A Survey of Hill's Estimator

by

Jan Henning Vollmer

(Under the direction of William P. McCormick)

Abstract

In this paper we will survey Hill's estimator, which is one of the most popular estimators for the tail index of heavy-tailed distributions. Applications are numerous and include, for example, insurance reliability theory, econometrics, geology and climatology. We will outline how Hill's estimator is constructed and summarize the developments of its properties like consistency and asymptotic normality. Therefore, we will introduce the concept of first- and second-order regular variation. Furthermore, we will give an overview of proposed methods for choosing the number of order statistics, a very crucial parameter in Hill's estimator. Graphical tools for the estimator will be illustrated on the basis of an example.

Index words: Hill's estimator, extreme value theory, tail index, regular variation, order statistics, asymptotic normality.

A Survey of Hill's Estimator

by

Jan Henning Vollmer

Vordiplom, Universität Osnabrück, Germany, 2001

A Thesis Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Master of Science

Athens, Georgia

2003

A SURVEY OF HILL'S ESTIMATOR

by

JAN HENNING VOLLMER

Approved:

Major Professor:	William P. McCormick
Committee:	Lynne Seymour
	Jaxk Reeves

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2003

Acknowledgments

I would like to take the opportunity to thank my major professor, Dr. William P. McCormick, for his great help and guidance during the work on my thesis.

Furthermore, I would like to thank Dr. Lynne Seymour and Dr. Jaxk Reeves for being on my committee and for being excellent teachers and great advisors throughout my entire exchange year.

Moreover, I would like to thank the DAAD (German Academic Exchange Service) and the Department of Statistics for supporting me and making my studies at the University of Georgia possible. Besides, I would also like to thank Professor Dr. Wulf Gaertner, Connie Durden and Dr. Robert Lund.

Last, but not least I would like to express my gratitude toward my families and friends in Germany and the United States for their support, especially during the harder times. They made this exchange year a valuable experience.

Abschließend möchte ich die Gelegenheit nutzen, meinen Dank an meine Familie und Freunde in Deutschland auszusprechen, die mich in all meinen Entscheidungen unterstützen und die immer für mich da sind.

Table of Contents

List of Figures

Chapter 1

Introduction

Extreme value theory can be applied to various fields of interest. Consider the following situations:

- ◇ Dams or sea dikes must be built high enough to exceed the maximum water height. The Dutch government specifies that the probability of a flood in a given year should be 1/10,000 (de Haan (1994)).
- ◇ Design strength of skyscrapers must be sufficient to withstand wind stresses from several directions (Resnick (1987)).
- ◇ Over a period of time a primary insurer receives k claims. The ECOMOR reinsurance contract binds the reinsurer to cover (for a specific premium) the excesses above the r-th largest claim (Teugels (1981a)).
- ◇ Portfolio managers and regulators are interested in the maximal limit on the potential losses of a given portfolio (Embrechts et al. (1997)).

The situations described above have a common feature: observational data already exists or can be collected, and the features of the observations of most interest depend on either the smallest or the largest values; i.e., the extremes. In many cases the events of interest even fall outside the range of the data. The challenge comprises of summarizing the data in appropriate models and making decisions on the basis of the behaviour of the extreme values.

In particular, many applied scientists are confronted with one of the fundamental problems in extreme value statistics: given a sample which one assumes to be independent and coming from an unknown distribution F, how can one

- estimate the endpoint of F,
- estimate large quantiles of F,
- or determine the probability of extreme events, i.e. the exceedance probability of high thresholds.

These problems may be considered special cases of the more general question: *What are the features of a tail of a distribution and how can these features be estimated?*

Hill (1975) introduced a simple, general approach to draw inference about the tail behaviour of a distribution. It is not required to specify the underlying distribution F globally, but merely the form of behaviour in the tail where it is desired to draw inference. In simpler words, for a given sample of size k we condition upon the $r+1$ upper (or lower) order statistics. Then, using the $r+1$ upper order statistics the conditional likelihood for the parameters that describe the tail of the distribution can be obtained. Based on the conditional likelihood function Hill's estimator can be constructed. A more detailed description will be provided in Chapter 2.

Techniques for drawing inference about the tail behaviour of a distribution are well developed, and alternative methods to Hill's that are also based on extreme order statistics have been proposed by Pickands (1975), Weissman (1978), and others. There are also numerous modifications of Hill's estimator; one of the best known is the so called moment estimator proposed by Dekkers et al. (1989). In general Hill's estimator compares favourably with other competitors, especially when the underlying distribution is of strict Pareto law, which is the special case Hill (1975) considered and that will also be covered in Chapter 2.

The goal of this paper is to compile the large amount of literature that studied Hill's estimator with respect to its asymptotic properties and the optimal choice of order statistics used in the estimation. Chapter 2 is a basic version of how Hill's estimator is constructed. In Chapter 3 we will present the developments on consistency and asymptotic normality of Hill's estimator, which can be useful for a comparison with alternative estimators (de Haan and Peng (1998)). Therefore we will introduce the concept of regular variation, which has proven very useful in establishing results on consistency and asymptotic normality. Chapter 4 addresses the difficult choice of the number of order statistics used in Hill's estimator. We will introduce graphical methods that have been proposed to minimize the difficulty of this choice. The graphical tools will be illustrated by considering an example.

Chapter 2

Constructing Hill's Estimator

This section shows step by step how the estimator is developed. It follows the original paper by Hill (1975). In addition, it provides proofs and explanations of the results. The following preliminary result on the distribution of order statistics will be needed.

2.1 Rényi Representation

Let $X_1, \ldots, X_k$ be a random sample of size k from a continuous strictly increasing distribution F, where $F(0) = 0$. Let $Z^{(1)} \geq Z^{(2)} \geq \cdots \geq Z^{(k)}$ be the corresponding order statistics, so that $Z^{(1)} = \max\{X_1, \ldots, X_k\}$, $Z^{(k)} = \min\{X_1, \ldots, X_k\}$, and $Z^{(i)}$ in general denotes the i-th order statistic. Note this ordering of the order statistics is not the usual one but is employed to maintain consistency with Hill's paper. Also, let $E_i \overset{iid}{\sim} \text{Exp}(1)$ for $i = 1, \ldots, k$ (i.e., the E_i's are independent expontentially distributed random variables, each of which has expectation 1).

Then, by the Rényi (1953) representation theorem

$$\begin{aligned} Z^{(i)} &\overset{d}{=} F^{-1}(\exp\{-[\frac{E_1}{k} + \frac{E_2}{k-1} + \cdots + \frac{E_i}{k-i+1}]\}) \\ &= F^{-1}(\exp\{-\sum_{m=1}^{i} \frac{E_m}{k-m+1}\}), \quad \text{for} \quad i = 1, 2, \ldots, k. \end{aligned} \tag{2.1}$$

Proof. First note that (2.1) is equivalent to

$$F(Z^{(i)}) \overset{d}{=} \exp\{-\sum_{m=1}^{i} \frac{E_m}{k-m+1}\}, \quad \text{for} \quad i = 1, 2, \ldots, k.$$

Let $U^{(1)} \geq U^{(2)} \geq \cdots \geq U^{(k)}$ be order statistics of an i.i.d. random sample from the uniform distribution on [0,1]. Then

$$(F(Z^{(1)}), \quad F(Z^{(2)}), \ldots, F(Z^{(k)})) \stackrel{d}{=} (U^{(1)}, U^{(2)}, \ldots, U^{(k)}),$$

where $A \stackrel{d}{=} B$ means equality in distribution of random variables or vectors A and B. Thus, we further obtain the equivalence

$$\begin{aligned} U^{(i)} &\stackrel{d}{=} \exp\{-\sum_{m=1}^{i} \frac{E_m}{k-m+1}\} \\ \Leftrightarrow -\log U^{(i)} &\stackrel{d}{=} \sum_{m=1}^{i} \frac{E_m}{k-m+1}, \quad \text{for} \quad i = 1, 2, \ldots, k. \end{aligned}$$

Since the random variable $V = -\log U$ has a standard exponential - or, Exp(1) - distribution with density $f(v) = e^{-v}$, $0 \leq v < \infty$ and also $-\log\ u$ is a monotonically decreasing function in u, the relation

$$V^{(i)} = -\log\ U^{(k-i+1)}, \quad 1 \leq i \leq k,$$

holds, where $V^{(i)}$ is the i-th order statistic from the standard exponential distribution. Hence, we need to show that

$$V^{(k-i+1)} \stackrel{d}{=} \sum_{m=1}^{i} \frac{E_m}{k-m+1}, \quad \text{for} \quad i = 1, 2, \ldots, k,$$

or equivalently that

$$(V^{(k)}, V^{(k-1)}, \ldots, V^{(1)}) \stackrel{d}{=} (\frac{E_1}{k}, \frac{E_1}{k} + \frac{E_2}{k-1}, \ldots, \frac{E_1}{k} + \frac{E_2}{k-1} + \cdots + \frac{E_k}{1}). \qquad (2.2)$$

It is sufficient to show that the two vectors have the same density function. Therefore, a general result on the joint density of all k order statistics $Z^{(1)}, Z^{(2)}, \ldots, Z^{(k)}$ is introduced:

$$f_{Z^{(1)}, Z^{(2)}, \ldots, Z^{(k)}}(z^{(1)}, z^{(2)}, \ldots, z^{(k)}) = k! \prod_{s=1}^{k} f(z^{(s)}) \qquad (2.3)$$

for $-\infty < z^{(k)} < z^{(k-1)} < \cdots < z^{(1)} < \infty$.

(Recall that we are employing Hill's nonstandard indexing for order statistics.)

When the random sample is from a standard exponential distribution, the joint density of all k order statistics in (2.3) takes the form

$$f_{V^{(1)},V^{(2)},\ldots,V^{(k)}}(v^{(1)},v^{(2)},\ldots,v^{(k)}) = k!\prod_{s=1}^{k} f(v^{(s)}) \tag{2.4}$$

$$= k!\prod_{s=1}^{k} e^{-v^{(s)}} = k!e^{-\sum_{s=1}^{k} v^{(s)}}, \quad 0 \le v^{(k)} \le v^{(k-1)} \le \cdots \le v^{(1)} \le \infty.$$

Hence, it is left to show that the vector $(\frac{E_1}{k}, \frac{E_1}{k}+\frac{E_2}{k-1}, \ldots, \frac{E_1}{k}+\frac{E_2}{k-1}+\cdots+\frac{E_k}{1})$ has the same density function as given in (2.4). Let

$$(Y_k, Y_{k-1}, \ldots, Y_1) = (\frac{E_1}{k}, \frac{E_1}{k}+\frac{E_2}{k-1}, \ldots, \frac{E_1}{k}+\frac{E_2}{k-1}+\cdots+\frac{E_k}{1})$$

Then

$$\begin{aligned}
Y_k &= \frac{E_1}{k} && \Leftrightarrow E_1 = kY_k, \\
Y_{k-1} &= \frac{E_1}{k}+\frac{E_2}{k-1} && \Leftrightarrow E_2 = (k-1)(Y_{k-1}-Y_k) \\
&\vdots && \quad \vdots \\
Y_1 &= \frac{E_1}{k}+\frac{E_2}{k-1}+\cdots+\frac{E_k}{1} && \Leftrightarrow E_k = Y_1 - Y_2,
\end{aligned}$$

or, more generally,

$$E_{k-j+1} = j(Y_j - Y_{j+1}), \quad \text{for} \quad j = 1, 2, \ldots, k,$$

where $Y_{k+1} = 0$. Note that $Y_k + Y_2 + \cdots + Y_k = E_1 + E_2 + \cdots + E_k$ and $Y_k \le Y_{k-1} \le \cdots \le Y_1$.

Then the density of $(Y_k, Y_{k-1}, \ldots, Y_1)$ can be determined by transformation method:

$$f_{Y_1,Y_{k-1},\ldots,Y_1}(y_k, y_{k-1}, \ldots, y_1) = f_{E_1,E_2,\ldots,E_k}(e_1, e_2, \ldots, e_k)|J|,$$

where $|J|$ is the determinant of the Jacobian and the E_j's are i.i.d. Exp(1).

Hence,

$$f_{E_1,E_2,\ldots,E_k}(e_1,e_2,\ldots,e_k)|J| = \prod_{s=1}^{k} e^{-e_s}|J| = e^{-\sum_{s=1}^{k} e_s}|J|,$$

where the Jacobian is

$$J = \begin{pmatrix} \frac{\partial E_1}{\partial Y_k} & \frac{\partial E_1}{\partial Y_{k-1}} & \frac{\partial E_1}{\partial Y_{k-2}} & \cdots & \frac{\partial E_1}{\partial Y_1} \\ \frac{\partial E_2}{\partial Y_k} & \frac{\partial E_2}{\partial Y_{k-1}} & \frac{\partial E_2}{\partial Y_{k-2}} & \cdots & \frac{\partial E_2}{\partial Y_1} \\ \vdots & & & & \vdots \\ \frac{\partial E_k}{\partial Y_k} & \frac{\partial E_k}{\partial Y_{k-1}} & \frac{\partial E_k}{\partial Y_{k-2}} & \cdots & \frac{\partial E_k}{\partial Y_1} \end{pmatrix}$$

$$J = \begin{pmatrix} k & 0 & 0 & \cdots & 0 \\ -(k-1) & k-1 & 0 & \cdots & 0 \\ 0 & -(k-2) & k-2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

So the Jacobian is a lower triangular matrix and its determinant is equal to the product of its diagonal elements. Thus,

$$|J| = k(k-1)(k-2)\ldots(1) = k!,$$

so that the joint density of $(Y_k, Y_{k-1}, \ldots, Y_1)$ is

$$f(y_k, y_{k-1}, \ldots, y_1) = k!e^{-\sum_{s=1}^{k} e_s} = k!e^{-\sum_{i=1}^{k} y_i}, \quad \text{for} \quad 0 \leq y_k \leq y_{k-1} \leq \cdots \leq y_1 \leq \infty.$$

Hence, the two vectors in equation (2.2) have the same density function. Thus, the Rényi representation theorem is proved. A similar proof can be found in Nevzorov (2001).

2.2 Conditional Likelihood Function

Based on the Rényi representation theorem in (2.1) one can obtain variables distributionally equivalent to an i.i.d. Exp(1) sample. Such an analysis forms the basis for calculating a conditional likelihood function.

From (2.1) it follows that,

$$\log\ F(Z^{(j-1)}) = -\sum_{m=1}^{j-1} \frac{E_m}{k-m+1}, \quad \text{and} \tag{2.5}$$

$$\log\ F(Z^{(j)}) = -\sum_{m=1}^{j} \frac{E_m}{k-m+1}. \tag{2.6}$$

Subtracting (2.6) from (2.5) we get

$$\begin{aligned} \log F(Z^{(j-1)}) - \log F(Z^{(j)}) &= -\sum_{m=1}^{j-1} \frac{E_m}{k-m+1} + \sum_{m=1}^{j} \frac{E_m}{k-m+1} \\ &= \frac{E_j}{k-j+1}. \end{aligned}$$

Hence

$$E_j = (k-j+1)[\log\ F(Z^{(j-1)}) - \log F(Z^{(j)})], \quad \text{for} \quad j = 1, 2, \ldots, k, \tag{2.7}$$

where by definition $F(Z^{(0)}) = 1$.

In the next step, it is assumed that for a certain range (in this case the lower tail) a specified function is valid.

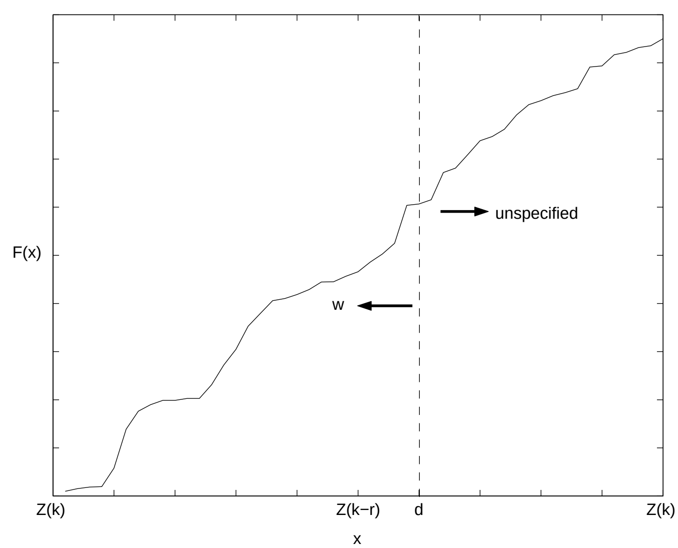

Figure 2.1: Example - lower tail

The range is defined by a known cutoff point d and the specified function of the form $F(x) = w(x; \theta)$ for $x \leq d$ includes an unknown parameter vector θ. The goal is to estimate θ based only on the portion of the data where the specified function is valid. See Figure 2.1.

Now, for $Z^{(k)} \leq Z^{(k-1)} \leq \cdots \leq Z^{(k-r)} \leq d$, i.e. the r lowest order statistics, the function $F(x) = w(x;\theta)$ is assumed to be valid, so that from (2.7) it follows that

$$E_j = (k-j+1)[\log\ w(Z^{(j-1)};\theta) - \log\ w(Z^{(j)},\theta)], \tag{2.8}$$

$$\text{for}\quad j = k-r+1,\quad k-r+2,\ldots,k,$$

and from (2.6) it follows that for $j = k-r$

$$\log\ w(Z^{(k-r)};\theta) = -\sum_{m=1}^{k-r} \frac{E_m}{k-m+1}.$$

Define

$$H = -k\ \log\ w(Z^{(k-x)};\theta) = k\sum_{m=1}^{k-r} \frac{E_m}{k-m+1}, \tag{2.9}$$

and let h be the observed values of H and denote g the corresponding density function. The r equations given by (2.8) are the basis for probability statements used in the conditional likelihood functions. For any event for which $Z^{(k-r)} \leq d$, let $z^{(j)}$ be the observed value of the j-th order statistic for $j = k-r,\ k-r+1,\ldots,k$. Then the conditional likelihood (conditional upon $Z^{(k-r)} \leq d$) is

$$\begin{aligned} L_1(\theta) &= P_\theta(Z^{(k)} = z^{(k)},\ldots,Z^{(k-r)} = z^{(k-r)} \mid Z^{(k-r)} \leq d) \\ &= \frac{P_\theta(Z^{(k)} = z^{(k)},\ldots,Z^{(k-r)} = z^{(k-r)}, Z^{(k-r)} \leq d)}{P_\theta(Z^{(k-r)} \leq d)} \end{aligned}$$

where we have taken the liberty to write the joint density in the form of a probability. Since $Z^{(k-r)} \leq d$ contains the event $Z^{(k-r)} = z^{(k-r)}$, we can write

$$\begin{aligned} L_1(\theta) &= \frac{P_\theta(Z^{(k)} = z^{(k)},\ldots,Z^{(k-r)} = z^{(k-r)})}{P_\theta(Z^{(k-r)} \leq d)} \\ &= f_{Z^{(k)},\ldots,Z^{(k-r)}}(z^{(k)},\ldots,z^{(k-r)})/p_d, \end{aligned}$$

where $p_d = P_\theta(Z^{(k-r)} \leq d)$.

Using (2.8) we can apply the transformation method to find the joint density of the $Z^{(j)}$'s.

$$\begin{aligned} L_1(\theta) &= \prod_{j=k-r}^{k} f_{Z^{(j)}}(z^{(j)})/p_d = \prod_{j=k-r}^{k} f_{E^{(j)}}(e^{(j)})|J|/p_d \\ &= \prod_{j=k-r+1}^{k} \exp\{-e^{(j)}\} \times g|J|/p_d \end{aligned}$$

where $e^{(j)} = e^{(j)}(z^{(j)}, z^{(j-1)})$ given in equation (2.8), J is the Jacobian of the transformation and g denotes the density function of H as defined in (2.9). Thus,

$$L_1(\theta) = \prod_{j=k-r+1}^{k} \exp\{-(k-j+1)[\log\ w(z^{(j-1)};\theta) - \log\ w(z^{(j)};\theta)]\} \times g|J|/p_d \quad (2.10)$$

Let $i = k - j + 1 \Leftrightarrow j = k - i + 1$. Then

$$L_1(\theta) = \exp\{-\sum_{i=1}^{r} i[\log\ w(z^{(k-i)},\theta) - \log\ w(z^{(k-i+1)};\theta)]\} \times g|J|/p_d$$

where

$$J = \begin{pmatrix} \frac{\partial e_{k-r+1}}{\partial z^{(k-r)}} & \frac{\partial e_{k-r+1}}{\partial z^{(k-r+1)}} & \cdots & \frac{\partial e_{k-r+1}}{\partial z^{(k)}} \\ \frac{\partial e_{k-r+2}}{\partial z^{(k-r)}} & \frac{\partial e_{k-r+2}}{\partial z^{(k-r+1)}} & \cdots & \frac{\partial e_{k-r+2}}{\partial z^{(k)}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial e_{k}}{\partial z^{(k-r)}} & \frac{\partial e_{k}}{\partial z^{(k-r+1)}} & \cdots & \frac{\partial e_{k}}{\partial z^{(k)}} \\ \frac{\partial h}{\partial z^{(k-r)}} & \frac{\partial h}{\partial z^{(k-r+1}} & \cdots & \frac{\partial h}{\partial z^{(k)}} \end{pmatrix}$$

So,

$$|J| = kr!\Big|\prod_{j=k-r}^{k} \frac{\partial \log\ w(z^{(j)};\theta)}{\partial z^{(j)}}\Big| = kr!\Big|\prod_{j=1}^{r+1} \frac{\partial \log\ w(z^{(k+1-j)};\theta)}{\partial z^{(k+1-j)}}\Big|.$$

To find the density function g of H consider the following: For any $i = 1, \ldots, k$ we have $w(z^{(i)};\theta) = F(Z^{(i)}) = U^{(i)} \sim$ order statistic from uniform (0,1), where $U^{(k)} \leq \cdots \leq U^{(i)} \leq \cdots \leq U^{(1)}$. The density of the i-th order statistic from a uniform distribution can be written as

$$f(u) = \frac{k!}{(k-i)!1!(i-1)!} u^{k-i}(1-u)^{i-1} \quad 0 \leq u \leq 1.$$

Let Beta (a, b) denote a Beta random variable with parameters a and b, and density function

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1}(1-u)^{b-1}.$$

Then it is easy to obtain that

$$U^{(i)} \sim \text{ Beta}(k-i+1, i)$$

and

$$U^{(k-r)} \sim \text{ Beta } (r+1, k-r).$$

So g is the density function of a random variable having the distribution of $Y :=$ $-k \log$ Beta $(r+1, k-r)$.

Using the transformation $U^{(k-r)} = e^{-Y/k}$ we get

$$\begin{aligned} f_Y(y) &= f_{U^{(k-r)}}(e^{-Y/k})|J| \\ &= \frac{\Gamma(r+1+k-r)}{\Gamma(r+1)\Gamma(k-r)}(e^{-Y/k})^{r+1-1}(1-e^{-Y/k})^{k-r-1}|J| \\ &= \frac{\Gamma(k+1)}{\Gamma(r+1)\Gamma(k-r)} e^{-rY/k}(1-e^{-Y/k})^{k-r-1}|J|, \end{aligned}$$

where the determinant of the Jacobian is

$$|J| = \left|\frac{de^{-Y/h}}{dY}\right| = \left|-\frac{1}{k}e^{-Y/k}\right| = \frac{1}{k}e^{-Y/k}.$$

Therefore,

$$f_Y(y) = \frac{\Gamma(k)}{\Gamma(r+1)\Gamma(k-r)} e^{-(r+1)Y/k}(1-e^{-Y/k})^{k-r+1}$$

is the density function g of H, so that the factor g in (2.10) reduces to

$$\begin{aligned} g &= \frac{\Gamma(k)}{\Gamma(r+1)\Gamma(k-r)} e^{-(r+1)(-k\log\ w(z^{(k-r)};\theta))/k}(1-e^{-k\log\ w(z^{(k-r)};\theta)/k})^{k-r-1} \\ &= \frac{\Gamma(k)}{\Gamma(r+1)\Gamma(k-r)}[w(z^{(k-r)};\theta)]^{(r+1)}[1-w(z^{(k-r)};\theta)]^{k-r-1}. \end{aligned}$$

Hence the conditional likelihood can be written as

$$\begin{aligned} L_1(\theta) &\propto \exp\{-\sum_{i=1}^{r} i[\log\, w(z^{(k-i)};\theta) - \log\, w(z^{(k-i+1)};\theta)] \\ &\times [w(z^{(k-r)};\theta)]^{(r+1)}[1-w(z^{(k-r)};\theta)]^{k-r-1}\left|\prod_{i=1}^{r+1}\frac{\partial\log\, w(z^{(k+1-j)};\theta)}{\partial z^{(k+1-j)}}\right|, \end{aligned}$$

where $\propto$ subsumes the constants and the probability p_d. The likelihood function can further be used to obtain conditional maximum likelihood estimates for θ or, if there exists a prior distribution of θ, for conditional posterior distributions for θ. The conditional likelihood function is for the general lower tail case where $F(x) = w(x;\theta)$ has no specific form yet. In the next sections, special cases are being presented for lower-tail and upper tail inference.

2.3 Special Case - Lower tail

Suppose for the region $x \leq d$ the specified function is of the form $w(x;\theta) = Cx^{\alpha}$, so the vector of unknown parameters is $\theta = (C, \alpha)$, with $\alpha > 0$, $C > 0$. Let $T_i = i[\log\, Z^{(k-i)} - \log\, Z^{(k-i+1)}]$, $i = 1, \ldots, r$. It follows from (2.8) that

$$E_{k+1-i} = i[\log\, w(Z^{(k-i)};\theta) - \log\, w(Z^{(k-i+1)};\theta)] \text{ for } i = 1, \ldots, r,$$

and after substituting the specified function,

$$\begin{aligned} E_{k+1-i} &= i[\log\, CZ^{(k-i)^{\alpha}} - \log\, CZ^{(k-i+1)^{\alpha}}] \\ &= i[\log\, C + \alpha\log\, Z^{(k-i)} - \log\, C - \alpha\log\, Z^{(k-i+1)}] \\ &= \alpha i[\log\, Z^{(k-i)} - \log\, Z^{(k-i+1)}] \\ &= \alpha T_i \sim \text{Exp}(1). \end{aligned}$$

Thus,

$$f_{T_1,\ldots,T_r}(t_1, \ldots, t_r, \theta) = f_{E_k,\ldots,E_{k+1-r}}(e_k, \ldots, e_{k+1-r})|J|,$$

where $e_k = \alpha t_1,\ e_{k-1} = \alpha t_2, \ldots,\ e_{k+1-r} = \alpha t_r$ and

$$|J| = \left| \prod_{j=1}^{r} \frac{\partial e_{k+1-j}}{\partial t_j} \right| = \left| \prod_{j=1}^{r} \alpha \right| = \alpha^r,$$

so that

$$f_{T_1,\ldots,T_r}(t_1, \ldots, t_r, \theta) = \alpha^r \prod_{i=1}^{r} \exp\{-\alpha t_i\} = \alpha^r \exp\{-\alpha \sum_{i=1}^{r} t_i\}$$

where t_i is the observed value of T_i. Hence, the conditional likelihood is given by

$$L_0(\alpha) \propto \alpha^r \exp\{-\alpha \sum_{i=1}^{r} t_i\}.$$

To find the maximum likelihood estimate for α, first take the natural log of $L_0(\alpha)$, then set the first derivative with respect to α equal to zero and finally solve for α:

$$\begin{aligned}
\log\ L_0(\alpha) &= r\log\ \alpha - \alpha \sum_{i=1}^{r} t_i \\
\frac{\partial \log\ L_0(\alpha)}{d\alpha} &= \frac{r}{\alpha} - \sum_{i=1}^{r} t_i \\
\left.\frac{\partial \log\ L_0(\alpha)}{d\alpha}\right|_{\alpha=\hat{\alpha}_0} &= 0 \\
&\Leftrightarrow\ \frac{r}{\hat{\alpha}_0} - \sum_{i=1}^{r} t_i = 0 \\
&\Leftrightarrow\ \hat{\alpha}_0 = \frac{r}{\sum_{i=1}^{r} t_i} \\
&\Leftrightarrow\ \hat{\alpha}_0 = [\log\ z^{(k-r)} - r^{-1} \sum_{i=1}^{r-1} \log\ z^{(k-i)}]^{-1} \qquad (2.11)
\end{aligned}$$

Moment properties of the maximum likelihood estimators $\hat{\alpha}_0$ can easily be derived

$$E[\hat{\alpha}_0 | Z^{(k-r)} \leq d] = E[\frac{r}{\sum_{i=1}^{r} t_i}] = rE[\frac{1}{\sum_{i=1}^{r} t_i}]$$

Since $\alpha T_i \sim \text{Exp}(1)$,

$$\begin{aligned}
\sum_{i=1}^{r} T_i &\sim Gamma(r, \frac{1}{\alpha}) \\
&\sim \frac{\alpha^r}{\Gamma(r)} t^{r-1} e^{-\alpha t}, \quad \text{for} \quad t > 0,
\end{aligned}$$

so that

$$\begin{aligned} E[\hat{\alpha}_0|Z^{(k-r)} \le d] &= r\int_0^\infty t^{-1}\frac{\alpha^r}{\Gamma(r)}t^{r-1}e^{-\alpha t}dt \\ &= \frac{r\alpha^r}{\Gamma(r)}\int_0^\infty t^{r-2}e^{-\alpha t}dt. \end{aligned}$$

Using the form of the gamma density, we get

$$\begin{aligned} E[\hat{\alpha}_0|Z^{(k-r)} \le d] &= \frac{r\alpha^r(r-2)!}{(r-1)!\alpha^{r-1}} \\ &= \frac{r\alpha}{r-1}. \end{aligned}$$

Thus, $\hat{\alpha}_0$ is biased and its bias is

$$\begin{aligned} BIAS_{\hat{\alpha}_0} &= \frac{r\alpha}{r-1} - \alpha \\ &= (\frac{r}{r-1} - 1)\alpha \\ &= \frac{\alpha}{r-1}. \end{aligned}$$

The variance of $\hat{\alpha}_0$ can be calculated with

$$Var[\hat{\alpha}_0|Z^{(k-r)} \le d] = E[\hat{\alpha}_0^2|Z^{(k-r)} \le d] - E[\hat{\alpha}_0|Z^{(k-r)} \le d]^2,$$

so that we first need

$$\begin{aligned} E[\hat{\alpha}_0^2|Z^{(k-r)} \le d] &= r^2\int_0^\infty t^{-2}\frac{\alpha^r}{\Gamma(r)}t^{r-1}e^{-\alpha t}dt \\ &= \frac{r^2\alpha^r}{\Gamma(r)}\int_0^\infty t^{r-3}e^{-\alpha t}dt. \end{aligned}$$

Again using the form of the gamma density, we get

$$\begin{aligned} E[\hat{\alpha}_0^2|Z^{(k-r)} \le d] &= \frac{r^2\alpha^r(r-3)!}{(r-1)!\alpha^{r-2}} \\ &= \frac{\alpha^2r^2}{(r-1)(r-2)}, \end{aligned}$$

so that the variance becomes

$$Var[\hat{\alpha}_0|Z^{(k-r)} \le d] = \frac{\alpha^2r^2}{(r-1)(r-2)} - \frac{\alpha^2r^2}{(r-1)^2}$$

$$\begin{aligned} &= \frac{\alpha^2 r^2 (r-1) - \alpha^2 r^2 (r-2)}{(r-1)^2 (r-2)} \\ &= \frac{\alpha^2 r^2}{(r-1)^2 (r-2)}. \end{aligned}$$

2.4 Special Case - Upper Tail

In order to deal with upper tails we can exploit the results from the lower tail special case by means of transformations of the random variables. Suppose $Y_1, \ldots, Y_k$ is a random sample with distribution function $G(y) = 1 - Cy^{-\alpha}$ for $y \geq D$, where D is a known constant. Inference about α is now based upon the largest order statistic for which $y \geq D$ holds. Note that, if the Y's are transformed, such that $X = Y^{-1}$ and $Z^{(i)} = Y^{(k-i+1)^{-1}}$ for $i = 1, \ldots, k$, then the theory developed for the lower tail case is directly applicable, so that

$$\begin{aligned} 1 - G(y) &= Cy^{-\alpha} \\ &= 1 - Pr\{Y \leq y\} \\ &= Pr\{Y \geq y\} \\ &= Pr\{Y^{-1} \leq y^{-1}\} \\ &= P\{X \leq x\} \\ &= Cx^{\alpha} \end{aligned}$$

for $x \leq d = D^{-1}$.

The maximum likelihood estimate for α can be obtained from equation (2.11) by simply expressing the $Z^{(i)}$ in terms of the $Y^{(i)}$.

$$\begin{aligned} \hat{\alpha}_0 &= r(\sum_{i=1}^{r} t_i)^{-1} \\ &= r(\sum_{i=1}^{r} i[\log z^{(k-i)} - \log z^{(k-i+1)}])^{-1} \end{aligned}$$

$$
\begin{aligned}
&= r(\sum_{i=1}^{r} i[\log\, y^{(i)} - \log\, y^{(i+1)}])^{-1} \\
&= r(\sum_{i=1}^{r} \log\, y^{(i)} - r\log\, y^{(r+1)})^{-1}. \qquad (2.12)
\end{aligned}
$$

A more general approach can also be made, where $G(y) = w(y;\theta)$ is assumed to be a valid distribution for an upper tail if $y \geq D$. Then, conditional upon $Y^{(r+1)} \geq D$, we obtain as in the lower tail case

$$E_i = (k - i + 1)[\log\, w(Y^{(i-1)};\theta) - \log\, w(Y^{(i)};\theta)] \qquad (2.13)$$

for $i = 2, \ldots, r+1$, and

$$E_1 = -k\log\, w(Y^{(1)};\theta).$$

Thus, the conditional likelihood function for θ is

$$
\begin{aligned}
L_1(\theta) &\propto |J| f_{E_1,\ldots,E_{r+1}}(e_1, \ldots, e_{r+1}) \\
&\propto |J| \exp\{k\log\, w(y^{(1)};\theta) - \sum_{i=1}^{r}(k-i)[\log\, w(y^{(i)};\theta) - \log\, w(y^{(i+1)};\theta)]\},
\end{aligned}
$$

where $|J|$ is proportional to $\prod_{i=1}^{r+1} \frac{\partial log\ w(y^{(i)};\theta)}{\partial y^{(i)}}$.

Chapter 3

Asymptotic Behaviour

Throughout this section we explore the results on asymptotic behaviour of Hill's estimator. Asymptotic behaviour of Hill's estimator can be used to

- compare Hill's estimator with its competitors,
- construct confidence intervals,
- and determine the optimal number of order statistics used in the estimation.

First we introduce the general class of extreme value distributions, then we will focus on Pareto-type distributions.

For the latter class of distributions Hill's estimator $\hat{\alpha}$ is a popular estimator for the unknown parameter α, often referred to as the tail index, and Mason (1982) and Deheuvels et al. (1988) established weak and strong consistency of $\hat{\alpha}$, respectively. To show asymptotic normality additional conditions on the underlying distribution function are necessary. However the different analytic conditions proposed in numerous papers on the problem of asymptotic normality do not lend themselves to easy comparison.

To be consistent with the notation used in the other sections, let $X_1, \ldots, X_k$ be a sequence of positive independent and identically distributed random variables from some distribution with distribution function F. As before $Z_k \leq \cdots \leq Z_1$ will denote the corresponding order statistics. Suppose for some constants $a_k > 0$ and $b_k \in \mathbf{R}$

and some $\gamma \in \mathbf{R}$

$$\lim_{k\to\infty} P\left(\frac{Z_1 - b_k}{a_k} \leq x\right) = G_\gamma(x) \tag{3.1}$$

for all x where $G_\gamma(x)$ is one of the extreme value distributions given by

$$G_\gamma(x) := \exp\{-(1+\gamma x)^{-1/\gamma}\}, \tag{3.2}$$

where γ is a real parameter and x is such that $1 + \gamma x > 0$. (For $\gamma = 0$, we interpret $(1+\gamma x)^{-1/\gamma}$ as e^{-x}.) Recall that $Z_1 = \max\{X_1, \ldots, X_k\}$. F is said to be in the domain of attraction of G_γ, the generalized extreme value distribution, if (3.1) holds [notation $F \in D(G_\gamma)$].

The question is how to estimate the extreme value index γ from the given sample. If γ is negative, then Hill's estimator can not be used. If, on the other hand, γ is positive, then the estimation of γ corresponds to the estimation of the tail index of a distribution. In this case $\hat{\alpha}$ is a popular estimator. Therefore, we introduce the Pareto-type distributions which possess typical heavy tails and which form the basis for investigation of asymptotic behaviour of $\hat{\alpha}$.

Without loss of generality, let us assume that $F(0) = 0$. We say that X has a heavy tailed distribution if

$$P(X > x) = 1 - F(x) = x^{-\alpha} L(x) \quad \text{for} \quad x > 0, \tag{3.3}$$

for some $0 < \alpha < \infty$ and some function L slowly varying at infinity:

$$\frac{L(\lambda x)}{L(x)} \to 1 \quad \text{when} \quad x \to \infty \quad \text{and} \quad \lambda > 0. \tag{3.4}$$

We say that F is of Pareto-type, and α is often referred to as the Pareto index.

Another way to express (3.3) is to say that $1 - F$ is regularly varying with index $-\alpha$. A distribution F concentrating on $[0, \infty)$ has a regularly varying tail with index $-\alpha$, $\alpha > 0$ (written $1 - F \in RV_{-\alpha}$) if

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(x)} = x^{-\alpha}, \quad x > 0. \tag{3.5}$$

The goal is to estimate the tail index α (or equivalently $1/\alpha$, which is the more common notation which we will follow) and the proposed estimator from (2.12) can be applied:

$$\hat{\alpha}_{k,r} := \frac{1}{r} \sum_{i=1}^{r} \log Z^{(i)} - \log Z^{(r+1)}. \tag{3.6}$$

We observe that the assumption of regular variation is sufficient to show consistency of $\hat{\alpha}_{k,r}$. However, in Section 3.2 we will need to introduce a second order refinement of (3.5) which has proven very useful in establishing asymptotic normality of Hill's estimator; i.e., for asymptotic normality we need a more stringent condition on F, often referred to as the second-order condition which specifies the rate of convergence in (3.3). In simpler terms, we need to condition on the tail behaviour beyond the defining condition (3.4).

Throughout Chapter 3 we will assume that r, the number of upper order statistics used in $\hat{\alpha}_{k,r}$, is a sequence of positive integers satisfying

$$1 \le r \le k-1, \quad r = r(k) \to \infty \quad \text{and} \quad r/k \to 0 \quad \text{as} \quad k \to \infty. \tag{3.7}$$

These mathematical conditions are explained intuitively, for example, in Embrechts et al. (1997):

$r(k) \to \infty$: use a sufficiently large number of order statistics, but

$r/k \to 0$: we should only concentrate on the upper order statistics, as we are interested in the tail property or, *let the tail speak for itself.*

In order to obtain statistical properties like consistency and asymptotic normality the assumption (4.7) and in some cases additional conditions on the sequence $\{r(k)\}$ are necessary. If r were held to be fixed as k increases, then $\hat{\alpha}_{k,r}$ converges in law to a gamma distribution; see for example Haeusler and Teugels (1985). Since the choice of r is crucial to the performance of Hill's estimator, we will address that problem in the subsequent chapter.

3.1 Consistency

Mason (1982) studied necessary and sufficient conditions for which $\hat{\alpha}_{k,r}$ converges in probability or almost surely to a finite positive constant α. He proved that for any $0 < \alpha < \infty$ the following three statements are equivalent:

(i) F has an upper tail of form given in (3.3);

(ii) $\hat{\alpha}_{k,r} \to \alpha$ in probability as $k \to \infty$ for all sequences $r(k)$ satisfying (3.7);

(iii) $\hat{\alpha}_{k,r} \to \alpha$ almost surely as $k \to \infty$ for any sequence of the form $r(k) = [k^c]$ with $0 < c < 1$, where $[x]$ denotes the integer part of x.

Statement (ii) implies that Hill's estimator is weakly consistent for all sequences satisfying (3.7) if and only if $1 - F \in RV_{-\alpha}$. If the sequence is of the special form $r(k) = [k^c]$ with $0 < c < 1$, then $\hat{\alpha}_{k,r}$ is a strongly consistent estimator of α (statement (iii)). From the equivalence of the first two statements it follows that the weak consistency of Hill's estimator characterizes fully the Pareto type distributions in general. Previous closely related work on the characterization of $RV_{-\alpha}$ can be found in de Haan (1983), de Haan and Resnick (1980) and Teugels (1981b).

Deheuvels et al. (1988) characterized those sequences $r(k)$ for which Hill's estimator is strongly consistent. Assuming that F satisfies (3.3) for some $0 < \alpha < \infty$, they prove that whenever $r(k)/\log\log k \to \infty$ as $k \to \infty$ and $r(k)/k \to 0$ as $k \to \infty$, then

$$\lim_{k\to\infty} \hat{\alpha}_{k,r} = \alpha \quad \text{a.s.}; \tag{3.8}$$

i.e., $\hat{\alpha}_{k,r}$ is a strongly consistent estimator of α. According to the characterization of Pareto type tails due to Mason (1982), it follows that F has an upper tail of the form given in (3.3) if and only if

(iv) $\hat{\alpha}_{k,r} \to \alpha$ almost surely as $k \to \infty$ for all sequences $r(k)$ satisfying (3.7) such that $r(k)/\log\log k \to \infty$ as $k \to \infty$.

3.2 Asymptotic Normality

In the preceding section it was stated that consistency of Hill's estimator is equivalent to regular variation of $1-F$. To establish asymptotic normality of $\hat{\alpha}_{k,r}$ second-order regular variation of the distribution tail is needed.

A distribution tail $1-F$ is second-order regularly varying with first-order parameter $-\alpha$ and second order parameter ρ (written $1-F \in 2RV_{-\alpha,\rho}$) if there exists a measurable function $A(t)$ with constant sign such that the following refinement of (3.3) holds:

$$\lim_{t\to\infty} \frac{\frac{1-F(tx)}{1-F(t)} - x^{-\alpha}}{A(t)} = cx^{-\alpha} \int_1^x u^{\rho-1} du, x > 0 \tag{3.9}$$

for $c \neq 0$. Note that for $\rho < 0$

$$\lim_{t\to\infty} \frac{\frac{1-F(tx)}{1-F(t)} - x^{-\alpha}}{A(t)} = cx^{-\alpha} \frac{x^{\rho}-1}{\rho}. \tag{3.10}$$

We will follow the common notation $\rho = -\beta$ when the second-order parameter is assumed to be constant. If $1-F$ satisfies (3.9), then it is known that

$$\sqrt{r(k)}(\hat{\alpha}_{k,r} - \alpha) \xrightarrow{D} N(0, \alpha^2) \tag{3.11}$$

provided the sequence $r(k)$ satisfies (3.7) and an additional restriction depending on the second-order condition (see, for example, de Haan and Resnick (1998)). Here, $\xrightarrow{D}$ denotes convergence in distribution and $N(0, \alpha^2)$ denotes a normally distributed random variable with expectation 0 and variance α^2.

The first result on asymptotic normality is due to Hall (1982) who restricted his attention to "smooth" distributions with slowly varying functions which converge to a constant at polynomial rate (a special case of (3.10)),

$$1 - F(x) = cx^{-\alpha}[1 + O(x^{-\beta})] \quad \text{as} \quad x \to \infty \tag{3.12}$$

where $\alpha > 0$ and c and β are positive constants. Hall established the existence of an optimal sequence $r(k)$, in particular

$$r \to \infty \quad \text{and} \quad r/k^{2\beta(2\beta+\alpha)} \to 0 \quad \text{as} \quad k \to \infty, \tag{3.13}$$

for which (3.11) holds. The sequence $r(k)$ is optimal in the sense that for any norming constant s_k, $s_k(\hat{\alpha}_{k,r} - \alpha)$ never converges in distribution to a non-degenerate limit if r tends to infinity faster than in (3.13).

Hall was partly motivated by Teugels (1981b) and de Haan and Resnick (1980), who proposed simple asymptotic estimates of the tail index as alternatives to Hill's estimator. The estimator they suggested can be used more generally assuming only first-order regular variation, but the price paid for this generality is a slow rate of convergence. Hence very large sample sizes would be necessary for the estimator to be reasonably accurate. For the same set of conditions Hall and Welsh (1984) proved that $\hat{\alpha}_{k,r}$ converges at a rate which is optimal in the class of all possible estimators for α. See also, Smith (1987) and Drees (1998).

Moreover Hall (1982) showed asymptotic normality for the special case of (3.9),

$$1 - F(x) = Cx^{-\alpha}[1 + Dx^{-\beta} + o(x^{-\beta})] \quad \text{as} \quad x \to \infty, \tag{3.14}$$

where $\alpha > 0$, $C > 0$, $\beta > 0$ and D is a real number. If (3.14) holds and $r_0 := r(k) \sim \lambda k^{2\beta/(2\beta+\alpha)}$ for a positive constant λ then

$$k^{\beta/(2\beta+\alpha)}(\hat{\alpha}_{k,r} - \hat{\alpha}) \xrightarrow{D} N(DC^{-\beta/\alpha}\alpha\beta(\alpha+\beta)^{-1}\lambda^{\beta/\alpha}, \alpha^2\lambda^{-1}). \tag{3.15}$$

For the same second-order condition on the underlying distribution function, Hall and Welsh (1985) constructed an estimate $\hat{r}$ of r and proved that the estimator $\hat{\alpha}_{k,\hat{r}}$ shares optimal convergence with $\hat{\alpha}_{k,r_0}$, the estimator based on the sequence proposed by Hall (1982). In particular, they proved that (3.15) holds if $r = r(k) = [\lambda k^{2\beta(2\beta+\alpha)}]$ and $\hat{r}/r \to 1$ in probability. No further assumptions about the random sequence $\hat{r}$ are required. We will get back to this result in Chapter 4.

Goldie and Smith (1987) investigated the slowly varying function as given in (3.4) and applied some of their important results to the estimation problem of the tail index parameter. In accordance with Hall (1982) they proved that if F has a regularly varying tail as in (3.14) then (3.15) holds. In addition, they extended Hall's result since less stringent conditions on the slowly varying part of the underlying distribution are necessary (see Theorem 4.3.2 in [24]).

Haeusler and Teugels (1985) derived a general condition for $1 - F \in RV_{-\alpha}$ from which it is possible to compute all sequences $r(k)$ for which (3.11) holds, i.e. for which Hill's estimator is asymptotically normal. The computation is only possible if some prior knowledge about the slowly varying function is available. For the basic condition and simpler, rather manageable forms under appropriate assumptions on L, the reader is referred to Sections 3 and 4 of [27]. We will concentrate on the discussion in Section 5, where the main results were applied in several examples. Suppose the model given in (3.3) satisfies

(i) $L(x) = C[1 + O(x^{-\beta})]$ as $x \to \infty$; $C, \alpha, \beta > 0$,

(ii) $L(x) = C[1 + Dx^{-\beta} + o(x^{-\beta})]$ as $x \to \infty$; $C, \alpha, \beta > 0$, $D \in \mathbf{R}$,

(iii) $L(x) = C(\log x)^{\beta}$, x large, $C, \alpha > 0$, $\beta \in \mathbf{R} \backslash \{0\}$, and

(iv) $L(x) = C[1 + O(\log x)^{-\beta}]$ as $x \to \infty$.

All these conditions (i)-(iv) satisfy the general condition proposed by Haeusler and Teugels, so that the sequences $r(k)$ can be computed for which $\sqrt{r(k)}(\hat{\alpha}_{k,r} - \alpha)$ is asymptotically normal:

(i) and (ii) $r(k) \to \infty$ such that $r(k) = o(k^{2\beta(2\beta+\alpha)})$,

(iii) $r(k) \to \infty$ such that $r(k) = o((\log k)^2)$,

(iv) $r(k) \to \infty$ such that $r(k) = o((\log k)^{2\beta})$.

Note that (i) is equivalent to (3.13) and (ii) corresponds to r_0, which were the two special cases examined by Hall (1982).

For further results on asymptotic normality of $\hat{\alpha}_{k,r}$ and the necessity of some sort of second-order condition, see Beirlant and Teugels (1986,1987) who extended the conditions for asymptotic normality to distribution functions in the domain of attraction of the limit gamma law. Under different sets of conditions on $L(x)$, Davis and Resnick (1984) also studied the problem of asymptotic behaviour where Hill's estimator was applied to estimate the survival function at great age, in case F is in the domain of attraction of an appropriate extreme value distribution.

Csörgő and Mason (1985) investigated the asymptotic distribution of $\hat{\alpha}_{k,r}$ in yet another way. First, they define the (left continuous) inverse or quantile function F^{-1} of F by

$$F^{-1}(s) = \inf\{x : F(x) \geq s\}, \quad 0 < s < 1. \tag{3.16}$$

Now $1 - F \in RV_{-\alpha}$ if and only if $F^{-1}(1-s) = s^{-\alpha} l(s)$, $0 < s < 1$, where l is a function slowly varying at zero. Using the Karamata's representation (see, for example, Bingham et al. (1987)), $F^{-1}(1-s)$ can be expressed as

$$F^{-1}(1-s) = s^{-\alpha} a(s) \exp\{\int_s^1 \frac{b(u)}{u} du\}, \quad 0 < s < 1, \tag{3.17}$$

where $\lim_{s\downarrow 0} a(s) = a_0$ with $0 < a_0 < \infty$ and $\lim_{s\downarrow 0} b(s) = 0$.

To prove asymptotic normality of Hill's estimator, Csörgő and Mason assumed $a(s)$ to be constantly a_0 in a non-degenerate right neighbourhood of zero. Furthermore let $RV^*_{-\alpha}$ denote the class of distribution functions associated with such $a(s)$.

One of their main theorems is that if $F \in RV^*_{-\alpha}$ and $r(k)$ satisfies (3.7) then

$$\sqrt{r(k)}(\hat{\alpha}_{k,r} - \alpha_k) \xrightarrow{D} N(0, \alpha^2), \tag{3.18}$$

where

$$\alpha_k = \frac{k}{r(k)} \int_{1-(\frac{r(k)}{k})}^1 (1-s) d \log F^{-1}(s).$$

In order to obtain asymptotic normality when the centering sequence α_k is replaced by the fixed asymptotic mean α, additional assumptions on the underlying distribution F are necessary.

Assume F is such that for some $0 < \alpha < \infty$, $0 < C < \infty$ and $0 < \beta < \infty$,

$$F(x) = Ce^{-x/\alpha}\{1 + O(e^{-\beta x})\} \quad \text{as} \quad x \to \infty. \tag{3.19}$$

Csörgő and Mason (1985) proved that for every sequence $r(k)$ of positive integers satisfying (3.7) $\sqrt{r(k)}(\hat{\alpha}_{k,r} - \alpha)$ is asymptotically normal. For further investigation of asymptotic normality of Hill's estimator following this approach, the reader is referred to Csörgő and Viharos (1995), who were able to obtain even more general results. To do so, they had to change the norming sequence $\sqrt{r(k)}$ to sequences depending also on the unknown slowly varying function $l(s)$ in complicated ways.

Additionally, Csörgő and Viharos showed that while the weak consistency of $\hat{\alpha}_{k,r}$ for all sequences $r(k)$ satisfying (3.7) fully characterizes Pareto-type distributions as stated earlier in this section as a result by Mason (1982), Hill's estimator is <u>not</u> universally asymptotically normal over $RV_{-\alpha}$. In particular they constructed distribution function $F \in RV_{-\alpha}$ for which $\hat{\alpha}_{k,[k^{2/3}]}$ does not have a non-degenerate asymptotic distribution for any centering and norming sequence (see also Csörgő and Viharos (1997)).

However, it is known that second-order regular variation plays an important role in establishing asymptotic normality of $\hat{\alpha}_{k,r}$. Geluk et al. (1997) showed that under a strengthening of (3.3) called the von Mises condition, namely,

$$\lim_{x \to \infty} \frac{xF'(x)}{1 - F(x)} = \alpha, \tag{3.20}$$

second-order regular variation is equivalent to asymptotic normality of Hill's estimator. De Haan and Resnick (1998) studied whether the von Mises condition for the distribution tail could be weakened. They proved that it is too strong in order to have asymptotic normality of $\sqrt{r(k)}(\hat{\alpha}_{k,r}-\bar{\alpha}_{k,r})$, where $\bar{\alpha}_{k,r}$ is a non-constant asymptotic mean; they also give minimal condition on the distribution so that normality holds. Nonetheless, if the non-constant asymptotic mean is replaced by a constant centering like α, a somewhat stronger assumption like second-order regular variation is necessary.

The asymptotic behaviour of Hill's estimator was and still is broadly studied. We have seen that in order to establish consistency and asymptotic normality, first-order and second-order regular variation are of particular importance, respectively. Furthermore, the properties and performance of Hill's estimator crucially depend on the number of order statistics used in estimation. The next chapter will address the question on how to choose r.

Chapter 4

Optimal Choice Of The Sample Fraction

As seen in Chapter 3 the performance of Hill's estimator depends crucially on the number of order statistics used in the estimation. Therefore considerable interest has been shown in methods for choosing r. For a special class of marginal distributions, adaptive methods for determining r were proposed by Hall (1982), Hall and Welsh (1985) and Hall (1990), among others. The criterion for an optimal choice of the number of order statistics is minimization of the asymptotic mean squared error (AMSE). Unfortunately, the optimal choice of r depends mainly on the unknown slowly varying part of the distribution tail. Thus, it is difficult to obtain a practical strategy for minimizing AMSE through an appropriate choice of r.

In practice, there exist several graphical tools which not only suggest the optimal number of order statistics used in estimation, but also can be used directly to estimate α. The most common graphical method is the Hill plot, which plots r against $\hat{\alpha}_{k,r}$. Resnick and Stărică (1997), for example, explored graphical methods based on this plot that can minimize the difficulty of choosing r. In order to visualize graphical possibilities to simplify the determination of r and also the choice of $\hat{\alpha}_{k,r}$, we will present an example.

We will conclude this introduction by giving an analytic method to choose r that was proposed in Hill's (1975) seminal paper and that was motivation for further investigation of this problem.

A subjective choice of the cutoff point d (or equivalently D), which corresponds to the choice of r, is often very difficult or inappropriate. In the original paper Hill

(1975) proposed some data analytic techniques which can be useful in the choice of r. Let us consider the special case for the lower tail as in Section 2.3, where the function of the form $w(x;\theta) = Cx^{\alpha}$ was assumed to be valid for the region $x \leq d$, but no longer assume that d is known. It follows from (2.8) that, conditional on $Z^{(k-r)} \leq d$, the αT_i's are Exp(1).

Now, if r has been chosen sufficiently small, so that in fact $z^{(k-r)} \leq d$, then the αT_i's should behave like a random sample from Exp(1), at least for $i = 1, 2, \ldots, r$. On the other hand, if r has been chosen too large, so that $z^{(k-r)}$ turns out to be $\geq d$, then the αT_i's should exhibit a behaviour particularly different from a standard exponential distribution. In this case, one ($Z^{(k-r)}$) or perhaps more order statistics have values where the approximation of $w(x;\theta)$ by Cx^{α} is poor. Based on this observation, one could test the hypothesis that the αT_i's have an exponential distribution for $i = 1, \ldots, r$, using, for example, the chi-square goodness-of-fit test. To determine the optimal r_{opt}, one could choose a particular (small) r, for which the hypothesis is accepted, and then increase r step by step until the hypothesis is being rejected.

Hall and Welsh (1985) showed that the simple and attractive sequential decision procedure proposed by Hill results in using too many order statistics. Hill's method is based on the fact that the αT_i's are approximately distributed as centered exponential variables. This is a very good approximation if r is close to 1, but worsens as r increases. The sequence of goodness-of-fit tests, as suggested by Hill, is stopped at r_{opt} which is equal to the largest value of r which provides a satisfactory exponential fit. The problem that arises is the following. Since the deterioration of the exponential approximation is very gradual as r increases past the optimal threshold, the hypothesis of exponentiality will still be accepted. Therefore a large number of non-exponentials must be added before the hypothesis will be rejected. By using this sequential procedure, r tends to be overestimated. Hall and Welsh extend this heuristic argument by showing the inappropriateness of this method for the special class of marginal distributions, namely,

$$1 - F(x) = Cx^{\alpha}[1 + Dx^{\beta} + o(x^{\beta})]$$

and investigated how to choose r minimizing the AMSE of $\hat{\alpha}$.

4.1 Minimizing Asymptotic Mean Squared Error

Minimizing the AMSE as a criterion to choose r is very intuitive. If too many order statistics are used in estimation, then the estimator might have a large bias. If, on the other hand, too few order statistics are included, then the variance is large. Hence, depending on the precise choice of r and on the slowly varying function L, there is an important trade-off between bias and variance possible. In particular the second order behaviour of the underlying distribution plays an important role; i.e., the asymptotic behaviour beyond the defining property $L(\lambda x)/L(x) \to 1$, as $x \to \infty$.

Balancing the variance and bias components will lead to an optimal choice of r. Hall (1982) showed that this balance can be established by minimizing the AMSE, in particular if one chooses $r(k)$ by

$$\begin{aligned} r_{opt}(k) &:= \arg\min_r AsyE(\hat{\alpha}_{k,r} - \alpha)^2 \qquad (4.1)\\ &:= \arg\min_r [AsyVar(\hat{\alpha}_{k,r}) + AsyBias^2(\hat{\alpha}_{k,r})]. \end{aligned}$$

Then

$$\sqrt{r_{opt}(k)}(\hat{\alpha}_k(r_0(k)) - \alpha) \xrightarrow{d} N(b, \alpha^2), \qquad (4.2)$$

so that the optimal sequence $r_{opt}(k)$ results in an asymptotic bias b. When the first- and second-order conditions of the underlying distribution are known, it is possible to evaluate $r_{opt}(k)$ asymptotically.

First we will follow de Haan and Peng (1998), who studied the asymptotically optimal value of r comparing Hill's estimator to other estimators of the tail index.

Apart from (3.3) we will assume a special case of the second-order condition as given in (3.10). Suppose there exists a function A of constant sign such that

$$\lim_{t\to\infty} \frac{\frac{1-F(tx)}{1-F(t)} - x^{-\alpha}}{A(t)} = x^{-\alpha}\frac{x^{\rho}-1}{\rho} \tag{4.3}$$

for $x > 0$, where ($\rho \leq 0$) is the second order parameter, governing the rate of convergence of $\frac{1-F(tx)}{1-F(t)}$ to $x^{-\alpha}$. We can rephrase (4.3) in terms of the inverse function of the distribution F. Let U be the right (or left) continuous inverse of the function $1/(1-F)$ and write $a(t) = \alpha^{-2}A(U(t))$. The function $|a(\cdot)|$ is regularly varying with index ρ, i.e. $|a(\cdot)| \in RV_{\rho}$. Relation (4.3) is equivalent to

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^{1/\alpha}}{a(t)} = x^{1/\alpha}\frac{x^{\rho/\alpha}-1}{\rho/\alpha} \tag{4.4}$$

locally uniformly for $x > 0$.

From here it is possible to determine the AMSE and then the asymptotic optimal value of $r(k)$.

Suppose (4.4) holds. Let $r = r(k)$ be a sequence of integers with $r(k) \to \infty$, $r(k)/k \to 0$ as $k \to \infty$. If

$$\lim_{k\to\infty} \sqrt{r}a(\frac{k}{r}) = \lambda \in (-\infty, \infty), \tag{4.5}$$

then we have

$$\sqrt{r}(\hat{\alpha}_{k,r} - \alpha) \xrightarrow{d} N\left(\frac{\alpha^3\lambda}{\rho-\alpha}, \alpha^2\right) \quad \text{as} \quad k \to \infty.$$

Hence the AMSE of $\hat{\alpha}_{k,r}$ equals

$$\frac{1}{k}\left(\alpha^2 + \frac{\alpha^6\lambda^2}{(\rho-\alpha)^2}\right)$$

From (4.5) it follows that for r tending to infinity sufficiently slowly, i.e. using a moderate number of order statistics in Hill's estimator, $\lambda = 0$ will follow. In this case $\hat{\alpha}_{k,r}$ is asymptoticaly unbiased. On the other hand, if we take r as large as possible, then the asymptotic variance will decrease, but when doing so, a bias might enter. Let us consider the special case

$$1 - F(x) = Cx^{-\alpha}(1 + x^{-\beta}),$$

where c, α and β are positive constants. Notice that $\rho = -\beta$. Then it follows from (4.5) that

$$r \sim Ck^{(2\beta)/(2\beta+\alpha)}$$

where C is a constant, depending on α, β, c and λ. Moreover, $\lambda = 0$ if and only if $C = 0$, hence $r = o(k^{(2\beta)/(2\beta+\alpha)})$ (see for example Embrechts et al. (1997)). A more general result can be found in de Haan and Peng (1998).

Hall and Welsh (1985) proved that the AMSE of Hill's estimator is minimal for

$$r_{opt}(k) = \left(\frac{C^{2\rho}(1-\rho)^2}{2D^2\rho^3} \right)^{1/(2\rho+1)} \times k^{2\rho/(2\rho+1)}$$

where $\rho = \beta/\alpha$, if the underlying distribution satisfies

$$1 - F(x) = Cx^{\alpha}[1 + Dx^{\beta} + o(x^{\beta})].$$

But since the parameters α, β (and hence ρ), $C > 0$ and $D \neq 0$ are unknown, this result can not be applied directly to determine the optimal number or order statistics for a given data set without additional assumptions on ρ. A preliminary result was presented by Hall (1982), who showed for the above mentioned model it is optimal to choose $r = r(k)$ tending to infinity at a rate of order $o(k^{(2\beta)/(2\beta+\alpha)})$.

More generally, Hall and Welsh (1985) showed that if one wants to estimate the optimal sequence $r_{opt}(k)$ solely on the basis of the sample, i.e. determine an estimator $\hat{r}_0$ such that

$$\sqrt{\hat{r}_{opt}(k)}(\hat{\alpha}_k(\hat{r}_{opt}(k)) - \alpha) \xrightarrow{d} N(b, \alpha^2) \tag{4.6}$$

then it is sufficient to prove

$$\frac{\hat{r}_{opt}(k)}{r_{opt}(k)} \to 1, \tag{4.7}$$

in probability. To find such a $\hat{r}_{opt}$ several authors have suggested bootstrap methods (Hall (1990), Gomes (2001), Danielson et al. (2001), among others). The basic idea of the bootstrap method to find the optimal number of order statistics adaptively is the following.

In the proposed methods above, the asymptotically optimal choice of r via minimization of the mean squared error depends on the unknown parameter α and the function $a(t)$ (see Dekkers and de Haan (1993)). To overcome this problem it is possible to estimate the

$$AMSE = [AsyVar(\hat{\alpha}_{k,r}) + AsyBias^2(\hat{\alpha}_{k,r})] \tag{4.8}$$

by a bootstrap procedure. Then one can minimize the estimated AMSE to find the optimal r. We will give a short overview of how to apply the bootstrap and state some important results of this procedure.

First, resamples $\{X_1^*, \ldots, X_{k_1}^*\}$ are drawn from the original sample $\{X_1, \ldots, X_k\}$ with replacement, where the resample size k_1 is of smaller order than k, i.e. $k_1 < k$. Let $Z_{(k_1)}^* \leq \cdots \leq Z_{(1)}^*$ denote the order statistics corresponding to the resample and define

$$\alpha_{k_1}^*(r_1) := \frac{1}{r_1} \sum_{i=1}^{r_1} \log Z_{(i)}^* - \log Z_{(r_1+1)}^*.$$

Hall (1990) assumed that the underlying distribution is of the special form $F(x) = Cx^\alpha$ (as in Hill's lower tail special case) and proposed the bootstrap estimate of AMSE

$$\widehat{\text{AMSE}}(k_1, r_1) = E((\alpha_{k_1}^*(r_1) - \alpha_k(r))^2 | X_1^*, \ldots, X_{k_1}^*). \tag{4.9}$$

However, in this setup r needs to be chosen such that $\alpha_k(r)$ is consistent, so Hall assumed that the asymptotically optimal r is of the form ck^γ, where $0 < \gamma < 1$ is a known constant but c is unknown. Now, if $\hat{r}_1$ is asymptotic to ck_1^γ then

$$\hat{r} = \hat{r}_1(k/k_1)^\gamma$$

is asymptotic to cn^γ.

The problem is that r, or its form, is generally unknown. Therefore Danielson et al. (2001) suggested to replace $\alpha_r(k)$ in (4.9) with a more suitable statistic. Define

$$M_k(r) = \frac{1}{r}\sum_{i=1}^{r}(\log Z^{(i)} - \log Z^{(r+1)})^2,$$

then it is known that $M_k(r)/(2\alpha_k(r))$ is a consistent estimator of α, which also balances the bias and variance components if $r \to \infty$ with optimal rate (see, for example, Gomes and Martins (2002)). Danielson et al. proposed the bootstrap estimate of AMSE

$$\widehat{\text{AMSE}}^*(k_1, r_1) := E((M^*_{k_1}(r_1) - 2(\alpha^*_{k_1}(r_1))^2)^2 | x_k),$$

where

$$M^*_{k_1}(r_1) = \frac{1}{r_1}\sum_{i=1}^{r_1}(\log Z^*_{(i)} - \log Z^*_{(r_1+1)})^2.$$

We will summarize the procedure of choosing the optimal r according to Danielson et al.(2001):

1. For a given choice of k_1 draw bootstrap resamples of size k_1.

2. Calculate $\widehat{\text{AMSE}}^*(k_1, r_1)$, i.e., the bootstrap AMSE, at each r_1.

3. Find the $r^*_{1,opt}(k_1)$ which minimizes this bootstrap AMSE.

4. Repeat this procedure for an even smaller resample size k_2, where $k_2 = k_1^2/k$. This yields $r^*_{2,opt}(k_2)$.

5. Subsequently, calculate $\hat{r}_{opt}(k)$ from the formula

$$\hat{r}_{opt}(k) = \frac{(r^*_{1,opt}(k_1))^2}{r^*_{2,opt}(k_2)}\left(\frac{(\log r^*_{1,opt}(k_1))^2}{(2\log k_1 - \log r^*_{1,opt}(k_1))^2}\right)^{(\log k_1 - \log r^*_{1,opt}(k_1))/\log k_1}.$$

6. Finally, estimate α by $\hat{\alpha}_k(\hat{r}_{opt}(k))$.

The number of bootstrap resamples is determined by the computational facilities. For further suggestions on how to choose the number of resamples and how to determine the bootstrap sample size, see Danielson et al. (2001).

The above mentioned results on how to choose r via minimization of AMSE or $\widehat{\text{AMSE}}$ are useful mainly from a methodological point of view. As seen in Chapter 3 the properties of Hill's estimator crucially depend on the higher order behaviour of the underlying distribution tail $1 - F$. However, in practice we rarely verify this behaviour, for example conditions like (4.3), which was assumed throughout this section. Hence, there is a need for useful tools in practice, which can be applied more generally to determine r. The next section will present some graphical methods to overcome the problem of choosing r.

4.2 Graphical Tools

The preceding section showed that under suitable second-order conditions, an optimal $r_{opt}(k)$ can be determined such that the AMSE of Hill's estimator is minimized. The practical usefulness of this theoretical method is limited. Asymptotic results as $k \to \infty$ provide little guidance about finite sample behaviour. Additionally, $r_{opt}(k)$ depends on the unknown parameters of F, like the second-order condition which is rarely verifiable in practice. Data-driven alternatives to estimate r by $\hat{r}_{opt}(k)$, like the presented bootstrap method, also require choices of certain parameters, and the choices are arbitrary. In the bootstrap procedure, for example, one has to choose the number of resamples and the resample size.

Hence, there is a need for computationally less challenging methods for a variety of applied purposes and for the purpose of checking whether the above mentioned procedures provide reasonable choices of r. Therefore, the analysis of the tail behaviour of a distribution function based on Hill's estimator is often times summarized graphically. In this section we will present a few graphical methods to determine the optimal r, and which, moreover, can be used to obtain $\hat{\alpha}$ directly.

We will illustrate the techniques and plotting strategies on a particular data set, the Danish data on large fire insurance losses [46], which was the basis for a very fundamental case study of extreme value techniques by McNeil (1997). Resnick (1997) pointed out several alternate statistical techniques and plotting devices that support McNeil's conclusions and that can be employed with similar data sets. The 2156 observations in the Danish data are large fire insurance losses of over one million Danish Krone (DKK) from the years 1980 to 1990, inclusive. See Figure 4.1. The loss figure is a total loss figure for the event concerned, and includes damage to buildings, furnishing and personal property, as well as loss of profits.

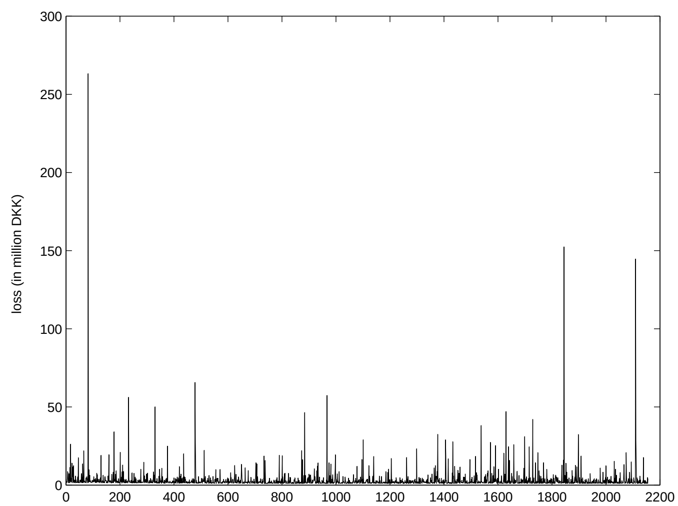

Figure 4.1: Danish Data on fire insurance losses

(1) The Hill plot. Let

$$\hat{\alpha}_{k,r} := H_{k,r} = \frac{1}{r} \sum_{i=1}^{r} \log \frac{Z^{(i)}}{Z^{(r+1)}}$$

denote Hill's estimator. The most basic instrument is the so called Hill plot, graphing

$$\{(r, H_{k,r}), \quad 1 \leq r \leq k-1\}. \tag{4.10}$$

In order to choose an appropriate number of order statistics on which the estimation of α will be based, one has to look for a stable region in the Hill plot. From the stable region of the plot one can infer a value of r and α.

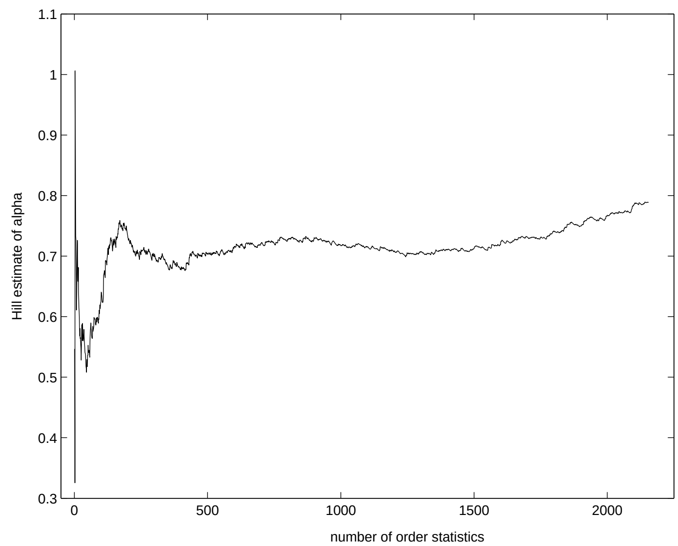

Figure 4.2: Hill plot

The example, Figure 4.2, shows a stable region for about $r \in [500, 1500]$. In this particular case we suggest about 500 order statistics to be an appropriate choice, since the graph is volatile when fewer order statistics are used and moderation is advised. The corresponding Hill estimate of α is approximately 0.7.

In general, it is known that if the underlying distribution is Pareto or close to Pareto, then then the Hill plot is a very powerful tool for determining α (see, for example Drees et al. (2000)). The Danish insurance data seem to follow a Pareto-type distribution (Embrechts et al (1997)). Whether F follows a Pareto-type distribution can be detected by Pareto quantile plots which are the basis for testing the goodness-of-fit hypothesis of strict Pareto behaviour. A broad discussion on Pareto-type models and quantile plots can be found in Beirlant et al. (1996) and its references.

In the example the choice of the stable region was rather obvious. However, in other cases the Hill plot can have a high volatility and hence the practitioner is confronted with two important and difficult decisions: first, determining a sensible range of r, and second, deciding on a specific value of r inside the range.

Resnick and Stărică (1997) introduced a computational as well as a graphical technique to reduce the difficulty of choosing the number of order statistics used to calculate Hill's estimator. The first method reduces the volatility of the Hill plot in the predetermined sensible region and hence assists the practitioner in deciding on the specific value of r. It is an averaging technique where values of Hill's estimator are "smoothed".

(2)Averaged or smoothed Hill plot. Let

$$avH_{k,r} := \frac{1}{(u-1)r} \sum_{p=r+1}^{ur} H_{k,p}, \tag{4.11}$$

where $u > 1$. Once one has determined a suitable range for r, say $[r_1, r_2]$, $avH_{r,k}$ can be calculated and the avHill plot

$$\{(r, avH_{k,r}), \quad r \in [r_1, r_2/u]\}$$

can be graphed.

In the selected region Hill's estimator values are smoothed, so that we have less volatility in our graph. Hence, it is less important to select the optimal r, since the estimate will not be as sensitive to the choice of r in comparison to the classical Hill plot. For a good choice of u Resnick and Stărică suggest taking a value between $k^{0.1}$ and $k^{0.2}$ in order to reach an equilibrium between variance reduction and a comfortable number of points used for the plot. We used $u = 3$ ($\approx k^{0.14}$) and decided to limit r not to be greater than 1800, so that $r_1 = 1$ and $r_2 = 1800$. Due to the fact that $u = 3$, the averaging stopped at $r = 600$. See Figure 4.3. Through averaging, the variance of Hill's estimator can be considerably reduced and the volatility of the plot tamed. The importance of selecting the optimal r diminishes and we can observe $\alpha \approx 0.7$

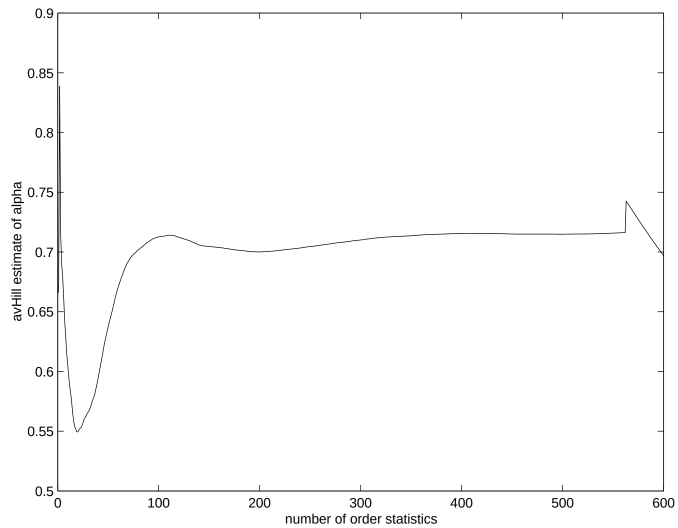

Figure 4.3: Smoothed Hill plot

(3) Alternative Hill plot. The second method to overcome the difficulty of choosing r is an alternative to the classical Hill plot, graphing

$$\{(\theta, H_{\lceil k^{\theta} \rceil, k}), \quad 0 \leq \theta \leq 1\},$$

where $\lceil y \rceil$ is written for the smallest integer greater or equal to $y \geq 0$. As with the classical Hill plot one tries to find a stable region of the graph.

The reason why the alternative Hill plot is more helpful in detecting such a region is simple. The significant part of the graph, i.e. the part corresponding to a relatively small number of order statistics, is displayed bigger in relation to the less important part of the graph. On the other hand, the part corresponding to a large number of order statistics gets rescaled, now covering less displayed space than in the traditional Hill plot. The region of interest is shown more precisely, thus making the interpretation of the graph easier and more accurate. See Figure 4.4. The stable region of the graph is $\theta \in [0.8, 0.9]$, which corresponds to $r \in [464, 1000]$. As before, we observe $\alpha \approx 0.7$. Note that the high volatility when too few order statistics are used now covers a larger portion of the graph.

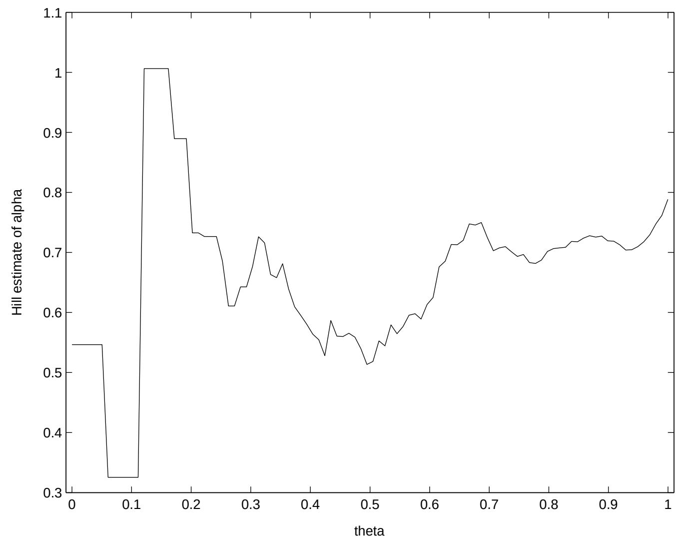

Figure 4.4: Alternative Hill plot

It is also possible to combine the two proposed methods: first, calculate the smoothed version of Hill's estimator and, second, plot the alternative Hill plot for avHill. Due to the fact that $u = 3$ the averaging stopped at $\theta = 0.8$. See figure 4.5.

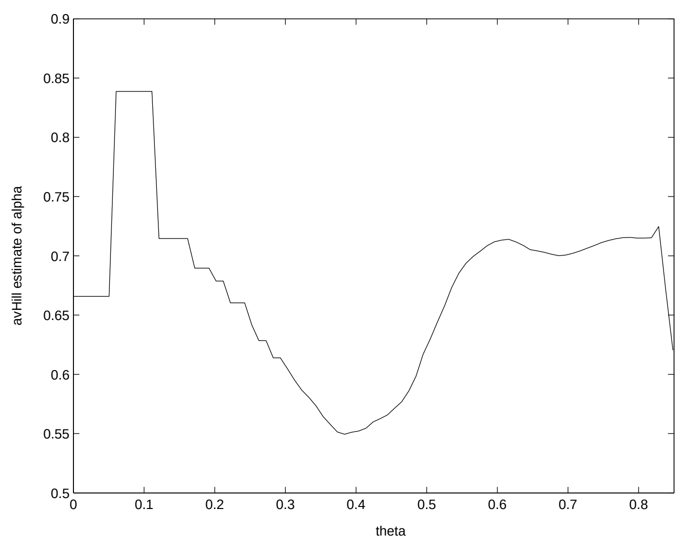

Figure 4.5: Alternative smoothed Hill plot

In spite of the relativly clear results, simulations have shown that there are finite sample cases in which the plots and Hill's estimator can not be trusted; see, for example, the "Hill horror plot" in Embrechts et al. (1997, fig. 4.1.13). If the slowly varying part of (3.3) happens to be constant, then we are dealing with the exact Pareto model $1 - F(x) = Cx^{-\alpha}$, the special case considered in 2.3. In this case Hill's estimator behaves well. However, if the ratio $L(tx)/L(x)$ converges to 1 at a slow rate, a large bias may be present.

Finally, we would like to note that we assumed throughout the discussion that the given sample is i.i.d. Hill's estimator can also be applied to dependent data, and its behaviour has been studied, for example, by Resnick and Stărică (1998) and its references.

Bibliography

[1] Beirlant, J. and Teugels, J. L. (1986). Asymptotics of Hill's estimator. *Theory Probab. Appl.* **31**, 530-536.

[2] Beirlant, J. and Teugels, J. L. (1987). Asymptotic normality of Hill's estimator. In *Extreme value theory, Oberwohlfach 1987*, eds. J. Hüsler and R. D. Reiss. Springer-Verlag, New York.

[3] Beirlant, J., Vynckier, P. and Teugels, J. L. (1996). *Practical analysis of extreme values*. Leuven University Press.

[4] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular variation*. Cambridge University Press, Cambridge.

[5] Csörgő, S. and Mason, D. (1985). Central limit theorems for sums of extreme values. *Math. Proc. Cambridge Philos. Soc.* **98**, 547-558.

[6] Csörgő, S. and Viharos, L. (1995). On the asymptotic normality of Hill's estimator. *Math. Proc. Cambridge Philos. Soc.* **118**, 375-382.

[7] Csörgő, S. and Viharos, L. (1997). On the asymptotic normality of Hill's estimator. *Bernoulli* **3**, 351-370.

[8] Danielsson, J., de Haan, L., Peng, L. and de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Anal.* **76**, 226-248.

[9] Davis, R. and Resnick, S. I. (1984). Tail estimates motivated by extreme value theory. *Ann. Statist.* **12**, 1467-1487.

[10] de Haan, L. and Resnick, S. I. (1980). A simple asymptotic estimate for the index of a stable distribution, *J. Roy. Statist. Soc. Ser. B* **42**, 83-87.

[11] de Haan, L. and Resnick, S. I. (1998). On asymptotic normality of the Hill estimator. *Comm. Statist. Stochastic Models* **14**, 849-866.

[12] de Haan, L. and Peng, L. (1998). Comparison of tail index estimators. *Statist. Neerl.* **52**, 60-70.

[13] de Haan, L. (1983). Corrigenda: "Estimation of the minimum of a function using order statistics" [J. Amer. Statist. Assoc. **76** (1981), 467-469]. *J. Amer. Statist. Assoc.* **78**, 1008.

[14] de Haan, L. (1994). Extreme value statistics. In *Extreme value theory and applications, Gaithersburg Maryland 1993*, eds. Galambos et al. Kluwer academic publishers.

[15] Deheuvels, P., Haeusler, E. and Mason, D. M. (1988). Almost sure convergence of the Hill estimator. *Math. Proc. Cambridge Philos. Soc.* **104**, 371-381.

[16] Dekkers, A. L. M. and de Haan, L. (1993). Optimal choice of sample fraction in extreme-value estimation. *J. Multivariate Anal.* **47**, 173-195.

[17] Dekkers, A. L. M., Einmahl, J. H. J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Ann. Statist.* **17**, 1833-1855.

[18] Drees, H., de Haan, L. and Resnick, S. I. (2000). How to make a Hill plot. *Ann. Statist.* **28**, 254-274.

[19] Drees, H. (1998). Optimal rates of convergence for estimates of the extreme value index. *Ann. Statist.* **26**, 434-448.

[20] Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic. Process. Appl.* **75**, 149-172.

[21] Drees, H., de Haan, L. and Resnick, S. I. (2000). How to make a Hill plot. *Ann. Statist.* **28**, 254-274.

[22] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling extremal events.* Springer-Verlag, Berlin.

[23] Geluk, J. L., de Haan, L., Resnick, S. and Stărică, C. (1997). Second-order regular variation, convolution and the central limit theorem. *Stochastic Process. Appl.* **69**, 139-159.

[24] Goldie, C. M. and Smith, R. L. (1987). Slow variation with remainder: theory and applications. *Quart. J. Math. Oxford Ser. (2)* **38**, 45-71.

[25] Gomes, M. I. and Oliviera, O. (2001). The bootstrap methodology in statistics of extremes — choice of the optimal sample fraction. *Extremes* **4**, 331-358.

[26] Gomes, M. I. and Martins, M. J. (2002). "Asymotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes* **5**, 5-31.

[27] Haeusler, E. and Teugels, J. L. (1985). On the asymptotic normality of Hill's estimator for the exponent of regular variation. *Ann. Statist.* **13**, 743-756.

[28] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B* **44**, 37-42.

[29] Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *J. Multivariate Anal.* **32**, 177-203.

[30] Hall, P. and Welsh, A. H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12**, 1079-1084.

[31] Hall, P. and Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *Ann. Statist.* **13**, 331-341.

[32] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163-1174.

[33] Mason, D. M. (1982). Laws of large numbers for sums of extreme values. *Ann. Probab.* **10**, 754-765.

[34] McNeil, A. (1997). Estimating the tails of loss severity distributions using extreme value theory. *Astin Bull.* **27**, 125-139.

[35] Nevzorov, V. B. (2001). *Records: mathematical theory.* Translated from the Russian manuscript by D. M. Chibisov. Translations of Mathematical Monographs, **194**. American Mathematical Society, Providence, RI.

[36] Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119-131.

[37] Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hungar.* **4**, 191-231.

[38] Resnick, S. I. and Stărică, C. (1997). Smoothing the Hill estimator. *Adv. in Appl. Probab.* **29**, 271-293.

[39] Resnick, S. I. and Stărică, C. (1998). Tail index estimation for dependent data. *Ann. Appl. Prob.* **8**, 1156-1183.

[40] Resnick, S. I. (1987). *Extreme values, regular variation, and point processes.* Apllied probability. A series of the applied probability trust **4**. Springer-Verlag, New York.

[41] Resnick, S. I. (1997). Discussion of the Danish data on large fire insurance losses. *Astin Bull.* **27**, 139-151.

[42] Smith, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15**, 1174-1207.

[43] Teugels, J. L. (1981a). Remarks on large claims. *Bull. Inst. Internat. Statist.* **49**, 1490-1500.

[44] Teugels, J. L. (1981b). Limit theorems on order statistics, *Ann. Probab.* **9**, 868-880.

[45] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73**, 812-815.

[46] *The Danish insurance data (univariate).* (n. d.). Retrieved June 6, 2003, from http://www.math.ethz.ch/∼mcneil/ftp/DanishData.txt