the distribution of this random variable is given by :

$$F_Z(y) = \frac{N\ !}{(N-i)!\ i\ !} i\ y^{i-1}(1-y)^{N-i}\ where \quad Z = F_Y(\acute{Y}_{i,n}),$$

note that the law of the variable $F_Y(\acute{Y}_{i,n})$ is independent of the variable Y and only depends on the order i.The random frequencies $F_Y(\acute{Y}_{i,n})$ are distributed around their mean values $E\left(F_Y(\acute{Y}_{i,n})\right)$. The mean value of the i^{th} frequency is given by :

$$E\left(F_Y(\acute{Y}_{i,n})\right) = \frac{i}{n+1},$$

this result leads to the statistical model

$$F_Y(\acute{Y}_{i,n}) = E\left(F_Y(\acute{Y}_{i,n})\right) + \varepsilon_{i,n} = \frac{i}{n+1} + \varepsilon_{i,n}, \tag{1.4}$$

the error term $\varepsilon_{i,n}$ has zero mean and is normally asymptotically distributed if the quotient i/N is not too close to zero and unity.

• Estimate equation 1.4 by transforming it by taking twice the logarithm of $F_Y(\acute{Y}_{i,n})$ and of $E\left(F_Y(\acute{Y}_{i,n})\right)$ and obtaining a non-linear model :

$$-\log\left[-\log\left(\frac{i}{n+1}\right)\right] = \frac{1}{\gamma}\log\sigma - \frac{1}{\gamma}\log\left[\sigma - \gamma\left(\acute{Y}_{i,n} - \mu\right)\right] + \upsilon_{i,n}\ , \tag{1.5}$$

the study is according to sign of the tail index γ, which determines the type of the asymptotic distribution.

The case $\gamma = 0$ (Gumbel), it is necessary to estimate a following model:

$$-\log\left[-\log\left(\frac{i}{n+1}\right)\right]=\frac{\acute{Y}_{i,n}-\mu}{\sigma}+\upsilon_{i,n} \tag{1.6}$$

The equations 1.5 and 1.6 are estimated by minimizing the sum of the squares of the residuals, under the assumption of normality and independence of the residuals, minimizing this function amounts to maximizing the likelihood of each model. Estimators of γ, μ and σ are relatively accurate although slightly biased.

For more details on this issue, one my consult please check (Gumbel [**28**] 1958, page 176 – 178) and (Kinnison [**42**] 1985,page 68 – 71).

Chapter 2

Regression conceptions

In this chapter, we will indeed the most important aspects of regression theory. The main definition and characteristics of this concept are presented. To make things easier, we start with a brief reminder of linear regression model. As this is a model with only one explanatory variable, we speak of simple regression or to explain a variable Y using a variable X. Then, focusing on the most important definitions the regression line to be estimated from the data of a sample by the least squares method [53]. One of the assumptions of the least squares estimation method is the assumption of constant variance, but in situations where the underlying distribution is continuous but skewed, constant variance cannot be assumed. This situation can best be solved by modifying least squares using a weighted least square, which allows the variance of the error term to be almost constant.

In this conceptions, we focus on methods which using linear regression based on a simple linear model.

2.1 The simple regression model

Generally, we consider the modelling between the dependent and one independent variable. When there is only one independent variable in the linear regression model, the model is generally termed as a simple linear regression model.

As a first approach, a natural idea is to suppose that the variable to be explained Y is a function ne of the explanatory variable x, that is to say of look for g in the set F of functions affine from $\mathbb{R}$ to $\mathbb{R}$. This is the principle of simple linear regression.

Before presenting the simple linear model, we define the regression function.

2.1.1 Definitions

Definition 2.1.1 *(**Regression function**) Let X and Y be two random variables such that $E(|Y|) < \infty$. The function $g : \mathbb{R} \to \mathbb{R}$ defined by :*

$$g(x) = E(Y|X = x)$$

is said to be the regression function of Y on X.

Definition 2.1.2 *(**Regression model** Y **on** X) We note Y the real random variable to be explained and X the explanatory variable. The model amounts to supposing, that on average $E(Y)$is an affine function of X (i.e, writing the model implicitly assumes a prior notion of causality in the sense that Y depends on X because the model is not symmetrical, see [10]). In the case where X is deterministic, the model is written:*

$$E(Y) = f(X) = a_0 + a_1 X$$

In the case where X is random, the model is then written conditionally on the observations of X :

$$E(Y|X = x) = a_0 + a_1 x$$

Definition 2.1.3 ***(Simple linear regression model)*** *A simple linear regression model is defined by an equation of the form ,*

$$Y_i = a_0 + a_1 x_i + \varepsilon_i \quad \forall i = \overline{1, n}$$

where a_0 *and* a_1 *are parameters and independent of the residuals* ε_i*.*

To make an inference of the model parameters, recall several main assumptions of simple linear regression :

- Y_i represents the i^{th} value of the response (dependent) random variable y.
- x_i represents the i^{th} value of the predictor (independent) deterministic variable x.
- a_0 and a_1 are the coefficients (represented by the intercept and slope of the model).
- The errors are uncorrelated $Cov(\varepsilon_i, \varepsilon_j) = 0$, $\forall i \neq j$.
- We'll model ε_i as being Gaussian, $\forall i = \overline{1, n}$:

$$\varepsilon_i \sim \mathcal{N}(0, \sigma^2_{\varepsilon_i}) \quad \forall i = \overline{1, n} \tag{2.1}$$

- σ^2 is constant throughout the range.
- Relationship is linear between X and Y , i.e., relation is a straight line.

The residual plot below suggests that : In the first case (left), all assumptions seem satisfied, but in the second graph (right), the relationship does not seem linear, the variance σ^2, is not constant throughout the range.

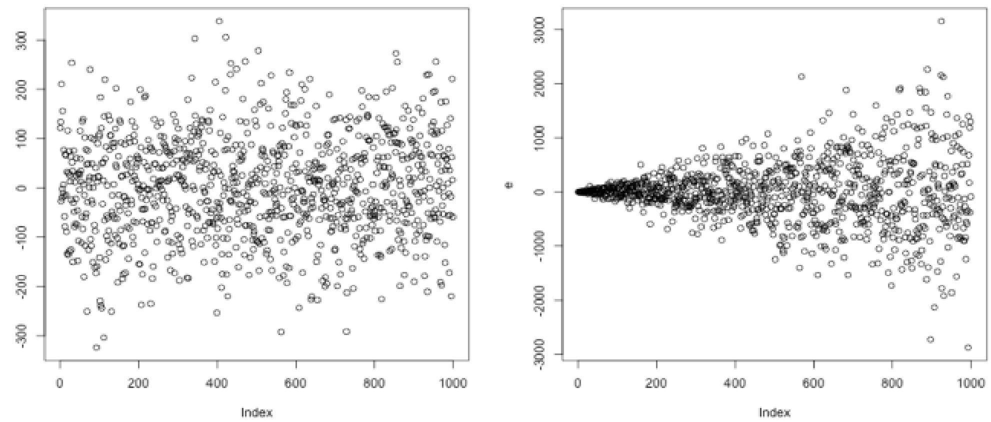

Figure 2.1: Examples of scatter plot the residuals

In addition, the simple linear regression model defined by 2.1.3 can be written in matrix form :

$$Y = Xa + \varepsilon \Leftrightarrow \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

The assumptions of this model can of course be relaxed. Note that the affine hypothesis is not as restrictive as it seems, variables can be changed. For example:

$Y_i = a_0 + a_1 \log(x_i) + \varepsilon_i$,

$Y_i^2 = a_0 + a_1 \exp(x_i) + \varepsilon_i$, and

$\log(Y_i/(1 - Y_i)) = a_0 + a_1 x_i + \varepsilon_i$ (logistic model) are also linear models.

Remark 2.1.1 *:*

Error distribution ε_i distributed according to 2.1 is used when we want to determine the distance of the estimators $(\hat{a}_0, \hat{a}_1)$, the confidence intervals and the hypothesis tests. So for each value of x_i (fixed),Y_i has a normal distribution of expectation $E(Y_i) = a_0 + a_1 x_i + E(\varepsilon_i)$ *and* $Var(Y_i) = Var(a_0 + a_1 x_i + \varepsilon_i) = \sigma^2$ *.The expectations of the different Y_i are thus aligned on the regression line that we have to estimate by the method of least squares,* *which we will present in the next section.*

2.2 Parameter estimation methods

In statistics, the maximum likelihood estimator for observations with additive gaussian noise is the least squares estimator, given a sample $(x_i)_{1\leq i\leq n}$, a simple regression model assumes that the results observed for y_i are related to x_i. We represent in a graph the set of observations (x_i, y_i). We can then propose a linear model, that is to say look for the line whose equation is $y_i = a_0 + a_1 x_i$ and which passes as close as possible to the points of the graph.

We recall here two notions of the methods :

2.2.1 The least squares method

To find good estimates for the coefficients a_0 and a_1, we employ the *Least Squares* (LS) method, which gives the line that minimizes the sum of the vertical distances from each point to the line.

Definition 2.2.1 *To estimate parameters $a_0$and a_1, by minimizing the sum of the squares of the differences between observations and model 2.1.3, Least squares are given by the following formulas :*

$$\widehat{a}_1 = \frac{S_{xy}}{S_x^2} \qquad \& \qquad \widehat{a}_0 = \overline{y} - \widehat{a}_1 \overline{x},$$

where :

$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$

$S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$, $S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}),^2$

$S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}).$

Statistical properties of LS estimates :

- These estimators are unbiased estimators : $E(\widehat{a}_1) = \widehat{a}_1$ *and* $E(\widehat{a}_0) = \widehat{a}_0$.
- $Var(\widehat{a}_1) = \dfrac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2} = \dfrac{\sigma^2}{nS_x^2}$
- $Var(\widehat{a}_0) = \dfrac{\sigma^2 \sum\limits_{i=1}^{n} x_i^2}{n\sum\limits_{i=1}^{n}(x_i - \overline{x})^2} = \dfrac{\sigma^2}{n}(1 + \dfrac{\overline{x}^2}{S_x^2})$
- $Cov(\widehat{a}_0, \widehat{a}_1) = Cov(\widehat{a}_1, \widehat{a}_0) = -\dfrac{\sigma^2\overline{x}}{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2} = -\dfrac{\overline{x}}{nS_x^2}.$

Remark 2.2.1 *An unbiased estimator of σ_ε^2 is given by :*

$$S_\varepsilon^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n-2}\sum_{i=1}^{n}\varepsilon_i^2$$

Fitting a straight line

The method of least squares is a procedure to determine the best fit line to data, the proof uses simple calculus and linear algebra. The basic problem is to find the best fit straight line $y = a_0 + a_1 x$ given that, for $i = \overline{1, n}$.

We consider that the points are on the line of equation, the vertical distance corresponding to the i^{th} observation is : $\varepsilon_i = y_i - a_0 - a_1 x_i$, these vertical distances are called the least squares residuals, the sum of squares of these distances can then be written as:

$$Q(a_0, a_1) = \sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)^2 \,; i = \overline{1, n}$$

The goal is then to find the right equation minimizing this error term, that is to say to determine a_0 and a_1 minimizing Q is minimal (see the definition 2.2.1), we calculate its derivative with respect to each of these two variables :

$$\frac{\partial Q}{\partial a_0} = 0 \quad \& \quad \frac{\partial Q}{\partial a_1} = 0$$

Finally, the least squares regression line is given by :

$$\widehat{y} = \widehat{a}_0 + \widehat{a}_1 x$$

Determination and correlation coefficients

The line of fit $\widehat{y} = \widehat{a}_0 + \widehat{a}_1 x$ is called the regression line or least squares, for graphical presentation and the calculation of the empirical linear correlation coefficient prompts one to try to fit a simple linear regression model. According to [49], we recall the definition of the coefficient of determination .

Definition 2.2.2 *(**The coefficient of determination**) The measure R-squared (R^2)this coefficient expresses the part of variation of Y explained by the variation of X. It expresses the ratio between the variance of Y explained by the model and the total variance. It is an indicator of the goodness of fit of the regression equation. It allows to have a global idea of the fit of the model is defined by the following relation :*

$$R^2 := \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

where :

$SSR = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2$ *(Regression sum of squares)*

$SST = \sum_{i=1}^{n} (y_i - \overline{y})^2$ *(Total sum of squares)*

$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)$ *(Error sum of squares.)*

Since $0 \leq SSE \leq SST$ *, we note that:*

$$0 \leq R^2 \leq 1$$

Definition 2.2.3 ***(The linear correlation coefficient)*** *A measure of linear association between Y and X when both Y and X are random is the coefficient of correlation ρ.This measure is the signed square root of R^2 :*

$$\rho(X,Y) = \pm\sqrt{R^2},$$

where $-1 \leq \rho(X,Y) \leq 1$ *, this means that we have a relation of the type $y = a_1x + a_0$ between the variables X and Y .*

Tests and confidence intervals

Following the estimation of the regression coefficients, the statistical test is the second step following the regression. It allows to quantify if an explanatory variable has a statistically significant effect, in this section we have seen how to perform hypothesis tests on the parameters a_0 and a_1 of the simple regression model, as well as on how to build confidence intervals.

a) Slope and intercept parameters test :

The hypothesis test concerning the slope a_1 and the intercept a_0 is of the form respectively :

$$\begin{cases} H_0 : a_1 = 0 \\ H_1 : a_1 \neq 0 \end{cases} \quad \& \quad \begin{cases} H_0 : a_0 = 0 \\ H_1 : a_0 \neq 0 \end{cases}$$

Since that $\widehat{a}_1$ and $\widehat{a}_0$ are a linear combination of the observation Y_i, so $\widehat{a}_1$ and $\widehat{a}_0$ will be normally distributed and can be expressed as follows :

$$\widehat{a}_1 \rightsquigarrow \mathcal{N}\left(a_1, \frac{\sigma^2}{nS_x^2}\right) \qquad \Leftrightarrow \qquad \frac{\widehat{a}_1 - a_1}{\frac{\sigma}{\sqrt{n}S_x}} \rightsquigarrow \mathcal{N}(0,1)$$

$$\widehat{a}_0 \rightsquigarrow \mathcal{N}\left(a_0, \frac{\sigma^2}{n}\left[1 + \frac{\overline{x}^2}{S_x^2}\right]\right) \qquad \Leftrightarrow \qquad \frac{\widehat{a}_0 - a_0}{\frac{\sigma}{\sqrt{n}}\sqrt{1 + \frac{\overline{x}^2}{S_x^2}}} \rightsquigarrow \mathcal{N}(0,1)$$

Since σ is unknown, we replace it by S , and therefor we obtain :

$$T_{a_1} = \frac{\widehat{a}_1 - a_1}{\frac{S}{\sqrt{n}S_x}} \rightsquigarrow t_{n-2} \quad \& \quad T_{a_0} = \frac{\widehat{a}_0 - a_0}{\frac{S}{\sqrt{n}}\sqrt{1 + \frac{\overline{x}^2}{S_x^2}}} \rightsquigarrow t_{n-2}$$

Under the null hypothesis we'll find:

$$T_{a_1} = \frac{\widehat{a}_1}{\frac{S}{\sqrt{n}S_x}} \rightsquigarrow t_{n-2} \quad \& \quad T_{a_0} = \frac{\widehat{a}_0}{\frac{S}{\sqrt{n}}\sqrt{1 + \frac{\overline{x}^2}{S_x^2}}} \rightsquigarrow t_{n-2}$$

Accordingly, at the level of significance $\alpha \in [0,1]$, H_0 is to be rejected if:

$$|T_{a_1}| > t_{1-\frac{\alpha}{2}}(n-2) \qquad \& \qquad |T_{a_0}| > t_{1-\frac{\alpha}{2}}(n-2)$$

where, $t_{1-\frac{\alpha}{2}}(n-2)$ is the $(1-\alpha/2)$ percentile of the student distribution with $(n-2)$ degrees of freedom.

b) Confidence interval

The point value of an estimator is generally insufficient and it is necessary to add a confidence interval to the significance level α (or to the confidence level $1-\alpha$). It is interesting to give the confidence intervals of level of confidence ($1-\alpha$), of the parameters

a_0 and a_1,we can make the following probability statement :

$$P\left(|T_{a_1}| < t_{n-2,1-\frac{\alpha}{2}}\right) = 1-\alpha$$
$$P\left(|T_{a_0}| < t_{n-2,1-\frac{\alpha}{2}}\right) = 1-\alpha$$

Therefore, the $(1-\alpha)$ confidence limits for a_0 , a_1 respectively are :

$$\widehat{a}_1 \pm t_{n-2,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}S_x},$$
$$\widehat{a}_0 \pm t_{n-2,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\sqrt{1+\frac{\overline{x}^2}{S_x^2}}\ .$$

Finally, the residual variance is constant over the studied domain, meaning that all distributions of Y (y_i) must have the same standard deviation. In another way, it is precisely in the hypotheses that it is necessary to check the equality of the variances of errors. If the variance is not constant over the entire range, we use "*Least squares*" with the variance potentially inverted for weighting. In the next section, we'll take a look at this method.

2.2.2 The weighted least squares method

Weighted Least Squares (WLS) is an estimation technique which weights the observations proportional to the reciprocal of the error variance for that observation and so overcomes the issue of non-constant variance. This term was originally used in Nelder and Wedderburn [57],1972.

Definition 2.2.4 *(**WLS in Simple Regression**) Consider the following model :*

$$Y_i = a_0 + a_1 X_i + \varepsilon_i$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma^2/w_i)$ for known constants $w_1, ..., w_n$. The weighted least squares estim-

ates of a_0 *and* a_1 *minimize the quantity*

$$Q_w(a_0, a_1) = \sum_{i=1}^{n} w_i \left(y_i - a_0 - a_1 x_i\right)^2.$$

Remark 2.2.2 *:*

• *Since each weight is inversely proportional to the error variance, it reflects the information in that observation. So, an observation with small error variance has a large weight since it contains relatively more information than an observation with large error variance (small weight).*

• *The weights have to be known (or more usually estimated) up to a proportionality constant.*

Definition 2.2.5 *To estimate parameters* a_0 *and* a_1*, the WLS estimates are then given as :*

$$\hat{a}_1 = \frac{\sum\limits_{i=1}^{n} w_i(y_i - \overline{y}_w)(x_i - \overline{x}_w)}{\sum\limits_{i=1}^{n} w_i(x_i - \overline{x}_w)^2} \qquad \& \qquad \hat{a}_0 = \overline{y}_w - \hat{a}_1 \overline{x}_w$$

where $\overline{x}_w$ *and* $\overline{y}_w$ *are the weighted means with ;*

$$\overline{x}_w = \frac{\sum\limits_{i=1}^{n} w_i x_i}{\sum\limits_{i=1}^{n} w_i} \qquad \& \qquad \overline{y}_w = \frac{\sum\limits_{i=1}^{n} w_i y_i}{\sum\limits_{i=1}^{n} w_i}$$

Statistical properties of WLS estimates

• These estimators are unbiased estimators

• $Var(\hat{a}_1) = \dfrac{\sigma^2}{\sum w_i(x_i - \overline{x}_w)^2}$

• $Var(\hat{a}_0) = \left[\dfrac{1}{\sum w_i} + \dfrac{\overline{x}_w^2}{\sum w_i(x_i - \overline{x}_w)^2}\right]$

• The weighted error mean square $Q_w(\widehat{a}_0, \widehat{a}_1)/(n-2)$ also gives us an unbiased estimator of σ^2.

Definition 2.2.6 ***(General WLS Solution)*** *Let W be a diagonal matrix with diagonal elements equal to $w_1, ..., w_n$. The weighted residual sum of squares is defined by*

$$\begin{aligned} Q_w(\beta) &= \sum_{i=1}^{n} w_i(y_i - x_i^t\beta)^2 \\ &= (Y - X\beta)^t W (Y - X\beta) \end{aligned}$$

The general solution to this is

$$\widehat{\beta} = \left(X^t W X\right)^{-1} X^t W Y$$

Definition 2.2.7 ***(WLS as a Transformation)*** *In general suppose we have the linear model*

$$Y = X\beta + \varepsilon$$

where $Var(\varepsilon) = W^{-1}\sigma^2$. Let $W^{1/2}$ be a diagonal matrix with diagonal entries equal to $\sqrt{w_i}$. Then we have $Var(W^{1/2}\varepsilon) = \sigma^2 I_n$. Hence we consider the transformation

$$\acute{Y} = W^{1/2}Y, \quad \acute{X} = W^{1/2}X \text{ and } \acute{\varepsilon} = W^{1/2}\varepsilon$$

This gives rise to the usual least squares model

$$\acute{Y} = \acute{X}\beta + \acute{\varepsilon}$$

using the results from regular least squares we then get the solution

$$\widehat{\beta} = \left((\acute{X})^t X\right)^{-1} (\acute{X})^t \acute{Y} = \left(X^t W X\right)^{-1} X^t W Y$$

hence this is the weighted least squares solution.

Example 2.2.1 *: Recall from the model* $y_i = a_0 + a_1 x_i + \varepsilon_i$ *where* $Var(\varepsilon_i) = x_i^2\sigma^2$*, we can transform this into a regular least squares problem by taking*

$$y_i^{'} = \frac{y_i}{x_i} \quad x_i^{'} = \frac{1}{x_i} \quad \varepsilon_i^{'} = \frac{\varepsilon_i}{x_i}$$

Then the model is

$$y_i^{'} = a_1 + a_0 x_i^{'} + \varepsilon_i^{'} \quad \textit{where } Var(\varepsilon_i^{'}) = \sigma^2$$

The residual sum of squares for the transformed model is

$$\begin{aligned} Q(a_0, a_1) &= \sum_{i=1}^{n} \left(y_i^{'} - a_1 - a_0 x_i^{'} \right)^2 \\ &= \sum_{i=1}^{n} \left(\frac{y_i}{x_i} - a_1 - a_0 \frac{1}{x_i} \right)^2 \\ &= \sum_{i=1}^{n} \left(\frac{1}{x_i} \right)^2 (y_i - a_0 - a_1 x_i)^2 \end{aligned}$$

This is the weighted residual sum of squares with $w_i = 1/x_i^2$*, hence the weighted least squares solution is the same as the regular least squares solution of the transformed model.*

Choice of the weights

How should we choose the weights ? Gauss considered differences in precision of β assuming a known variance (σ^2) and generalized his method of least squares with weights as inverses of the square root of variances (Plackett, [48]1949).

In general, we will choose the weights w_i in simple regression as follows :

$$w_i = \frac{\sigma^2}{Var(\varepsilon_i)},$$

it is therefore necessary to evaluate the form of $Var(\varepsilon_i)$ as a function of x_i to know the weights to use.

• To define this form, we generally draw the graph of the residuals or of the variance of the residuals as a function of the explanatory variable x. We can then make assumptions about the form of $Var(\varepsilon_i)$ as a function of the x_i in order to make a weighted regression with each of the proposed forms and choose the one for which the graph of the residuals as a function of x is the best. We can also take as weight the inverse of the variances:

$$w_i = \frac{1}{Var(x_i)}.$$

In cases where the variance of ε_i is proportional to x_i, then

$$w_i = \frac{1}{x_i}.$$

In cases where the variance of ε_i is proportional to x_i^2, then

$$w_i = \frac{1}{x_i^2}.$$

Remark 2.2.3 *:*

1. *Another common case is where each observation is not a single measure but an average of n_i actual measures and the original measures each have variance σ^2.In that case, standard results tell us that*

$$Var(\varepsilon_i) = Var(y_i) = \frac{\sigma^2}{n_i},$$

thus we would use weighted least squares with weights $w_i = n_i$.

2. *In many real-life situations, the weights are not known, in such cases we need to estimate the weights in order to use weighted least squares.*

Example 2.2.2 *The data taken from Tomassone et al.,* [55], 1998.

We consider data comprising 10 observations with the explanatory variable X. The variable Y is generated using the following model :

$$y_i = 3 + 2x_i + \varepsilon_i,$$

where the ε_i are normally distributed $E(\varepsilon_i) = 0, et \;\; Var(\varepsilon_i) = (0.2x_i)^2$, we present the data thus generated in the following table :

x_i	1	2	3	4	5	6	7	8	9	10
Y_i	4.90	6.55	8.67	12.59	17.38	13.81	14.60	32.46	18.73	20.27

Table 2.1: Values x_i and Y_i generated by the model studied

A simple regression study always begins with a plot of the observations (x_i, y_i), $i = \overline{1, 10}$. This first representation makes it possible to know if the linear model is relevant.

Graphic Representation : in figure 2.2, we plot Y_i and individuals x_i.

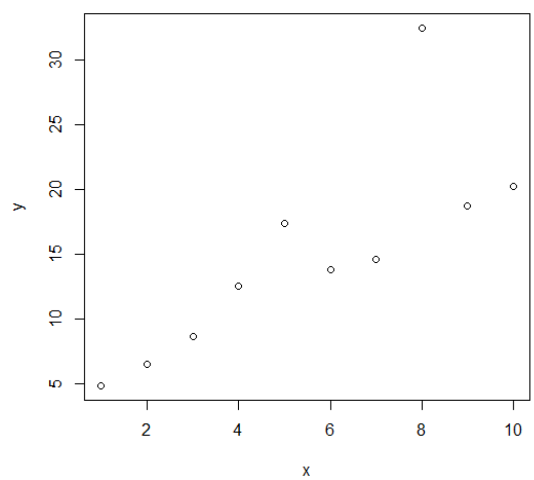

Figure 2.2: plot Y_i and individuals x_i

1. The least squares method :

The least squares method provides the following estimated coefficients on the example, for all $i = \overline{1, 10}$. The regression equation is ,

$$\hat{Y}_i = 3.49 + 2.09x_i$$

The estimated slope of the line : $\hat{a}_1 = 2.09$

The estimated y-intercept : $\hat{a}_0 = 3.49$

Least squares regression line :

We are looking for the line for which the sum of the squares of the vertical deviations of the points from the line is minimum. On the graph, we have drawn any line through the data and we represent the errors for some points, figure 2.3 below illustrates the regression line by least squares.

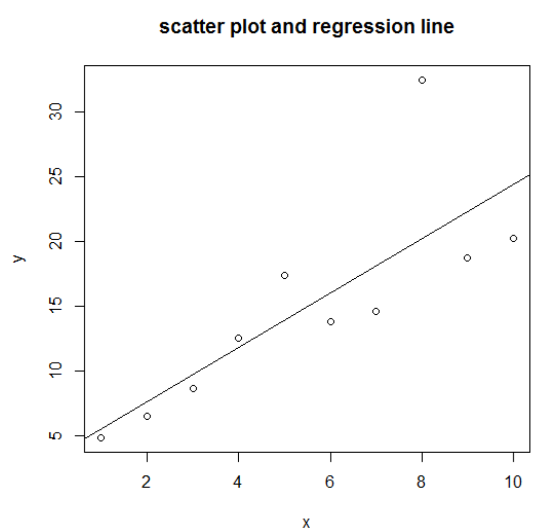

Figure 2.3: Linear regression line and scatter plot.

We first compute the residuals, ε_i, the basic regression is shown in table 2.2 :

the regression model explains 62.94% of the total variation.

$SSE = \sum_{i=1}^{10} \varepsilon_i^2$	212.39
$\widehat{\sigma}^2 = SSE\ /n - 2$	26.55
$R^2 = SSR\ /SST$	0.6294

Table 2.2: Regression results for the LS method

2. The weighted least squares method :

A weighted regression study, using the values $(1/x^2)$ as weights. These weights are known since they must be proportional to the true variances, the occurrence equal to $(0.2x_i)^2$. The weighted least squares method provides the following estimated coefficients on the example :

$$\hat{Y}_i = 2.53 + 2.28x_i.$$

In table 2.3, the regression using the method of weighted least squares :

$SSE = \sum_{i=1}^{10} \varepsilon_i^2$	3.75
$\widehat{\sigma}^2 = SCR\ /n - 2$	0.47
$R^2 = SSR\ /SST$	0.8611

Table 2.3: Regression results for the WLS method

the regression model explains 86.11 % of the total variation.

- The residuals are always immediately available, so we can graph them :

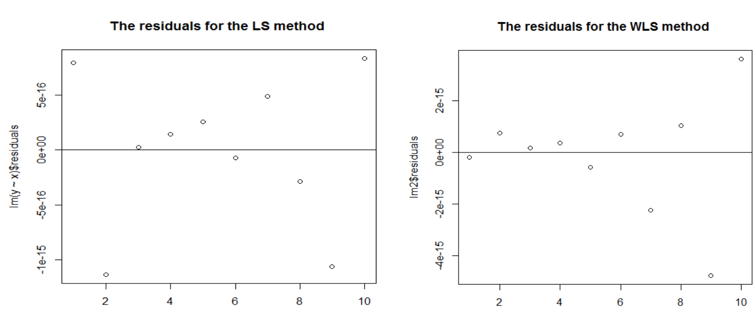

Figure 2.4: Scatter plot the residuals for the LS and WLS methods.

Remark 2.2.4 *On this basis, the following comments can be made:*

1. *All quantities related to the sum of the squares of the dependent variable are assigned by weights and are not comparable to those obtained by the least squares regression.*

2. *The estimated coefficients are relatively close to those of the least squares regression.*

Generally, the weighted least squares method, like the other least squares methods, is also sensitive to extreme values. We can find out in the following parts.

Chapter 3

A weighted least-squares estimation method for distributional parameters

In the fields of mathematics and statistics, regression procedures are often used for estimating distributional parameters. In this procedure, the distribution function is transformed to a linear regression model. The aim of the current chapter, we consider weighted least squares WLS estimation method, based on an different expressions of weight, for distributional parameters. The considered estimation method is then applied to the estimation of parameters of different distributions, such as Weibull, Gumbel and Pareto. We also extract the approximate results and explain the performance of these estimates in a simulation study.

3.1 Preliminary

A linear regression model was obtained, in which the dependent variable is a nonparametric estimate of the value of the distribution function at the ranked sample. Then, the estimates of least squares LS of the coefficients of the regression model become the estimates of the parameters of the statistical distribution. However, heteroscedasticity (non-constant variance) is present in the used regression model, whereby LS estimates lose the efficiency property. In such cases, the use of WLS regression to estimate the parameters of some distributions, such as the parameters of the Weibull, Gumbel and Pareto distributions, have been studied by regression estimation methods (Bergman [5]1986, Hossain and Howlader, [37]1996, Zhang et al., [59] 2007, [60] 2008, Zyl [64] 2012, Zyl and Schall [65] 2012, Kantar and Arik, [39] 2014, Kantar and Yildirim, [40] 2015, Lu and Tao [45] 2007.)

When performing a WLS, the variances of the dependent variables are unknown and must be estimated to perform of this method. Hung [33] 2001, Lu et al., [44] 2004, Zyl and Schall [65] 2012 emphasize that a weight function should be used when performing regression methods, and propose different weights using large sample properties of the empirical distribution function or order statistics, to stabilize the variance in order to perform the WLS estimation method.

In this chapter, we propose a weighted least squares WLS estimation method for distributional parameters. Also, knowing how to calculate the weights with two ideas : a weight function proposed by Bergman [5] 1986 and the idea of Zyl and schall [65] 2012. Simulation results showed that this method performs well with respect to some other existing methods .

3.2 Estimation of distributional parameters by regression models

In estimate the parameters of the considered distributions, the distribution functions are transformed into a linear regression model.

To motivate our methodology, the Weibull distribution is one of the widely used distributions in technical practice. This distribution was first introduced by Walodi Weibull

(1887−1979), who used it in the theory of reliability. We consider that the cdf is presented:

$$F(x, \lambda, \alpha) = 1 - e^{-(\lambda x)^{\alpha}};\ \text{for } x > 0, \tag{3.1}$$

where : λ is the scale parameter, α is the shape parameter. After some algebraic manipulation, equation 3.1 can be expressed as follows :

$$\ln\left[-\ln(1 - F(x, \lambda, \alpha))\right] = \alpha \ln \lambda + \alpha \ln x, \tag{3.2}$$

For a sample of size n and $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$, equation 3.2 the regression model can be rewritten as follows :

$$\ln\left[-\ln(1 - F(x_{(i)}))\right] = \alpha \ln \lambda + \alpha \ln x_{(i)}, \tag{3.3}$$

where i the order number .

For estimates of $F(x_{(i)})$, Bernard and Bosi-Levenbach [6],1953 using the following methods of estimation summary in table 3.1, where $\widehat{F}_i$ is some non-parametric estimate of $F(x_{(i)})$:

Method	$\widehat{F}_i$
Mean Rank	$\dfrac{i}{(n+1)}$
Median Rank	$\dfrac{i-0.3}{(n+0.4)}$
Symmetric CDF	$\dfrac{i-0.5}{n}$

Table 3.1: Methods of estimation

For complete samples, $\dfrac{i}{(n+1)}$ and $\dfrac{i-0.3}{(n+0.4)}$ are generally used (Zyl [64] 2012,, Zyl and Schall [65], 2012).

If we replace $\ln(-\ln(1-\widehat{F}_i))$ with $Y_{(i)}$, $\alpha \ln \lambda$ with a, α with b, and $\ln x_{(i)}$ with $X_{(i)}$, the regression model with error term occurs as:

$$Y_{(i)} = a + bX_{(i)} + \varepsilon_{(i)} \tag{3.4}$$

For the Gumbel or extreme value distribution type **I**, the cdf is given by

$$F(x, \mu, \beta) = \exp\left(-e^{\frac{-(x-\mu)}{\beta}}\right), \tag{3.5}$$

where β is the scale parameter and μ is the shape parameter. Equation 3.5 can be linearized as follows:

$$-\ln\left[-\ln(F(x, \mu, \beta))\right] = \frac{x}{\beta} - \frac{\mu}{\beta} \tag{3.6}$$

Equation 3.6 may be written as:

$$-\ln\left[-\ln(F(x_{(i)}))\right] = \frac{x_{(i)}}{\beta} - \frac{\mu}{\beta} \tag{3.7}$$

If we replace $x_{(i)}$ with X_i, $-\frac{\mu}{\beta}$ with a, $\frac{1}{\beta}$ with b and $-\ln\left[-\ln(\widehat{F}_i)\right]$ with Y_i, the linear regression model is obtained for the Gumbel distribution.

The cdf of the Pareto random variable is given as follows:

$$F(x, \alpha, \beta) = 1 - \left(\frac{\beta}{x}\right)^{\alpha}, \tag{3.8}$$

where β is the scale parameter and α is the shape parameter.

The Pareto distribution, which is generally used to model extreme values, is skewed and heavy-tailed.

Similar to the Weibull and Gumbel distributions, the obtained regression model for the

Pareto distribution is presented as follows:

$$\ln(1 - F(x)) = \alpha \ln x - \alpha \ln \beta \tag{3.9}$$

For the ordered sample, the regression model is rewritten as:

$$\ln\left(1 - F(x_{(i)})\right) = \alpha \ln x_{(i)} - \alpha \ln \beta, \tag{3.10}$$

If we replace $\ln x_{(i)}$ with X_r, $-\alpha \ln \beta$ with a, α with b and $\ln\left(1 - F(x_{(i)})\right)$ with Y_i, the linear regression model is obtained for the Pareto distribution.

Generally, using the regression model given in 3.4, we can easily use LS and other regression estimation methods to estimate distribution parameters.

3.3 Expressions of the weights

Weights expressions in regression are required in probability plotting type regression, there are several expressions performing a weighted regression. In this section, we focus on expressions of weights proposed by *Bergman* [5] 1986, in order to, estimating parameters using a weight function. Also, expressions *Zyl and Schall* [65] 2012 using large sample properties of order statistics.

Next, we will present the different expressions for weights :

3.3.1 Expressions using a weights functions

One problem with the linear regression is that each datum point has been given the same weight. It has been shown that this assumption is erroneous Bergman [5], 1986. If a linear regression is to be performed in a correct way it is obvious that a weight function should be used, as he proposed an analytic expression for the appropriate weight function by using the theory of propagation of errors. It is the intention of this section to find exact expressions for estimate the parameters of the Pareto distribution proposed by *Lu and Tao* [45] 2007, using this weight function.

Suppose that random variables $X_1, X_2, ..., X_n$ are independent and identically distributed as 3.9. If the regression model is in the form of equation 3.4. In the following:

- The weighted sum is given by

$$Q = \sum_{i=1}^{n} w_i (Y_{(i)} - Y(x_{(i)}))^2, \; Y(x_{(i)}) = \alpha \ln x_{(i)} - \alpha \ln \beta$$

- Bergman[5] emphasized that it is unreasonable for $x_{(i)}$ to have the same weight in Equation 3.10 and proposed that a weight function should be used in performing the linear regression. The weight factor Bergman proposed is

$$w_i = \left[\left(1 - \widehat{F}(x_{(i)})\right) \ln \left(1 - \widehat{F}(x_{(i)})\right) \right]^2, \quad i = \overline{1, n}$$

- After minimizing Q, we obtain the WLS estimators of $\hat{\beta}$ and $\hat{\alpha}$ which are respectively as follows :

$$\hat{\beta} = -\frac{\sum_{i=1}^{n} w_i Y_i \sum_{i=1}^{n} w_i X_i - \sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i Y_i X_i}{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i X_i^2 - \left(\sum_{i=1}^{n} w_i X_i\right)^2}$$

$$\hat{\alpha} = \exp \left[\frac{\sum_{i=1}^{n} w_i Y_i - \hat{\beta} \sum_{i=1}^{n} w_i X_i}{\hat{\beta} \sum_{i=1}^{n} w_i} \right]$$

Remark 3.3.1 *From the property of Pareto distribution, we learn that* $Y_i = -\ln(1 - F_X(x_{(i)}))$ *is standard exponential distribution (i.e.:* $E(Y_i) = \sum_{j=1}^{i} \frac{1}{(n-j+1)}$ *and* $Var(Y_i) = \sum_{j=1}^{i} \frac{1}{(n-j+1)^2}$ *) and the variance of the order statistics does not satisfy the condition of being constant (according to, Balakrishnan and Cohen [3], 1991).That is, the weight of each point is not identical.*

3.3.2 Expressions using derivation of weights for least-squares

The expressions for weights used in least squares regression are from the large sample variances, by deriving the inverse of the approximate variance of the scalar function Λ of the order statistic. It is assumed that the derivative of $\wedge$ is continuous at the expected value of the order statistic.

Below we mention the basic properties of this expression, and then continue with some applications for estimating distributional parameters.

Proposition 3.3.1 ***(Zyl & Schall,2012)*** *Let $x_1, x_2, ..., x_n$ be an i.i.d sample from a distribution F with corresponding o.s $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$. The WLS expression to minimize with respect to the parameters is*

$$\sum_{i=1}^{n} w_i \left[E(\Lambda(x_{(i)})) - \Lambda(x_{(i)})\right]^2,$$

where the weight for the i^{th} squared residual $u_i^2 = \left[\Lambda(X_i) - \Lambda(x_{(i)})\right]^2$ is

$$w_i = 1/Var(\Lambda(x_{(i)})), \qquad i = \overline{1,n}$$

Corollary 3.3.1 ***(Zyl & Schall,2012)*** *The statistics $F(x_{(1)}), ..., F(x_{(n)})$ are beta distributed with $F(x_{(i)}) \sim Beta(i, n-i+1)$*

$$E(F(x_{(i)})) = \frac{i}{(n+1)} = m_i$$
$$Var(F(x_{(i)})) = \frac{i(n-i+1)}{(n+2)(n+1)^2} = \frac{m_i(1-m_i)}{(n+2)}$$

Let X_i be such that $F^{-1}(X_i) = i/(n+1)$ Asymptotically , for $i = \overline{1,n}$

$$\sqrt{n}\left(x_{(i)} - X_i\right) \xrightarrow{d} \mathcal{N}(0, \sigma_i^2) \text{ with } \sigma_i^2 = \frac{m_i(1-m_i)}{\left(F'(X_i)\right)^2}$$

provided $F^{'}(m_i) = f(m_i)$ *exists. If the first derivative of* Λ *is continuous at* X_i *and* $\Lambda^{'}(X_i) \neq 0$. *Then*

$$\sqrt{n}\left(\Lambda(x_{(i)}) - \Lambda(X_i)\right) \xrightarrow{d} \mathcal{N}\left(0, Var(x_{(i)})\left(\frac{d\Lambda(x_{(i)})}{dx_{(i)}}\right)^2\right)$$

It follows that

$$Var(\Lambda(x_{(i)})) \approx \frac{m_i(1-m_i)}{(n+2)\,(f(X_i))^2}\left[\frac{d\Lambda(x_{(i)})}{dx_{(i)}}\right]^2_{x_{(i)}=x_i} \tag{3.11}$$

Remark 3.3.2

1. *The function* $\wedge$ *need not be a linear function of the order statistics.*

2. *Order statistics and thus also functions of order statistics are asymptotically independently distributed . In this work we treat the residuals,* $u_i = \Lambda(X_i) - \Lambda(x_{(i)})$ *of the least squares regression as if they were independent.*

Finally, readers interested on this properties of this expression can refer to [DasGupta [14], 2008 page 93,Kendall, Stuart and Ord [41], 1987 page 462] and references therein.

Applying this expression for estimation of parameters of Weibull and Gumbel distributions proposed by *Zyl and Schall* [65] 2012, yield the following results:

According to Weibull distribution ;

Let $\Lambda(x_{(i)}) = \ln\left[-\ln(1 - F(x_{(i)}))\right]$ and $\mu_i = E(\Lambda(x_{(i)}))$. Then

$$\begin{aligned}\Lambda(x_{(i)}) &= \alpha \ln \lambda + \alpha \ln x_{(i)} \\ \Lambda(x_{(i)}) + (\mu_i - \mu_i) &= \alpha \ln \lambda + \alpha \ln x_{(i)} \\ \mu_i &= \alpha \ln \lambda + \alpha \ln x_{(i)} + (\mu_i - \Lambda(x_{(i)})) \\ \mu_i &= \alpha \ln \lambda + \alpha \ln x_{(i)} + u_i\end{aligned}$$

where $u_i = \mu_i - \alpha \ln \lambda - \alpha \ln x_{(i)}$, $i = \overline{1, n}$, are the residuals for the regression and the weights are the inverses of the variances of the residuals.

The approximate variance of $\ln\left[-\ln(1-F(x_{(i)}))\right]$ by 3.11, is :

$$\begin{aligned} Var(\ln\left[-\ln(1-F(x_{(i)}))\right]) &\approx \frac{m_i(1-m_i)}{(n+2)\left(f(X_i)\right)^2}\left[\frac{d\ln\left[-\ln(1-F(x_{(i)}))\right]}{dx_{(i)}}\right]^2_{x_{(i)}=x_i} \\ &\approx \frac{m_i(1-m_i)}{(n+2)\left(\ln(1-m_i)\right)^2(1-m_i)^2} \\ &\approx \frac{i}{(n+2)\left(\ln(\frac{n-i+1}{n+1})\right)^2(n-i+1)} \end{aligned}$$

For this reason, the WLS regression equation is solved by letting : $\widehat{\theta}_{WLS} = \left(X^tWX\right)^{-1}X^tWY$,

where W matrix is diagonal, $X = \begin{pmatrix} 1 & \ln(x_{(1)}) \\ : & : \\ 1 & \ln(x_{(i)}) \end{pmatrix}$,

$Y^t = \left(\ln(-\ln(1-\widehat{F}_1)), ..., \ln(-\ln(1-\widehat{F}_i))\right)$ and $\widehat{\theta} = \begin{pmatrix} \widehat{\alpha}\ln\widehat{\lambda} \\ \widehat{\alpha} \end{pmatrix}$; $\widehat{\lambda} = \exp\left(-\widehat{\theta}_1/\widehat{\theta}_2\right)$, $\widehat{\alpha} = \widehat{\theta}_2$.

Then it follows that,

$$\widehat{\theta}_2 := \frac{\sum_i w_i(x_i-\bar{x})(y_i-\bar{y})}{\sum_i w_i(x_i-\bar{x})}, \quad \widehat{\theta}_1 := \bar{y} - \widehat{\theta}_2\bar{x} \quad \text{with } w_i = 1/Var(\Lambda(x_{(i)}))$$

According to Gumbel distribution ;

In order to find the expressions for estimate $\widehat{\beta} = \left(\frac{1}{\widehat{\theta}_2}\right)$ and $\widehat{\mu} = \left(\frac{\widehat{\theta}_1}{\widehat{\theta}_2}\right)$ by WLS. Using similar arguments as for the Weibull regression the equation used to estimate the parameters with $\mu_i = E(\Lambda(x_{(i)}))$ is :

$$\mu_i = u_i - \frac{x_{(i)}}{\beta} + \frac{\mu}{\beta} \quad \text{where } u_i = -\ln(-\ln F(x_{(i)})) - \frac{x_{(i)}}{\beta} + \frac{\mu}{\beta}$$

The weights are the inverses of the variances of the residuals, is defined by :

$$\begin{aligned} Var(-\ln\left[-\ln(F(x_{(i)}))\right]) &\approx \frac{1-m_i}{(n+2)\left(\ln(m_i)\right)^2} \\ &\approx \frac{n-i+1}{(n+1)(n+2)\left(\ln(\frac{i}{n+1})\right)^2} \end{aligned}$$

For more details and proofs, the reader can refer to the following work: Gradshteyn and Ryzhik [30]1980, Zyl [63] 2016.

Remark 3.3.3

• *The variances of the residual values are functions of the order statistics and are independent of the parameters of the distribution can easily be found by using simulation by comparing the approximation of the variance 3.11 with to the true variance.*

•• *As theoretical results, looking the work of Zyl and Schall [65] 2012, Zyl[63] 2016,* $m_i = \frac{i}{(n+1)}$ *and the Bernard median ranks were used in the approximation of the variance, where the approximation was good even for a relatively small sample size, and that Barnard's median ranks result in better approximations of the variances for any parameters of the two distributions considered, the Weibull and the Gumbel distributions.*

3.4 Simulation results

In this section, we present some simulation results which are designed to evaluate the feasibility of the proposed WLS (*Zyl and Schall*) estimation method, by comparing with the maximum likelihood estimation (MLE) for the parameters of the Weibull and Gumbel distributions.

In addition, we compare the performance of the proposed WLS (*Lu and Tao*) with LS estimation (regression of Y on X) and MLE for the parameters of Pareto distribution.

In the following tables (3.2 , 3.3 and 3.4), the estimation was performed on simulated samples based on the mean square error (MSE) and bias.

In table 3.2 samples were generated from the Weibull distribution with $\alpha = 1.5$, $\lambda = 1$. It can be seen that : for all samples sizes, the weighted regression method outperforms MLE especially with respect to bias in the estimation of the shape and scale parameters, and the use of the $Bernard$ weights decreased the bias. Also, the MSE of the MLE outperforms the weighted regression method with respect to the estimation of the shape parameter. In addition, the MSE is small for the weighted methods of estimating the scale parameter.

Methods	MSE (α)	Bias(α)	MSE(λ)	Bias(λ)
$n = 10$				
MLE	0.0231	0.0519	0.3224	0.2213
WLS	0.0529	0.0230	0.2264	−0.0316
WLS $(Bernard)$	0.1856	0.0219	0.0510	−0.0185
$n = 30$				
MLE	0.0310	0.0040	0.1321	0.1024
WLS	0.0312	0.0039	0.1134	−0.0521
WLS $(Bernard)$	0.0529	0.0034	0.0170	−0.0496
$n = 50$				
MLE	0.0152	0.0039	0.0734	0.0594
WLS	0.0156	0.0042	0.0677	−0.0223
WLS $(Bernard)$	0.0409	0.0038	0.0131	−0.0199
$n = 100$				
MLE	0.0074	0.0067	0.0296	0.0315
WLS	0.0075	0.0069	0.0292	−0.0079
WLS $(Bernard)$	0.0159	0.0065	0.0052	−0.0070

Table 3.2: Bias and MSE of estimated parameters of the Weibull distribution(10000 simulated samples).

From the simulation results presented in table 3.3 samples were generated from a Gumbel distribution with $\mu = 0.5$ and $\beta = 2$. It was found that, with respect to bias, the weighted estimate outperforms the ML estimation, in estimating shape and scale parameters.

Also, it can be seen a small MSE for the weighted regression methods for both shape and scale parameters and it is better to use $Bernard$'s median ranks when calculating the weights.

Methods	MSE (μ)	Bias(μ)	MSE (β)	Bias(β)
$n = 10$				
MLE	0.4812	0.0353	0.7622	0.2213
WLS	0.4675	0.0230	0.7333	−0.0356
WLS ($Bernard$)	0.4755	0.0209	0.4724	−0.0329
$n = 30$				
MLE	0.1463	0.0045	0.1583	0.1024
WLS	0.1444	0.0039	0.1526	−0.0622
WLS ($Bernard$)	0.1455	0.0035	0.1186	−0.0610
$n = 50$				
MLE	0.0956	0.0040	0.1284	0.0597
WLS	0.0913	0.0037	0.1196	−0.0249
WLS ($Bernard$)	0.0908	0.0034	0.0824	−0.0208
$n = 100$				
MLE	0.0498	0.0076	0.0397	0.0415
WLS	0.0465	0.0069	0.0340	−0.0083
WLS ($Bernard$)	0.0467	0.0063	0.0309	−0.0079

Table 3.3: Bias and MSE of estimated parameters of the Gumblel distribution(10000 simulated samples).

The performance the considered WLS is evaluated for the shape parameter of the Pareto distribution, which is summarized in Table 3.4. Moreover the proposed WLS estimation shows better performance next to LSE and MLE for most of the considered sample sizes and shape parameter cases.

In conclusion, the results of the simulations demonstrate that the considered WLS better performance than certain alternative estimation methods in terms of MSE and bias for most of the considered sample sizes, scale and shape cases.

Methods	shape			
	0.5		1	
	MSE	Bias	MSE	Bias
$n = 10$				
MLE	0.37793	−0.23616	0.64652	−0.36161
LSE	0.33463	0.12155	0.55351	0.13296
WLS	0.20697	−0.00887	0.40388	−0.02669
$n = 30$				
MLE	0.21024	−0.04675	0.32873	−0.08123
LSE	0.22036	0.03511	0.34982	0.05828
WLS	0.10046	0.00373	0.20983	0.00735
$n = 50$				
MLE	0.08017	−0.03157	0.2688	−0.05252
LSE	0.01483	0.03255	0.29834	0.05296
WLS	0.08730	0.00329	0.2636	0.00510
$n = 100$				
MLE	0.06421	−0.02110	0.21689	−0.02091
LSE	0.07876	0.02781	0.24596	0.04425
WLS	0.06487	0.00210	0.20061	0.00384

Table 3.4: Bias and MSE of the estimated shape parameters of the Pareto distribution.(10000 simulated samples).

Part II

Main results

Chapter 4

Heavy tail index estimator through weighted least-squares rank regression

The main aim of this chapter is to propose a weighted least square estimator based method to estimate the shape parameter of the Frechet distribution by deriving approximate weights to stabilize the variances . A simulation study was performed to evaluate the behavior of the proposed estimator, it is found that the considered weighted estimation method shows better performance than other methods in terms of bias and root mean square error, and in the case of real data.

4.1 Introduction

In many theoretical concepts, the parametric estimating distribution methods have received great interest, among them are : Maximum likelihood estimation (MLE) method which has good theoretical properties for large sample sizes and is often preferred. On the other hand, the use of regression depends on a probability plot to estimate the parameters of statistical distributions because the procedure for its implementation is simple in cases of complete and censoring data. Where it represents the linear regression model, and its dependent variable is the nonparametric estimate for the value of the distribution function at the ranked sample, is obtained. From it, the estimates of the least squares of the parameters of the resulting regression model become the estimates of the parameters of the studied statistical distribution.

The Frechet (extreme value type **II**) distribution is one of the probability distributions used to model extreme events. The extreme value distribution is becoming increasingly important in engineering statistics as a suitable distribution to represent phenomena with usually large maximum observations, introduced by French mathematician Maurice Frechet in 1927.

Let $(X_1, X_2, ..., X_n)$ denotes a sample of size n from a Frechet distribution F. The probability density function (pdf) with shape parameter $\alpha > 0$ is,

$$f(x; \alpha) = \alpha x^{-(\alpha+1)} \exp(-x^{-\alpha}), \quad for\ x > 0 \tag{4.1}$$

the cumulative distribution function (cdf) is given by :

$$F(x; \alpha) = exp(-x^{-\alpha}) \tag{4.2}$$

The principles of least squares estimation (LSE) are independently discovered by Gauss [25], 1795; Legendre [43], 1805 and Adrain [1] 1808, based on the relationship between the empirical cumulative distribution function (cdf) and the order statistics are frequently used to estimate parameters of distributions. The distribution function can be transformed to a linear regression model, if it can be written as an explicit function.

The weighted least squares method (WLS) is applied for parameter estimation, this method is comparatively concise and easy to perceive. In the literature, WLS estimation can be a better alternative that is superior to the existing methods : some research has been conducted on the Frechet distribution where Annasaheb and Girish [2], 2018 studied the performance of three different estimation methods of scale parameter LS, WLS and MLE, for two parameters Frechet distribution, where the weights proposed by Bergman [5], 1986 and the results of the Monte Carlo simulation were show that the MLE method was the best as compared to LS and WLS method in terms of bias as well as mean square error.

The rest of this chapter is organized as follows : in the second section, we state our estimators and main results. This is followed by a simulation study of our proposed estimator where we discuss its behavior with a illustrative example from Danish data.

4.2 Estimators and main results

In this section, we describe the methods of estimation for the shape parameter Frechet distribution.

4.2.1 Least squares method

Least squares, or least sum of squares, requires that a straight line be fitted to a set of data points, such that the sum of the squares of the distance of the points to the fitted line is minimized.

Suppose that random variables $X_1, X_2, ..., X_n$ are independent and identically distributed from the Frechet distribution. After algebraic manipulation, Equation 4.2 can be linearized as follows :

$$-\log(-\log(F(x)) = \alpha \log(x)$$

We consider that random variables $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ be the order statistics of $x_1, x_2, \ldots, x_n$, the regression model is rewritten as:

$$-\log(-\log(F(x_{(i)}; \alpha))) = \alpha \log(x_{(i)}) \tag{4.3}$$

Comparing equation 4.3 with $Y_i = \alpha X_i$, we get $Y_i = -\log(-\log(F(x_{(i)};\alpha)))$ and $X_i = \log(x_{(i)})$, the regression model with error term occurs as :

$$Y_{(i)} = \alpha x_{(i)} + \varepsilon_{(i)} \tag{4.4}$$

On the other hand, the error term of the model given in equation 4.4 is not identically distributed as mentioned model have no equal variance. This situation may adversely affect the LSE. In such cases, alternative estimation approaches to stabilize variances should be used.

In estimation, the sum of the squares of the errors, which is defined below, should be minimized

$$\min_{\alpha} \sum_{i=1}^{n} (Y_i - \alpha \log(x_{(i)}))^2 \quad \text{with} \quad Y_i = -\log(-\log(F(x_{(i)};\alpha))) \tag{4.5}$$

The estimator of $F(x_{(i)})$ can be considered to follow the mean rank estimator :

$$\widehat{F}(x_{(i)}) = \frac{i}{n+1}$$

where i is the rank of the data point in the sample in ascending order and $\widehat{F}_i$ is non-parametric estimate of $F(x_{(i)};\alpha)$. See Barnard [6].

Therefore, the estimate the parameter α is given by differentiating equation 4.5 partially α and equaling to zero, to estimate of the EVI_γ : $(\gamma = 1/\alpha)$ by LSE is :

$$\widehat{\gamma}_{LSE} = \frac{\sum_{i=1}^{n} -\log(-\log \widehat{F}(x_{(i)})) \log(x_{(i)})}{\sum_{i=1}^{n} \left(\log(x_{(i)})\right)^2}$$

4.2.2 Weighted Least Squares method

One of the main advantages of using regression procedure for estimating parameter is that its implementation is simple for Frechet distribution, the order statistics $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ denotes a sample of size n from a Frechet distribution F, the regression model based on equation

$$Y_{_{(i)}} = \alpha X_{_{(i)}}, \tag{4.6}$$

called regression of Y on X by Zhang [59], knowing that the order statistics $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ do not have constant variance, nor do the log transformed order statistics X, so that the regression model 4.6 is non-homogeneous.

Equation 4.3 with error term yield the following equation and replacing $F(x_{(i)})$ by its estimate, called $\hat{F}_i$, we obtain the equation

$$-\log(-\log(\hat{F}(x_{(i)})) = \alpha\log(x_{(i)}) + \varepsilon_i \tag{4.7}$$

To estimate $\widehat{\alpha}$ of the regression parameter α, than the regression model can be expression to minimize the function

$$\min_{\alpha}\sum_{i=1}^{n} w_i\left[Y_{(i)} - \alpha\log(x_{(i)})\right]^2,$$

where w_i is the weight factor $i = 1, ..., n$.

Next to, in order to calculate weights, using the large sample properties of the empirical distribution function or order statistics, and by deriving least squares weights from the large sample variances, using the approximate inverse of the variance of the scalar function Λ, to stabilize the variances in order to perform the WLS estimation method suggested by Zyl and Schall [65] 2012, specified in the following formula

$$Var(\Lambda(x_{(i)})) \approx \frac{m_i(1-m_i)}{(n+2)\left(f(x_{(i)})\right)^2}\left[\frac{d\Lambda(x_{(i)})}{dx_{(i)}}\right]^2_{x_{(i)}=x_i} \tag{4.8}$$

Furthermore if $\Lambda(x_{(i)})$ is of the form as $\Lambda(x_{(i)}) = \Lambda(F(x_{(i)}))$, it can be seen that

$$\begin{aligned} \left[\frac{d\Lambda(x_{(i)})}{dx_{(i)}}\right]^2_{x_{(i)}=x_i} &= \left[\frac{d\Lambda(F(x_{(i)}))}{dF(x_{(i)})}\frac{dF(x_{(i)})}{dx_{(i)}}\right]^2_{x_{(i)}=x_i} \\ &= f\left(x_{(i)}\right)\left[\frac{d\Lambda(F(x_{(i)}))}{dF(x_{(i)})}\right]^2_{x_{(i)}=x_i}, \end{aligned}$$

So, the relationship $-\log(-\log(F(x)) = \alpha\log(x)$ is used to perform rank regression, the approximate variance of $-\log(-\log(F(x_{(i)};\alpha))) = \alpha\log(x_{(i)})$ using the formula 4.8 is

$$\begin{aligned} Var(-\log(-\log(F(x_{(i)})))) &\approx \frac{m_i(1-m_i)}{(n+2)}\left[\frac{d\left[-\log(-\log(F(x_{(i)};\alpha)))\right]}{d\left(x_{(i)}\right)}\right]^2 \\ &\approx \frac{m_i(1-m_i)}{(n+2)}\frac{1}{m_i^2}\ ,\ \ m_i = \frac{i}{n+1} \\ &\approx \frac{i}{(n+1-i)^2}, \end{aligned}$$

therefore, we get the weights are independent of the parameter of the considered distribution.

In addition, the linear regression model given in 4.6, The weighted least-squares regression equation is solved by letting

$$\begin{aligned} &Y_i = Y^t = (-\log(-\log(\hat{F}_1)), ..., -\log(-\log(\hat{F}_n)), \\ &X_i = X^t = (\log(x_{(1)}), ..., \log(x_{(n)}))\text{ and} \\ &w = diag\left(w_1, w_2, ..., w_n\right),\ w_i = \tfrac{(n+1-i)^2}{i},\ i = 1, ..., n \end{aligned}$$

which is solved by

$$\widehat{\alpha} = (X^t w X)^{-1} X^t w Y.$$

Finally, we build our estimator $\hat{\gamma}_{WLS}$ as follow :

$$\hat{\gamma}_{WLS} := \frac{\sum\limits_{i=1}^{n} -w_i\log(x_{(i)})\log(-\log\hat{F}(x_{(i)}))}{\sum\limits_{i=1}^{n} w_i(\log(x_{(i)}))^2},\ w_i \approx 1/Var(-\log(-\log(F(x_{(i)})))).$$

In Table 4.1, we will show values $\hat{\gamma}_{LSE}$ and $\hat{\gamma}_{WLS}$ by changing values of $\gamma = (1.67,\ 1.11,\ 0.5)$ and sample size $n = (10;\ 20;\ 30;\ 50;\ 100;\ 200;\ 500;\ 1000;\ 2000)$.

4.3 Simulation study and application

4.3.1 Performance of the estimator

In this section, we examines the performance of our estimators $\hat{\gamma}_{LSE}$ and $\hat{\gamma}_{WLS}$ against the maximum likelihood estimator $\hat{\gamma}_{MLE}$ and Maximum product of spacing estimation $\hat{\gamma}_{MPSE}$, by simulation studies. A common approach to select the best method is the Monte Carlo simulation by using appropriate criteria: bias and mean squared error MSE [39].

We propose a Monte Carlo study of 10000 randomly generated samples, for each sample sizes ranging from $n = 10, 20, 30, 50, 100, 200, 500, 1000$ to 2000 for Frechet distribution and the shape parameters are considered as $\gamma = (1.67;\ 1.11;\ 0.5)$. The performance of this new estimator named by $\hat{\gamma}_{WLS}$ is evaluated in terms of bias and root mean squared error (RMSE) which are summarized in table 4.2.

The Bias of an estimator is $Bias(\hat{\gamma}) = E(\hat{\gamma}) - \gamma$. The $RMSE$ is defined as root of the sum of the variance and the squared bias of an estimator.

4.3.2 Results and discussion

According to bias criterion :

We evaluate the estimator WLS the proposed in this study in term of bias criterion, is best for the small sample size $n = 10$ and it is the best performer next to the LSE, MLE and $MPSE$. For other size $n > 10$ and in all cases of shape parameters we shows that in general the estimator WLS is clearly the best estimator in term of bias next to the MLE, $MPSE$ and LSE. In addition, bias decreases with increasing sample size and shape parameters cases.

According to the $RMSE$ criterion :

For the sample size $n = 10$ and for $\gamma = (1.67; 0.5)$, the proposed WLS shows smaller than $RMSE$ of the LSE, $MPSE$ and MLE, also for $\gamma = 1.11$ the $RMSE$ of MLE it's smaller than $RMSE$ of the $MPSE$, LSE and the WLS.

For $n > 10$ we show the $RMSE$ of LSE it's larger than $MPSE$, MLE and WLS for each shape parameters cases. Since the $RMSE$ of the WLS is asymptotically the best, it can be seen from analysis that MLE and $MPSE$ have better performance as the sample size increases the $RMSE$ decreases in each methods and shape parameters cases, thus we conclude that there are accurate increments of the parameters.

4.3.3 Real data example

As a real application, We take 2167 observations from the Danish data that describe large fire insurance claims in Denmark from Thursday 3^{rd} January 1980 until Monday 31^{st} December 1990 available in "*evir*" package of the Rsoftware [31].This data has been used by many value theories in an important application context.

In this section, we are concerned performance of the proposed estimator in weekly and monthly maximum losses during the mentioned period. There are 310 weekly maxima and 132 monthly maxima from the given 2167 observations which would provide an excellent example of the use of extreme as all studies confirm that the Danish data show a heavy tail with an index between 1 and 2.

This allows us to fit the data to heavy-tailed models with the proposed estimator which meets the objective of this study and compare it with new bias-reduced estimator for μ in the case of infinite second moment proposed by Brahimi et al., [8], 2013 (see table 7) defined by the following formula

$$\hat{\mu} := (k/n)(n\hat{c}/k)^{1/\widehat{\alpha}}\left(\frac{\widehat{\alpha}}{\widehat{\alpha}-1}+\frac{\hat{d}\hat{c}^{-\hat{\beta}/\widehat{\alpha}}(k/n)^{\hat{\beta}/\widehat{\alpha}-1}}{\hat{\beta}-1}\right)+\frac{1}{n}\sum_{i=k+1}^{n}X_{n-i+1,n}$$

Our case study is mostly based on samples from the Frechet distribution 4.2 with shape parameter $\alpha = 1.5$ $(\gamma = 1/\alpha)$ we then calculate estimate of shape parameter using the previously mentioned estimation method in this study, see Table 4.3.

n	Methods	$\gamma = 1.67$	$\gamma = 1.11$	$\gamma = 0.5$
10	MLE	1.32995	0.95437	0.45015
	$MPSE$	1.30107	1.89420	0.4126
	LSE	2.49297	1.61866	0.49807
	WLS	2.06358	1.48917	0.47880
20	MLE	1.49870	0.99107	0.52681
	$MPSE$	1.46761	0.95218	0.49972
	LSE	2.09857	1.66033	0.70631
	WLS	2.04806	1.34121	0.68271
30	MLE	1.60355	1.07210	0.56824
	$MPSE$	1.54872	1.01028	0.49935
	LSE	2.16820	1.35582	0.70790
	WLS	2.06776	1.28703	0.60878
50	MLE	1.44972	1.05178	0.47361
	$MPSE$	1.42337	1.01293	0.44725
	LSE	1.48879	1.37017	0.47361
	WLS	1.44812	1.04935	0.45039
100	MLE	1.52108	1.04215	0.45006
	$MPSE$	1.41076	1.03052	0.43118
	LSE	1.38191	1.13933	0.47727
	WLS	1.19045	1.09802	0.46916
200	MLE	1.70824	1.22987	0.53599
	$MPSE$	1.69816	1.21880	0.54960
	LSE	1.82353	1.22692	0.55239
	WLS	1.77158	1.19479	0.53146
500	MLE	1.65717	1.12576	0.49174
	$MPSE$	1.52088	1.02215	0.47583
	LSE	1.79043	1.13586	0.47122
	WLS	1.66917	1.12096	0.49279
1000	MLE	1.65780	1.12194	0.49958
	$MPSE$	1.63754	1.08834	0.46457
	LSE	1.73552	1.14054	0.51225
	WLS	1.68209	1.11442	0.51928
2000	MLE	1.63653	1.08665	0.51535
	$MPSE$	1.60534	1.06778	0.50646
	LSE	1.64144	1.11822	0.52656
	WLS	1.65295	1.11291	0.51404

Table 4.1: The estimation of $\widehat{\gamma}$ by different estimators at true value $\gamma = (1/0.6; 1/0.9; 0.5)$ (note: the value of each entry is mean, and results are re-scaled by the factor 0.00001)

		$\gamma = 1.666$		$\gamma = 1.111$		$\gamma = 0.5$	
n	Methods	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
10	*MLE*	−0.42880	0.57890	−0.95252	1.03148	−0.42863	0.47367
	MPSE	−0.44991	0.55032	−0.97364	1.04259	−0.43001	0.46928
	LSE	−0.33831	0.54721	−0.89221	1.05260	−0.40149	0.46417
	WLS	−0.33348	0.53601	−0.88898	1.05734	−0.40004	0.41580
20	*MLE*	−0.32511	0.28018	−0.21651	0.58715	−0.09743	0.18422
	MPSE	−0.33076	0.30029	−0.22108	0.60739	−0.09956	0.19070
	LSE	0.33339	0.59442	0.22226	0.69628	0.10002	0.19833
	WLS	0.32463	0.20644	0.21442	0.50429	0.09239	0.18193
30	*MLE*	−0.21613	0.25724	−0.11059	0.55177	−0.07176	0.16830
	MPSE	−0.23182	0.30835	−0.14268	0.57266	−0.73280	0.18033
	LSE	0.25063	0.46406	0.16709	0.60937	0.07519	0.19922
	WLS	0.20260	0.26453	0.15507	0.50969	0.06978	0.13936
50	*MLE*	−0.11182	0.17517	−0.11782	0.41691	−0.04952	0.08261
	MPSE	−0.14393	0.27124	−0.11813	0.42014	−0.05096	0.09907
	LSE	0.17100	0.33949	0.11953	0.42633	0.05136	0.10185
	WLS	0.10558	0.14598	0.10372	0.33065	0.04367	0.07379
100	*MLE*	−0.09519	0.12317	−0.05947	0.32213	−0.03155	0.04696
	MPSE	−0.09938	0.16086	−0.06458	0.35150	−0.03479	0.05707
	LSE	0.10359	0.22574	0.06906	0.39049	0.03696	0.06772
	WLS	0.08920	0.11128	0.04345	0.28419	0.02676	0.03939
200	*MLE*	−0.05917	0.08742	−0.04145	0.30828	−0.01665	0.02923
	MPSE	−0.06157	0.10553	−0.04194	0.33039	−0.01778	0.03045
	LSE	0.06246	0.15187	0.04203	0.35125	0.01874	0.04556
	WLS	0.05417	0.06031	0.03611	0.30687	0.01525	0.02809
500	*MLE*	−0.02649	0.05554	−0.01999	0.13703	−0.00744	0.01966
	MPSE	−0.02988	0.07663	−0.20984	0.15184	−0.01032	0.02104
	LSE	0.03100	0.09109	0.02067	0.17072	0.01182	0.02733
	WLS	0.02109	0.05095	0.01739	0.11530	0.00703	0.01838
1000	*MLE*	−0.01663	0.03994	−0.01091	0.02627	−0.00619	0.01182
	MPSE	−0.01475	0.05597	−0.01109	0.03230	−0.00428	0.01187
	LSE	0.01767	0.07299	0.01178	0.04199	0.00945	0.01189
	WLS	0.01599	0.03091	0.01066	0.01727	0.00480	0.01127
2000	*MLE*	−0.00940	0.02776	−0.00626	0.01850	−0.00302	0.00833
	MPSE	−0.01002	0.03387	−0.00657	0.02521	−0.00306	0.00945
	LSE	0.01027	0.04365	0.00685	0.02910	0.00308	0.01309
	WLS	0.00899	0.01911	0.00599	0.01274	0.00270	0.00473

Table 4.2: Simulated bias and RMSE when $\gamma = (1.667; 1.111; 0.5)$, and results are re-scaled by the factor 0.00001

Monthly

N	$\hat{\gamma}_{MLE}$	$\hat{\gamma}_{MPSE}$	$\hat{\gamma}_{LS}$	$\hat{\gamma}_{WLS}$	$\hat{\mu}$
132	0.63622	0.67531	0.71649	0.68363	0.466853

Weekly

N	$\hat{\gamma}_{MLE}$	$\hat{\gamma}_{MPSE}$	$\hat{\gamma}_{LS}$	$\hat{\gamma}_{WLS}$	$\hat{\mu}$
310	0.67842	0.65912	0.69471	0.67593	0.408663

Table 4.3: Parameter estimate for Frechet distribution of the weekly and monthly maxima of the Danish fire losses.

Conclusion & discussion

In this thesis, we aim to use a class of weighted least squares estimators for the tail index of a distribution function. Different weight functions and tail index to compare the WLS *and the least squares* LS *estimators show that in some cases the use of the weights makes the asymptotic variance smaller, by derivation of weights for least-squares from large sample variances.*

Our approach is based on the method of WLS *where the weights are inspired from the ideas of* Zyl & $schall, 2012$ *and independent of the parameters the distribution .*

A simulation study is carried out to evaluate the performance of the proposed estimator, and the efficiency of the method with the proposed weights, it has been shown that our newly estimator of Frechet distribution is perform better than other considered methods estimators based on the order statistics in all the shape parameters and sample cases, and for real data set of danish fire.

Moreover, it is also emphasized that the considered estimation methods can be applied to $BurrXII$, $Cauchy$ *and other distributions, which have explicit cumulative distribution functions, after calculating the inverse of the approximate variance them, and estimating the variances in the* WLS *estimation.*

In future research, we plan to investigate the performance of the WLS *estimation method in the case of right censored data and contaminated data.*

Bibliography

[1] Adrain, R. 1808. *Research concerning the probabilities of the errors which happen in making observations, &c*. The analyst, or mathematical museum. Vol: 1, No : 4, page : $93 - 109$.

[2] Annasaheb, M.S., Girish, C.B. 2018. *Comparison of three estimation methods for frechet distribution*. Department of statistics, India. Volume 5 Issue 1.

[3] Balakrishnan, N., Cohen, A.C. 1991. *Order statistics and inference. Estimation methods*. Statistical modeling and decision science. Academic press, Inc., Boston, MA.

[4] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. 2004. *Statistics of extremes : theory and applications*. Wiley series in probability and statistics. Chichester.

[5] Bergman, B. 1986. *Estimation of weibull parameters using a weight function*. Journal of materials science letters, No:5, page:$611 - 614$.

[6] Bernard, A., Bosi-Levenbach, E.C. 1953. *The plotting of observations on probability paper*. Stat. Neederlandica, No :7, page :$163 - 173$.

[7] Bingham, N., Goldie, C., Teugels, J. 1987. *Regular variation*. Cambridge university press.

[8] Brahimi, B., Meraghni, D., Necir, A., Yahia, D. 2013. *A bias-reduced estimator for the mean of a heavy-tailed distribution with an infinite second moment*. Journal of statistical planning and inference No :143, page : $1064 - 1081$.

[9] Coles, S. 2001. *An introduction to statistical modeling of extremes values*. Springer series in statistics. Springer-Verlag. London.

[10] Cullagh., Nelder. 1992. *Introduction au modèle de régression linéaire simple*. Université de toulouse.

[11] Csörgö, S., Deheuvels, P., Mason, D. 1985. *Kernel estimates of the tail index of a distribution.*The analysis of statistics, page:$1050 - 1077$

[12] Danielsson, J., de Haan, L., Peng, L., de Vries, C. G. 2001. *Using a bootstrap method to choose the sample fraction in tail index estimation*. Journal of multivariate analysis, No:76, page:$226 - 248$.

[13] Danielsson, J., Ergun, L. M., De Haan, L., de Vries, C. G. 2016. *Tail index estimation: quantile driven threshold selection*. Available at SSRN: https://ssrn.com/abstract=2717478.

[14] DasGupta, A. 2008. *Asymptotic theory of statistics and probability*. Springer texts in statistics, NY.

[15] David, H.A. 1970. *Order statistics*. John Wiley & Sons, Inc., New York-London Sydney.

[16] De Haan, L., Ferreira, A. 2006. *Extreme value theory : An introduction.* Springer., Vol: 21.

[17] De Haan, L., Stadtmüller, U. 1996. *Generalized regular variation of second order*. Jornal of australian math. Soc. (Series A) No:61, page :$381 - 395$.

[18] Deheuvels, P., Haeusler, E., Mason, D.M. 1988. *Almost sure convergence of the hill estimator*. Math. Proc. Cambridge philos. Soc., Vol:104,No:2, page:$371 - 381$.

[19] Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L. 1989. *A moment estimator for the index of an extreme-value distribution*. Analysis of statistics. No:17, page:$1833 - 1855$.

[20] Dombry, C. 2013. *Maximum likelihood estimators for the extreme value index based on the block maxima method*. http ://hal.archives-ouvertes.fr/hal−00780279.

[21] Drees, H. 1995. *Refined pickands estimators of the extreme value index*. The analysis of statistics, page: 2059 − 2080.

[22] Embrechts, P., Kluppelberg, C., Mikosch, T. 1997. *Modelling extremal events for insurance and finance*. Springer-Verlag. Berlin.

[23] Falk, M. 1995. *Some best parameter estimates for distributions with finite endpoint*. Statistics : A journal of theoretical and applied statistics. Vol:27, No:2, page:115−125.

[24] Fisher, R., Tippet, L. 1928. *Limiting forms of the frequency distribution of the largest or smallest member of a sample*. Proceedings of the cambridge, philosophical society, No:24, page: 180 − 190.

[25] Gauss, C.F,. Stewart, G.W., 1795. *Theory of the combination of observations least subject to errors: part one, part two, supplement*. Society for industrial and applied mathematics. books.google.com.

[26] Galton,F. 1886. *Regression towards mediocrity in hereditary stature*. Journal of the anthropological institute, No:15, page: 246 − 63.

[27] Gardes, L., Girard, S., Lekina, A. 2010. *Functional nonparametric estimation of conditional extreme quantiles*. Journal of multivariate Anal. Vol:101, No: 2, page:419 − 433.

[28] Gumbel, E. 1958. *Statistical of extreme*. Colombia university press, New york.

[29] Gnedenko, B. 1943. *Sur la distribution limite du terme maximum d'une série aléatoire*. The annals of mathematics, Vol:44, No:3, page: 423 − 453.

[30] Gradshteyn, I.S, Ryzhik, M.Y. 1980. *Table of integrals, series and products*. Academic press, NY. page:573 − 575.

[31] Ihaka, R., Gentleman, R. 1996. *R: a language for data analysis and graphics*. Journal of computational and graphical statistics, No:5, page :299 − 314.

[32] Haan, L. 1976. *Sample extremes : an elementary introduction*. Statistica neerlandica, Vol: 4, No:304, page: 161 − 172.

[33] Hung, W.L. 2001. *Weighted least squares estimation of the shape parameter of the weibull distribution*. Quality and reliability engineering international Vol:6, No:17, page: 467 − 469.

[34] Hill, B.M. 1975. *A simple general approach to inference about the tail of a distribution*. Ann. Statist.,No:3, page:1163 − 1174.

[35] Hosking, J. 1985. *Algorithm as 215 : maximum-likelihood estimation of the parameters of the generalized extreme-value distribution*. Journal of the royal statistical society. Series C (Applied statistics), Vol:3 ,No:34, page: 301 − 310.

[36] Hosking, J., Wallis, J. 1987. *Parameter and quantile estimation for the generalized pareto distribution*. Technometrics 29, page: 339 − 349.

[37] Hossain, A. Howlader, H.A. 1996. *Unweighted least squares estimation of weibull parameters*. Journal of statistical computation and simulation, No:54, page:265 − 271.

[38] Jenkinson, A. 1955. *The frequency distribution of the annual maximum (or minimum) values of meteorological elements*. Quarterly journal of the royal meteorological society, Vol:81, No: 348, page: 158 − 171

[39] Kantar, Y.M., Arik, I. 2014. *M-estimation of log-logistic distribution parameters with outliers*. International journal of agricultural and statistical sciences.

[40] Kantar, Y.M., Yildirim, V. 2015. *Robust estimation for parameters of the extended burr type III distribution*. Communications in statistics – theory and methods, simulation and computation,Vol:44, No:7, page: 1901 − 1930.

[41] Kendall, M., Stuart, A., Ord, J.K. 1987. *Kendall's advanced theory of statistics*. Charles griffin and company, London.

[42] Kinnison, R. 1985. *Applied extreme value*. Battele press. New york.

[43] Legendre, AM. 1805. *Nouvelles méthodes pour la détermination des orbites des comètes*. Paris, Appendix, "Sur la Méthode des moindres quarrés", page. $72 - 80$.

[44] Lu, H.L., Chen, C.H., Wu, J.W. 2004. *A Note on weighted least-squares estimation of the shape parameter of the weibull distribution*. Qual. Reliab. Eng. Int., Vol:6, No:20, page :$579 - 586$.

[45] Lu, H.L., Tao, S.H. 2007. *The estimation of pareto distribution by a weighted least square method*. Quality & Quantity, No:41, page: $913 - 926$. Springer.

[46] Mason, D.M. 1982. *Laws of large numbers for sums of extreme values*. Ann. Probab,Vol:10, No:3, page:$754 - 764$.

[47] Necir, A. 2006. *A functional law of the Iterated logarithm for kernel type estimators of the tail index*. Journal of statistical planning and inference No: 136, page:$780 - 802$.

[48] Plackett, R. L. 1949. *A historical note on the method of least squares*. Biometrika, Vol:36, No:4, page :$458 - 460$.

[49] Picard,F. 1996. *Première notions de statistique régression linéaire*. Université lyon1. $UMRCNRS - 5558$, Laboratoire de biométrie et biologie evolutive.

[50] Pickands, J. 1975. *Statistical inference using extreme order statistics*. The annals of statistics, Vol :3, No: 1, page: $119 - 131$.

[51] Reiss, R., Thomas, M. 2001. *Statistical analysis of extreme values with applications to assurance, finance, hydrology and other fields*. Basel. Birkhauser. Verlag.

[52] Resnick, S. 1987. Extreme values, regular variation, and point processes. Springer verlag.

[53] Rousson.V. 2004. *Analyse de régression appliquée*. $2^{ème}$ édition, DUNOD,Paris.

[54] Smith, R. 1987. *Estimating tails of probability distributions*. The annals of statistics, Vol: 15, No: 3, page :$1174-1207$.

[55] Tomassone R., Charles-Bajard S., Bellanger L. 1998. *Introduction à la planification expérimentale, DEA" Analyse et modélisation des systèmes biologiques"* .Fiche pratique splus 2 /05/01/99.

[56] Von Mises, R. 1954. *La distribution de la plus grande de n valeurs*. American mathematical society, providence, RI, II(In Selected Papers), page : $271-294$.

[57] Nelder, J. A., Wedderburn,R.W.M . 1972. *Generalized linear models*. Journal of the royal statistical society, series A. No: 135, page: $370-384$.

[58] Yun, S. 2002. *On a generalized pickands estimator of the extreme value index*. Journal of statistical planning and inference. No: 102, page:$389-409$.

[59] Zhang, L.F., Xie, M., Tang, L.C. 2007. *A study of two estimation approaches for parameters of weibull distribution based on WPP, Reliab*. Eng. Syst. Safe.,No: 92, page:$360-368$.

[60] Zhang, L.F., Xie, M., Tang, L.C. 2008. *On weighted least squares estimation for the parameters of weibull distribution*. Recent advances in reliability and quality in design springer series in reliability engineering, $2, 11$, Chapter $57-84$.

[61] Zhou, C. 2009. *Existence and consistency of the maximum likelihood estimator for the extreme value index*. Journal of multivariate analysis, Vol: 100, No: 4, page : $794-815$.

[62] Zhou, C. 2010. *The extent of the maximum likelihood estimator for the extreme value index*. Journal of multivariate analysis, Vol : 101, No: 4, page : $971 - 983$.

[63] Zyl, J.M.V. 2016. *Exact expressions for the weights used in least-squares regression estimation for the log-logistic and weibull distribution*. University of the free state, PO Box 339, Bloemfontein, South Africa.

[64] Zyl, J.M.V. 2012. *A median regression model to estimate the parameters of the three-parameter generalized pareto distribution*. Commun. Stat-Simul. C, No: 41, page: $544 - 553$.

[65] Zyl, J.M.V., Schall, R. 2012. *Parameter estimation through weighted least-squares rank regression with specific reference to the weibull and gumbel distributions*. Communications in statistics-simul. No: 41, page: $1654 - 1666$.

Résumé

Cette thèse est consacrée à l'étude d'un estimateur par la régression pour estimer l'indice de queue de la distribution à queue lourde. En particulier, il est montré que l'estimateur considéré est en général basé sur la méthode des moindres carrés pondérés.

L'objectif principal de la thèse de prolonger les travaux de Zyl et schall, 2012 ; pour estimer le paramètre de forme de la distribution de Fréchet. En dérivant les grandes variances de l'échantillon et en utilisant l'inverse de la variance approximative pour calculer les poids de cet estimateur.

Une étude de simulation à l'aide du logiciel statistique **R** est réalisée pour évaluer les performances du nouvel estimateur qui s'est avéré plus performant que les autres estimateurs de méthodes considérés sur la base de statistiques d'ordre pour des échantillons de petite et grande taille, et en cas de données réelles.

ملخص

هذه الأطروحة مخصصة لدراسة مقدّر الانحدار لتقدير مؤشر الذيل لتوزيع الذيل الثقيل. على وجه الخصوص ، يتضح أن المقدر المدروس يعتمد بشكل عام على طريقة المربعات الصغرى الموزونة.

الهدف الرئيسي من الأطروحة هو تمديد عمل زيل و شال 2012 لتقدير معامل الشكل لتوزيع فريشيه من خلال اشتقاق تباينات العينة الكبيرة واستخدام معكوس التباين التقريبي لحساب أوزان هذا المقدّر.

تم إجراء دراسة محاكاة باستخدام البرنامج الإحصائي R لتقييم أداء المقدّر الجديد الذي ثبت أنه يعمل بشكل أفضل من مقدرات الطرق الأخرى بناءً على إحصائيات الطلب لحجم العينة الصغير والكبير، وفي حالة البيانات الحقيقية.