On Extreme Value Statistics

maximum likelihood	portfolio optimization
extremal rainfall	Internet auctions

ISBN: 978 90 5170 912 4

Cover design: Crasborn Graphics Designers bno, Valkenburg a.d. Geul

This book is no. 432 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

On Extreme Value Statistics

maximum likelihood   portfolio optimization
extremal rainfall   Internet auctions

Over extreme waarden statistiek

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the rector magnificus

Prof.dr. S.W.J. Lamberts

and in accordance with the decision of Doctoral Board.

The public defence shall be held on

Thursday 13 November 2008 at 13:30 hrs

by

CHEN ZHOU
born in Qingdao, China

ERASMUS UNIVERSITEIT ROTTERDAM

Doctoral Committee

Promoter:	Prof.dr. C. G. de Vries
Other members:	Prof.dr. P. P. Wakker
	Prof.dr. J. Hüsler
	Prof.dr. M. C. W. Janssen

Dedicated to Prof. Shihong Cheng

(1939-2007)

Preface

The minimal function of a PhD thesis is to obtain a PhD degree. After achieving this, the thesis might not be read any more except the preface. It is my honor that you are reading the preface now. I would appreciate more if you could at least turn one more page to read the table of content. From there you will find the structured picture of my 4-year PhD path.

In 2001, when I was an undergraduate student in Peking University, Prof. Shihong Cheng passed a PhD thesis to me for reading. I benefited a lot from reading not only the preface but its entire text. This experience lead me to the field Extreme Value Theory, and inspired me to do a PhD on it. There is no doubt that Prof. Shihong Cheng was expecting the current book a lot. Unfortunately, it is no longer possible for him to read it. I dedicate this book to him for having faith on me all the time.

Statistically, 100% of the past PhD theses has confirmed that: "there are so many people helping me during my PhD life". It also applies to me. To review all the stories about all the helpful people may result in a book as thick as the thesis itself. On the other hand, a simple list of names may include most of them but lose the details. A Chinese solution is to pick up a few significant representatives, while the missed names are by no means less important.

Talking about the thesis that inspired me to study Extreme Value Theory, it was published in 1970 and read by many scientists in this field. One could imagine my excitement when I noticed that I was going to work with the author for four years as my PhD life. The excitement was further enhanced after I met Laurens de Haan and found out that he is such a nice person and helpful supervisor. From the first time we met at Schiphol airport, Laurens started to carry the heaviest suitcase of mine. Helping on the top trouble is what Laurens always did in the past four years. With his great ideas and well-organized working schedule, I was never in short of creativity or concrete results during my PhD work. Besides, Laurens offered me much more in other aspects: history courses during lunch are always for free; tours around Lisbon are the best reward for hard working. In addition: the coffee in Laurens' office is the magic power for studying Extreme Value Theory.

Inspiring a mathematician with economic theory is one of the most difficult jobs in the world. Casper, you successfully achieved this by constantly turning the formulas into fun. Walking between 9th floor and 8th floor is always a good refreshment for me after working on a complicated calculation. On my way back, I always carry brilliant economic explanations from Casper that reflects the formulas. I thank him for helping me to understand the economic potential of my mathematical knowledge.

Life in Rotterdam is wonderful thanks to all my friends. To find one representative is rather difficult, since there are at least two names on the very top: Francesco and Michiel. I am quite sure that I can always count you on when something can not be done by myself. That could be solving a problem in research, fixing bicycles, holding a great party, or finishing half of the beer in De Smitse. Here the word "half" does not mean lacking of capacities from we three. It is simply because of our three "competitors": Chris, Jeroen and Joop. The entire "six-party discussions" became more and more interesting thanks to your participation.

Tinbergen Institute offers a great atmosphere for PhD students. I would thank to all TI staff for your patience in helping me and my fellow colleagues in 9th floor for your kindness.

As living in a long distance from my hometown, it is very lucky for me to have so many great Chinese friends who make my life here more home feeling. Helps from Chinese friends dated back to the day I landed here. I would like to thank Li Deyuan for hosting me at that moment. It is an important help for me to start my PhD smoothly.

My deepest thank should certainly goes to my mum. Four years ago, you decided not to keep me by your side but let me do what I want. That results in an incredible amount of difficulties for you during these years. You never asked anything back from me, instead, always supported me from distance. This is already more than any other help.

Normally, the last paragraph of the preface is devoted to thank the other half of the author with words like "to my love, without your patience and help, this thesis will never be finished." To my dearest Yijing, our case is slightly different: I am actually quiet and less disturbing when I was concentrating on my PhD work. Considering that I am not hard working, in fact, you have spent more patience than regular. Without you being aside, I may still finish the thesis, but I will definitely lose the entire colorful life. Your endless love is the fundamental resource of my energy for being happy all the time.

There is a last thank I reserve in my heart, for keeping me brave and confident to face any challenge in my life.

Zhou Chen, 28-08-2008

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the 18th century, statisticians sometimes worked as consultants to gamblers. In order to answer questions like "If a fair coin is flipped 100 times, what is the probability of getting 60 or more heads?", Abraham de Moivre discovered the so-called "normal curve". Independently, Pierre-Simon Laplace derived the central limit theorem, where the normal distribution acts as the limit for the distribution of the sample mean.

Nowadays, statisticians sometimes work as consultants for economists, to whom the normal distribution is far from a satisfactory model. For example, one may need to model large-impact financial events in order to to answer questions like "What is the probability of getting into a crisis period similar to the credit squeeze in 2007 in the coming 10 years?". At first glance, estimating the chances of events that rarely happen or even have never happened before sounds like a "mission impossible". The development of Extreme Value Theory (EVT) shows that it is in fact possible to achieve this goal.

Different from the central limit theorem, Extreme Value Theory starts from the limit distribution of the sample maximum. Initiated by M. Fréchet, R. Fisher and R. von Mises, the limit theory completed by B. Gnedenko, gave the fundamental assumption in EVT, the *"extreme value condition"*. Statistically, the extreme value condition provides a semi-parametric model for the tails of distribution functions. Therefore it can be applied to evaluate the rare events. On the other hand, since the assumption is rather general and natural, the semi-parametric model can have extensive applications in numerous fields.

Starting from J. Pickands, one-dimensional extreme value statistics solves the estimation of rare events regarding a single random variable. The one-dimensional extreme value statistics considers the tail of a distribution function as a specific parametric model. This allows us to estimate the parameters by extremal observations from data. This idea is different from a parametric approach in the sense that only the tail is parameterized without imposing any assumption at the moderate level. Correspondingly, in estimating the parameters, only the observations in the tail are used. Therefore, it provides a more accurate fit for the tail compared to the regular parametric approaches that also take the data at moderate levels into consideration. On the other hand, by using such a semi-parametric model, it is possible to evaluate a rare event which is more extreme than the events that we have observed. This is beyond the reach of non-parametric statistics.

Besides the one-dimensional problem, extreme value statistics can deal with multivariate rare events, i.e. rare events associated to random vectors. In fact, the world we are living in is rather complicated due to dependence: most of the rare events are characterized by a few (or even infinitely many) dependent random variables. This requires the development of multi-dimensional and infinite-dimensional EVT and their corresponding statistical methodologies. From the 1970s onwards, multivariate EVT has been gradually established. The basic idea is similar to the one-dimensional case: parameterize the tail of a multivariate distribution rather than the entire distribution function. However, the dependence structure is now an issue. The model of the dependence structure is important and can be non-parametric. Besides statistical analysis for rare events based on random vectors, the concept *"tail dependence"* arises in multivariate extreme value statistics. Tail dependence characterizes the relation among a few random variables for being simultaneously extreme. It is rather independent from the regular dependent concept. In practice, it successfully explains why "bad luck never comes alone".

Compared to one-dimensional and multi-dimensional EVT which are well developed and applied, the infinite-dimensional EVT is still a growing field. As the theory is getting more and more attention, it might capture more and more interest for application in the future. Infinite-dimensional extreme value statistics considers rare events characterized by stochastic processes that are commonly used to model randomness in time or space. The difficulty compared to finite-dimensional case is that the dependence structure of infinite-dimensional EVT has to be characterized on the functional space which is rather abstract.

In this thesis, we contribute to extreme value statistics by making theoretical improvement as well as applying it to finance, meteorology and economics. This chapter firstly reviews the fundamentals of EVT and its corresponding statistical inference in Section 1.1. Then, in Section 1.2, we briefly introduce the main content of each chapter in this thesis.

1.1 Extreme Value Theory

1.1.1 One-dimensional EVT

Let $X_1, X_2, \cdots$ be independent and identical distributed (i.i.d.) with distribution function F. Suppose that the distribution function F is in the domain of attraction of an extreme value distribution, i.e. there are a positive function a and a function b, such that

$$\lim_{n\to\infty} P\left(\max_{1\leq i\leq n} \frac{X_i - b(n)}{a(n)} \leq x\right) = G(x),$$

for each continuous point x of G, where G is a non-degenerate distribution function. We denote this by $F \in \mathcal{D}$. Then a and b can be chosen such that

$$G(x) = G_\gamma(x) := \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\}$$

for all x with $1 + \gamma x > 0$, where γ is a real constant. We call γ as the *extreme value index*[1]. Then we also say $F \in \mathcal{D}(G_\gamma)$.

One way to characterize the necessary and sufficient condition of being in the domain of attraction is the following property. The condition $F \in \mathcal{D}(G_\gamma)$ holds, if and only if there exists a positive function a_0 such that for all x with $1 + \gamma x > 0$,

$$\lim_{t\uparrow x^*} P\left(\frac{X-t}{a_0(t)} > x | X > t\right) = (1+\gamma x)^{-1/\gamma} =: 1 - Q_\gamma(x).$$

Here $x^* := \sup\{x : F(x) < 1\}$ is the right endpoint of F. This means that the larger observations in a sample approximately follow the probability distribution Q_γ - the generalized Pareto distribution (GPD). Hence, it is possible to extrapolate data in the tail of F from a GPD. Since the GPD is a parametric model, as soon as the parameter γ is precisely estimated, it is possible to calculate the probability of rare events. This is the essential idea of extreme value statistics.

Estimating the extreme value index, though, is by no means an easy task. One possible approach is to extrapolate data from the empirical excesses and to fit these to a GPD by using maximum likelihood procedure. This approach was initiated by Smith (1987) and results in the so-called maximum likelihood estimator for the extreme value index. This estimator performs reasonably well in practice. Theoretically, the asymptotic properties of the estimator had been investigated by Drees *et al.* (2004).

1.1.2 Multi-dimensional EVT

Let us now consider the multi-dimensional case, or rather the two-dimensional case for simplicity. Let (X, Y) be a random vector with distribution function F. Suppose $F \in \mathcal{D}$,

[1] When $\gamma > 0$, $\alpha = 1/\gamma$ is called the tail index.

i.e. if $(X_1, Y_1), (X_2, Y_2), \cdots$ are i.i.d. random vectors with distribution function F, there are positive functions a and c and functions b and d, such that

$$\lim_{n\to\infty} P\left(\max_{1\le i\le n} \frac{X_i - b(n)}{a(n)} \le x, \max_{1\le i\le n} \frac{Y_i - d(n)}{c(n)} \le y\right) = G(x,y),$$

where G is a two-dimensional distribution function with non-degenerate marginals. Then we also say that $F \in \mathcal{D}(G)$ and G is a (multivariate) extreme value distribution.

Similar to the one-dimensional case, there exists a related two-dimensional GPD function Q_H, obtained for example as follows:

$$\begin{aligned}&\lim_{t\to\infty} P\left(\frac{X-b(t)}{a(t)} > \frac{x^{\gamma_1}-1}{\gamma_1} \quad\text{or}\quad \frac{Y-d(t)}{c(t)} > \frac{y^{\gamma_2}-1}{\gamma_2} | X > b(t) \quad\text{or}\quad Y > d(t)\right)\\ =&2\int_0^1 \max\left(\frac{s}{x}, \frac{1-s}{y}\right) H(ds) =: 1 - Q_H(x,y),\end{aligned}$$

for $(x,y) \in D_H = \left\{(x,y) : 2\int_0^1 \max\left(\frac{s}{x}, \frac{1-s}{y}\right) H(ds) \le 1\right\} \supset \{(x,y) : x, y \ge 2\}$, where γ_1 and γ_2 are the marginal extreme value indices, and H is a probability distribution function on $[0,1]$ with mean $1/2$. Here H characterizes the dependence in the tail. Notice that H is quite different from the traditional dependence measures such as the correlation coefficient. The latter measures the dependence at the moderate level. Hence it is possible to have low correlation coefficient with a strong tail dependence and vice versa.

In multi-variate extreme value statistics, one needs to estimate the marginal parameters as well as the dependence measure H. For the statistics under multivariate EVT framework, we refer to Huang (1992). With estimations for the marginal parameters and the dependence structure, it is possible to study rare events characterized by (X, Y) belonging to a certain set, see de Haan and Ferreira (2006, Chapter 8).

1.1.3 Infinite-dimensional EVT

Infinite-dimensional EVT studies extremes on stochastic processes. Equivalently speaking, we consider extremes of random elements in a functional space. Since in application one usually uses the continuous stochastic process, correspondingly, we consider extremes in $C[0,1]$, the space of continuous functions defined on the unit interval. The setup is as follows. Let $\{X(s)\}_{s\in[0,1]}$ be a stochastic process in $C[0,1]$. Consider independent copies $X_1, X_2, \cdots$ of the process X. Compose for each n a continuous stochastic process

$$\left\{\max_{1\le i\le n} X_i(s)\right\}_{s\in[0,1]}.$$

Suppose that for some positive functions $a_s(n)$ and real functions $b_s(n)$, the sequence of processes

$$\left\{\max_{1\le i\le n} \frac{X_i(s) - b_s(n)}{a_s(n)}\right\}_{s\in[0,1]}$$

converges in $C[0,1]$. If this is the case, we say $X \in \mathcal{D}$. Let us call the limiting process $\{U(s)\}_{s\in[0,1]}$. Then we also say $X \in \mathcal{D}(U)$.

Similar to the multi-dimensional case, the extremes on stochastic processes can also be studied by separating the marginal parameter functions and the dependence structure. However, the dependence structure is more complicate. It is defined on a function space. To characterize the dependence structure via a measure on some function space is possible but not convenient for statistical applications. Recent developments of infinite-dimensional EVT provide statistically applicable representations of the dependence structure. These approaches create possibilities to parameterize the dependence structure. By estimating a few parameters in the parametric dependence structure, it is possible to study the rare events characterized by the original stochastic process.

1.2 Outline

In this thesis, we study extreme value statistics from the theoretical development to its applications, from the one-dimensional case to the infinite-dimensional case. The thesis bundles eight papers and is partitioned into four parts.

Part I considers one-dimensional EVT which consists of three chapters. We focus on the maximum likelihood estimator of the extreme value index initiated by Smith (1987). In the literature of estimating the extreme value index, a minimal requirement is that any estimator should be consistent under the extreme value condition. It has been proved that for most known estimators a more restrictive but natural condition–the second order condition–leads to the asymptotic normality. Roughly speaking, the second order condition specifies the speed of convergence in the extreme value condition, see de Haan and Stadtmüller (1996).

Since for the maximum likelihood estimator no explicit form is known, the study of its asymptotic properties is much more difficult than those cases in which the estimators are explicit functions of the observations. Smith (1987) proved the asymptotic normality under conditions that are rather more restrictive than the standard setup, i.e. the extreme value condition and the second order condition. Under the second order condition, Drees *et al.* (2004) proved the asymptotic normality in case the extreme value index is higher than -1/2 but they needed to use the assumption that the estimator is not far off its real value.

Although the maximum likelihood estimator of the extreme value index has been widely used in applications, there are still a number of questions:

1) Does the estimator always exist? In particular, does an estimator satisfying the assumptions in Drees *et al.* (2004) really exist? Note that the latter question is important

for completing the proof of the asymptotic normality.
2) Does the statement "the extreme value condition implies consistency" hold for the maximum likelihood estimator?
3) Is it possible to extend the maximum likelihood estimator to a larger set of the extreme value index rather than $(-1/2, +\infty)$ and prove similar asymptotic properties?
4) In practice, is it possible to construct an explicit estimator that approximates closely the maximum likelihood estimator?

In Part I, we answer all the above questions. Chapter 2 is based on Zhou (2008b), which gives positive answers to the first two questions. In this paper, the existence is proved for all extreme value indices higher than -1. It is also proved that the estimator is consistent under only the extreme value condition. Furthermore, when the second order condition is valid, following the proof in this chapter, the existence of a maximum likelihood estimator satisfying the restrictions in Drees *et al.* (2004) is a direct consequence. Hence, our result completes the proof of the asymptotic normality in Drees *et al.* (2004).

The result in Chapter 2 also opens a question about the asymptotic property for the extreme value index lying in between -1 and -1/2 when the second order condition holds. Chapter 3 studies this problem and shows that the asymptotic normality still holds in this region, i.e. the valid interval for the maximum likelihood estimator is now extended to $(-1, +\infty)$. This answers question 3 above. This chapter is based on Zhou (2007).

The last chapter in Part I intends to answer question 4 above. In this chapter, we build a 2-step estimator that is very close to the maximum likelihood estimator. The two have the same asymptotic properties. Since the 2-step estimator can be calculated easily as a function of the observations, it is much simpler to use in practice. This chapter is based on Zhou (2008a).

In all, Part I fills the theoretical gap in the theory of the maximum likelihood estimator for the extreme value index, and provides new statistical methodology for application. It can be seen as the theoretical part of this thesis. The other three parts are devoted to apply EVT in different fields.

Part II contains a single chapter, Chapter 5, which considers an application of multivariate EVT in finance. For financial investment, a well-known adagium is "do not put all your eggs in the same basket". In other words, holding diversified portfolio lowers the portfolio risk. The optimal construction of the portfolio depends on the risk of individual securities as well as their dependence structure. When one considers extremal risks, i.e. big losses, it is reasonable to model the individual security returns by an EVT model and consider Value at Risk (VaR) as the risk criterion. With this setup, the risk of a big loss is usually evaluated by the primary risk indicator–the tail index. However, in reality, the tail indices across different securities in a specific market are quite close to each other, possibly due to arbitrage. Therefore, it is necessary to consider a secondary risk indicator. A natural choice for this is the scale parameter in the EVT model. The risk of a portfolio is then determined by both individual risks and the dependence structure. In this chapter, we study the portfolio selection problem and the diversification effects via the secondary risk indicator in a multivariate EVT framework without assuming any parametric dependence structure. We show that diversification does not always lead to the optimal portfolio. In case diversification has a positive effect, we propose a portfolio selection procedure to construct the optimal portfolio. In case the diversification has a negative effect, we propose to select the individual security with the minimum secondary risk indicator. When the secondary risk indicators are also at the same level across all individual securities, we propose to use a third risk indicator namely the *"probability of dominance"* to be the criterion to select the optimal individual security. Then the optimal individual security is the one that has the minimum connection to the systematic risk. An empirical study illustrates the entire portfolio selection procedure.

Part III considers an application of infinite-dimensional EVT in meteorology. This part consists of two chapters which are based on Buishand *et al.* (2008) and de Haan and Zhou (2008) respectively. In Chapter 6, we focus on the statistical problem of extremely heavy rainfall. We consider daily rainfall observations at 32 stations in the province of North Holland (The Netherlands) during 30 years. Let T be the total rainfall in this area on one day. An important question is: what is the amount of rainfall T that is exceeded once in 100 years? This is clearly a problem belonging to EVT. Also it is a genuinely spatial problem, i.e. extremes on stochastic processes. We use a parametric model in infinite-dimensional EVT to handle the dependence structure. Then it is possible to use simulations to come up with a reasonable answer to the question above. Chapter 6 solves the problem by a combination of resampling and simulation. One of the difficulties in this approach is how to estimate the parameter in the parametric dependence model. The estimation requires technical calculation to connect the marginal distributions of the original process to the dependence parameter. The detail of this calculation is presented in Chapter 7.

Part IV studies an extreme-value-type problem in Internet auctions. Roughly speaking, the main difference between an Internet auction and the classical Dutch or English auction is that we do not have a specific room for auction participants to sit in, i.e. we do not know the real number of participants of a specific Internet auction. This difference is due to the fact that, on Internet, no one could exactly observe all those potential bidders who checked the website with or without placing a bid. However, the strategic theory of classical auctions always requires that the number of participants is known. This missing information creates a big difficulty in the study of Internet auctions. The following two

questions arises:
1) Is it possible to connect the real bidding activities with the number of potential bidders?
2) Is it possible to investigate the valuations of bidders? More precisely, can we estimate the expected final payoff to a specific Internet auction by only observing the bidding history?

The two chapters in Part IV answer these two questions respectively. Chapter 8 is based on de Haan *et al.* (2008b). In this chapter, we first show that under the independent private value paradigm (IPVP) the valuations of the active bidders form a specific record sequence. This fact implies that if the number n of potential bidders is large, the number of active bidders is approximately $2 \log n$ which explains the relative inactivity in Internet auctions. Empirical evidence for the $2 \log n$ rule is provided. Furthermore, this evidence can also be interpreted as a weak test of the IPVP.

Chapter 9 is based on de Haan *et al.* (2008a), which approaches the second question above. In this chapter, we continue with modeling the valuations of the active bidders as a specific record sequence. We study the difference between the observed final payoff, i.e. the price of the final deal and the expected final payoff. We use the EVT model to model the distribution function of the bidder's valuation. In order to have a realistic model, we should at least have a finite expected final payoff. We prove that this requires the assumption that the extreme value index is lower than 2. With this assumption we turn to study whether the observed final payoff precisely estimates the expected final payoff. For non-zero extreme value indicies, the observed final payoff is never a satisfactory estimator for its expectation. However, when the extreme value index equals to 0, we propose a subclass model and its generalized version, for which the observed final payoff consistently estimates the expected final payoff. The chapter concludes that by modeling the distribution function of the bidder's valuation as an EVT model, only for zero extreme value index, the final price of an Internet auction may reflect what the seller deserves, while for positive or negative extreme value index, an underestimation or overestimation case can be expected.

To sum up, besides contributing to the literature of the maximum likelihood estimator of the extreme value index, the thesis is devoted to show different applications of extreme value statistics in different fields: we apply one-dimensional, multi-dimensional and infinite-dimensional Extreme Value Theories to microeconomics, finance and meteorology. From those applications, it is shown that as a natural and precise model for rare events, extreme value statistics exhibits its strong potential in applied statistics.

Part I

Maximum Likelihood

Chapter 2

Existence and Consistency of the Maximum Likelihood Estimator for the Extreme Value Index

2.1 Introduction

Let $X_1, X_2, \cdots$ be independent and identically distributed (i.i.d.) random variables from a distribution function F. Suppose that F is in the domain of attraction of an extreme value distribution, i.e. there exist constants $a_n > 0$ and b_n, such that

$$F^n(a_n x + b_n) \to G_\gamma(x), \quad \text{for all } 1 + \gamma x > 0$$

where $G_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma})$ is the corresponding extreme value distribution function and $\gamma \in \mathbb{R}$ is the extreme value index (Gnedenko (1943)). Commonly, that is denoted by $F \in D(G_\gamma)$.

The necessary and sufficient condition of $F \in D(G_\gamma)$ can be represented in different ways. We state the following criterion, see e.g. de Haan (1984a).

Theorem 2.1.1 *Let* $U := \left(\frac{1}{1-F}\right)^{\leftarrow}$ *be the left-continuous inverse function of* $1/(1-F)$. *Then* $F \in D(G_\gamma)$ *if and only if there exists a function* $a(t) > 0$ *such that*

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \tag{2.1}$$

for all $x > 0$.

The condition (2.1) is called the *extreme value condition*.

Under this setup, a major issue for estimating extremal events is the estimation of the extreme value index γ. For $\gamma > 0$, an estimator, so-called the *Hill estimator*, was suggested by Hill (1975) as follows

$$\hat{\gamma}_H = \frac{1}{k}\sum_{i=0}^{k-1} \log X_{n,n-i} - \log X_{n,n-k},$$

where k is a suitable sequence such that $k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$, and $X_{n,1} \le \cdots \le X_{n,n}$ denote the order statistics of $X_1, X_2, \cdots, X_n$.

There are some other estimators for general $\gamma \in \mathbb{R}$, such as the *Pickands' estimator* suggested by Pickands III (1975), the *moment estimator* suggested by Dekkers *et al.* (1989), and the *UH estimator* suggested by Beirlant *et al.* (1996).

An alternative characterization of the necessary and sufficient condition for a distribution function F belonging to the domain of attraction is via the "excess distribution function" as in Balkema and de Haan (1974). Denote the excess distribution function

$$F_t(x) := P(X - t \le x | X > t) = \frac{F(t+x) - F(t)}{1 - F(t)}.$$

Then $F \in D(G_\gamma)$ is equivalent to

$$\lim_{t \to x^*} F_t(x\sigma(t)) = H_\gamma(x) := 1 - (1 + \gamma x)^{-1/\gamma},$$

for all $1 + \gamma x > 0$, where $\sigma(t)$ is a positive function and x^* is the right endpoint of F, i.e. $x^* = \sup\{x | F(x) < 1\}$. H_γ is the so-called *generalized Pareto distribution (GPD) function*. Intuitively, the distribution function F is in the domain of attraction if and only if the excesses above a high threshold are asymptotically generalized Pareto distributed.

Smith (1987) introduced a maximum likelihood estimator (MLE) of the extreme value index by fitting the GPD with the empirical excesses. The maximum likelihood estimators for the extreme value index and the scale, $\hat{\gamma}_{ML}$ and $\hat{\sigma}_{ML}$, are obtained by solving the likelihood equations. The likelihood equations are (c.f. Drees *et al.* (2004))

$$\begin{aligned} \sum_{i=1}^{k} \frac{1}{\gamma^2} \log\Big(1 + \frac{\gamma}{\sigma}(X_{n,n-i+1} - X_{n,n-k})\Big) & \\ - \left(\frac{1}{\gamma} + 1\right) \frac{(1/\sigma)(X_{n,n-i+1} - X_{n,n-k})}{1 + (\gamma/\sigma)(X_{n,n-i+1} - X_{n,n-k})} &= 0 \\ \sum_{i=1}^{k} \left(\frac{1}{\gamma} + 1\right) \frac{(\gamma/\sigma)(X_{n,n-i+1} - X_{n,n-k})}{1 + (\gamma/\sigma)(X_{n,n-i+1} - X_{n,n-k})} &= k, \end{aligned} \tag{2.2}$$

(the equations for $\gamma = 0$ are defined by continuity). Excluding $\gamma = 0$ as a solution, (2.2) can be simplified as

$$\begin{aligned} \frac{1}{k} \sum_{i=1}^{k} \log\Big(1 + \frac{\gamma}{\sigma}(X_{n,n-i+1} - X_{n,n-k})\Big) &= \gamma \\ \frac{1}{k} \sum_{i=1}^{k} \frac{1}{1 + (\gamma/\sigma)(X_{n,n-i+1} - X_{n,n-k})} &= \frac{1}{\gamma + 1}. \end{aligned} \tag{2.3}$$

The equations are based on excesses $Y_i := X_{n,n-i+1} - X_{n,n-k}$, where $i = 1, \cdots, k$ and k is a suitable sequence of integers as in the Hill estimator.

Grimshaw (1993) discussed a numerical way to solve the likelihood equations as follows. From the equations (2.3), with the notation Y_i, it is derived that,

$$\left(\frac{1}{k}\sum_{i=1}^{k}\log\left(1+\frac{\gamma}{\sigma}Y_i\right)+1\right)\cdot\frac{1}{k}\sum_{i=1}^{k}\frac{1}{1+(\gamma/\sigma)Y_i}=1. \tag{2.4}$$

In order to write this in short hand, denote the two parts in (2.4) as functions

$$\begin{aligned} f_n(t) &:= \frac{1}{k}\sum_{i=1}^{k}\log(1+tY_i)+1, \\ g_n(t) &:= \frac{1}{k}\sum_{i=1}^{k}\frac{1}{1+tY_i}, \\ h_n(t) &:= f_n(t)g_n(t)-1. \end{aligned}$$

Then, it is clear that any root $(\hat{\gamma},\hat{\sigma})$ of (2.3) satisfies $h_n(\hat{\gamma}/\hat{\sigma})=0$. Conversely, if t^* is a non-zero root of $h_n(t)=0$, we obtain $(\hat{\gamma},\hat{\sigma})=(f_n(t^*)-1,(f_n(t^*)-1)/t^*)$ as the solution of (2.3). With this idea, the maximum likelihood estimator can be calculated in the following procedure:

1. find the root t_n^* of $h_n(t)=0$;
2. $\hat{\gamma}_{ML}=f_n(t_n^*)-1$;
3. $\hat{\sigma}_{ML}=\hat{\gamma}_{ML}/t_n^*$.

The first step was solved in a numerical way in Grimshaw (1993). After that, the maximum likelihood estimators of γ and σ were calculated based on the numerical root of $h_n(t)=0$.

Note that (2.3) is the simplified version of (2.2) by assuming $\gamma\neq 0$, although $t_n^*=0$ is always a solution of $h_n(t)=0$, we should not take it as the proper solution. Hence, in solving (2.3) we disregard the solution $\hat{\gamma}_{ML}=0$, even if in reality $\gamma=0$.

The asymptotic properties of the mentioned estimators, including the maximum likelihood estimator, have been discussed in the literature. For all of the above estimators except the maximum likelihood estimator, it is proved that, they are consistent under the extreme value condition (2.1). In order to get the asymptotic normality, de Haan and Stadtmüller (1996) introduced the *second order condition* as

$$\lim_{t\to\infty}\frac{\frac{U(tx)-U(t)}{a(t)}-\frac{x^\gamma-1}{\gamma}}{A(t)}=H(x), \tag{2.5}$$

for all $x>0$, where $H(x)$ is assumed not to be a multiple of $\frac{x^\gamma-1}{\gamma}$, and $A(t)\to 0$ as $t\to\infty$. With the second order condition, the asymptotic normality of all the above estimators has been proved. Compared to this condition, we call the extreme value condition (2.1), the *first order condition*.

In general the following two statements hold for most of the extreme value index estimators in literature.
1) The first order condition implies the (weak) convergence;
2) The second order condition implies the asymptotic normality.

Since most of the estimators have an explicit formula, the proof of the two statements are given by the expansion of the explicit estimator. The first statement is always proved by the first order expansion. To obtain the second statement, further expansion based on the second order condition is normally required.

For the maximum likelihood estimator, because it is only given by solving the likelihood equations instead of an explicit formula, its asymptotic properties have to be proved in a different way. In case $\gamma > -1/2$, Smith (1987) sketched the proof of the consistency and asymptotic normality assuming a few extra conditions. A different proof of the second statement i.e. the second order condition implies the asymptotic normality is provided in Drees *et al.* (2004) by assuming that there exists a solution of the likelihood equations not too far off the real value. In Proposition 3.1 of this paper, it is stated that *"Any solution* $(\tilde{\gamma}, \tilde{\sigma})$ *of (20) satisfying (21) and* $\log \tilde{\sigma} = O_P(1)$ *admits the approximation"*. Here two more requirements on the solution are assumed. These two requirements can be equivalently presented within the notations in the current chapter as follows: the asymptotic normality result holds for any solution $(\hat{\gamma}, \hat{\sigma})$ satisfying

$$\left| \frac{\hat{\gamma}}{\hat{\sigma}/a(n/k)} - \gamma \right| = O_p(k^{-1/2}) \quad \text{and} \quad \log \frac{\hat{\sigma}}{a(n/k)} = O_p(1).$$

Although at the end of the Proposition 3.1 in Drees *et al.* (2004), it is stated that *"Conversely, there exists a solution of (20) which satisfies (28), respectively (29), and hence also (21)."* Along the lines of the proof, we can not find a clear evidence to support this conclusion.

Therefore, the existence of the solution of the likelihood equations is still an open question. And the existence of a solution satisfying the restrictions in Drees *et al.* (2004) is particularly important to make the proof of the asymptotic normality complete. Meanwhile, it is still unclear whether the first statement holds for the maximum likelihood estimator, i.e. whether the first order condition implies the consistency.

In this chapter, under only the first order condition, we are going to give positive answers for the existence and consistency by proving that the following statement holds almost surely for $\gamma > -1$: the likelihood equations are eventually solvable, and a suitable solution sequence converges to the real γ almost surely, when the sample size goes to infinity.

The proof of the existence of the maximum likelihood estimator is new. With second order condition, the same proof leads to stronger asymptotic properties of the solution which fulfill the restrictions in Drees *et al.* (2004).

In Section 2.2, the main theorem will be given. The proof for the positive case is given in Section 2.3. Since the idea of the proofs for the negative and zero cases are essentially the same as that of the positive case, only with more detailed calculation, the proofs are postponed to Appendix 2.A. In Section 2.4, the result under second order condition is discussed. Again, only the proof for the positive case is given, while the proofs for the negative and zero case are postponed to Appendix 2.B. Section 2.5 concludes this chapter.

2.2 Main theorem

Similar to Grimshaw's numerical way, we use the simplified version of the likelihood equations, (2.3). The main results on the existence and consistency are given as the following theorems.

Theorem 2.2.1 *Suppose the first order condition (2.1) holds for the extreme value index* $\gamma > -1$ *and* $\gamma \neq 0$*. If the sequence* $k = k(n)$ *satisfies* $k(n) \to \infty$*,* $k(n)/n \to 0$*, and* $k(n)/\log n \to \infty$*, then*

$$P(\{\textit{The MLE does not exist for infinitely many } n\}) = 0.$$

Or, equivalently,

$$P(\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\{\textit{The MLE exists for sample size } n\}) = 1.$$

On this probability 1 set, there exists a random integer N*, such that for any sample size* $n > N$*, there exists a suitable solution of the likelihood equations,* $(\hat{\gamma}_n, \hat{\sigma}_n)$*, and this solution satisfies*

$$\hat{\gamma}_n \xrightarrow{a.s.} \gamma$$

and

$$\frac{\hat{\sigma}_n}{a(n/k)} \xrightarrow{a.s.} 1$$

as $n \to \infty$*, where* $a(t)$ *is the auxiliary function in (2.1)*

If we skip the extra condition $k(n)/\log n \to \infty$, but only keep $k(n) \to \infty$ and $k(n)/n \to 0$, Theorem 2.2.1 still holds in the sense of "in probability" instead of "almost surely" as in Theorem 2.2.2.

Theorem 2.2.2 *Suppose the first order condition (2.1) holds for the extreme value index* $\gamma > -1$ *and* $\gamma \neq 0$*. If the sequence* $k = k(n)$ *satisfies* $k(n) \to \infty$*,* $k(n)/n \to 0$*, then*

$$P(\{\textit{The MLE exists for infinitely many } n\}) = 1.$$

On this probability 1 set, the likelihood equations is solvable for infinitely many sample sizes $\{n_i \in \mathbb{N} : 1 \le n_1 < n_2 < \cdots\}_{i=1}^{\infty}$. *There exist suitable solutions* $(\hat{\gamma}_{n_i}, \hat{\sigma}_{n_i})$ *for each related sample size, which satisfies*

$$\hat{\gamma}_{n_i} \xrightarrow{P} \gamma \quad \textit{and} \quad \frac{\hat{\sigma}_{n_i}}{a(n_i/k(n_i))} \xrightarrow{P} 1$$

as $i \to \infty$.

The proof of this weak form is similar to the proof of Theorem 2.2.1. The extra condition on k in Theorem 2.2.1 is only used to obtain the almost sure convergence. Actually, a similar condition is also required to get almost sure convergence for other estimators. For instance, for the Hill estimator, see Deheuvels *et al.* (1988). To ensure the convergence in probability, this condition can be skipped, and the proof is simpler.

Because our purpose is to obtain the consistent solution of the likelihood equations from a certain sample size onwards, we insist on proving the strong form.

As discussed before, the zero solution of the likelihood equations is always disregarded. Therefore, the theorem under the case $\gamma = 0$ is different because we should prove the existence of a non-zero solution, and it converges to zero almost surely when the sample size goes to infinity. The theorem is stated as follows.

Theorem 2.2.3 *Suppose the first order condition (2.1) holds for the extreme value index* $\gamma = 0$. *Suppose that with probability 1, the following relation does not hold for sufficiently large* n,

$$\frac{1}{2k}\sum_{i=1}^{k}(X_{n,n-i+1} - X_{n,n-k})^2 = \left(\frac{1}{k}\sum_{i=1}^{k}(X_{n,n-i+1} - X_{n,n-k})\right)^2 \tag{2.6}$$

If the sequence $k = k(n)$ *satisfies* $k(n) \to \infty$, $k(n)/n \to 0$, *and* $k(n)/(\log n)^c \to \infty$ *for some* $c > 1$, *then*

$P(\{$*A non-zero solution of the likelihood equations does not exist for infinitely many* $n\}) = 0$.

Or, equivalently,

$P(\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\{$*There exists a non-zero solution of the likelihood equations for sample size* $n\}) = 1$.

On this probability 1 set, there exists a random integer N, *such that for any sample size* $n > N$, *there exists a suitable solution of the likelihood equations,* $(\hat{\gamma}_n, \hat{\sigma}_n)$, *and this solution satisfies*

$$\hat{\gamma}_n \xrightarrow{a.s.} 0$$

as $n \to \infty$.

Remark 2.2.1 *The extra condition in Theorem 2.2.3 that (2.6) does not hold almost surely ensures that zero is not the proper solution of the likelihood equations. This is the same condition as in Drees* et al. *(2004), Remark 2.2. The condistion is not very restrictive, for example, it holds when F possesses a density.*

Remark 2.2.2 *The condition on the sequence k is stronger than in Theorem 2.2.1. However, it is still relatively weak.*

Combining Theorem 2.2.1 and Theorem 2.2.3, the existence and strong convergence of the maximum likelihood estimator has been proved for $\gamma > -1$. This extends the scope of the asymptotic properties as stated in the following remark.

Remark 2.2.3 *The asymptotic normality of the maximum likelihood estimator is proved only for $\gamma > -1/2$. So, up to now, the maximum likelihood estimator is only used for $\gamma > -1/2$. Our result extends the scope of γ to $\gamma > -1$ in the sense of consistency. It means that, although we have not obtained enough information on the asymptotic distribution of the maximum likelihood estimator, when $-1 < \gamma \leq -1/2$, we can still use it in the sense of strong convergence.*

2.3 Proof

With the notation $U := \left(\frac{1}{1-F}\right)^{\leftarrow}$, the i.i.d random variables can be rewritten as $\{X_n\}_{n=1}^{\infty} \stackrel{d}{=} \{U(Z_n)\}_{n=1}^{\infty}$, where $\{Z_n\}_{n=1}^{\infty}$ are i.i.d. random variables with distribution function $1-1/x$, $x \geq 1$. Some useful lemmas about the sequence $\{Z_n\}_{n=1}^{\infty}$ will be proved in Subsection 2.3.1. After that, we prove Theorem 2.2.1 for $\gamma > 0$ in Subsection 2.3.2. The proofs for $-1 < \gamma < 0$ and Theorem 2.2.3 ($\gamma = 0$) are in Appendix 2.A.

2.3.1 Lemmas

Let $Z_1, Z_2, \cdots$ be i.i.d. random variables with distribution function $1 - 1/x$, $x \geq 1$ and let $Z_{n,1} \leq Z_{n,2} \leq \cdots \leq Z_{n,n}$ be the order statistics.

Lemma 2.3.1 *Let $\phi : [1, +\infty) \to \mathbb{R}$ be such that $\int_0^1 \phi\left(\frac{1}{s}\right) ds < \infty$. Suppose $\phi(1/s)$ is uniformly continuous on $(0,1]$. Then for a sequence $k = k(n) \to \infty$, $k/n \to 0$ and $k/\log n \to \infty$, we have*

$$\frac{1}{k}\sum_{i=0}^{k-1} \phi\left(\frac{Z_{n,n-i}}{Z_{n,n-k}}\right) \xrightarrow{a.s.} \int_0^1 \phi\left(\frac{1}{s}\right) ds,$$

as $n \to \infty$.

Proof of Lemma 2.3.1

It is clear that $1/Z_i$ is a random variable with uniform distribution on $(0,1]$. So $1/Z_{n,1} \geq 1/Z_{n,2} \geq \cdots \geq 1/Z_{n,n}$ are the order statistics of an i.i.d. uniform sample. By applying Theorem 3(III) in Einmahl and Mason (1988), when the sequence $k(n)$ satisfies the above conditions,

$$\sup_{0<s\leq 1} \frac{\left|\frac{n}{Z_{n,n-\lceil sk\rceil+1}} - sk\right|}{(2k\log\log n)^{1/2}} \leq M \quad a.s. \tag{2.7}$$

where $M > 1$ is a fixed number. This implies that,

$$\frac{n}{kZ_{n,n-\lceil sk\rceil+1}} \to s \quad a.s. \tag{2.8}$$

holds uniformly for $s \in (0,1]$, where $\lceil t\rceil$ is the smallest integer which is greater or equal to t. By taking $s = 1$ in (2.8), it becomes

$$\frac{n}{kZ_{n,n-k+1}} \to 1 \quad a.s.$$

Then, replacing k with $k+1$, we get that

$$\frac{n}{kZ_{n,n-k}} \to 1 \quad a.s. \tag{2.9}$$

Hence,

$$\frac{Z_{n,n-k}}{Z_{n,n-\lceil sk\rceil+1}} \to s \quad a.s. \tag{2.10}$$

holds uniformly for $s \in (0,1]$. If $\phi(1/s)$ is uniformly continuous on $(0,1]$, then

$$\phi\left(\frac{Z_{n,n-\lceil sk\rceil+1}}{Z_{n,n-k}}\right) \to \phi\left(\frac{1}{s}\right) \quad a.s.$$

holds uniformly on $s \in (0,1]$. It leads to

$$\int_0^1 \phi\left(\frac{Z_{n,n-\lceil sk\rceil+1}}{Z_{n,n-k}}\right) ds \to \int_0^1 \phi\left(\frac{1}{s}\right) ds \quad a.s.$$

which completes the proof of Lemma 2.3.1. □

Lemma 2.3.2 *Suppose a positive function V is regularly varying at infinity with index $\gamma \neq 0$, i.e.*

$$\lim_{t\to\infty} \frac{V(tx)}{V(t)} = x^\gamma,$$

and $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ is a monotone function, such that

$$\lim_{\varepsilon\to 0} \int_0^1 \psi((1+\varepsilon)t^{-\gamma-\varepsilon})dt = \int_0^1 \psi(t^{-\gamma})dt < \infty. \tag{2.11}$$

Suppose there exists a positive number E, such that for all $|\varepsilon| < E$, $\psi((1+\varepsilon)t^{-\gamma-\varepsilon})$ is uniformly continuous on $(0,1]$. Then, for a sequence $k = k(n) \to \infty$, $k/n \to 0$ and $k/\log n \to \infty$, we have that

$$\frac{1}{k}\sum_{i=0}^{k-1}\psi\left(\frac{V(Z_{n,n-i})}{V(Z_{n,n-k})}\right) \xrightarrow{a.s.} \int_0^1 \psi(t^{-\gamma})dt,$$

as $n \to \infty$.

Proof of Lemma 2.3.2

From Proposition 1.7(5) in Geluk and de Haan (1987), we have the following inequality. Given any $\varepsilon > 0$, there exists t_0, such that, for $t \geq t_0$ and $x \geq 1$

$$(1-\varepsilon)x^{\gamma-\varepsilon} < \frac{V(tx)}{V(t)} < (1+\varepsilon)x^{\gamma+\varepsilon}.$$

From the proof of Lemma 2.3.1, with the same sequence $k(n)$, we have that $Z_{n,n-k} \to \infty$ almost surely as $n \to \infty$. Hence, eventually,

$$(1-\varepsilon)\left(\frac{Z_{n,n-i}}{Z_{n,n-k}}\right)^{\gamma-\varepsilon} < \frac{V(Z_{n,n-i})}{V(Z_{n,n-k})} < (1+\varepsilon)\left(\frac{Z_{n,n-i}}{Z_{n,n-k}}\right)^{\gamma+\varepsilon}.$$

Because ψ is a monotone function, so $\frac{1}{k}\sum_{i=0}^{k-1}\psi\left(\frac{V(Z_{n,n-i})}{V(Z_{n,n-k})}\right)$ is between

$$\frac{1}{k}\sum_{i=0}^{k-1}\psi\left((1-\varepsilon)\left(\frac{Z_{n,n-i}}{Z_{n,n-k}}\right)^{\gamma-\varepsilon}\right) \quad \text{and} \quad \frac{1}{k}\sum_{i=0}^{k-1}\psi\left((1+\varepsilon)\left(\frac{Z_{n,n-i}}{Z_{n,n-k}}\right)^{\gamma+\varepsilon}\right).$$

Note that, according to the condition (2.11)

$$\int_0^1 \psi((1+\varepsilon)t^{-\gamma-\varepsilon})dt < \infty,$$

when $|\varepsilon|$ is small enough. By applying Lemma 2.3.1, as $n \to \infty$,

$$\frac{1}{k}\sum_{i=0}^{k-1}\psi\left((1+\varepsilon)\left(\frac{Z_{n,n-i}}{Z_{n,n-k}}\right)^{\gamma+\varepsilon}\right) \xrightarrow{a.s.} \int_0^1 \psi((1+\varepsilon)t^{-\gamma-\varepsilon})dt$$

for small $|\varepsilon| < E$. Considering the condition (2.11), when $\varepsilon \to 0$, both of the two boundaries convergence to $\int_0^1 \psi(t^{-\gamma})dt$, which completes the proof of this lemma. □

2.3.2 Positive case: proof for $\gamma > 0$

When $\gamma > 0$, the auxiliary function $a(t)$ in (2.1) satisfies $a(t) \sim \gamma U(t)$. Therefore, an approximate solution of $h_n(t) = 0$ is $t^{(0)} := \gamma/(\gamma U(Z_{n,n-k})) = 1/U(Z_{n,n-k})$. Note that $f_n(t^{(0)}) - 1$ is the Hill estimator, and

$$g_n(t^{(0)}) = \frac{1}{k}\sum_{i=1}^{k}\frac{U(Z_{n,n-k})}{U(Z_{n,n-i+1})}.$$

To prove the existence, we disturb the approximate solution by a small increment as $t^{(\delta)} = \frac{1+\delta}{U(Z_{n,n-k})}$ for $|\delta| < \frac{1}{2}$. We are going to find a sequence $\delta_n > 0$ such that, for sufficiently large n, $h_n(t^{(-\delta_n)})$ and $h_n(t^{(\delta_n)})$ have different signs. This ensures that there exists a root of $h_n(t) = 0$ between $t^{(-\delta_n)}$ and $t^{(\delta_n)}$.

The following lemma studies the asymptotic behavior of f_n and g_n at $t = t^{(0)}$.

Lemma 2.3.3 *Suppose (2.1) holds for $\gamma > 0$, the sequence k satisfies $k(n)/n \to 0$, and $k(n)/\log n \to \infty$ as $n \to \infty$, we have that*

$$f_n(t^{(0)}) \xrightarrow{a.s.} \gamma + 1, \tag{2.12}$$

$$g_n(t^{(0)}) \xrightarrow{a.s.} \frac{1}{\gamma + 1}, \tag{2.13}$$

as $n \to \infty$. Furthermore, the statistic

$$\tilde{g}_n := \frac{1}{k} \sum_{i=1}^{k} \left(\frac{U(Z_{n,n-k})}{U(Z_{n,n-i+1})} \right)^2,$$

satisfies that for $n \to \infty$,

$$\tilde{g}_n \xrightarrow{a.s.} \frac{1}{2\gamma + 1} \tag{2.14}$$

Proof of Lemma 2.3.3

When $\gamma > 0$, $f_n(t^{(0)}) - 1$ is the Hill estimator. According to the almost sure convergence of the Hill estimator in Deheuvels *et al.* (1988), (2.12) holds.

For the relation (2.13), note that (2.1) implies that U is regularly varying at infinity with index γ. By checking that $\Psi(x) = 1/x$ satisfies condition (2.11) and $t^{\gamma+\varepsilon}/(1+\varepsilon)$ is uniformly continuous on $[0, 1]$ for all $|\varepsilon| < \gamma$, we can apply Lemma 2.3.2 to obtain (2.13). The proof of (2.14) is similar. □

Now we turn to find a suitable δ_n that serves our purpose. Given $\delta > 0$, we first calculate the upper bound of $f_n(t^{(\delta_n)})$ and $g_n(t^{(\delta_n)})$ for any $0 < \delta_n < \delta$ as

$$\begin{aligned} f_n(t^{(\delta_n)}) - f_n(t^{(0)}) &= \frac{1}{k} \sum_{i=1}^{k} \log \left(1 + \frac{\delta_n (U(Z_{n,n-i+1})/U(Z_{n,n-k}) - 1)}{U(Z_{n,n-i+1})/U(Z_{n,n-k})} \right) \\ &\le \frac{1}{k} \sum_{i=1}^{k} \delta_n \left(1 - \frac{U(Z_{n,n-k})}{U(Z_{n,n-i+1})} \right) \\ &= \delta_n (1 - g_n(t^{(0)})) \end{aligned} \tag{2.15}$$

and

$$g_n(t^{(\delta_n)}) - g_n(t^{(0)}) = \frac{1}{k} \sum_{i=1}^{k} \frac{-\delta_n (U(Z_{n,n-i+1})/U(Z_{n,n-k}) - 1)}{\frac{U(Z_{n,n-i+1})}{U(Z_{n,n-k})} \left(\frac{U(Z_{n,n-i+1})}{U(Z_{n,n-k})} (1 + \delta_n) - \delta_n \right)}$$

$$\leq \frac{1}{k}\sum_{i=1}^{k}\frac{-\delta_n}{1+\delta}\left(\frac{U(Z_{n,n-k})}{U(Z_{n,n-i+1})}-\left(\frac{U(Z_{n,n-k})}{U(Z_{n,n-i+1})}\right)^2\right)$$
$$= -\frac{\delta_n}{1+\delta}(g_n(t^{(0)})-\tilde{g}_n).$$

Hence,

$$\begin{aligned}h_n(t^{(\delta_n)}) &= f_n(t^{(\delta_n)})g_n(t^{(\delta_n)})-1\\ &< f_n(t^{(0)})g_n(t^{(0)})-1+\delta_n\left(g_n(t^{(0)})(1-g_n(t^{(0)}))-f_n(t^{(0)})\frac{g_n(t^{(0)})-\tilde{g}_n}{1+\delta}\right)\\ &:= f_n(t^{(0)})g_n(t^{(0)})-1+\delta_n A_n.\end{aligned}$$

Since the lower bound of $f_n(t^{(-\delta_n)})$ and $g_n(t^{(-\delta_n)})$ for $0<\delta_n<\delta$ follows similar calculation, we only present the result as follows

$$f_n(t^{(-\delta_n)})-f_n(t^{(0)}) \geq \frac{\log(1-\delta)}{\delta}\delta_n(1-g_n(t^{(0)})) \tag{2.16}$$

and

$$g_n(t^{(-\delta_n)})-g_n(t^{(0)}) \geq \delta_n(g_n(t^{(0)})-\tilde{g}_n).$$

Hence,

$$\begin{aligned}h_n(t^{(-\delta_n)}) &> f_n(t^{(0)})g_n(t^{(0)})-1+\delta_n\left(\frac{\log(1-\delta)}{\delta}g_n(t^{(0)})(1-g_n(t^{(0)}))+f_n(t^{(0)})(g_n(t^{(0)})-\tilde{g}_n)\right)\\ &+\delta_n^2\frac{\log(1-\delta)}{\delta}(1-g_n(t^{(0)}))(g_n(t^{(0)})-\tilde{g}_n)\\ &:= f_n(t^{(0)})g_n(t^{(0)})-1+\delta_n B_n.\end{aligned}$$

We are going to choose suitable δ_n, such that $\delta_n \to 0$ almost surely as $n\to\infty$. Then, by Lemma 2.3.3, as $n\to\infty$, we have that

$$\begin{aligned}A_n &\xrightarrow{a.s.} A(\delta) := \tfrac{\gamma}{(1+\gamma)^2}-\tfrac{\gamma}{(2\gamma+1)(1+\delta)}\\ B_n &\xrightarrow{a.s.} B(\delta) := \tfrac{\log(1-\delta)}{\delta}\tfrac{\gamma}{(1+\gamma)^2}+\tfrac{\gamma}{2\gamma+1}.\end{aligned}$$

By taking $\delta\to 0$, we get that

$$A(\delta) \to -\frac{\gamma^3}{(1+\gamma)^2(2\gamma+1)} < 0$$

and

$$B(\delta) \to \frac{\gamma^3}{(1+\gamma)^2(2\gamma+1)} > 0.$$

Hence, we use the following procedure to choose suitable δ_n.
1) Choose a suitable $\delta>0$, such that $A(\delta)<0$ and $B(\delta)>0$.
2) As $n\to\infty$, we have eventually, $A_n<0$ and $B_n>0$. Denote

$$\delta_n := \left|f_n(t^{(0)})g_n(t^{(0)})-1\right|\left(\left(-\frac{1}{A_n}\right)\bigvee\frac{1}{B_n}\right) > 0.$$

From Lemma 2.3.3, it is clear that $\delta_n \xrightarrow{a.s.} 0$ as $n \to \infty$. Hence, eventually, $\delta_n < \delta$. By employing the inequalities on the upper bound of $h_n(t^{(\delta_n)})$ and the lower bound of $h_n(t^{(-\delta_n)})$, we have that

$$h_n(t^{(\delta_n)}) < f_n(t^{(0)})g_n(t^{(0)}) - 1 + \delta_n A_n \le f_n(t^{(0)})g_n(t^{(0)}) - 1 - \left|f_n(t^{(0)})g_n(t^{(0)}) - 1\right| \le 0$$

and

$$h_n(t^{(-\delta_n)}) > f_n(t^{(0)})g_n(t^{(0)}) - 1 + \delta_n B_n \ge f_n(t^{(0)})g_n(t^{(0)}) - 1 + \left|f_n(t^{(0)})g_n(t^{(0)}) - 1\right| \ge 0.$$

Hence, we proved that there exists a root of $h_n(t) = 0$ on the interval $[t^{(-\delta_n)}, t^{(\delta_n)}]$.

Now we are going to prove the consistency. Denote the root as t_n^*. Because f_n is an increasing function,

$$f_n(t^{(-\delta_n)}) \le f_n(t_n^*) \le f_n(t^{(\delta_n)}).$$

From (2.15) and (2.16), we have that

$$f_n(t^{(0)}) + \frac{\log(1-\delta)}{\delta}\delta_n(1 - g_n(t^{(0)})) \le f_n(t_n^*) \le f_n(t^{(0)}) + \delta_n(1 - g_n(t^{(0)})).$$

By relation (2.12) and considering the fact that $\delta_n \to 0$ almost surely as $n \to \infty$, we get that $f_n(t_n^*) \xrightarrow{a.s.} \gamma + 1$. Hence $\hat{\gamma}_n^* = f_n(t_n^*) - 1 \xrightarrow{a.s.} \gamma$, i.e. $\hat{\gamma}_n^*$ is strongly consistent.

To prove the almost sure convergence of $\hat{\sigma}_n^*$, we use the fact that, as $t \to \infty$, $a(t) \sim \gamma U(t)$. So, as $n \to \infty$,

$$\frac{a(n/k)}{\gamma X_{n,n-k}} \sim \frac{U(n/k)}{U(Z_{n,n-k})}.$$

Since U is regularly varying at infinity with index $\gamma > 0$, from (2.9) we get that

$$\lim_{n\to\infty} \frac{a(n/k)}{\gamma X_{n,n-k}} = 1 \quad a.s.$$

Together with the fact that $\delta_n \to 0$ almost surely as $n \to \infty$, it implies that

$$\lim_{n\to\infty} \frac{a(n/k)t^{(\pm\delta_n)}}{\gamma} = 1 \quad a.s.$$

Since t_n^* lies on the interval $[t^{(-\delta_n)}, t^{(\delta_n)}]$, we get that

$$\lim_{n\to\infty} \frac{a(n/k)t_n^*}{\gamma} = 1 \quad a.s.$$

Combining this with the almost sure convergence of $\hat{\gamma}_n^*$, the consistency of $\hat{\sigma}_n^* = \frac{\hat{\gamma}_n^*}{t_n^*}$ is proved for γ positive.

Furthermore, an interesting remark for the positive case finishes this subsection.

Remark 2.3.1 *The proof of the positive case suggested that the root of $h_n(t) = 0$ lies on the interval $[t^{(-\delta_n)}, t^{(\delta_n)}]$ and $t^{(0)}$ can be an approximate root of $h_n(t) = 0$. This argument leads to two observations. Practically, we can search the root of $h_n(t) = 0$ only on the interval $[t^{(-\delta_n)}, t^{(\delta_n)}]$ when applying Grimshaw's numerical method. Theoretically, based on this approximate root, we get an approximate extreme value index estimator, $f_n(t_n^{(0)}) - 1$, which is exactly the Hill estimator.*

2.4 Extension under the second order condition

As discussed in Section 2.1, in order to complete the proof of asymptotic normality of the maximum likelihood estimator for the extreme value index in Drees *et al.* (2004), it is necessary to prove that under the second order condition there exists a solution $(\hat{\gamma}, \hat{\sigma})$ such that

$$\left|\frac{\hat{\gamma}}{\hat{\sigma}/a(n/k)} - \gamma\right| = O_p(k^{-1/2}) \quad \text{and} \quad \log \frac{\hat{\sigma}}{a(n/k)} = O_p(1).$$

Notice that the consistency of the scale estimator ensures the second required relation even under the first order condition. Hence, only the existence of a solution satisfying the first relation should be proved. According to Grimshaw's numerical method, the solution $(\hat{\gamma}, \hat{\sigma})$ is derived from the root t_n^* of $h_n(t) = 0$, by the relations $\hat{\gamma} = f_n(t_n^*) - 1$ and $\hat{\sigma} = \hat{\gamma}/t_n^*$. Using this notation, the first required relation is simplified as

$$|t^* a(n/k) - \gamma| = O_p(k^{-1/2}). \tag{2.17}$$

We shall prove the following proposition.

Proposition 2.4.1 *Under the second order condition with $\gamma > -1/2$, there exists a root t_n^* of $h_n(t) = 0$ satisfying (2.17).*

Here we present the proof only for $\gamma > 0$. For $-1/2 < \gamma \leq 0$, the proof is postponed to Appendix 2.B.

Proof of Proposition 2.4.1

When $\gamma > 0$, from the proof in Subsection 2.3.2, we get that when n is sufficiently large, there exists a root t_n^* of $h_n(t) = 0$ lying between $t^{(-\delta_n)}$ and $t^{(\delta_n)}$, where $t^{(\delta)}$ is defined as

$$t^{(\delta)} = \frac{1 + \delta}{U(Z_{n,n-k})},$$

and

$$\delta_n = \left|f_n(t^{(0)}) g_n(t^{(0)}) - 1\right| \left(\left(-\frac{1}{A_n}\right) \bigvee \frac{1}{B_n}\right)$$

is a positive sequence that goes to 0 almost surely as $n \to \infty$. Notice that in the positive case one may take $a(t) = \gamma U(t)$. Hence,

$$t^{(\delta)} a(n/k) = \gamma(1+\delta)\frac{U(n/k)}{U(Z_{n,n-k})}.$$

With the second order condition, Lemma 2.3.3 can be extended to get the speed of convergence, i.e. to prove that $\sqrt{k}(f_n(t^{(0)}) - (\gamma+1))$ and $\sqrt{k}(g_n(t^{(0)}) - 1/(\gamma+1))$ are both asymptotically normally distributed as $n \to \infty$. The asymptotic normality of f_n is in fact the asymptotic normality of Hill estimator. Since both the sequences A_n and B_n converge to non-zero constant as $n \to \infty$, it is proved that $\sqrt{k}\delta_n = O_p(1)$.

Under the second order condition, $\sqrt{k}\left(\frac{U(n/k)}{U(Z_{n,n-k})} - 1\right)$ is asymptotically normally distributed, see, e.g. de Haan and Ferreira (2006) Theorem 2.4.1. Therefore, combining the above two asymptotic relations, we get that $\sqrt{k}(t^{(\delta_n)}a(n/k) - \gamma) = O_p(1)$.

A similar relation holds for $t^{(-\delta_n)}$. Because t_n^* lies in between $t^{(-\delta_n)}$ and $t^{(\delta_n)}$, it is proved that

$$\sqrt{k}(t_n^* a(n/k) - \gamma) = O_p(1),$$

which is equivalent to relation (2.17). Hence, under the second order condition, there exists a suitable root t^* verifying (2.17). □

From Proposition 2.4.1, we get that the extra conditions in the proof of asymptotic normality in Drees *et al.* (2004) are in fact fulfilled.

2.5 Conclusion

This chapter studies the existence and consistency of the maximum likelihood estimator of the extreme value index. Under only the first order condition, it is proved that for $\gamma > -1$, as $n \to \infty$, a solution of the likelihood equations eventually exists. The estimators are consistent in the sense that $\hat{\gamma} \to \gamma$ and $\hat{\sigma}/a(n/k) \to 1$ almost surely as $n \to \infty$.

The asymptotic normality under second order condition has been proved by Drees *et al.* (2004) for $\gamma > -1/2$. The consistency result in this chapter illustrates that the maximum likelihood estimator can also be applied for $-1 < \gamma \leq -1/2$.

From the proof, it is suggested that the solution lies in a specific interval in each case ($\gamma > 0$, $-1 < \gamma < 0$ and $\gamma = 0$). Those intervals can be used in numerically solving the likelihood equations.

The paper Drees *et al.* (2004) on the asymptotic normality starts from the assumption that a sequence of solutions exist and converge to the real value with a certain speed of convergence. The proofs in this chapter can be extended to show that such a sequence of solutions does exist. Hence the two studies together offer exactly the asymptotic normality result that is needed for applications.

2.A Appendix A

Proof of existence and consistency when $-1 < \gamma < 0$
When $-1 < \gamma < 0$, (2.1) implies that $U(\infty) < \infty$, and the function $U(\infty) - U(x)$ is regularly varying at infinity with index γ. Similar to the positive case, now define the sequence $t^{(\delta)}$ as $t^{(\delta)} = -\frac{1+\delta}{U(\infty)-X_{n,n-k}}$. But for this case, it can only be defined for $\delta \in (-1/2, 0)$, because for $\delta > 0$, it is not ensured that $1 + t^{(\delta)}Y_i > 0$ for $i = 1, 2, \cdots, k$. So the trick in the positive case can only be used for one side. For the other side, we are going to introduce another way to build up inequality.

Compared to the positive case, we do not find a sequence $\delta_n \to 0$ such that $h_n(t^{(\delta_n)}) < 0$. Instead, we first prove that for some fixed $\delta < 0$, when n is sufficiently large $h_n(t^{(\delta)}) < 0$ holds. The following lemma studies the asymptotic behavior of f_n and g_n at $t = t^{(\delta)}$.

Lemma 2.A.1 *Suppose (2.1) holds for $-1 < \gamma < 0$, and the sequence $k(n)$ satisfies $k(n) \to \infty$, $k(n)/n \to 0$ and $k(n)/\log n \to \infty$ as $n \to \infty$, we have that, for any $\delta \in (-1/2, 0]$, as $n \to \infty$, the following relations hold,*

$$\begin{aligned} f_n(t^{(\delta)}) &\xrightarrow{a.s.} f(\delta) := 1 + \int_0^1 \log((1+\delta)s^{-\gamma} - \delta)ds \\ g_n(t^{(\delta)}) &\xrightarrow{a.s.} g(\delta) := \int_0^1 \frac{ds}{(1+\delta)s^{-\gamma} - \delta} \\ h_n(t^{(\delta)}) &\xrightarrow{a.s.} h(\delta) := f(\delta)g(\delta) - 1. \end{aligned}$$

Proof of Lemma 2.A.1
They can be proved by applying Lemma 2.3.2. When $\gamma < 0$, the uniformly continuity required by Lemma 2.3.2 is fulfilled for each relation. $\square$

It is clear that $f(0) = \gamma + 1$, $g(0) = \frac{1}{\gamma+1}$ and $h(0) = 0$, and all three functions are left continuous at 0. Meanwhile, by calculating the left derivative of f and g at 0, we get the left derivatives of h at 0 as

$$h'(0-) = \begin{cases} -\frac{\gamma^3}{(\gamma+1)(2\gamma+1)} > 0 & -1/2 < \gamma < 0 \\ +\infty & -1 < \gamma \le -1/2 \end{cases}$$

So, it is sufficient to conclude that, there exists a $\delta_0 < 0$, for any $\delta_0 < \delta < 0$, when n is sufficiently large,

$$h_n(t_n^{(\delta)}) < 0 \quad a.s. \tag{2.18}$$

For the other side, define a different sequence

$$s_n := -\frac{1 - 1/k}{X_{n,n} - X_{n,n-k}}. \tag{2.19}$$

Our purpose is to prove that, for sufficiently large n,

$$h_n(s_n) > 0 \quad a.s. \tag{2.20}$$

The following lemma is useful in the later proof of the theorem.

Lemma 2.A.2 *With the same condition on k in Lemma 2.A.1, suppose $\lambda > 0$ is a fixed constant, then as $n \to \infty$,*

$$\int_{k^{-1/5}}^{1} \log \frac{X_{n,n} - X_{n,n-\lceil sk\rceil+1}}{X_{n,n} - X_{n,n-k}} ds \quad \xrightarrow{a.s.} \quad \gamma, \tag{2.21}$$

$$\int_{k^{-1/5}}^{1} \frac{1}{\frac{X_{n,n}-X_{n,n-\lceil sk\rceil+1}}{X_{n,n}-X_{n,n-k}} + \lambda} ds \quad \xrightarrow{a.s.} \quad \int_0^1 \frac{dt}{t^{-\gamma} + \lambda}. \tag{2.22}$$

Proof of Lemma 2.A.2

For $s \in (k^{-1/5}, 1]$,

$$0 < \frac{U(\infty) - X_{n,n}}{U(\infty) - X_{n,n-\lceil sk\rceil+1}} \le \frac{U(\infty) - X_{n,n}}{U(\infty) - X_{n,n-\lceil k^{4/5}\rceil+1}} \le (1+\varepsilon)\left(\frac{Z_{n,n}}{Z_{n,n-\lceil k^{4/5}\rceil+1}}\right)^{\gamma+\varepsilon}.$$

By using (2.10), the right side goes to 0 almost surely. So, for any given $\tau > 0$, we have eventually

$$U(\infty) - X_{n,n} < \tau(U(\infty) - X_{n,n-\lceil sk\rceil+1}).$$

Next, we use this to construct the bounds for $\frac{X_{n,n}-X_{n,n-\lceil sk\rceil+1}}{X_{n,n}-X_{n,n-k}}$ as

$$\begin{aligned}
\frac{U(\infty) - X_{n,n-\lceil sk\rceil+1}}{U(\infty) - X_{n,n-k}} &\ge \frac{X_{n,n} - X_{n,n-\lceil sk\rceil+1}}{X_{n,n} - X_{n,n-k}} \ge \frac{X_{n,n} - X_{n,n-\lceil sk\rceil+1}}{U(\infty) - X_{n,n-k}} \\
=\frac{U(\infty) - X_{n,n-\lceil sk\rceil+1}}{U(\infty) - X_{n,n-k}} &- \frac{U(\infty) - X_{n,n}}{U(\infty) - X_{n,n-k}} > (1-\tau)\frac{U(\infty) - X_{n,n-\lceil sk\rceil+1}}{U(\infty) - X_{n,n-k}}
\end{aligned}$$

which leads to

$$\begin{aligned}
\int_{k^{-1/5}}^{1} \log \frac{U(\infty) - X_{n,n-\lceil sk\rceil+1}}{U(\infty) - X_{n,n-k}} ds &\ge \int_{k^{-1/5}}^{1} \log \frac{X_{n,n} - X_{n,n-\lceil sk\rceil+1}}{X_{n,n} - X_{n,n-k}} ds \\
&> \int_{k^{-1/5}}^{1} \log(1-\tau) + \log \frac{U(\infty) - X_{n,n-\lceil sk\rceil+1}}{U(\infty) - X_{n,n-k}} ds.
\end{aligned} \tag{2.23}$$

By checking that $\log((1+\varepsilon)t^{\gamma+\varepsilon})$ is uniformly continuous on $[1, \infty)$, for $|\varepsilon| < 1$, we have

$$\int_{k^{-1/5}}^{1} \log \frac{U(\infty) - X_{n,n-\lceil sk\rceil+1}}{U(\infty) - X_{n,n-k}} ds - \int_{k^{-1/5}}^{1} \log s^{-\gamma} ds \to 0 \quad a.s.$$

which leads to

$$\int_{k^{-1/5}}^{1} \log \frac{U(\infty) - X_{n,n-\lceil sk\rceil+1}}{U(\infty) - X_{n,n-k}} ds \xrightarrow{a.s.} \gamma$$

With this result, by first taking $n \to \infty$, and then taking $\tau \to 0$ in (2.23), (2.21) is proved. The proof of (2.22) is essentially the same. □

With the above lemma, the lower bound of $f_n(s_n)$ can be calculated as follows,

$$\begin{aligned} f_n(s_n) &= 1 + \int_0^1 \log\left(1 - \frac{1-1/k}{X_{n,n} - X_{n,n-k}}(X_{n,n-\lceil sk\rceil+1} - X_{n,n-k})\right) ds \\ &> 1 + k^{-1/5} \log\left(1 - \frac{1-1/k}{X_{n,n} - X_{n,n-k}}(X_{n,n} - X_{n,n-k})\right) \\ &\quad + \int_{k^{-1/5}}^1 \log\left(1 - \frac{1}{X_{n,n} - X_{n,n-k}}(X_{n,n-\lceil sk\rceil+1} - X_{n,n-k})\right) ds \\ &= 1 - \frac{\log k}{k^{1/5}} + \int_{k^{-1/5}}^1 \log \frac{X_{n,n} - X_{n,n-\lceil sk\rceil+1}}{X_{n,n} - X_{n,n-k}} ds \xrightarrow{a.s.} 1 + \gamma > 0 \qquad (2.24) \end{aligned}$$

Similarly, for given $\lambda > 0$ when $k > 1/\lambda$,

$$\begin{aligned} g_n(s_n) &= \int_0^1 \frac{1}{1 - \frac{1-1/k}{X_{n,n}-X_{n,n-k}}(X_{n,n-\lceil sk\rceil+1} - X_{n,n-k})} \\ &= \left(\int_0^{1/k} + \int_{1/k}^{k^{-1/5}} + \int_{k^{-1/5}}^1\right) \frac{1}{1 - \frac{1-1/k}{X_{n,n}-X_{n,n-k}}(X_{n,n-\lceil sk\rceil+1} - X_{n,n-k})} ds \\ &> 1 + 0 + \int_{k^{-1/5}}^1 \frac{1}{1 - \frac{1-\lambda}{X_{n,n}-X_{n,n-k}}(X_{n,n-\lceil sk\rceil+1} - X_{n,n-k})} ds \\ &\geq 1 + \int_{k^{-1/5}}^1 \frac{1}{\frac{X_{n,n}-X_{n,n-\lceil sk\rceil+1}}{X_{n,n}-X_{n,n-k}} + \lambda} ds \xrightarrow{a.s.} 1 + \int_0^1 \frac{dt}{t^{-\gamma} + \lambda} > 0. \end{aligned}$$

So, for sufficiently large n, we have eventually

$$h_n(s_n) = f_n(s_n)g_n(s_n) - 1 > (1 + \gamma - \varepsilon)(1 + \int_0^1 \frac{dt}{t^{-\gamma} + \lambda} - \varepsilon) - 1.$$

where $\varepsilon > 0$ is small enough. Taking $\lambda \to 0$ and $\varepsilon \to 0$,

$$(1 + \gamma - \varepsilon)(\int_0^1 \frac{dt}{t^{-\gamma} + \lambda} + 1 - \varepsilon) \to \gamma + 2 > 1$$

It implies that, by choosing suitable λ, for sufficient large n, (2.20) holds.

Finally, combining (2.18) and (2.20), there must exist a root of $h_n(t) = 0$ between s_n and $t^{(\delta)}$, when n is sufficiently large. Denote the root as t_n^* again. Similar to the positive case, from relation (2.24), we have

$$\gamma + 1 \leq \liminf f_n(s_n) \leq \liminf f_n(t_n^*) \leq \limsup f_n(t_n^*) \leq \limsup f_n(t^{(\delta)}) = f(\delta) \quad a.s.$$

Note that $f(\delta) \to \gamma + 1$ as $\delta \to 0$. By taking $\delta \to 0$, the consistency of $\hat{\gamma}^* = f_n(t_n^*) - 1$ is proved. Notice that when $\gamma < 0$, $a(t) \sim -\gamma(U(\infty) - U(t))$, the proof of the consistency of $\hat{\sigma}^*$ is essentially the same as in the positive case.

Proof of existence and consistency when $\gamma = 0$

When (2.1) holds for $\gamma = 0$, there are two cases: $U(\infty) = \infty$ or $U(\infty) < \infty$. Since the maximum likelihood estimator is shift invariant, we can always assume that $0 < U(\infty) \leq \infty$ without loss of generality. Under this assumption, considering that $Z_{n,n-k} \to \infty$, we assume that the considered order statistics in the likelihood equations, $X_{n,n-k}, \cdots, X_{n,n}$, are all positive.

From Proposition B.2.17 in de Haan and Ferreira (2006), we have the following inequalities. Given any $\varepsilon > 0$, there exists t_0, such that, for $tx \geq t_0$ and $x > 0$,

$$-\varepsilon \max(x^{\varepsilon}, x^{-\varepsilon}) < \frac{U(tx) - U(t)}{a_0(t)} - \log x < \varepsilon \max(x^{\varepsilon}, x^{-\varepsilon}), \tag{2.25}$$

where a_0 is a specific auxiliary function such that $a_0(t)$ is a positive function and $a_0(t) \sim a(t)$ as $t \to \infty$. Without loss of generality, we use the notation $a(t)$ for this specific function $a_0(t)$ in the proof.

Suppose $Z_{n,1} \leq Z_{n,2} \leq \cdots \leq Z_{n,n}$ are the order statistics defined in Subsection 2.3.1. The following lemmas are useful in the proof.

Lemma 2.A.3 *Suppose the sequence $k = k(n)$ satisfies that $k(n) \to \infty$, $k(n)/n \to 0$, and $k(n)/\log n \to \infty$ as $n \to \infty$, then for any $p > 0$, as $n \to \infty$*

$$\frac{\log Z_{n,n} - \log Z_{n,n-k}}{k^p} \xrightarrow{a.s.} 0.$$

Proof of Lemma 2.A.3

From (2.9), we get that, as $n \to \infty$,

$$\begin{aligned} \frac{n}{kZ_{n,n-k}} \to 1 \quad a.s. \quad &\Rightarrow \quad \log n - \log k - \log Z_{n,n-k} \to 0 \quad a.s. \\ &\Rightarrow \quad \frac{\log n - \log Z_{n,n-k}}{k^p} \to 0 \quad a.s. \end{aligned}$$

By applying the law of iterated logarithm for sample maxima, (see, e.g. Shorack and Wellner (1986) p.408),

$$\limsup_{n\to\infty} \frac{\log Z_{n,n} - \log n}{\log\log n} = 1 \quad a.s. \Rightarrow \limsup_{n\to\infty} \frac{\log Z_{n,n} - \log n}{k^p} = 0 \quad a.s. \tag{2.26}$$

where the last step is provided by $k(n)/\log n \to \infty$.

Combining these two equations above, we get that

$$\limsup_{n\to\infty} \frac{\log Z_{n,n} - \log Z_{n,n-k}}{k^p} = 0 \quad a.s.$$

Since $\log Z_{n,n} - \log Z_{n,n-k} \geq 0$ always holds, the lemma is proved. □

Lemma 2.A.4 *Suppose (2.1) holds for $\gamma = 0$, if the sequence $k = k(n)$ satisfies $k(n) \to \infty$, $k(n)/n \to 0$, and $k(n)/(\log n)^c \to \infty$ for some $c > 1$, then*

$$M_l := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{U(Z_{n,n-i+1}) - U(Z_{n,n-k})}{a(Z_{n,n-k})}\right)^l \xrightarrow{a.s.} l!,$$

for all $l \in \mathbb{N}$.

Proof of Lemma 2.A.4

We first prove the lemma for the case $l = 1$. From (2.25) we have that

$$\begin{aligned}
&\frac{1}{k}\sum_{i=1}^{k}\frac{U(Z_{n,n-i+1}) - U(Z_{n,n-k})}{a(Z_{n,n-k})} = \int_0^1 \frac{U(Z_{n,n-\lceil sk\rceil+1}) - U(Z_{n,n-k})}{a(Z_{n,n-k})}ds \\
\le&\frac{1}{k^{1/5}}\log\frac{Z_{n,n}}{Z_{n,n-k}} + \frac{1}{k^{1/5}}\varepsilon\left(\frac{Z_{n,n}}{Z_{n,n-k}}\right)^{\varepsilon} + \int_{k^{-1/5}}^1\left(\log\frac{Z_{n,n-\lceil sk\rceil+1}}{Z_{n,k}} + \varepsilon\left(\frac{Z_{n,n-\lceil sk\rceil+1}}{Z_{n,n-k}}\right)^{\varepsilon}\right)ds \\
=:&I_1 + I_2 + I_3 \qquad (2.27)
\end{aligned}$$

First of all, from Lemma 2.A.3, $I_1 \to 0$ almost surely as $n \to \infty$. Secondly, from (2.26), we have that for any $\tau > 1$ when n is sufficiently large,

$$\log Z_{n,n} - \log n \le \tau \log\log n$$

holds almost surely. Hence, eventually, we have that $Z_{n,n}/n \le (\log n)^{\tau}$. Considering that $k(n)/(\log n)^c \to \infty$ for some $c > 1$, by taking $\tau < c$, we have that $Z_{n,n}/(nk) \to 0$ holds almost surely as $n \to \infty$. Combining with the fact that $kZ_{n,n-k}/n \to 1$ almost surely, we have that, as $n \to \infty$,

$$\frac{Z_{n,n}}{k^2 Z_{n,n-k}} \xrightarrow{a.s.} 0.$$

By taking $\varepsilon < 1/10$ in (2.27), we proved that $I_2 \to 0$ almost surely as $n \to \infty$.

Thirdly, we consider I_3. From (2.7), we get that for all $s \in (k^{-1/5}, 1]$ and sufficiently large n,

$$\left|\frac{n/k}{sZ_{n,n-\lceil sk\rceil+1}} - 1\right| \le M\frac{(2\log\log n)^{1/2}}{k^{1/2-1/5}} \quad a.s.$$

Since $k/(\log n)^c \to \infty$ as $n \to \infty$, we get that

$$\frac{n/k}{sZ_{n,n-\lceil sk\rceil+1}} \to 1 \quad a.s.$$

holds uniformly for all $s \in (k^{-1/5}, 1]$. Together with (2.9), we have that

$$\frac{Z_{n,n-k}}{sZ_{n,n-\lceil sk\rceil+1}} \to 1 \quad a.s.$$

holds uniformly for all $s \in (k^{-1/5}, 1]$.

Thus, as $n \to \infty$,

$$\int_{k^{-1/5}}^{1} \log \frac{Z_{n,n-k}}{sZ_{n,n-\lceil sk \rceil+1}} ds \to 0 \ \ a.s.$$

It implies that

$$\int_{k^{-1/5}}^{1} \log \frac{Z_{n,n-\lceil sk \rceil+1}}{Z_{n,n-k}} ds - \int_{k^{-1/5}}^{1} \log \left(\frac{1}{s} \right) ds \to 0 \ \ a.s.$$

as $n \to \infty$. A similar result holds for function $\varepsilon \left(\frac{1}{s}\right)^{\varepsilon}$. Hence we have that

$$I_3 - \int_{k^{-1/5}}^{1} \left(\log \left(\frac{1}{s} \right) + \varepsilon \left(\frac{1}{s} \right)^{\varepsilon} \right) ds \xrightarrow{a.s.} 0,$$

which leads to

$$I_3 \to 1 + \frac{\varepsilon}{1-\varepsilon},$$

as $n \to \infty$. Combining these three parts, we have that

$$\limsup_{n\to\infty} \frac{1}{k} \sum_{i=1}^{k} \frac{U(Z_{n,n-i+1}) - U(Z_{n,n-k})}{a_0(Z_{n,n-k})} \leq 1 + \frac{\varepsilon}{1-\varepsilon} \ \ a.s.$$

A similar lower bound applies. Then by taking $\varepsilon \to 0$, the lemma under $l = 1$ is proved. The proofs of the other cases are similar, the only difference in the proofs is that the dividing point $k^{-1/5}$ has to be changed into some smaller power function of k. That is not a hurdle to pass if a correspondingly smaller ε is chosen. □

Lemma 2.A.5 *Suppose the conditions in Lemma 2.A.4 hold. Then*

$$\lim_{n\to\infty} \int_{k^{-1/5}}^{1} \log \frac{U(Z_{n,n}) - U(Z_{n,n-\lceil sk \rceil+1})}{U(Z_{n,n}) - U(Z_{n,n-k})} ds = 0 \quad a.s.$$

Proof of Lemma 2.A.5

Since

$$\frac{U(Z_{n,n}) - U(Z_{n,n-\lceil sk \rceil+1})}{U(Z_{n,n}) - U(Z_{n,n-k})} < 1,$$

for all $s \in [k^{-1/5}, 1]$, it is only necessary to prove that

$$\liminf_{n\to\infty} \int_{k^{-1/5}}^{1} \log \frac{U(Z_{n,n}) - U(Z_{n,n-\lceil sk \rceil+1})}{U(Z_{n,n}) - U(Z_{n,n-k})} ds \geq 0.$$

Since $Z_{n,n}/Z_{n,n-\lceil k^{4/5} \rceil+1} \to \infty$ as $n \to \infty$, for any $C > 0$, when n is sufficiently large,

$$Z_{n,n} > e^{2C} \cdot Z_{n,n-\lceil k^{4/5} \rceil+1} \geq e^{2C} \cdot Z_{n,n-\lceil sk \rceil+1},$$

for all $s \in [k^{-1/5}, 1]$. Considering that U is an increasing function, we get that

$$\frac{U(Z_{n,n}) - U(Z_{n,n-\lceil sk \rceil+1})}{a(Z_{n,n-\lceil sk \rceil+1})} \geq \frac{U(e^{2C} \cdot Z_{n,n-\lceil sk \rceil+1}) - U(Z_{n,n-\lceil sk \rceil+1})}{a(Z_{n,n-\lceil sk \rceil+1})} \to 2C \quad a.s.$$

Hence, for sufficiently large n,

$$\frac{U(Z_{n,n}) - U(Z_{n,n-\lceil sk \rceil+1})}{a(Z_{n,n-\lceil sk \rceil+1})} > C.$$

On the other hand, inequality (2.25) implies that

$$\frac{U(Z_{n,n-k}) - U(Z_{n,n-\lceil sk \rceil+1})}{a(Z_{n,n-\lceil sk \rceil+1})} \geq \log \frac{Z_{n,n-k}}{Z_{n,n-\lceil sk \rceil+1}} - \varepsilon \left(\frac{Z_{n,n-\lceil sk \rceil+1}}{Z_{n,n-k}} \right)^{\varepsilon} \geq -\left(\frac{1}{\varepsilon} + \varepsilon \right) \left(\frac{Z_{n,n-\lceil sk \rceil+1}}{Z_{n,n-k}} \right)^{\varepsilon}.$$

The last step is a direct consequence of the fact that for all $x > 1$ and $\varepsilon > 0$, $\log x \leq \frac{1}{\varepsilon} x^{\varepsilon}$. Combining the two inequalities above, we get that for all $s \in [k^{-1/5}, 1]$,

$$\begin{aligned}
\frac{U(Z_{n,n}) - U(Z_{n,n-\lceil sk \rceil+1})}{U(Z_{n,n}) - U(Z_{n,n-k})} &= \frac{1}{1 + \frac{U(Z_{n,n-\lceil sk \rceil+1}) - U(Z_{n,n-k})}{U(Z_{n,n}) - U(Z_{n,n-\lceil sk \rceil+1})}} \\
&> \frac{1}{1 + \frac{1}{C} \left(\frac{1}{\varepsilon} + \varepsilon \right) \left(\frac{Z_{n,n-\lceil sk \rceil+1}}{Z_{n,n-k}} \right)^{\varepsilon}} \\
&> \frac{1}{1 + \frac{1}{C} \left(\frac{1}{\varepsilon} + \varepsilon \right)} \cdot \frac{1}{\left(\frac{Z_{n,n-\lceil sk \rceil+1}}{Z_{n,n-k}} \right)^{\varepsilon}} \\
&=: L(C, \varepsilon) \cdot \left(\frac{Z_{n,n-k}}{Z_{n,n-\lceil sk \rceil+1}} \right)^{\varepsilon}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\int_{k^{-1/5}}^{1} \log \frac{U(Z_{n,n}) - U(Z_{n,n-\lceil sk \rceil+1})}{U(Z_{n,n}) - U(Z_{n,n-k})} ds &\geq \int_{k^{-1/5}}^{1} \left(\log L(C, \varepsilon) + \varepsilon \log \frac{Z_{n,n-k}}{Z_{n,n-\lceil sk \rceil+1}} \right) ds \\
&= (1 - k^{-1/5}) \log L(C, \varepsilon) + \varepsilon \int_{k^{-1/5}}^{1} \log \frac{Z_{n,n-k}}{Z_{n,n-\lceil sk \rceil+1}} ds \\
&:= I_4 + I_5.
\end{aligned}$$

As discussed in the proof of Lemma 2.A.4, we have that, as $n \to \infty$

$$\int_{k^{-1/5}}^{1} \log \frac{Z_{n,n-k}}{Z_{n,n-\lceil sk \rceil+1}} ds - \int_{k^{-1/5}}^{1} \log s ds \to 0 \quad a.s.$$

Therefore $I_5 \to -1$ almost surely as $n \to \infty$. Together with the fact that $I_4 \to \log L(C, \varepsilon)$, we get

$$\liminf_{n \to \infty} \int_{k^{-1/5}}^{1} \log \frac{U(Z_{n,n}) - U(Z_{n,n-\lceil sk \rceil+1})}{U(Z_{n,n}) - U(Z_{n,n-k})} ds \geq \log L(C, \varepsilon) - \varepsilon.$$

By taking $\varepsilon \to 0$ and $C\varepsilon \to \infty$, we proved the lemma. □

Now we shall prove Theorem 2.2.3. Similar to the case γ positive we shall find a sequence $\delta_n > 0$ such that $h_n(t^{(\delta_n)}) < 0$ for sufficiently large n, where $t^{(\delta)}$ is defined as $t^{(\delta)} = \frac{\delta}{a(Z_{n,n-k})}$. Meanwhile, similar to the case γ negative, we shall prove that for a sequence s_n, $h(s_n) > 0$ holds for sufficiently large n.

We start with the sequence δ_n by calculating the upper bounds of $f_n(t^{(\delta_n)})$ and $g_n(t^{(\delta_n)})$ as follows. From the inequalities $\log(1+x) < 1 + x - x^2 + x^3$ and $1/(1+x) < 1 - x + x^2 - x^3 + x^4$ for all $x > 0$, we get

$$\begin{aligned} f_n(t^{(\delta_n)}) - 1 &= \frac{1}{k}\sum_{i=1}^{k} \log\left(1 + \frac{\delta_n(U(Z_{n,n-i+1}) - U(Z_{n,n-k}))}{a(Z_{n,n-k})}\right) \\ &\le \delta_n M_1 - \frac{\delta_n^2}{2} M_2 + \frac{\delta_n^3}{3} M_3 \end{aligned}$$

$$\begin{aligned} g_n(t^{(\delta_n)}) - 1 &= \frac{1}{k}\sum_{i=1}^{k} \frac{1}{1 + \frac{\delta_n(U(Z_{n,n-i+1}) - U(Z_{n,n-k}))}{a(Z_{n,n-k})}} - 1 \\ &\le -\delta_n M_1 + \delta_n^2 M_2 - \delta_n^3 M_3 + \delta_n^4 M_4. \end{aligned}$$

Combining the above two, we get that, for $\delta_n < 1$

$$\begin{aligned} h_n(t^{(\delta_n)}) &\le \delta_n^2\left(\frac{M_2}{2} - M_1^2\right) + \delta_n^3\left(\frac{3M_1M_2}{2} - \frac{2M_3}{3}\right) + \delta_n^4\left(M_4 - \frac{M_2^2}{2} - \frac{4M_1M_3}{3}\right) \\ &\quad + \delta_n^5\left(\frac{5M_2M_3}{6} + M_1M_4\right) - \delta_n^6\left(\frac{M_3^2}{3} + \frac{M_2M_4}{2}\right) + \delta_n^7\frac{M_3M_4}{3} \\ &\le \delta_n^2\left(\frac{M_2}{2} - M_1^2\right) + \delta_n^3\left(\frac{3M_1M_2}{2} - \frac{2M_3}{3}\right) \\ &\quad + \delta_n^4\left(M_4 - \frac{M_2^2}{2} - \frac{4M_1M_3}{3} + \frac{5M_2M_3}{6} + M_1M_4 + \frac{M_3M_4}{3}\right) \\ &:= \delta_n^2(P + Q\delta_n + R\delta_n^2), \end{aligned}$$

where $P = \frac{M_2}{2} - M_1^2 \to 0$, $Q = \frac{3M_1M_2}{2} - \frac{2M_3}{3} \to -1$, and $R = M_4 - \frac{M_2^2}{2} - \frac{4M_1M_3}{3} + \frac{5M_2M_3}{6} + M_1M_4 + \frac{M_3M_4}{3} \to 96$ almost surely as $n \to \infty$.

Choose

$$\delta_n := \frac{-Q - \sqrt{Q^2 - 3\,|PR|}}{2R} > 0.$$

Then $\delta_n \to 0$ almost surely as $n \to \infty$. Hence, for sufficiently large n, $0 < \delta_n < 1$. Meanwhile, δ_n is always on the interval $(\frac{-Q-\sqrt{Q^2-4PR}}{2R}, \frac{-Q+\sqrt{Q^2-4PR}}{2R})$, which ensures that $P + Q\delta_n + R\delta_n^2 < 0$. Therefore, continuing with the upper bound of $h_n(t^{(\delta_n)})$, we have that, for sufficiently large n, $h_n(t^{(\delta_n)}) < 0$.

For the other side, we use the s_n defined in (2.19) and prove that for sufficiently large n, $h_n(s_n) > 0$. Similar to the proof for $-1 < \gamma < 0$, we have that

$$f_n(s_n) > 1 - \frac{\log k}{k^{1/5}} + \int_{k^{-1/5}}^{1} \log\frac{U(Z_{n,n}) - U(Z_{n,n-\lceil sk\rceil+1})}{U(Z_{n,n}) - U(Z_{n,n-k})} ds \xrightarrow{a.s.} 1.$$

Notice that the last step comes from Lemma 2.A.5. For $g_n(s_n)$, similar to the proof for $-1 < \gamma < 0$, by fixing $\lambda > 0$, when $k > 1/\lambda$, we have that

$$g_n(s_n) \ge 1 + \int_{k^{-1/5}}^{1} \frac{1}{\frac{X_{n,n}-X_{n,n-\lceil sk\rceil+1}}{X_{n,n}-X_{n,n-k}} + \lambda} \ge 1 + \frac{1 - k^{-1/5}}{1+\lambda} \to 1 + \frac{1}{1+\lambda} > 0.$$

Therefore, for sufficiently large n, $h_n(s_n) > (1-\varepsilon)(1+\frac{1}{1+\lambda}-\varepsilon)-1>0$, where $\varepsilon>0$ is small enough.

Although $h_n(t^{(\delta_n)})$ and $h_n(s_n)$ eventually have different signs, it is so far not sufficient to conclude the existence of a non-zero root of $h_n(t)=0$ between $t^{(\delta_n)}$ and s_n. The reason is that 0 is always a root of $h_n(t)$ lying between they two. In order to pass this hurdle, we study the behavior of $h_n(t)$ in the neighborhood of 0. Notice that $h_n'(0)=0$ and

$$h_n''(0) = \frac{1}{k}\sum_{i=1}^{k}(X_{n,n-i+1}-X_{n,n-k})^2 - 2\left(\frac{1}{k}\sum_{i=1}^{k}(X_{n,n-i+1}-X_{n,n-k})\right)^2.$$

The extra condition (2.6) ensures that on a probability 1 set, $h_n''(0)\neq 0$. Together with $h_n'(0)=0$, it is confirmed that $h_n(t)$ has the same sign at the two sides of 0, in a close neighborhood. Now, considering the fact that $h_n(t^{(\delta_n)})$ and $h_n(s_n)$ have different signs, it is proved that, for sufficiently large n, with probability 1, there exists a non-zero root t_n^* of $h_n(t)=0$ on the interval $(s_n, t^{(\delta_n)})$.

Since $\delta_n \to 0$ almost surely as $n\to\infty$, from the upper bound of $f_n(t^{(\delta_n)})$, it is not difficult to verify that $\limsup_{n\to\infty} f_n(t^{(\delta_n)}) \leq 1$ almost surely. On the other hand, Lemma 2.A.5 ensures that $\liminf_{n\to\infty} f_n(s_n)\geq 1$ almost surely. Considering the fact that f_n is an increasing function and t_n^* lies in between s_n and $t^{(\delta_n)}$, we conclude that $\hat{\gamma}_n^* = f_n(t_n^*)-1 \xrightarrow{a.s.} 0$, i.e. $\hat{\gamma}_n^*$ is strongly consistent.

2.B Appendix B

Proof of Proposition 2.4.1 for $-1/2 < \gamma < 0$
The proof of existence for γ negative showed that

$$t^{(\delta)} = -\frac{1+\delta}{U(\infty) - X_{n,n-k}}$$

is the upper bound of the root t_n^*, where δ is a fixed negative number. To use a fixed $\delta < 0$ instead of a negative sequence $\delta_n \to 0$ as shown in the positive case is a compromise to have a unified proof for all $-1 < \gamma < 0$. Actually, in case $-1/2 < \gamma < 0$, it is still possible to use a negative sequence $\delta_n \to 0$ to prove that $h_n(t^{(\delta_n)}) < 0$. The proof is parallel to the positive case. Thus we omit its detail but only sketch the main instruments.

Denote

$$\bar{g}_n := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{U(\infty) - U(Z_{n,n-k})}{U(\infty) - U(Z_{n,n-i+1})}\right)^2.$$

Similar to Lemma 2.3.3, the following three relations hold, as $n \to \infty$,

$$\begin{aligned} f_n(t^{(0)}) &\xrightarrow{a.s.} \gamma + 1, \\ g_n(t^{(0)}) &\xrightarrow{a.s.} \frac{1}{\gamma+1}, \\ \bar{g}_n &\xrightarrow{a.s.} \frac{1}{2\gamma+1}. \end{aligned}$$

Note that the last relation requires $\gamma > -1/2$.

The upper bounds of $f_n(t^{(\delta_n)})$ and $g_n(t^{(\delta_n)})$ are calculated as

$$\begin{aligned} f_n(t^{(\delta_n)}) - f_n(t^{(0)}) &< \delta_n(1 - g_n(t^{(0)})) \\ g_n(t^{(\delta_n)}) - g_n(t^{(0)}) &< -\delta_n(g_n(t^{(0)}) - \bar{g}_n) + \delta_n^2(g_n(t^{(0)}) - \bar{g}_n)^2. \end{aligned}$$

Hence,

$$h_n(t^{(\delta_n)}) < f_n(t^{(0)})g_n(t^{(0)}) - 1 + \delta_n C_n.$$

where $C_n \to -\frac{\gamma^3}{(1+\gamma)^2(2\gamma+1)} > 0$ almost surely as $n \to \infty$. Therefore, by taking

$$\delta_n := \left|f_n(t^{(0)})g_n(t^{(0)}) - 1\right|\left(-\frac{1}{C_n}\right),$$

for sufficiently large n, $\delta_n < 0$ and $h_n(t^{(\delta_n)}) < 0$.

In the proof of the existence for γ negative, we did not define $t^{(\delta)}$ for $\delta > 0$ because $1 + t^{(\delta)}Y_1 > 0$ is not always ensured for positive δ. In particular, when $\gamma < -1/2$, as $n \to \infty$, it can be proved that eventually $1 + t^{(\delta)}Y_1 < 0$ for $\delta > 0$. Thus $f_n(t^{(\delta)})$ is not well defined. However, when $-1/2 < \gamma < 0$, with the second order condition, it is still possible to define $t^{(-\delta_n)}$, where δ_n is the same as above. In this case we first verify that $1+t^{(-\delta_n)}Y_1>0$ holds for sufficiently large n. This inequality is equivalent to

$$-\delta_n < \frac{U(\infty)-U(Z_{n,n})}{U(Z_{n,n})-U(Z_{n,n-k})}. \tag{2.28}$$

Similar to the proof in Section 2.4, by multiplying $\sqrt{k}$, the left side is $O_p(1)$. The right side can be bounded as

$$\begin{aligned}\sqrt{k}\frac{U(\infty)-U(Z_{n,n})}{U(Z_{n,n})-U(Z_{n,n-k})} &> \sqrt{k}\frac{U(\infty)-U(Z_{n,n})}{U(\infty)-U(Z_{n,n-k})}\\ &\geq \sqrt{k}(1-\varepsilon)\left(\frac{Z_{n,n}}{Z_{n,n-k}}\right)^{\gamma-\varepsilon},\end{aligned}$$

where ε is a given positive number. Since $\frac{Z_{n,n}}{k^{1-\varepsilon}Z_{n,n-k}} \to \infty$ almost surely as $n\to\infty$, we have for any $M>0$

$$\begin{aligned}\sqrt{k}\frac{U(\infty)-U(Z_{n,n})}{U(Z_{n,n})-U(Z_{n,n-k})} &> \sqrt{k}(1-\varepsilon)\left(Mk^{1-\varepsilon}\right)^{\gamma-\varepsilon}\\ &= (1-\varepsilon)M^{\gamma-\varepsilon}k^{(1-\varepsilon)(\gamma-\varepsilon)+1/2}\end{aligned}$$

Because $\gamma>-1/2$, for sufficiently small ε, $(1-\varepsilon)(\gamma-\varepsilon)+1/2>0$. Thus, the right side goes to infinity, which verifies the inequality (2.28).

Now we turn to the lower bounds of $f_n(t^{(-\delta_n)})$ and $g_n(t^{(-\delta_n)})$. First of all, from the proof of the inequality (2.28), it is observed that the inequality can be improved as that

$$-\delta_n < \tau\left(\frac{U(\infty)-U(Z_{n,n})}{U(Z_{n,n})-U(Z_{n,n-k})}\right)$$

eventually holds for any $\tau>0$. It implies that, eventually, we have

$$-\delta_n\frac{U(Z_{n,n})-U(Z_{n,n-k})}{U(\infty)-U(Z_{n,n})} < \tau.$$

From the inequality that $\log(1-x) > \frac{-x}{1-x}$ for all $0<x<1$, the lower bound of $f_n(t^{(-\delta_n)})$ is given as

$$\begin{aligned}f_n(t^{(-\delta_n)})-f_n(t^{(0)}) &\geq \frac{\delta_n}{1+\delta_n\frac{U(Z_{n,n})-U(Z_{n,n-k})}{U(\infty)-U(Z_{n,n})}}(g_n(t^{(0)})-1)\\ &> \frac{\delta_n}{1-\tau}(g_n(t^{(0)})-1).\end{aligned}$$

Meanwhile, we have that

$$g_n(t^{(-\delta_n)})-g_n(t^{(0)}) \geq -\delta_n(\bar{g}_n-g_n(t^{(0)})).$$

Hence, combining these two,

$$h_n(t^{(-\delta_n)}) > f_n(t^{(0)})g_n(t^{(0)}) - 1 + \delta_n D_n,$$

where $D_n \to -\frac{1}{1-\tau}\frac{\gamma}{(1+\gamma)^2} + \frac{\gamma}{1+2\gamma} < 0$ for sufficiently small τ. Meanwhile, since

$$\frac{1}{1-\tau}\frac{\gamma}{(1+\gamma)^2} - \frac{\gamma}{1+2\gamma} < -\frac{\gamma^3}{(1+\gamma)^2(2\gamma+1)},$$

it implies that, eventually, $-D_n < C_n$. Therefore, by modifying δ_n as

$$\delta'_n := \left|f_n(t^{(0)})g_n(t^{(0)}) - 1\right|\left(\frac{1}{D_n}\right),$$

it still serves the purpose that $h_n(t^{(\delta'_n)}) < 0$ and $\sqrt{k}\delta'_n = O_p(1)$, while $h_n(t^{(-\delta'_n)}) > 0$ is now ensured.

Finally, we conclude that fop $-1/2 < \gamma < 0$, with the second order condition, there exists a root t^*_n of $h_n(t) = 0$ lying between $t^{(\delta'_n)}$ and $t^{(-\delta'_n)}$.

Similar to the positive case, we first study the asymptotic behavior of $t^{(\pm\delta'_n)}a(n/k)$. Notice that in the negative case one may take $a(t) = -\gamma(U(\infty) - U(t))$. It is clear that

$$t^{(\delta'_n)}a(n/k) = \gamma(1-\delta'_n)\frac{U(\infty) - U(n/k)}{U(\infty) - U(Z_{n,n-k})}.$$

Under the second order condition, the property of location estimation ensures that

$$\sqrt{k}\left(\frac{U(\infty) - U(n/k)}{U(\infty) - U(Z_{n,n-k})} - 1\right) = O_p(1).$$

Together with the fact that $\sqrt{k}\delta'_n = O_p(1)$, we get $\sqrt{k}t^{(\delta'_n)}a(n/k) = O_p(1)$. Similar result holds for $t^{(-\delta'_n)}a(n/k)$, hence, also t^*_n. This completes the proof of (2.17) for $-1/2 < \gamma < 0$.

Proof of Proposition 2.4.1 for $\gamma = 0$

The proof of existence for $\gamma = 0$ showed that by defining $t^{(\delta)} = \frac{\delta}{a(Z_{n,n-k})}$ and choosing a suitable positive sequence $\delta_n \to 0$ as $n \to \infty$, we obtained the upper bound of the root t^*_n as $t^{(\delta_n)}$. For the lower bound, with only the first order condition, we could not choose a corresponding negative sequence because it is not ensured that $1 + t^{(\delta)}Y_1 > 0$ for negative δ. As a compromise, we used s_n defined in (2.19). Now, under the second order condition, we can pass the hurdle and return to a similar setup as the upper bound. The reason is the following lemma.

Lemma 2.B.1 *Suppose the second order condition in (2.5) holds with $\gamma = 0$. Suppose the sequence k satisfies the condition in Lemma 2.A.3. Then, for any $p > 0$, as $n \to \infty$*

$$k^{-p}\frac{U(Z_{n,n}) - U(Z_{n,n-k})}{a(Z_{n,n-k})} \xrightarrow{a.s.} 0.$$

Proof of Lemma 2.B.1

It is only necessary to prove that

$$\limsup_{n\to\infty} k^{-p}\frac{U(Z_{n,n})-U(Z_{n,n-k})}{a(Z_{n,n-k})}=0 \quad a.s.$$

Under the second order condition with $\gamma=0$, Theorem 2.3.6 in de Haan and Ferreira (2006) provides the following inequality: for any $\varepsilon>0$, there exists t_0 such that for all $t\geq t_0$ and $x\geq 1$,

$$\left|\frac{\frac{U(tx)-U(t)}{a_0(t)}-\log x}{A_0(t)}-\Psi_\rho(x)\right|\leq \varepsilon x^{\rho+\varepsilon},$$

where $\rho\leq 0$, a_0 and A_0 are specific choices of a and A in (2.5), and

$$\Psi_\rho(x)=\begin{cases}\frac{x^\rho-1}{\rho}, & \rho<0;\\ \frac{1}{2}(\log x)^2, & \rho=0.\end{cases}$$

By applying this inequality with $t=Z_{n,n-k}$ and $x=Z_{n,n}/Z_{n,n-k}$, we get that

$$\frac{U(Z_{n,n})-U(Z_{n,n-k})}{a_0(Z_{n,n-k})}\leq \log\frac{Z_{n,n}}{Z_{n,n-k}}+A_0(Z_{n,n-k})\left(\Psi_\rho\left(\frac{Z_{n,n}}{Z_{n,n-k}}\right)+\varepsilon\left(\frac{Z_{n,n}}{Z_{n,n-k}}\right)^{\rho+\varepsilon}\right):=I_6+I_7.$$

Firstly, Lemma 2.A.3 ensures that $I_6/k^p\to 0$ as $n\to\infty$. Secondly, by choosing $\varepsilon<p$, it is not difficult to verify that as $n\to\infty$,

$$\frac{1}{k^p}\varepsilon\left(\frac{Z_{n,n}}{Z_{n,n-k}}\right)^{\rho+\varepsilon}\to 0.$$

When $\rho<0$, since $\lim_{x\to\infty}\Psi_\rho(x)=-1/\rho$, we have that as $n\to\infty$,

$$\frac{1}{k^p}\Psi_\rho\left(\frac{Z_{n,n}}{Z_{n,n-k}}\right)\to 0.$$

For $\rho=0$, by applying Lemma 2.A.3, the above relation still holds. Together with the fact that $A_0(Z_{n,n-k})\to 0$ as $n\to\infty$, we have that $I_7/k^p\to 0$ as $n\to\infty$. Thus we proved the lemma by substituting a for a_0. Notice that $a_0(t)\sim a(t)$ as $t\to\infty$. The lemma holds for any a function satisfying the second order condition. □

By taking $p=1/2$ in Lemma 2.B.1 we get that for a negative sequence δ_n' satisfying that $\sqrt{k}\delta_n'=O_p(1)$ as $n\to\infty$, $t^{(\delta_n')}Y_1\to 0$ almost surely. Thus $1+t^{(\delta_n')}Y_1>0$ holds for sufficiently large n. We shall find such a sequence ensuring the inequality $h_n(t^{(\delta_n')})>0$. From the discussion above, we get that, when n is sufficiently large, $0>t^{(\delta_n')}Y_1>-1/6$. Since $\log(1+x)>x-x^2/2+x^3/(3(1-1/6))$ for all $-1/6<x<0$, we get that

$$f_n(t^{(\delta_n')})\geq M_1\delta_n'-\frac{M_2}{2}(\delta_n')^2+\frac{2M_3}{5}(\delta_n')^3,$$

where M_l is defined in Lemma 2.A.4 for $l \in \mathbb{N}$. Together with the inequality that $g_n(t^{(\delta'_n)}) \geq 1 - M_1\delta'_n + M_2(\delta'_n)^2 - M_3(\delta'_n)^3$, for $-1 < \delta'_n < 0$, the lower bound of $h_n(t^{(\delta'_n)})$ is given as

$$\begin{aligned} h_n(t^{(\delta'_n)}) \geq & (\delta'_n)^2 \left(\frac{M_2}{2} - M_1^2\right) + (\delta'_n)^3 \left(-\frac{3M_3}{5} + \frac{3M_1M_2}{2}\right) \\ & - (\delta'_n)^4 \left(\frac{7M_3M_1}{5} + \frac{M_2^2}{2}\right) + (\delta'_n)^5 \frac{9M_2M_3}{10} - (\delta'_n)^6 \frac{2M_3^2}{5} \\ \geq & (\delta'_n)^2 \left(\frac{M_2}{2} - M_1^2\right) + (\delta'_n)^3 \left(-\frac{3M_3}{5} + \frac{3M_1M_2}{2}\right) \\ & - (\delta'_n)^4 \left(\frac{7M_3M_1}{5} + \frac{M_2^2}{2}\right) - (\delta'_n)^4 \frac{9M_2M_3}{10} - (\delta'_n)^4 \frac{2M_3^2}{5} \\ := & (\delta'_n)^2 P + (\delta'_n)^3 Q_1 + (\delta'_n)^4 R_1, \end{aligned}$$

where $P = \frac{M_2}{2} - M_1^2 \to 0$, $Q_1 = -\frac{3M_3}{5} + \frac{3M_1M_2}{2} \to -3/5$, and $R_1 = -\left(\frac{7M_3M_1}{5} + \frac{M_2^2}{2} + \frac{9M_2M_3}{10} + \frac{2M_3^2}{5}\right) \to -\frac{178}{5}$ as $n \to \infty$.

Denote

$$\delta'_n := \frac{-Q_1 - \sqrt{Q_1^2 - 3\,|PR_1|}}{2R_1} < 0.$$

Then $\delta'_n \to 0$ almost surely as $n \to \infty$. Hence, for sufficiently large n, $-1 < \delta'_n < 0$. Meanwhile, δ'_n is always on the interval $(\frac{-Q_1+\sqrt{Q_1^2-4PR_1}}{2R_1}, \frac{-Q_1-\sqrt{Q_1^2-4PR_1}}{2R_1})$, which ensures that $P + Q_1\delta'_n + R_1(\delta'_n)^2 > 0$. Therefore, continuing with the lower bound of $h_n(t^{(\delta'_n)})$, we have that, for sufficiently large n, $h_n(t^{(\delta'_n)}) > 0$.

Hence, similar to the proof of the existence for $\gamma = 0$, we conclude that for sufficiently large n, there exists a non-zero root t_n^* of $h_n(t) = 0$ lying between $t^{(\delta'_n)}$ and $t^{(\delta_n)}$.

Now we turn to consider the speed of convergence under the second order condition, i.e. we change from almost sure convergence to convergence in probability. It is not difficult to verify that with the second order condition, for M_l in Lemma 2.A.4, the speeds of convergence are at the level $1/\sqrt{k}$, which implies the same speeds of convergence for P, Q_1 and R_1, thus also δ'_n, i.e. $\sqrt{k}\delta'_n = O_p(1)$. Because a is a regularly varying function with index 0 and $Z_{n,n-k}/(n/k) \to 1$ almost surely as $n \to \infty$, it is a direct consequence that $a(n/k)/a(Z_{n,n-k}) \to 1$ almost surely. Since $t^{(\delta'_n)}a(n/k) = \delta_n \cdot \frac{a(n/k)}{a(Z_{n,n-k})}$, we get that that $\sqrt{k}t^{(\delta'_n)}a(n/k) = O_p(1)$, i.e. (2.17) holds for $t^{(\delta'_n)}$.

We recall the definition of δ_n as

$$\delta_n := \frac{-Q - \sqrt{Q^2 - 3\,|PS|}}{2S} > 0.$$

For the definitions of Q, P and S, see the proof of existence in case $\gamma = 0$. Similar to the discussion above, we obtain the result for $t^{(\delta_n)}$ similar to the one for $t^{(\delta'_n)}$. Thus (2.17) holds for t_n^* which lies in between $t^{(\delta'_n)}$ and $t^{(\delta_n)}$. $\square$

Chapter 3

Extending the Maximum Likelihood Estimator for the Extreme Value Index

3.1 Introduction

Let $X_1, X_2, \cdots$ be independent and identically distributed (i.i.d.) random variables from a distribution function F. Suppose that F is in the domain of attraction of an extreme value distribution, i.e. there exist constants $a_n > 0$ and b_n, such that

$$F^n(a_n x + b_n) \to G_\gamma(x), \quad \text{for all } 1 + \gamma x > 0$$

where $G_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma})$ is the corresponding extreme value distribution function and $\gamma \in \mathbb{R}$ is the extreme value index (Gnedenko (1943)). This is commonly denoted by $F \in D(G_\gamma)$.

To estimate the extreme value index γ is a major issue in extreme value statistics. For positive γ, Hill (1975) proposed the so-called the *Hill estimator*. For general γ, we have the *Pickands' estimator* suggested by Pickands III (1975), the *moment estimator* suggested by Dekkers *et al.* (1989), and the *UH estimator* suggested by Beirlant *et al.* (1996). In Falk (1995), an estimator for $\gamma < -1/2$ is proposed. Since the estimator has a similar construction as the Hill estimator, it is usually called the *negative Hill estimator*. The construction is as follows. Denote the order statistics of $X_1, X_2, \cdots, X_n$ as $X_{n,1} \leq \cdots \leq X_{n,n}$. Then

$$\hat{\gamma}_F = \frac{1}{k} \sum_{i=2}^{k} \log(X_{n,n} - X_{n,n-i+1}) - \log(X_{n,n} - X_{n,n-k})$$

gives the negative Hill estimator, where k is a suitable sequence such that $k(n) \to \infty$ and $k(n)/n \to 0$ as $n \to \infty$.

To study the asymptotic properties of those estimators, we need the necessary and sufficient conditions of $F \in D(G_\gamma)$. One of the commonly used criterion is as follows, see e.g. de Haan (1984a).

Theorem 3.1.1 *Let* $U := \left(\frac{1}{1-F}\right)^{\leftarrow}$ *be the generalized inverse function of* $1/(1-F)$. *Then* $F \in D(G_\gamma)$ *if and only if there exists a function* $a(t) > 0$ *such that*

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \tag{3.1}$$

for all $x > 0$.

The condition (3.1) is called the *extreme value condition*. For all of the above estimators, it is known that they are consistent under the extreme value condition.

In order to get the asymptotic normality, de Haan and Stadtmüller (1996) introduced the *second order condition* as

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma-1}{\gamma}}{A(t)} = H(x), \tag{3.2}$$

for all $x > 0$, where $H(x)$ is assumed not to be a multiple of $\frac{x^\gamma-1}{\gamma}$, and $A(t) \to 0$ as $t \to \infty$. With the second order condition, the asymptotic normality of all the above estimators has been proved.[1] Compared to this condition, we call the extreme value condition (3.1), the *first order condition*.

An alternative characterization of the necessary and sufficient condition for a distribution function F belonging to the domain of attraction is via the "excess distribution function" as in Balkema and de Haan (1974). Denote the excess distribution function

$$F_t(x) := P(X - t \le x | X > t) = \frac{F(t+x) - F(t)}{1 - F(t)}.$$

Then $F \in D(G_\gamma)$ is equivalent to

$$\lim_{t\to x^*} F_t(x\sigma(t)) = H_\gamma(x) := 1 - (1 + \gamma x)^{-1/\gamma},$$

for all $1 + \gamma x > 0$, where $\sigma(t)$ is a positive function and x^* is the right endpoint of F, i.e. $x^* = \sup\{x | F(x) < 1\}$. H_γ is the so-called *generalized Pareto distribution (GPD) function*. Intuitively, the distribution function F is in the domain of attraction if and only if the excesses above a high threshold are asymptotically generalized Pareto distributed.

Smith (1987) introduced a maximum likelihood estimator (MLE) of the extreme value index by fitting the GPD with the empirical excesses. The MLE for the extreme value

[1] For the negative Hill estimator, it is proved for $-1 < \gamma < -1/2$.

index and the scale, $\hat{\gamma}_{ML}$ and $\hat{\sigma}_{ML}$, are obtained by solving the likelihood equations. The likelihood equations are (c.f. Drees *et al.* (2004))

$$\begin{aligned}\sum_{i=1}^{k} \frac{1}{\gamma^2} \log\Big(1+\frac{\gamma}{\sigma}(X_{n,n-i+1}-X_{n,n-k})\Big) & \\ -\left(\frac{1}{\gamma}+1\right)\frac{(1/\sigma)(X_{n,n-i+1}-X_{n,n-k})}{1+(\gamma/\sigma)(X_{n,n-i+1}-X_{n,n-k})} &= 0 \\ \sum_{i=1}^{k}\left(\frac{1}{\gamma}+1\right)\frac{(\gamma/\sigma)(X_{n,n-i+1}-X_{n,n-k})}{1+(\gamma/\sigma)(X_{n,n-i+1}-X_{n,n-k})} &= k,\end{aligned} \tag{3.3}$$

(the equations for $\gamma = 0$ are defined by continuity). Excluding $\gamma = 0$ as a solution, (3.3) can be simplified as

$$\begin{aligned}\frac{1}{k}\sum_{i=1}^{k} \log\Big(1+\frac{\gamma}{\sigma}(X_{n,n-i+1}-X_{n,n-k})\Big) &= \gamma \\ \frac{1}{k}\sum_{i=1}^{k}\frac{1}{1+(\gamma/\sigma)(X_{n,n-i+1}-X_{n,n-k})} &= \frac{1}{\gamma+1}.\end{aligned} \tag{3.4}$$

The equations are based on excesses $Y_i := X_{n,n-i+1} - X_{n,n-k}$, where $i = 1, \cdots, k$ and k is a suitable sequence of integers as in the Hill estimator.

For the MLE, because it is only given by solving the likelihood equations instead of an explicit formula, its asymptotic properties have to be proved in a different way. In case $\gamma > -1/2$, Smith (1987) sketched the proof of the consistency and asymptotic normality assuming a few extra conditions. Alternatively, Zhou (2008b) proved that the first order condition implies the consistency of the MLE for $\gamma > -1$. Drees *et al.* (2004) proved that the second order condition implies the asymptotic normality for $\gamma > -1/2$. In the proof, they assume that there exists a solution of the likelihood equation not too far off the real value. The existence of such a solution has been proved in Zhou (2008b). The combination of these two studies completes the proof of the asymptotic normality for the MLE in the case $\gamma > -1/2$.

Since the consistency is proved for $\gamma > -1$, a natural question arises: what are the asymptotic properties for $-1 < \gamma \leq -1/2$? In this chapter, we prove the asymptotic normality for $-1 < \gamma \leq -1/2$ under the second order condition.

The chapter is organized as follows. Section 3.2 sketches the idea of the proof and presents the main theorem. Section 3.3 provides the detail of the proof.

3.2 Main theorem

With the notation $U := \left(\frac{1}{1-F}\right)^{\leftarrow}$, the i.i.d random variables can be rewritten as $\{X_n\}_{n=1}^{\infty} \stackrel{d}{=} \{U(Z_n)\}_{n=1}^{\infty}$, where $\{Z_n\}_{n=1}^{\infty}$ are i.i.d. random variables with distribution function $1 -$

$1/x,\ x \geq 1$. Correspondingly, let $Z_{n,1} \leq Z_{n,2} \leq \cdots \leq Z_{n,n}$ be the order statistics of $Z_1, Z_2, \cdots, Z_n$. Then we can write $X_{n,n-i} = U(Z_{n,n-i})$ for $i = 0, 1, \cdots, k$.

To investigate the MLE, the starting point is how to solve the likelihood equations. Grimshaw (1993) discussed a numerical way as follows. From the equations (3.4), with the notation $Y_i := X_{n,n-i+1} - X_{n,n-k}$, it follows that,

$$\left(\frac{1}{k}\sum_{i=1}^{k}\log\left(1+\frac{\gamma}{\sigma}Y_i\right)+1\right)\cdot\frac{1}{k}\sum_{i=1}^{k}\frac{1}{1+(\gamma/\sigma)Y_i}=1. \tag{3.5}$$

In order to write this in short hand, denote the two parts in (3.5) as functions:

$$\begin{aligned}
f_n(t) &:= \frac{1}{k}\sum_{i=1}^{k}\log(1+tY_i)+1,\\
g_n(t) &:= \frac{1}{k}\sum_{i=1}^{k}\frac{1}{1+tY_i},\\
h_n(t) &:= f_n(t)g_n(t)-1.
\end{aligned}$$

Then, it is clear that any root $(\hat\gamma, \hat\sigma)$ of (3.4) satisfies $h_n(\hat\gamma/\hat\sigma) = 0$. Conversely, if t^* is a non-zero root of $h_n(t) = 0$, we obtain $(\hat\gamma, \hat\sigma) = (f_n(t^*) - 1, (f_n(t^*) - 1)/t^*)$ as the solution of (3.4). With this idea, the MLE can be calculated by the following procedure:
1. find the root t_n^* of $h_n(t) = 0$;
2. $\hat\gamma_{ML} = f_n(t_n^*) - 1$;
3. $\hat\sigma_{ML} = \hat\gamma_{ML}/t_n^*$.

The first step was solved in a numerical way in Grimshaw (1993). After that, the MLE of $(\hat\gamma_{ML}, \hat\sigma_{ML})$ will be calculated based on the numerical root of $h_n(t) = 0$.

Zhou (2008b) gives bounds for the root of $h_n(t) = 0$ for sufficiently large n. When $-1 < \gamma < 0$, we have that $U(\infty) < \infty$ is the finite right endpoint of F. It is stated in Zhou (2008b) that under only the first order condition, for any $\delta > 0$, when n is sufficiently large, the root t_n^* of $h_n(t) = 0$ lies between s_n and $t_n^{(\delta)}$ where

$$s_n := -\frac{1-1/k}{X_{n,n}-X_{n,n-k}}$$

and

$$t_n^{(\delta)} = -\frac{1-\delta}{U(\infty)-X_{n,n-k}}.$$

Under the second order condition, a stronger result can be obtained. Firstly, we refer to the inequality in Theorem 2.3.6, de Haan and Ferreira (2006) in case $-1 < \gamma < -1/2$ as the following proposition.

Proposition 3.2.1 *Suppose (3.2) holds for* $-1 < \gamma < -1/2$ *and* $\rho \le 0$. *Then there are functions* a_0 *and* A_0 *satisfying the following property: for any* $\varepsilon, \delta > 0$, *there exists* $t_0 = t_0(\varepsilon, \delta)$ *such that for all* t, $tx \ge t_0$,

$$\left| \frac{\frac{U(tx)-U(t)}{a_0(t)} - \frac{x^\gamma - 1}{\gamma}}{A_0(t)} - \Psi_{\gamma,\rho}(x) \right| \le \varepsilon x^{\gamma+\rho} \max(x^\delta, x^{-\delta}).$$

Here

$$\Psi_{\gamma,\rho}(x) = \begin{cases} \frac{x^{\gamma+\rho}-1}{\gamma+\rho}, & \rho < 0 \\ \frac{x^\gamma}{\gamma} \log x, & \rho = 0. \end{cases}$$

In the rest of this chapter, when we write A and a, we mean the specific choice A_0 and a_0.

We first locate the root of $h_n(t) = 0$. With $r_0 := U(\infty) - U(Z_{n,n-k})$, we introduce

$$f_n(-1/r_0) - 1 = \frac{1}{k} \sum_{i=1}^{k} \log \frac{U(\infty) - U(Z_{n,n-i+1})}{U(\infty) - U(Z_{n,n-k})} \tag{3.6}$$

as the *pseudo negative Hill estimator*, because we use the real endpoint $U(\infty)$ instead of its estimation $X_{n,n}$ as in $\hat{\gamma}_F$. Intuitively, since $X_{n,n}$ converges to $U(\infty)$ very fast in the case $\gamma < -1/2$, the pseudo negative Hill estimator should have the same asymptotic behavior as the negative Hill estimator $\hat{\gamma}_F$, as we shall prove later. Similarly, we have

$$g_n(-1/r_0) = \frac{1}{k} \sum_{i=1}^{k} \frac{U(\infty) - U(Z_{n,n-k})}{U(\infty) - U(Z_{n,n-i+1})}. \tag{3.7}$$

We shall prove that, the root of $h_n(t) = 0$ is not far off $-1/r_0$. In order to do that, we introduce the following sequences.

$$\begin{aligned} \delta_n &:= \frac{U(\infty) - U(Z_{n,n})}{U(\infty) - U(Z_{n,n-k})}, \\ q_n &:= \frac{U(Z_{n,n}) - U(Z_{n,n-k})}{1 - 1/k} = \frac{r_0}{1 - 1/k}(1 - \delta_n), \\ p_n &:= \frac{U(\infty) - U(Z_{n,n-k})}{1 - k^{\gamma+\varepsilon}} = \frac{r_0}{1 - k^{\gamma+\varepsilon}}, \end{aligned}$$

where ε is a fixed positive number.

The following proposition provides bounds for the location of the root of $h_n(t) = 0$.

Proposition 3.2.2 *For* $-1 < \gamma < -1/2$, *we choose a sufficiently small* $\varepsilon > 0$ *such that* $\gamma + \varepsilon < -1/2$. *Suppose the sequence* $k = k(n)$ *satisfies* $k \to \infty$, $k/n \to 0$ *and* $k^{-\gamma} A(n/k) \to 0$ *as* $n \to \infty$. *Under the second order condition with* $-1 < \gamma < -1/2$, *we have that for sufficiently large* n,

$$h_n\left(-\frac{1}{q_n}\right) > 0, \tag{3.8}$$

$$h_n\left(-\frac{1}{p_n}\right) < 0, \tag{3.9}$$

$$q_n < p_n. \tag{3.10}$$

Proposition 3.2.2 implies that eventually there exists a root t_n^* of $h_n(t) = 0$ lying between $-1/q_n$ and $-1/p_n$.

For the next step, we study the asymptotic properties of p_n and q_n. Denote Ω_γ as a random variable with distribution function $\exp\left(-(\gamma x)^{-1/\gamma}\right)$, for $\gamma x > 0$. Notice that Ω_1 follows the standard Frechét distribution, while for negative γ, Ω_γ has finite right endpoint 0. We study the asymptotic properties of δ_n, q_n and p_n in the following proposition.

Proposition 3.2.3 *Under the condition of Proposition 3.2.2, we have that as $n \to \infty$,*

$$k^{-\gamma}\frac{\delta_n}{\gamma} \xrightarrow{d} \Omega_\gamma, \tag{3.11}$$

$$k^{-\gamma}\left(\frac{q_n}{a(Z_{n,n-k})} + 1/\gamma\right) = O_p(1), \tag{3.12}$$

$$k^{-\gamma-\varepsilon}\left(\frac{p_n}{a(Z_{n,n-k})} + 1/\gamma\right) = O_p(1). \tag{3.13}$$

Since $-1/t_n^*$ is in between p_n and q_n, from Proposition 3.2.3, we get that

$$k^{-\gamma-\varepsilon}\left(\frac{\gamma/a(Z_{n,n-k})}{t_n^*} - 1\right) = O_p(1), \tag{3.14}$$

as $n \to \infty$. Finally, since $\hat{\gamma} = f_n(t_n^*) - 1$, the asymptotic normality of $\hat{\gamma}$ is proved by studying the asymptotic behavior of $f_n(-1/q_n) - 1$ and $f_n(-1/p_n) - 1$. To obtain the asymptotic normality of $\hat{\sigma}$, the proof starts from

$$\frac{\hat{\sigma}}{a(n/k)} = \frac{\hat{\gamma}/t_n^*}{a(n/k)} = \frac{\hat{\gamma}}{\gamma} \cdot \frac{\gamma/a(Z_{n,n-k})}{t_n^*} \cdot \frac{a(Z_{n,n-k})}{a(n/k)}. \tag{3.15}$$

Notice that as $n \to \infty$, the first and third factors go to 1 at speed $k^{1/2}$, the second factor goes to 1 at a faster speed $k^{-\gamma-\varepsilon}$. Since $\gamma + \varepsilon < -1/2$, we get that $\hat{\sigma}/a(n/k)$ goes to 1 at speed $k^{1/2}$ with a limit distribution dominated by limit distributions of the first and third factors. The final result is given as follows.

Theorem 3.2.1 *Suppose the second order condition holds for the extreme value index $-1 < \gamma \leq -1/2$. If the sequence $k = k(n)$ satisfies $k \to \infty$, $k/n \to 0$ and $k^{-\gamma}A(n/k) \to 0$ as $n \to \infty$, then for sufficiently large n, there exist a sequence of solution $(\hat{\gamma}_n, \hat{\sigma}_n)$ of the likelihood equations satisfying*

$$\sqrt{k}\left(\hat{\gamma}_n - \gamma, \frac{\hat{\sigma}_n}{a(n/k)} - 1\right) \xrightarrow{d} (W_1, W_2)$$

as $n \to \infty$, where $(W_1, W_2)^T$ follows a two-dimensional normal distribution with mean $(0,0)^T$ and covariance matrix

$$\begin{pmatrix} \gamma^2 & \gamma \\ \gamma & 1+\gamma^2 \end{pmatrix}.$$

Theorem 3.2.1 contains the case $\gamma = -1/2$. The proof of this case is similar to $-1 < \gamma < -1/2$ with some minor changes. It will be sketched in Subsection 3.3.3 without providing details.

Remark 3.2.1 *For $\gamma > -1/2$, Drees* et al. *(2004) provided the covariance matrix of the limit law for the maximum likelihood estimator as*

$$\begin{pmatrix} (1+\gamma)^2 & -(1+\gamma) \\ -(1+\gamma) & 1+(1+\gamma)^2 \end{pmatrix}.$$

Together with the covariance matrix for $-1 < \gamma \leq -1/2$, we observe that the asymptotic variances of the shape and scale components and the asymptotic covariance between them are all continuous functions of γ. However, they are not differentiable at the point $\gamma = -1/2$.

3.3 Proof

Proposition 3.2.2 locates the solution of the likelihood equation. Hence, if the speeds of convergence and the limit laws of the left and right boundaries are the same, the MLE must have the same asymptotic behavior. Since the boundaries have explicit representation, we can directly study their asymptotic properties. They are presented and proved as lemmas in Subsection 3.3.1. Section 3.3.2 gives the proof of the propositions and Theorem 3.2.1 for $-1 < \gamma < -1/2$. Section 3.3.3 sketches the proof of the case $\gamma = -1/2$.

3.3.1 Lemmas

First of all, we state Lemma 2.4.10 in de Haan and Ferreira (2006) here as follows.

Lemma 3.3.1 *With the notation $\{Z_{n,i}\}_{i=1}^n$ as defined in the beginning of Section 3.2 and a sequence $k = k(n)$ such that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$, there exists a sequence of Brownian motions $\{W_n(s)\}_{s\geq 0}$ such that for each $\varepsilon > 0$,*

$$\sup_{k^{-1}\leq s\leq 1} s^{\gamma+1/2+\varepsilon} \left| \sqrt{k}\left(\frac{\left(\frac{k}{n}Z_{n,n-[ks]}\right)^{\gamma}-1}{\gamma} - \frac{s^{-\gamma}-1}{\gamma} \right) - s^{-\gamma-1}W_n(s) \right| = o_p(1). \quad (3.16)$$

Using the notation in Lemma 3.3.1, we define

$$W_3 = \int_0^1 s^{-1}W_n(s) - W_n(1)ds$$

and $W_4 = W_n(1)$. One can see that both W_3 and W_4 follow a standard normal distribution and they are independent. Notice that γW_3 is the limit law of the negative Hill as presented in the following lemma.

Lemma 3.3.2 *Assume the conditions on the sequence k in Proposition 3.2.2 hold. Under the second order condition with $-1 < \gamma < -1/2$, the negative Hill estimator $\hat{\gamma}_F$ satisfies*

$$\sqrt{k}(\hat{\gamma}_F - \gamma) \xrightarrow{d} \gamma W_3,$$

as $n \to \infty$.

Proof of Lemma 3.3.2

Since $-1 < \gamma < -1/2$, $k^{-\gamma} A(n/k) \to 0$ implies that $k^{1/2} A(n/k) \to 0$. Then from Theorem 3.6.4 in de Haan and Ferreira (2006), we get the result in the lemma. □

W_4 plays a role as the limit law for the scale function as shown the following lemma.

Lemma 3.3.3 *Suppose the sequence k satisfies the conditions in Proposition 3.2.2. Under the second order condition, we have that*

$$\sqrt{k}\left(\frac{a(Z_{n,n-k})}{a(n/k)} - 1\right) \xrightarrow{d} \gamma W_4,$$

as $n \to \infty$.

Proof of Lemma 3.3.3

From the second order condition, according to Theorem 2.3.6 in de Haan and Ferreira (2006), we get that for any $\varepsilon, \delta > 0$, there exists t_0 such that for all $t, tx \geq t_0$,

$$\left| \frac{\frac{a(tx)}{a(t)} - x^\gamma}{A(t)} - x^\gamma \frac{x^\rho - 1}{\rho} \right| \leq \varepsilon x^{\gamma+\rho} \max(x^\delta, x^{-\delta}). \tag{3.17}$$

We apply this with $t = n/k$ and $x = \frac{Z_{n,n-k}}{n/k}$. Notice that as $n \to \infty$ $Z_{n,n-k} \to \infty$, and $x \xrightarrow{P} 1$. Moreover,

$$\lim_{x \to 1} x^\gamma \frac{x^\rho - 1}{\rho} \pm \varepsilon x^{\gamma+\rho} \max(x^\delta, x^{-\delta}) = \pm\varepsilon.$$

We get that

$$\sqrt{k}\left(\frac{a(Z_{n,n-k})}{a(n/k)} - 1\right) = \sqrt{k}\left(\left(\frac{Z_{n,n-k}}{n/k}\right)^\gamma - 1\right) + \sqrt{k} A(n/k) O_p(1).$$

By taking $s = 1$ in (3.16), and considering that $k^{1/2} A(n/k) \to 0$, the lemma is proved. □

As discussed in Section 3.2, the asymptotic behavior of the pseudo negative Hill estimator is the same as that of the negative Hill estimator. This is shown in the following lemma.

Lemma 3.3.4 *Assume the conditions of the sequence k in Proposition 3.2.2 hold. Under the second order condition with $-1 < \gamma < -1/2$, we have that*

$$\sqrt{k}\left(f_n(-1/r_0) - 1 - \gamma\right) \xrightarrow{d} \gamma W_3$$

as $n \to \infty$.

Proof of Lemma 3.3.4

Recall (3.6) for the definition of $f_n(-1/r_0)$. We start with the following relation from Lemma 4.5.4 in de Haan and Ferreira (2006),

$$\lim_{t\to\infty} \frac{\frac{U(\infty)-U(t)}{a(t)} + \frac{1}{\gamma}}{A(t)} = \frac{1}{\gamma(\gamma+\rho)}.$$

We rewrite it as

$$U(\infty) - U(t) = a(t)\left(-\frac{1}{\gamma} + A(t)\frac{1}{\gamma(\gamma+\rho)}(1+o(1))\right).$$

Thus, we have that, for fixed x as $t \to \infty$

$$\frac{U(\infty)-U(tx)}{U(\infty)-U(t)} = \frac{a(tx)}{a(t)} \frac{1-\gamma A(tx)\frac{1}{\gamma(\gamma+\rho)}(1+o(1))}{1-\gamma A(t)\frac{1}{\gamma(\gamma+\rho)}(1+o(1))}.$$

The inequality (3.17) implies that

$$\frac{a(tx)}{a(t)} = x^\gamma + A(t)x^\gamma\frac{x^\rho-1}{\rho}(1+o(1)).$$

for any fixed $x > 0$. Thus, we get that

$$\begin{aligned}
\frac{U(\infty)-U(tx)}{U(\infty)-U(t)} &= x^\gamma\left(1 + A(t)\frac{x^\rho-1}{\rho}(1+o(1))\right)\frac{1-\gamma A(tx)\frac{1}{\gamma(\gamma+\rho)}(1+o(1))}{1-\gamma A(t)\frac{1}{\gamma(\gamma+\rho)}(1+o(1))} \\
&= x^\gamma\left\{1 + A(t)\frac{x^\rho-1}{\rho}(1+o(1)) - \gamma A(tx)\frac{1}{\gamma(\gamma+\rho)}(1+o(1))\right. \\
&\quad \left. + \gamma A(t)\frac{1}{\gamma(\gamma+\rho)}(1+o(1))\right\}(1+o(1)) \\
&= x^\gamma + x^\gamma A(t)\left(\frac{x^\rho-1}{\rho} - \gamma\frac{A(tx)}{A(t)}\frac{1}{\gamma(\gamma+\rho)} + \gamma\frac{1}{\gamma(\gamma+\rho)}\right)(1+o(1)) \\
&= x^\gamma + x^\gamma A(t)\left(\frac{x^\rho-1}{\rho} - \frac{A(tx)}{A(t)}\frac{1}{\gamma+\rho} + \frac{1}{\gamma+\rho}\right)(1+o(1)) \\
&= x^\gamma + x^\gamma A(t)\frac{x^\rho-1}{\rho}\left(1 - \frac{\rho}{\gamma+\rho}\right)(1+(1)) \\
&= x^\gamma + x^\gamma A(t)\frac{x^\rho-1}{\rho}\frac{\gamma}{\gamma+\rho}(1+o(1))
\end{aligned}$$

Denote $\tilde{A}(t) = A(t)\frac{\gamma}{\gamma+\rho}$. Now, we have that for any fixed $x > 0$,

$$\lim_{t\to\infty} \frac{\frac{U(\infty)-U(tx)}{U(\infty)-U(t)} - x^{\gamma}}{\tilde{A}(t)} = x^{\gamma}\frac{x^{\rho}-1}{\rho}.$$

Therefore, according to Theorem 2.3.9 in de Haan and Ferreira (2006)), for any $\varepsilon, \delta > 0$, there exists $t_1 > 0$ such that for all $t, tx > t_1$

$$\left| \frac{\frac{U(\infty)-U(tx)}{U(\infty)-U(t)} - x^{\gamma}}{\tilde{A}(t)} - x^{\gamma}\frac{x^{\rho}-1}{\rho} \right| \leq \varepsilon x^{\gamma+\rho} \max(x^{\delta}, x^{-\delta}). \tag{3.18}$$

Since $Z_{n,n-[ks]} \to \infty$ as $n \to \infty$ uniformly for all $s \in [0,1]$, we get that

$$\begin{aligned} \left(\frac{Z_{n,n-[ks]}}{n/k}\right)^{\gamma} + \tilde{A}(n/k)R_1\left(\frac{Z_{n,n-[ks]}}{n/k}\right) &\leq \frac{U(\infty) - U(Z_{n,n-[ks]})}{U(\infty) - U(n/k)} \\ &\leq \left(\frac{Z_{n,n-[ks]}}{n/k}\right)^{\gamma} + \tilde{A}(n/k)R_2\left(\frac{Z_{n,n-[ks]}}{n/k}\right), \end{aligned} \tag{3.19}$$

where $R_1(x) := x^{\gamma}\frac{x^{\rho}-1}{\rho} - \varepsilon x^{\gamma+\rho}\max(x^{\delta}, x^{-\delta})$ and $R_2(x) := x^{\gamma}\frac{x^{\rho}-1}{\rho} + \varepsilon x^{\gamma+\rho}\max(x^{\delta}, x^{-\delta})$. Notice that both R_1 and R_2 are bounded on $[1/2, \infty)$. Since $\frac{Z_{n,n-[ks]}}{n/k} \geq \frac{Z_{n,n-k}}{n/k} \xrightarrow{P} 1$ as $n \to \infty$ which implies that $\frac{Z_{n,n-[ks]}}{n/k}$ is eventually larger than 1/2. Combining (3.16), (3.19) and the fact that $k^{1/2}\tilde{A}(n/k) \to 0$, we have that for any $\varepsilon > 0$,

$$\sqrt{k}\left(\frac{U(\infty) - U(Z_{n,n-[ks]})}{U(\infty) - U(n/k)} - s^{-\gamma}\right) = \gamma s^{-\gamma-1}W_n(s) + o_p(1)s^{-\gamma-1/2-\varepsilon} \tag{3.20}$$

holds uniformly for $s \in [1/k, 1]$. Thus

$$\log\left(s^{\gamma}\frac{U(\infty) - U(Z_{n,n-[ks]})}{U(\infty) - U(n/k)}\right) = \log\left(1 + k^{-1/2}\left(\gamma s^{-1}W_n(s) + o_p(1)s^{-1/2-\varepsilon}\right)\right).$$

Since

$$\limsup_{s\to 0}\left|\frac{W_n(s)}{s^{1/2-\varepsilon}}\right| < \infty \quad a.s.$$

we have that $k^{-1/2}\left(\gamma s^{-1}W_n(s) + o_p(1)s^{-1/2-\varepsilon}\right) = o_p(1)$ holds uniformly for $s \in [k^{-1+3\varepsilon}, 1]$. Hence

$$\log\left(s^{\gamma}\frac{U(\infty) - U(Z_{n,n-[ks]})}{U(\infty) - U(n/k)}\right) = k^{-1/2}\left(\gamma s^{-1}W_n(s) + o_p(1)s^{-1/2-\varepsilon}\right)(1 + o_p(1)).$$

holds uniformly for $s \in [k^{-1+3\varepsilon}, 1]$. Therefore, we get the following relation,

$$\begin{aligned} \sqrt{k}\int_{k^{-1+3\varepsilon}}^{1} \log\left(s^{\gamma}\frac{U(\infty) - U(Z_{n,n-[ks]})}{U(\infty) - U(n/k)}\right) ds &= \int_{k^{-1+3\varepsilon}}^{1} \gamma s^{-1}W_n(s)ds + o_p(1) \\ &= \int_0^1 \gamma s^{-1}W_n(s)ds + o_p(1) \end{aligned}$$

Notice that

$$0 \geq \sqrt{k}\int_0^{k^{-1+3\varepsilon}} \log\left(\frac{U(\infty)-U(Z_{n,n-[ks]})}{U(\infty)-U(n/k)}\right) ds \geq k^{-1/2+3\varepsilon}\log\left(\frac{U(\infty)-U(Z_{n,n})}{U(\infty)-U(n/k)}\right).$$

By taking $s=0$ in (3.19) and considering that $Z_{n,n}/n = O_p(1)$, we get $\left(\frac{U(\infty)-U(Z_{n,n})}{U(\infty)-U(n/k)}\right) = k^{-\gamma}O_p(1)$ Hence the term $\sqrt{k}\int_0^{k^{-1+3\varepsilon}} \log\left(\frac{U(\infty)-U(Z_{n,n-[ks]})}{U(\infty)-U(n/k)}\right) ds$ is negligible.

By verifying that $\sqrt{k}\int_0^{k^{-1+3\varepsilon}} \log s ds \to 0$ as $n \to \infty$, we conclude that

$$\sqrt{k}\left(\int_0^1 \log\left(\frac{U(\infty)-U(Z_{n,n-[ks]})}{U(\infty)-U(n/k)}\right) ds - \gamma\right) = \int_0^1 \gamma s^{-1} W_n(s) ds + o_p(1).$$

From (3.19), similar to Lemma 3.3.3, we get that

$$\sqrt{k}\log\left(\frac{U(\infty)-U(Z_{n,n-k})}{U(\infty)-U(n/k)}\right) = \gamma W_n(1) + o_p(1).$$

Hence

$$\sqrt{k}\left(f_n(-1/r_0) - 1 - \gamma\right) \xrightarrow{d} \gamma W_3$$

is proved. $\square$.

The last lemma studies the asymptotic behavior of $g_n(-1/r_0)$.

Lemma 3.3.5 *Assume the conditions of the sequence k in Proposition 3.2.2 hold. Under the second order condition with $-1 < \gamma < -1/2$, we have that as $n \to \infty$*

$$k^{\gamma+1}\left(g_n(-1/r_0) - \frac{1}{1+\gamma}\right) \xrightarrow{d} S_1,$$

where S_1 is a random variable with stable distribution.

Proof of Lemma 3.3.5

Recall (3.7) for the definition of $g_n(-1/r_0)$. Similar to the proof of Lemma 3.3.4, considering (3.18) and $k^{-\gamma}\tilde{A}(Z_{n,n-k}) \to 0$, we get that for $1 \leq i \leq k$

$$\frac{U(\infty)-U(Z_{n,n-i+1})}{U(\infty)-U(Z_{n,n-k})} \leq \left(\frac{Z_{n,n-i+1}}{Z_{n,n-k}}\right)^{\gamma}\left(1+\tilde{A}(Z_{n,n-k})O_p(1)\right) = \left(\frac{Z_{n,n-i+1}}{Z_{n,n-k}}\right)^{\gamma}\left(1+k^{\gamma}o_p(1)\right).$$

A similar inequality holds for the lower bound of $\frac{U(\infty)-U(Z_{n,n-i+1})}{U(\infty)-U(Z_{n,n-k})}$. Hence as $n \to \infty$, we have the following relation uniformly for all $1 \leq i \leq k$:

$$\frac{U(\infty)-U(Z_{n,n-i+1})}{U(\infty)-U(Z_{n,n-k})} = \left(\frac{Z_{n,n-i+1}}{Z_{n,n-k}}\right)^{\gamma}(1+o_p(1)k^{\gamma}).$$

Hence

$$g_n(-1/r_0) = \frac{1}{k}\sum_{i=1}^{k}\frac{U(\infty)-U(Z_{n,n-k})}{U(\infty)-U(Z_{n,n-i+1})} = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{Z_{n,n-i+1}}{Z_{n,n-k}}\right)^{-\gamma}(1+o_p(1)k^{\gamma}).$$

Denote $V_i = \frac{Z_{n,n-i+1}}{Z_{n,n-k}}$ for $1 \le i \le k$. Then V_i can be recognized as the order statistics of a sample generated from the standard Pareto distribution $F(x) = 1 - 1/x, x > 1$, with sample size k. From the central limit theory with stable limit law, we know that

$$k^{\gamma+1}\left(\frac{1}{k}\sum_{i=1}^{k} V_i^{-\gamma} - \frac{1}{\gamma+1}\right) \xrightarrow{d} S_1,$$

where S_1 is a random variable with a stable distribution, index $\alpha = -1/\gamma$.

Since $\gamma < -1/2$, we have that $k^{\gamma+1}(k^{\gamma}o_p(1)) = k^{2\gamma+1}o_p(1) \to 0$ as $n \to \infty$. Therefore, we complete the proof of the lemma. □

3.3.2 Proof of Theorem 3.2.1 with $-1 < \gamma < -1/2$

We first prove Proposition 3.2.3, and then Proposition 3.2.2. In the end, we shall prove the main theorem.

Proof of Proposition 3.2.3

Notice that as $n \to \infty$, $\frac{Z_{n,n}}{n} \xrightarrow{d} \Omega_1$ where Ω_1 is the standard Fréchet distribution. Together with the fact that $\frac{Z_{n,n-k}}{n/k} \xrightarrow{P} 1$, we get that as $n \to \infty$, $\frac{Z_{n,n}}{kZ_{n,n-k}} \xrightarrow{d} \Omega_1$, which implies that

$$\frac{1}{\gamma}\left(\frac{Z_{n,n}}{kZ_{n,n-k}}\right)^{\gamma} \xrightarrow{d} \Omega_{\gamma}.$$

From (3.18), since the functions R_i are bounded and $k^{-\gamma}\tilde{A}(n/k) \to 0$, we get that

$$\delta_n = \frac{U(\infty) - U(Z_{n,n})}{U(\infty) - U(Z_{n,n-k})} = \left(\frac{Z_{n,n}}{Z_{n,n-k}}\right)^{\gamma} + o_p(1)k^{\gamma}.$$

Hence, the limit distribution of $k^{-\gamma}\frac{\delta_n}{\gamma}$ is the same as the limit distribution of $\frac{1}{\gamma}\left(\frac{Z_{n,n}}{kZ_{n,n-k}}\right)^{\gamma}$. Thus, (3.11) is proved.

Because of the negative γ, we can take $a(t) = -\gamma(U(\infty) - U(t))$. Then we rewrite q_n and p_n as

$$\frac{q_n}{a(Z_{n,n-k})} = \frac{1}{-\gamma}\frac{1}{1-1/k}(1-\delta_n),$$

and

$$\frac{p_n}{a(Z_{n,n-k})} = \frac{1}{-\gamma}\frac{1}{1-k^{\gamma+\varepsilon}}.$$

Therefore, the asymptotic behavior of δ_n directly leads to the asymptotic relation (3.12), while (3.13) is obvious. □

Proof of Proposition 3.2.2

From Proposition 3.2.3, we have that as $n \to \infty$, $\delta_n \to 0$ and $k\delta_n \to \infty$ Furthermore, since $\varepsilon < -\gamma - 1/2 < -\gamma$, we have that as $n \to \infty$, $k^{\gamma+\varepsilon} \to 0$ and $k^{1+\gamma+\varepsilon} \to \infty$. Therefore, eventually, $(1 - k^{\gamma+\varepsilon})(1 - \delta_n) - 1 \sim -k^{\gamma+\varepsilon} - \delta_n < -1/k$, which leads to

$$\frac{1}{1 - 1/k}(1 - \delta_n) < \frac{1}{1 - k^{\gamma+\varepsilon}}.$$

It implies that eventually, $q_n < p_n$ as stated in (3.10).

To prove the positivity of $h_n(-1/q_n)$, we start from the following inequality.

$$\begin{aligned}
f_n(-1/q_n) &= 1 + \frac{1}{k}\sum_{i=1}^{k} \log\left(1 - \frac{X_{n,n-i+1} - X_{n,n-k}}{q_n}\right) \\
&= 1 + \frac{-\log k}{k} + \frac{1}{k}\sum_{i=2}^{k} \log\left(1 - \frac{X_{n,n-i+1} - X_{n,n-k}}{q_n}\right) \\
&\geq 1 + \frac{-\log k}{k} + \frac{1}{k}\sum_{i=2}^{k} \log\left(1 - \frac{X_{n,n-i+1} - X_{n,n-k}}{X_{n,n} - X_{n,n-k}}\right) \\
&= 1 + \frac{-\log k}{k} + \frac{1}{k}\sum_{i=2}^{k} \log \frac{X_{n,n} - X_{n,n-i+1}}{X_{n,n} - X_{n,n-k}} \\
&:= 1 + I_1 + I_2.
\end{aligned}$$

Notice that I_2 is exactly the negative Hill estimator $\hat{\gamma}_F$ and $I_1 \to 0$ at a speed higher than $k^{1/2}$. We get that

$$\liminf_{n\to\infty} f_n(-1/q_n) \geq 1 + \gamma. \tag{3.21}$$

For $g_n(-1/q_n)$, the calculation is similar. Suppose $1/k < \Delta$ for a fixed $\Delta > 0$, then

$$\begin{aligned}
g_n(-1/q_n) &= \frac{1}{k}\sum_{i=1}^{k} \frac{1}{1 - \frac{X_{n,n-i+1} - X_{n,n-k}}{q_n}} \\
&= 1 + \frac{1}{k}\sum_{i=2}^{k} \frac{1}{1 - \frac{X_{n,n-i+1} - X_{n,n-k}}{q_n}} \\
&\geq 1 + \frac{1}{k}\sum_{i=2}^{k} \frac{1}{1 - (1-\Delta)\frac{X_{n,n-i+1} - X_{n,n-k}}{X_{n,n} - X_{n,n-k}}} \\
&\geq 1 + \frac{1}{k}\sum_{i=2}^{k} \frac{1}{\frac{X_{n,n} - X_{n,n-i+1}}{X_{n,n} - X_{n,n-k}} + \Delta} \\
&:= 1 + I_3.
\end{aligned}$$

Notice that $\frac{X_{n,n} - X_{n,n-i+1}}{X_{n,n} - X_{n,n-k}}$ is bounded as

$$\frac{U(\infty) - X_{n,n-i+1}}{U(\infty) - X_{n,n-k}} - \delta_n < \frac{X_{n,n} - X_{n,n-i+1}}{X_{n,n} - X_{n,n-k}} < \frac{U(\infty) - X_{n,n-i+1}}{U(\infty) - X_{n,n-k}}.$$

From (3.20), we get that, as $n \to \infty$,

$$\frac{U(\infty) - X_{n,n-[ks]}}{U(\infty) - X_{n,n-k}} \to s^{-\gamma}$$

holds uniformly for $s \in [1/k, 1]$. Considering that $\delta_n \to 0$, we have

$$\begin{aligned} I_3 &= \int_{1/k}^{1} \frac{ds}{\frac{X_{n,n}-X_{n,n-[ks]}}{X_{n,n}-X_{n,n-k}} + \Delta} \\ &\geq \int_{1/k}^{1} \frac{ds}{\frac{U(\infty)-X_{n,n-[ks]}}{U(\infty)-X_{n,n-k}} + \Delta} \\ &\to \int_{0}^{1} \frac{ds}{s^{-\gamma} + \Delta} \end{aligned}$$

and

$$\begin{aligned} I_3 &= \int_{1/k}^{1} \frac{ds}{\frac{X_{n,n}-X_{n,n-[ks]}}{X_{n,n}-X_{n,n-k}} + \Delta} \\ &\leq \int_{1/k}^{1} \frac{ds}{\frac{U(\infty)-X_{n,n-[ks]}}{U(\infty)-X_{n,n-k}} - \delta_n + \Delta} \\ &\leq \int_{1/k}^{1} \frac{ds}{\frac{U(\infty)-X_{n,n-[ks]}}{U(\infty)-X_{n,n-k}} + \Delta} + \frac{\delta_n}{\Delta^2} \\ &\to \int_{0}^{1} \frac{ds}{s^{-\gamma} + \Delta}. \end{aligned}$$

Thus it is proved that $I_3 \to \int_0^1 \frac{dt}{t^{-\gamma}+\Delta}$. For any $\varepsilon > 0$, by taking Δ small enough, $\liminf_{n\to\infty} I_3 > \frac{1}{\gamma+1} - \varepsilon$. Therefore,

$$\liminf_{n\to\infty} g_n(-1/q_n) \geq 1 + \frac{1}{\gamma+1} - \varepsilon. \tag{3.22}$$

Combining (3.21) and (3.22), we get that

$$\liminf_{n\to\infty} h_n(-1/q_n) \geq (1+\gamma)\left(1 + \frac{1}{\gamma+1} - \varepsilon\right) - 1 = (1+\gamma)(1-\varepsilon) > 0,$$

which completes the proof of (3.8).

Now we turn to prove (3.9). Since $-1/p_n = -(1-k^{\gamma+\varepsilon})\frac{1}{r_0}$, $f_n(-1/p_n)$ can be bounded as

$$\begin{aligned} f_n(-1/p_n) &= f_n\left(-\frac{1}{r_0} + \frac{k^{\gamma+\varepsilon}}{r_0}\right) \\ &= f_n\left(-\frac{1}{r_0}\right) + \frac{1}{k}\sum_{i=1}^{k} \log\left(\frac{1 - \frac{1}{p_n}(X_{n,n-i+1} - X_{n,n-k})}{1 - \frac{1}{r_0}(X_{n,n-i+1} - X_{n,n-k})}\right) \end{aligned}$$

$$
\begin{aligned}
&= f_n\left(-\frac{1}{r_0}\right) + \frac{1}{k}\sum_{i=1}^{k}\log\left(1 + \frac{\frac{k^{\gamma+\varepsilon}}{r_0}(X_{n,n-i+1} - X_{n,n-k})}{1 - \frac{1}{r_0}(X_{n,n-i+1} - X_{n,n-k})}\right) \\
&\le f_n\left(-\frac{1}{r_0}\right) + \frac{1}{k}\sum_{i=1}^{k}\frac{\frac{k^{\gamma+\varepsilon}}{r_0}(X_{n,n-i+1} - X_{n,n-k})}{1 - \frac{1}{r_0}(X_{n,n-i+1} - X_{n,n-k})} \\
&= f_n\left(-\frac{1}{r_0}\right) + k^{\gamma+\varepsilon}\left(g_n\left(-\frac{1}{r_0}\right) - 1\right). \qquad (3.23)
\end{aligned}
$$

Next, for $g_n(-1/p_n)$, we rewrite it as

$$
\begin{aligned}
g_n(-1/p_n) &= \frac{1}{k}\sum_{i=1}^{k}\frac{1}{1 - \frac{1}{p_n}(X_{n,n-i+1} - X_{n,n-k})} \\
&= \int_0^1 \frac{1}{1 - \frac{1}{p_n}(X_{n,n-[sk]+1} - X_{n,n-k})}ds \\
&= \int_0^1 \frac{1}{1 - \frac{1}{r_0}(X_{n,n-[sk]+1} - X_{n,n-k})} \cdot \frac{1}{1 + \frac{\frac{k^{\gamma+\varepsilon}}{r_0}(X_{n,n-[ks]+1} - X_{n,n-k})}{1 - \frac{1}{r_0}(X_{n,n-[ks]+1} - X_{n,n-k})}}ds \\
&:= \int_0^1 \frac{1}{1 - \frac{1}{r_0}(X_{n,n-[sk]+1} - X_{n,n-k})} \cdot \frac{1}{1 + \theta_n(s)}ds,
\end{aligned}
$$

where

$$
\theta_n(s) = \frac{\frac{k^{\gamma+\varepsilon}}{r_0}(X_{n,n-[ks]+1} - X_{n,n-k})}{1 - \frac{1}{r_0}(X_{n,n-[ks]+1} - X_{n,n-k})}.
$$

Denote $s_k = 2k^{-1-\varepsilon/\gamma} > 1/k$. From (3.18), we get that

$$
\lim_{n\to\infty} k^{-\gamma-\varepsilon}\left(1 - \frac{X_{n,n-[s_k k]+1} - X_{n,n-k}}{r_0}\right) = \lim_{n\to\infty} k^{-\gamma-\varepsilon}(s_k)^{-\gamma} = 2^{-\gamma} > 1,
$$

which implies that

$$
\lim_{n\to\infty}\frac{X_{n,n-[s_k k]+1} - X_{n,n-k}}{r_0} = 1.
$$

Thus, for sufficiently large n,

$$
k^{-\gamma-\varepsilon}\left(1 - \frac{X_{n,n-[sk]+1} - X_{n,n-k}}{r_0}\right) \ge \frac{X_{n,n-[sk]+1} - X_{n,n-k}}{r_0} \Rightarrow \theta_n(s) \le 1
$$

holds for all $s_k \le s < 1$.

Notice that $1/(1+x) < 1$ for all $x > 0$ and $1/(1+x) < 1 - x/2$ for $0 < x \le 1$. We have that

$$
\begin{aligned}
g_n(-1/p_n) \le& \int_0^{s_k}\frac{1}{1 - \frac{1}{r_0}(X_{n,n-[sk]+1} - X_{n,n-k})}ds \\
&+ \int_{s_k}^1 \frac{1}{1 - \frac{1}{r_0}(X_{n,n-[sk]+1} - X_{n,n-k})} \cdot \left(1 - \frac{\theta_n(s)}{2}\right)ds
\end{aligned}
$$

$$= g_n(-1/r_0) - \frac{1}{2}\int_{s_k}^{1} \frac{1}{1-\frac{1}{r_0}(X_{n,n-[sk]+1}-X_{n,n-k})}\theta_n(s)ds$$

$$= g_n(-1/r_0) - \frac{k^{\gamma+\varepsilon}}{2}\int_{s_k}^{1} \frac{\frac{1}{r_0}(X_{n,n-[sk]+1}-X_{n,n-k})}{\left(1-\frac{1}{r_0}(X_{n,n-[sk]+1}-X_{n,n-k})\right)^2}ds$$

$$:= g_n(-1/r_0) - \frac{k^{\gamma+\varepsilon}}{2}I_4$$

From (3.20), we have that as $n\to\infty$,

$$s^{\gamma}\left(1-\frac{1}{r_0}(X_{n,n-[sk]+1}-X_{n,n-k})\right) = s^{\gamma}\left(\frac{U(\infty)-X_{n,n-[sk]+1}}{U(\infty)-X_{n,n-k}}\right) = 1+o_p(1)$$

holds uniformly for all $s\in[s_k,1]$. Hence, we get that

$$I_4 = \int_{s_k}^{1}\frac{1-s^{-\gamma}}{(s^{-2\gamma})}ds + o_p(1) = s_k^{2\gamma+1} + o_p(1) = k^{(-1-\varepsilon/\gamma)(2\gamma+1)}O_p(1).$$

Now we turn to $h_n(-1/p_n)$. By employing the boundaries of $f_n(-1/p_n)$ and $g_n(-1/p_n)$, we have

$$h_n(-1/p_n) \le \left(f_n\left(-\frac{1}{r_0}\right)g_n\left(-\frac{1}{r_0}\right)-1\right)$$
$$- k^{\gamma+\varepsilon}\frac{I_4}{2}f_n\left(-\frac{1}{r_0}\right) + k^{\gamma+\varepsilon}\left(g_n\left(-\frac{1}{r_0}\right)-1\right)g_n\left(-\frac{1}{r_0}\right)$$
$$:= I_5 - I_6 + I_7.$$

First consider I_5. From Lemma 3.3.4 and Lemma 3.3.5, since $\gamma+1<1/2$, we get $k^{\gamma+1}I_5 = O_p(1)$. Next consider I_6. The asymptotic property of I_4 ensures that

$$k^{\gamma+1}I_6 = k^{\gamma+1}k^{\gamma+\varepsilon}k^{(-1-\varepsilon/\gamma)(2\gamma+1)}O_p(1) = k^{-\varepsilon(\gamma+1)/\gamma}O_p(1)\to\infty.$$

Thirdly, consider I_7. Since $2\gamma+1+\varepsilon<\gamma+1/2+\varepsilon<0$, I_7 has the following asymptotic property

$$k^{\gamma+1}I_7 = k^{\gamma+1}k^{\gamma+\varepsilon}O_p(1) = k^{2\gamma+1+\varepsilon}O_p(1) = o_p(1).$$

Finally, combining all these three parts, it is proved that $k^{\gamma+1}h_n(-1/p_n)\to-\infty$ as $n\to\infty$, which implies (3.9). □

Proof of Theorem 3.2.1

From the proof of Proposition 3.2.2 and Lemma 3.3.2, we get that

$$f_n\left(-\frac{1}{q_n}\right) - 1 - \gamma \ge I_1 + (I_2-\gamma),$$

where $I_1 = o_p(k^{-1/2})$ and $\sqrt{k}(I_2-\gamma)\xrightarrow{d}\gamma W_3$ as $n\to\infty$. Meanwhile, combining (3.23) and Lemma 3.3.4, we have that

$$f_n\left(-\frac{1}{p_n}\right) - 1 - \gamma \le \left(f_n\left(-\frac{1}{r_0}\right)-1-\gamma\right) + k^{\gamma+\varepsilon}O_p(1),$$

where $k^{\gamma+\varepsilon}O_p(1) = o_p(k^{-1/2})$ and $\sqrt{k}(f_n(-1/r_0) - 1 - \gamma) \xrightarrow{d} \gamma W_3$ as $n \to \infty$.

Since the root t_n^* lies between $-1/q_n$ and $-1/p_n$, and f_n is an increasing function, we get that $\sqrt{k}(f_n(t_n^*) - 1 - \gamma) \xrightarrow{d} \gamma W_3$, i.e. $\sqrt{k}(\hat{\gamma} - \gamma) \xrightarrow{d} \gamma W_3 := W_1$ as $n \to \infty$.

For the scale estimation, considering (3.15), the asymptotic normality of $\hat{\gamma}$ and Lemma 3.3.3, the limit law of the scale estimation is given as

$$\begin{aligned}\sqrt{k}\left(\frac{\hat{\sigma}}{a(n/k)} - 1\right) &= \sqrt{k}\left(\frac{\hat{\gamma}}{\gamma} - 1\right) + \sqrt{k}\left(\frac{a(Z_{n,n-k})}{a(n/k)} - 1\right) + o_p(1) \\ &\xrightarrow{d} W_3 + \gamma W_4 \\ &:= W_2,\end{aligned}$$

as $n \to \infty$. Thus, it is proved that

$$\sqrt{k}\left(\hat{\gamma}_n - \gamma, \frac{\hat{\sigma}}{a(n/k)} - 1\right) \xrightarrow{d} (W_1, W_2)$$

as $n \to \infty$. Here $(W_1, W_2)^T$ follows the two-dimensional normal distribution with mean $(0, 0)^T$ and covariance matrix

$$\begin{pmatrix} \gamma^2 & \gamma \\ \gamma & 1+\gamma^2 \end{pmatrix}. \square$$

Remark 3.3.1 *In Theorem 3.2.1, we requires that $k^{-\gamma}A(n/k) \to 0$ as $n \to \infty$ for simplicity. In fact, when $-1 < \gamma < -1/2$, if the condition is relaxed as $k^{-\gamma}A(n/k) \to \lambda$ as $n \to \infty$ for some real number λ, we will get a bias part for S_1 defined in Lemma 3.5, and a bias part for the limit of $k^{-\gamma}\frac{\delta_n}{\gamma}$ as in (3.11). However, both of them are still $O_p(1)$. Thus (3.12), (3.13) and (3.14) still holds, which implies that the result in Theorem 3.2.1 remains.*

3.3.3 Proof for the case $\gamma = -1/2$

In this subsection we sketch the proof for the case $\gamma = -1/2$. Similar to Lemma 3.3.4 and 3.3.5, the following asymptotic properties on $f_n(-1/r_0)$ and $g_n(-1/r_0)$ hold for $\gamma = -1/2$.

Lemma 3.3.6 *Suppose the second order condition holds with $\gamma = -1/2$ and the sequence k satisfies the condition in Theorem 3.2.1. Write*

$$\bar{g}_n := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{U(\infty) - U(Z_{n,n-k})}{U(\infty) - U(Z_{n,n-i+1})}\right)^2.$$

Then we have that as $n \to \infty$,

$$\sqrt{k}(f_n(-1/r_0) - 1/2) \xrightarrow{d} -\frac{1}{2}W_3, \tag{3.24}$$

$$\sqrt{\frac{k}{\log k}}\left(g_n(-1/r_0)-2\right) \xrightarrow{d} S_2, \tag{3.25}$$

$$\frac{1}{\log k}\bar{g}_n \xrightarrow{P} 1, \tag{3.26}$$

where W_3 is the same as in Lemma 3.3.2, and S_2 has a normal distribution.

Then we locate the root of $h_n(t) = 0$ by the following proposition.

Proposition 3.3.1 *For any δ close to 0, write $t_n^{(\delta)} = -(1+\delta)/r_0$. For sufficiently large n, there exists two random sequences π_n and ω_n such that*

$$\pi_n < 0 < \omega_n \tag{3.27}$$

$$\sqrt{k\log k}\pi_n = O_p(1) \quad \textit{and} \quad \sqrt{k\log k}\omega_n = O_p(1) \tag{3.28}$$

$$h_n(t_n^{(\pi_n)}) < 0 \quad \textit{and} \quad h_n(t_n^{(\omega_n)}) > 0 \tag{3.29}$$

Proposition 3.3.1 implies that there exists a root t_n^* of $h_n(t) = 0$ lies between $t_n^{(\omega_n)}$ and $t_n^{(\pi_n)}$. Hence, similar to (3.14), we get that

$$\sqrt{k}(r_0 t_n^{(\omega_n)} + 1) = o_p(1).$$

By verifying that both $f_n(t_n^{(\pi_n)})$ and $f_n(t_n^{(\omega_n)})$ converge to $1/2$ with the same speed of convergence $1/\sqrt{k}$ and share the same asymptotic limit as $f_n(-1/r_0)$, Theorem 3.2.1 is then proved for $\gamma = 1/2$.

Chapter 4

A 2-Step Estimator of the Extreme Value Index

4.1 Introduction

Let $X_1, X_2, \cdots$ be independent and identically distributed (i.i.d.) random variables from a distribution function F. Suppose F is in the domain of attraction of an extreme value distribution, i.e. there exist constants $a_n > 0$ and b_n, such that,

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x),$$

for all $1 + \gamma x > 0$, where $G_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma})$ is the corresponding extreme value distribution and $\gamma \in \mathbb{R}$ is called the *extreme value index* (see Gnedenko (1943)). Commonly, it is denoted as $F \in D(G_\gamma)$.

There are a few characterizations of the necessary and sufficient condition for a distribution function F belonging to the domain of attraction. One of them is via the "excess distribution function" as in Balkema and de Haan (1974). Denote the excess distribution function as

$$F_t(x) := P(X - t \le x | X > t) = \frac{F(t+x) - F(t)}{1 - F(t)}.$$

Then $F \in D(G_\gamma)$ is equivalent to

$$\lim_{t\to x^*} F_t(x\sigma(t)) = H_\gamma(x) := 1 - (1+\gamma x)^{-1/\gamma},$$

for all $1 + \gamma x > 0$, where $\sigma(t)$ is a positive function and x^* is the right endpoint of F, i.e. $x^* = \sup\{x | F(x) < 1\}$. The distribution function H_γ is the so-called *generalized Pareto distribution (GPD) function*. Intuitively, the distribution function F belongs to the domain of attraction if and only if the excesses above a high threshold are asymptotically generalized Pareto distributed.

This characterization creates several possible ways to deal with a major issue in Extreme Value Theory: estimating the extreme value index γ.

Denote $X_{n,1} \leq \cdots \leq X_{n,n}$ as the order statistics of $X_1, X_2, \cdots, X_n$. For a suitable sequence such that $k_n \to \infty$, $k_n/n \to 0$ as $n \to \infty$, the k_nth upper order statistic $X_{n,n-k_n}$ may take the place of the "high threshold". Then $X_{n,n} - X_{n,n-k_n}, \cdots X_{n,n-k_n+1} - X_{n,n-k_n}$ can be recognized as the order statistics of the empirical excesses above the high threshold. Thus, together as a new sample, they are asymptotically generalized Pareto distributed. In the rest of this chapter, without declaration, we briefly use k instead of k_n.

Theoretically, the 1/2 and 3/4-quantiles of the GPD can be calculated as $(2^\gamma - 1)/\gamma$ and $(4^\gamma - 1)/\gamma$; empirically, they can be estimated as $X_{n,n-[k/2]} - X_{n,n-k}$ and $X_{n,n-[k/4]} - X_{n,n-k}$ respectively. This creates the quantile estimator, suggested by Pickands III (1975), as follows.

$$\hat{\gamma}_P = \frac{1}{\log 2} \log \frac{X_{n,n-[k/4]} - X_{n,n-[k/2]}}{X_{n,n-[k/2]} - X_{n,n-k}}.$$

When $\gamma > 0$, the function $\sigma(t)$ can be chosen as $\sigma(t) = \gamma t$ and $x^* = +\infty$. Thus, the condition on the excess distribution function can be rewritten as

$$\lim_{t \to +\infty} P(\frac{X}{t} \leq x | X > t) = 1 - x^{-1/\gamma}.$$

Therefore, similar to the above intuition, by taking $X_{n,n-k}$ as the "high threshold", we get that, as $n \to \infty$, the excess ratios $X_{n,n}/X_{n,n-k}, \cdots X_{n,n-k+1}/X_{n,n-k}$ form a sample of order statistics from a Pareto distribution. By fitting the Pareto distribution with the maximum likelihood procedure, Hill (1975) suggested the so-called Hill estimator as

$$\hat{\gamma}_H = \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n,n-i} - \log X_{n,n-k}.$$

The Hill estimator is only applied for positive γ. In order to deal with a general $\gamma \in \mathbb{R}$, Dekkers *et al.* (1989) introduced the moment estimator,

$$\hat{\gamma}_M = \hat{\gamma}_H + 1 - \frac{1}{2} \left(1 - \frac{\hat{\gamma}_H^2}{M_n^{(2)}} \right)^{-1},$$

where

$$M_n^{(2)} = \frac{1}{k} \sum_{i=0}^{k-1} (\log X_{n,n-i} - \log X_{n,n-k})^2.$$

An alternative way to have a general estimator was suggested by Beirlant *et al.* (1996) as the UH estimator. By denoting $UH_{n,n-i} = X_{n,n-i}\hat{\gamma}_H$ for $i = 0, 1, \cdots, k$, the estimator

$$\hat{\gamma}_{UH} = \frac{1}{k} \sum_{i=0}^{k-1} \log UH_{n,n-i} - \log UH_{n,n-k}$$

is valid for all $\gamma \in \mathbb{R}$.

Although $\hat{\gamma}_H$ and $\hat{\gamma}_M$ perform reasonably well for γ positive and $\gamma \in \mathbb{R}$ respectively, they both have the disadvantage that they are not shift invariant. The estimator $\hat{\gamma}_P$ is a shift and scale invariant estimator, but according to de Haan and Peng (1998), it does not perform as well as the other two in most cases.

A shift invariant estimator needs to be constructed from the excesses instead of the excess ratios. Hosking and Wallis (1987) proposed the probability-weighted moment (PWM) estimator by assigning different weights to the excesses. It is defined as

$$\hat{\gamma}_{PWM} = \frac{P_n - 4R_n}{P_n - 2R_n},$$

where

$$P_n = \frac{1}{k}\sum_{i=0}^{k-1} X_{n,n-i} - X_{n,n-k},$$

and

$$R_n = \frac{1}{k}\sum_{i=0}^{k-1} \frac{i}{k}(X_{n,n-i} - X_{n,n-k}).$$

In order to have consistency, the PWM estimator can only be applied for $\gamma < 1$. To obtain the asymptotic normality, γ should be further restricted as $\gamma < 1/2$. Note that the PWM estimator is shift and scale invariant.

Similar to the idea of the Hill estimator, Smith (1987) applied the maximum likelihood procedure to fit the GPD with a general γ, which leads to the maximum likelihood estimator of the extreme value index. The likelihood equations are as follows:

$$\begin{aligned} \sum_{i=1}^{k} \frac{1}{\gamma^2}\log\Big(1 + \frac{\gamma}{\sigma}(X_{n,n-i+1} - X_{n,n-k})\Big) & \\ - \left(\frac{1}{\gamma}+1\right)\frac{(1/\sigma)(X_{n,n-i+1} - X_{n,n-k})}{1 + (\gamma/\sigma)(X_{n,n-i+1} - X_{n,n-k})} &= 0, \\ \sum_{i=1}^{k}\left(\frac{1}{\gamma}+1\right)\frac{(\gamma/\sigma)(X_{n,n-i+1} - X_{n,n-k})}{1 + (\gamma/\sigma)(X_{n,n-i+1} - X_{n,n-k})} &= k, \end{aligned} \tag{4.1}$$

(the equations for $\gamma = 0$ should be interpreted as the limit when $\gamma \to 0$). For $\gamma \neq 0$, they can be simplified to

$$\begin{aligned} \frac{1}{k}\sum_{i=1}^{k}\log\Big(1 + \frac{\gamma}{\sigma}(X_{n,n-i+1} - X_{n,n-k})\Big) &= \gamma, \\ \frac{1}{k}\sum_{i=1}^{k}\frac{1}{1 + (\gamma/\sigma)(X_{n,n-i+1} - X_{n,n-k})} &= \frac{1}{\gamma+1}. \end{aligned}$$

When $\gamma > -1/2$, the maximum likelihood estimators for the extreme value index and the scale, $\hat{\gamma}_{ML}$ and $\hat{\sigma}_{ML}$, are obtained by solving these equations.

In order to obtain the asymptotic normality for most of the estimators of the extreme value index, further restrictive condition on F is required. de Haan and Stadtmüller (1996) proposed the generalized second order condition as follows. Denote $F^{\leftarrow}$ as the generalized inverse of F. Assume that there exist measurable, locally bounded functions $a, \Phi : (0,1) \to (0,\infty)$ and $\Psi : (0,\infty) \to \mathbb{R}$, such that for all $x > 0$

$$\lim_{t \downarrow 0} \frac{(F^{\leftarrow}(1-tx) - F^{\leftarrow}(1-t))/a(t) - (x^{-\gamma}-1)/\gamma}{\Phi(t)} = \Psi(x). \tag{4.2}$$

According to de Haan and Stadtmüller (1996), $|\Phi|$ is $-\rho$-varying at 0 for some $\rho \leq 0$, and

$$\Psi(x) = \begin{cases} (x^{-(\gamma+\rho)} - 1)/(\gamma+\rho), & \rho < 0 \\ -x^{-\gamma}\log(x)/\gamma, & \gamma \neq 0, \rho = 0 \\ \log^2(x), & \gamma = \rho = 0. \end{cases}$$

Under the generalized second order condition, for $\gamma > -1/2$, Drees *et al.* (2004) proved asymptotic normality of the maximum likelihood estimator by assuming that the sequence k_n satisfies

$$\Phi(k_n/n) = O(k_n^{-1/2}), \tag{4.3}$$

as $n \to \infty$. The asymptotic normality is a direct consequence of the following theorem (Theorem 2.1 in Drees *et al.* (2004)).

Theorem 4.1.1 *Assume condition (4.2) holds for some $\gamma > -1/2$, and the sequence k_n satisfies (4.3). Then the system of likelihood equations (4.1) has a sequence of solutions $(\hat{\gamma}_n, \hat{\sigma}_n)$ that verifies*

$$\begin{aligned} & k^{1/2}(\hat{\gamma}_n - \gamma) \\ & -\frac{(\gamma+1)^2}{\gamma} k^{1/2}\Phi\left(\frac{k}{n}\right)\int_0^1 (t^{\gamma} - (2\gamma+1)t^{2\gamma})\Psi(t)dt \\ \xrightarrow{d} \; & \frac{(\gamma+1)^2}{\gamma}\int_0^1 (t^{\gamma} - (2\gamma+1)t^{2\gamma})(W(1) - t^{-(\gamma+1)}W(t))dt, \end{aligned} \tag{4.4}$$

$$\begin{aligned} & k^{1/2}\left(\frac{\hat{\sigma}_n}{a(k/n)} - 1\right) \\ & -\frac{(\gamma+1)^2}{\gamma} k^{1/2}\Phi\left(\frac{k}{n}\right)\int_0^1 ((\gamma+1)(2\gamma+1)t^{2\gamma} - t^{\gamma})\Psi(t)dt \\ \xrightarrow{d} \; & \frac{(\gamma+1)^2}{\gamma}\int_0^1 ((\gamma+1)(2\gamma+1)t^{2\gamma} - t^{\gamma})(W(1) - t^{-(\gamma+1)}W(t))dt, \end{aligned} \tag{4.5}$$

as $n \to \infty$, and the convergence holds jointly with the same standard Brownian motion W. For $\gamma = 0$ these equations should be interpreted as their limits when $\gamma \to 0$.

From this theorem, (4.4) can be rewritten as

$$k^{1/2}(\hat{\gamma}_{ML}-\gamma)=\frac{(\gamma+1)^2}{\gamma}\int_0^1(t^\gamma-(2\gamma+1)t^{2\gamma})L_n(t)dt+o_p(1), \qquad (4.6)$$

where

$$\begin{aligned} L_n(t) &= W_n(1)-t^{-(\gamma+1)}W_n(t)+k^{1/2}\tilde{\Phi}\left(\frac{k}{n}\right)\Psi(t), \qquad (4.7)\\ W_n(t) &= k^{-1/2}W(kt), \end{aligned}$$

$\tilde{\Phi}(k/n)\sim\Phi(k/n)$ as $n\to\infty$ and W is a standard Brownian motion which implies that W_n is also a standard Brownian motion. Then the integral of the two parts $W_n(1)-t^{-(\gamma+1)}W_n(t)$ and $k^{1/2}\tilde{\Phi}\left(\frac{k}{n}\right)\Psi(t)$ lead to a mean-zero normal distribution and the asymptotic bias respectively, which completes the proof of the asymptotic normality. Notice that the asymptotic bias depends on the second order parameter ρ and the asymptotic variance can be calculated as shown in Remark 2.1 and Corollary 2.1 in Drees *et al.* (2004).

It is clear that the maximum likelihood estimator is shift and scale invariant. Meanwhile, it performs well for $\gamma>-1/2$. But it still has a disadvantage: there is no explicit formula for this estimator. It is always given by solving the likelihood equations, but there is even no guarantee for the existence of a solution. The existence was stated in Drees *et al.* (2004) but there is no proof of that statement in the paper. The numerical way to find a solution of these equations had been discussed in Grimshaw (1993).

An alternative way to deal with this problem is to find an approximate solution for the likelihood equations, i.e. an explicit estimator such that the difference between the maximum likelihood estimator and the alternative estimator is approximately 0. As an example, Theorem 2.2 and Remark 2.4 in Drees *et al.* (2004) proved that, when $\gamma=0$, with the generalized second order condition and assumption on the sequence k as in (4.3), we have that

$$k^{1/2}(\hat{\gamma}_*-\hat{\gamma}_{ML})\xrightarrow{P}0,$$

where

$$\hat{\gamma}_*=1-\frac{1}{2}\left(1-\frac{(m_n^{(1)})^2}{m_n^{(2)}}\right)^{-1},$$

and

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

In this case, $\hat{\gamma}_*$, a shift and scale invariant estimator with explicit formula, is close enough to the maximum likelihood estimator. But this is only for a special case $\gamma=0$. Can we find such kind of estimator in general case? In this chapter, a 2-step estimator is established which gives a positive answer to this question. The idea is similar to the PWM estimator which is based on the weighted sum of the excesses. However, in the 2-step estimator, the weights are determined in prior according to a pre-estimation of the extreme value index. This is similar to the UH estimator where the extreme value index is pre-estimated by the Hill estimator.

In Section 4.2, Theorem 4.2.2 shows that, the 2-step estimator is close enough to the maximum likelihood estimator. By suitable choice in the first step, we may get a shift and scale invariant estimator. Simulations are given in Section 4.3. Section 4.4 concludes this chapter.

4.2 Result and proof

We start with stating the following theorem in Drees (1998).

Theorem 4.2.1 *Given (4.2) with $\gamma > -1/2$ and (4.3), one can find a probability space and define on that space a Brownian Motion W and a sequence of stochastic processes Q_n such that*
(i) for each n, $(Q_n(t))_{t\in[0,1]} \stackrel{d}{=} (X_{n,n-[kt]})_{t\in[0,1]}$;
(ii) there exist functions $\tilde{a}(k/n) = a(k/n)(1 + o(\Phi(k/n)))$ and $\tilde{\Phi}(k/n) \sim \Phi(k/n)$ such that, for all $\varepsilon > 0$,

$$\begin{aligned}\sup_{t\in[0,1]} t^{\gamma+1/2+\varepsilon} \Bigg| & \frac{Q_n(t) - F^{\leftarrow}(1-k/n)}{\tilde{a}(k/n)} \\ & - \left(\frac{t^{-\gamma}-1}{\gamma} - t^{-(\gamma+1)}\frac{W(kt)}{k} + \tilde{\Phi}\left(\frac{k}{n}\right)\Psi(t)\right)\Bigg| \\ & = o_p(k^{-1/2}) + o_p\left(\tilde{\Phi}\left(\frac{k}{n}\right)\right),\end{aligned} \tag{4.8}$$

as $n \to \infty$.

The following notation is introduced in order to shorten the proof in the rest of this chapter.

$$Y_n(t) = k^{1/2}\left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(k/n)} - \frac{t^{-\gamma}-1}{\gamma}\right). \tag{4.9}$$

When $\gamma = 0$, $\frac{t^{-\gamma}-1}{\gamma}$ should be read as $-\log t$.

With the notations in (4.7) and (4.9), a direct consequence of Theorem 4.2.1 is the following lemma.

Lemma 4.2.1 *Suppose (4.2) and (4.3) hold. Then for all $\varepsilon > 0$,*

$$Y_n(t) = L_n(t) + o_p(1)t^{-(\gamma+1/2+\varepsilon)}, \tag{4.10}$$

as $n \to \infty$, where the o_p-term is uniform for $t \in [0,1]$.

Our purpose is to find an estimator which is close enough to the maximum likelihood estimator. Hence, it should have the same asymptotic structure as the right side of (4.6). In order to do so, we should connect $L_n(t)$ with the observations. From Lemma 4.2.1, intuitively, we can substitute $L_n(t)$ by $Y_n(t)$, which is partially based on the observations.

There are still two remaining difficulties. First, the asymptotic structure in (4.6) is an integral of the product of $L_n(t)$ and $t^\gamma-(2\gamma+1)t^{2\gamma}$. To replace $L_n(t)$ by $Y_n(t)$, we have to study the functional approximation between them, i.e. whether the asymptotic structure is close to the integral of the product of $Y_n(t)$ and such kind of function. Secondly, there is still the parameter γ unknown. We solve this problem by using a first step estimator of γ, and show that it is still close enough.

To deal with the first difficulty, we study the weighted integral of the process $Y_n(t)$ on $[0,1]$. For the weight function, we focus on (pseudo) power functions. Suppose a continuous function $f:(0,1]\to\mathbb{R}$ satisfies

$$|f(t)| = O(t^{\gamma-\delta}) \qquad \text{when } \ t\to 0^+, \tag{4.11}$$

for some $0<\delta<1/2$. Then, we can choose a positive ε such that $\varepsilon+\delta<1/2$. By applying (4.10) for this ε, we get

$$\int_0^1 f(t)(Y_n(t)-L_n(t))dt = o_p(1).$$

By checking

$$\int_0^1 f(t)\Psi(t)dt<\infty$$

for $f(t)$ satisfying condition (4.11), it is insured that,

$$\int_0^1 f(t)L_n(t)dt$$

is bounded in probability as $n\to\infty$. Hence we have the following corollary.

Corollary 4.2.1 *With the same conditions as in Theorem 4.2.1, given a continuous function $f:(0,1]\to\mathbb{R}$ satisfying (4.11) for some $0<\delta<1/2$, we have that*

$$k^{1/2}\left(\int_0^1 f(t)\frac{Q_n(t)-Q_n(1)}{\tilde a(k/n)}dt-\int_0^1 f(t)\frac{t^{-\gamma}-1}{\gamma}dt\right)=\int_0^1 f(t)L_n(t)dt+o_p(1).$$

Next let us consider (for $\gamma>-1/2$) a continuous function $g:(0,1]\to\mathbb{R}$ satisfying

$$|g(t)| = O(t^{2\gamma-\delta}) \qquad \text{when } \ t\to 0^+, \tag{4.12}$$

for some $0<\delta<(\gamma\wedge 0)+1/2$. We can find positive numbers ε and δ such that $2\varepsilon+\delta<(\gamma\wedge 0)+1/2$. We write

$$k^{1/2}\left(\int_0^1 g(t)\left(\frac{Q_n(t)-Q_n(1)}{\tilde a(k/n)}\right)^2dt-\int_0^1 g(t)\left(\frac{t^{-\gamma}-1}{\gamma}\right)^2dt\right)$$

$$
\begin{aligned}
&= \left(\int_0^{k^{-1}} + \int_{k^{-1}}^1\right) g(t) \left(k^{-1/2} Y_n(t) + 2\frac{t^{-\gamma}-1}{\gamma}\right) Y_n(t) dt \\
&= I_1 + I_2.
\end{aligned}
$$

Because $k^{-1/2}Y_n(t) = o_p(t^{-(\gamma+\varepsilon)})$ uniformly for all $t \in [k^{-1}, 1]$, and $\int_0^1 g(t)t^{-(2\gamma+1/2+2\varepsilon)}dt$ is finite, by applying (4.10) for this ε, we get

$$
\begin{aligned}
I_2 &= \int_{k^{-1}}^1 g(t) \left(2\frac{t^{-\gamma}-1}{\gamma} + o_p(t^{-(\gamma+\varepsilon)})\right) (L_n(t) + o_p(1)t^{-(\gamma+1/2+\varepsilon)}) dt \\
&= \int_{k^{-1}}^1 2g(t)\frac{t^{-\gamma}-1}{\gamma} L_n(t) dt + o_p(1) \\
&= \int_0^1 2g(t)\frac{t^{-\gamma}-1}{\gamma} L_n(t) dt + o_p(1).
\end{aligned}
$$

The last equality comes from that the final integration is bounded in probability as $n \to \infty$.

For the rest part, I_1, it is going to be proved that $I_1 = o_p(1)$. On the interval $[0, k^{-1})$, for any $0 < \eta < 1$

$$
\left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(k/n)}\right)^2 = \left(\frac{Q_n(\eta k^{-1}) - Q_n(1)}{\tilde{a}(k/n)}\right)^2 = o_p(k^{2(\gamma+\varepsilon)}).
$$

Note that

$$
\delta < 2\gamma + 1 \Rightarrow \int_0^{k^{-1}} g(t) dt = O(k^{-(2\gamma-\delta+1)}).
$$

Finally, we have that

$$
k^{1/2} \int_0^{k^{-1}} g(t) \left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(k/n)}\right)^2 dt = o_p(k^{2\varepsilon+\delta-1/2}) = o_p(1).
$$

Meanwhile,

$$
k^{1/2} \int_0^{k^{-1}} g(t) \left(\frac{t^{-\gamma}-1}{\gamma}\right)^2 dt = O(k^{\delta-1/2}) = o_p(1),
$$

which completes the proof of $I_1 = o_p(1)$.

This conclusion is rewritten as the following corollary.

Corollary 4.2.2 *With the same conditions as in Theorem 4.2.1, given a continuous function $g : (0, 1] \to \mathbb{R}$ satisfying (4.12), for some $0 < \delta < 1/2 + \gamma$, we have that*

$$
\begin{aligned}
&k^{1/2}\left(\int_0^1 g(t) \left(\frac{Q_n(t) - Q_n(1)}{\tilde{a}(k/n)}\right)^2 dt - \int_0^1 g(t) \left(\frac{t^{-\gamma}-1}{\gamma}\right)^2 dt\right) \\
&= \int_0^1 2g(t)\frac{t^{-\gamma}-1}{\gamma} L_n(t) dt + o_p(1).
\end{aligned}
$$

In order to obtain the right side of (4.4), we introduce the following functions to apply Corollary 4.2.1 and Corollary 4.2.2.

$$\begin{aligned} f_1(\gamma,t) &= \frac{1}{\sqrt{2\gamma+1}}t^\gamma, \\ f_2(\gamma,t) &= \frac{\partial f_1(\gamma,t)}{\partial\gamma}, \\ f_3(\gamma,t) &= t^{2\gamma}, \\ f_4(\gamma,t) &= \frac{\partial f_3(\gamma,t)}{\partial\gamma}. \end{aligned}$$

Denote weighted moments with true γ as

$$\begin{aligned} h_n^{(1)}(\gamma) &= \int_0^1 f_1(\gamma,t)(Q_n(t)-Q_n(1))dt, \\ h_n^{(2)}(\gamma) &= \int_0^1 f_3(\gamma,t)(Q_n(t)-Q_n(1))^2 dt. \end{aligned}$$

The asymptotic behavior of $h_n^{(1)}(\gamma)$ and $h_n^{(2)}(\gamma)$ can be obtained by applying Corollary 4.2.1 and Corollary 4.2.2 for f_1 and f_3 as follows

$$k^{1/2}\left(\frac{h_n^{(1)}(\gamma)}{\tilde{a}(k/n)} - \frac{1}{\sqrt{2\gamma+1}(\gamma+1)}\right) = \int_0^1 f_1(\gamma,t)L_n(t)dt + o_p(1), \tag{4.13}$$

$$k^{1/2}\left(\frac{h_n^{(2)}(\gamma)}{(\tilde{a}(k/n))^2} - \frac{2}{(2\gamma+1)(\gamma+1)}\right) = \int_0^1 2f_3(\gamma,t)\frac{t^{-\gamma}-1}{\gamma}L_n(t)dt + o_p(1). \tag{4.14}$$

They lead to the asymptotic behavior of their combination as

$$\begin{aligned} & k^{1/2}\left(\frac{(h_n^{(1)}(\gamma))^2}{h_n^{(2)}(\gamma)} - \frac{1}{2(\gamma+1)}\right) \\ = \ & k^{1/2}\left(\frac{h_n^{(1)}(\gamma)}{\tilde{a}(k/n)} - \frac{1}{\sqrt{2\gamma+1}(\gamma+1)}\right)\cdot 2\frac{\sqrt{2\gamma+1}}{2} \\ & +k^{1/2}\left(\frac{h_n^{(2)}(\gamma)}{(\tilde{a}(k/n))^2} - \frac{2}{(2\gamma+1)(\gamma+1)}\right)\cdot(-1)\frac{2\gamma+1}{4} + o_p(1) \\ = \ & \sqrt{2\gamma+1}\int_0^1 f_1(\gamma,t)L_n(t)dt - \frac{2\gamma+1}{2}\int_0^1 f_3(\gamma,t)\frac{t^{-\gamma}-1}{\gamma}L_n(t)dt + o_p(1) \\ = \ & -\frac{1}{2\gamma}\int_0^1 (t^\gamma-(2\gamma+1)t^{2\gamma})L_n(t)dt + o_p(1). \end{aligned} \tag{4.15}$$

Define an auxiliary random variable as

$$\varphi(\gamma) := \frac{1}{2}\frac{h_n^{(2)}(\gamma)}{(h_n^{(1)}(\gamma))^2} - 1.$$

From (4.15) and

$$\frac{1}{2(\varphi(\gamma)+1)} = \frac{(h_n^{(1)}(\gamma))^2}{h_n^{(2)}(\gamma)},$$

we get the asymptotic behavior of $\varphi(\gamma)$ as

$$k^{1/2}(\varphi(\gamma)-\gamma) = \frac{(\gamma+1)^2}{\gamma}\int_0^1 (t^\gamma - (2\gamma+1)t^{2\gamma})L_n(t)dt + o_p(1).$$

Compared to (4.6), we proved that

$$k^{1/2}(\varphi(\gamma) - \hat{\gamma}_{ML}) = o_p(1).$$

Now the only problem is that, the real parameter γ is still a part of the auxiliary random variable $\varphi(\gamma)$. We introduce a first step estimator to replace it and try to keep the asymptotic property at the same time. By rewriting the final estimator in explicit form, we define the 2-step estimator as follows.

Definition 4.2.1 *Suppose a first step estimator of the extreme value index, $\hat{\gamma}^{(1)}$, is given, which uses the largest k order statistics. Assume the first step estimator approaches γ in speed $1/\sqrt{k}$, i.e.*

$$k^{1/2}\left(\hat{\gamma}^{(1)} - \gamma\right) \xrightarrow{d} N, \tag{4.16}$$

where N is a random variable with a suitable distribution. For all of the suggested estimators above, such a limit exists and follows a normal distribution.

If $\hat{\gamma}^{(1)} > -1/2$, define the weights $w_i^{(j)}$ as

$$w_i^{(j)} = \int_{\frac{i-1}{k}}^{\frac{i}{k}} t^{j\hat{\gamma}^{(1)}}dt = \frac{1}{j\hat{\gamma}^{(1)}+1}\left(\left(\frac{i}{k}\right)^{j\hat{\gamma}^{(1)}+1} - \left(\frac{i-1}{k}\right)^{j\hat{\gamma}^{(1)}+1}\right), \tag{4.17}$$

for $j = 1, 2$ and $i = 1, \cdots, k$. Then, define the weighted moments as

$$WM_n^{(j)} = \sum_{i=1}^{k} w_i^{(j)}(X_{n,n-i+1} - X_{n,n-k})^j, \qquad j = 1, 2. \tag{4.18}$$

Finally, define the estimator

$$\hat{\gamma}_{STEP} = \frac{2\hat{\gamma}^{(1)}+1}{2}\frac{WM_n^{(2)}}{(WM_n^{(1)})^2} - 1, \tag{4.19}$$

as the **2-step estimator** *of the extreme value index.*

The following theorem shows that this estimator is close enough to the maximum likelihood estimator.

Theorem 4.2.2 *Assume (4.2) holds and the sequence k satisfies (4.3). If $\gamma > -1/2$, then*

$$k^{1/2}(\hat{\gamma}_{STEP} - \hat{\gamma}_{ML}) \xrightarrow{P} 0.$$

Proof of Theorem 4.2.2

We have already proved that the auxiliary random variable $\varphi(\gamma)$ is close enough to the maximum likelihood estimator. Since the 2-step estimator is in fact $\varphi(\hat{\gamma}^{(1)})$, in order to complete the proof of the theorem, we only need to show that the difference between the 2-step estimator and the auxiliary random variable is also negligible, i.e.

$$\sqrt{k}(\varphi(\gamma) - \hat{\gamma}_{STEP}) = o_p(1). \tag{4.20}$$

From (4.18) and the definition of $Q_n(t)$ in Theorem 4.2.1, by changing γ into its estimator $\hat{\gamma}^{(1)}$ in $h_n^{(1)}(\gamma)$ and $h_n^{(2)}(\gamma)$, we can rewrite the weighted moments in the definition as

$$(WM_n^{(1)}, WM_n^{(2)}) \stackrel{d}{=} (h_n^{(1)}(\hat{\gamma}^{(1)})\sqrt{2\hat{\gamma}^{(1)}+1}, h_n^{(2)}(\hat{\gamma}^{(1)})).$$

According to the definition of $\hat{\gamma}_{STEP}$ in (4.19), it is clear that

$$\frac{1}{2(\hat{\gamma}_{STEP}+1)} = \frac{1}{2\hat{\gamma}^{(1)}+1}\frac{(WM_n^{(1)})^2}{WM_n^{(2)}} \stackrel{d}{=} \frac{(h_n^{(1)}(\hat{\gamma}^{(1)}))^2}{h_n^{(2)}(\hat{\gamma}^{(1)})}. \tag{4.21}$$

This is a slight change from $\varphi(\gamma)$ change in sense of the following lemma.

Lemma 4.2.2 *Under the conditions of Theorem 4.2.2, we have*

$$k^{1/2}\left(\frac{(h_n^{(1)}(\hat{\gamma}^{(1)}))^2}{h_n^{(2)}(\hat{\gamma}^{(1)})} - \frac{(h_n^{(1)}(\gamma))^2}{h_n^{(2)}(\gamma)}\right) = o_p(1). \tag{4.22}$$

Proof of Lemma 4.2.2

We start with the Taylor expansion of $f_1(\hat{\gamma}^{(1)}, t)$,

$$f_1(\hat{\gamma}^{(1)}, t) = f_1(\gamma, t) + (\hat{\gamma}^{(1)} - \gamma)f_2(\gamma, t) + \frac{(\hat{\gamma}^{(1)} - \gamma)^2}{2}\frac{\partial f_2(s,t)}{\partial s}|_{s=\eta_n(t)},$$

where $\eta_n(t)$ is a random variable depending on n and t, but always between γ and $\hat{\gamma}^{(1)}$. Since $\hat{\gamma}^{(1)} \xrightarrow{P} \gamma$ as $n \to \infty$, we have $\eta_n(t) \xrightarrow{P} \gamma$ uniformly for all $t \in (0, 1]$. Then, for any $\delta > 0$,

$$\frac{\partial f_2(s,t)}{\partial s}|_{s=\eta_n(t)} = O_p(t^{\eta_n(t)}(\log t)^2) = t^{\gamma-\delta}O_p(1),$$

when $n \to \infty$, the O_p-term is uniform for all $t \in (0, 1]$. So we can continue with the Taylor expansion as follows,

$$f_1(\hat{\gamma}^{(1)}, t) \quad = \quad f_1(\gamma, t) + (\hat{\gamma}^{(1)} - \gamma)f_2(\gamma, t) + \frac{(\hat{\gamma}^{(1)} - \gamma)^2}{2}t^{\gamma-\delta}O_p(1)$$

$$= \quad f_1(\gamma,t)+(\hat\gamma^{(1)}-\gamma)f_2(\gamma,t)+(\hat\gamma^{(1)}-\gamma)t^{\gamma-\delta}o_p(1),$$

as $n\to\infty$. In this expansion, the o_p-term is also uniform for $t\in(0,1]$. Then, by using this expansion and applying Corollary 4.2.1 for f_2 satisfying (4.12), we get that

$$\begin{aligned}
&k^{1/2}\left(\frac{h_n^{(1)}(\hat\gamma^{(1)})-h_n^{(1)}(\gamma)}{\tilde a(k/n)}+(\hat\gamma^{(1)}-\gamma)\frac{2\gamma^2+6\gamma+3}{(\gamma+1)^2(2\gamma+1)^{3/2}}\right)\\
=&k^{1/2}\left(\int_0^1 (f_1(\hat\gamma^{(1)},t)-f_1(\gamma,t))\frac{Q_n(t)-Q_n(1)}{\tilde a(k/n)}dt-(\hat\gamma^{(1)}-\gamma)\int_0^1 f_2(\gamma,t)\frac{t^{-\gamma}-1}{\gamma}dt\right)\\
=&k^{1/2}\left\{\int_0^1\left[f_2(\gamma,t)(\hat\gamma^{(1)}-\gamma)+(\hat\gamma^{(1)}-\gamma)t^{\gamma-\delta}o_p(1)\right]\frac{Q_n(t)-Q_n(1)}{\tilde a(k/n)}dt\right.\\
&\left.-(\hat\gamma^{(1)}-\gamma)\int_0^1 f_2(\gamma,t)\frac{t^{-\gamma}-1}{\gamma}dt\right\}\\
=&(\hat\gamma^{(1)}-\gamma)\left(\int_0^1 f_2(\gamma,t)L_n(t)dt\right)+k^{1/2}(\hat\gamma^{(1)}-\gamma)o_p(1)+o_p(1)\\
=&o_p(1). \qquad (4.23)
\end{aligned}$$

A similar relationship between $h_n^{(2)}(\hat\gamma^{(1)})$ and $h_n^{(2)}(\gamma)$ is given as

$$k^{1/2}\left(\frac{h_n^{(2)}(\hat\gamma^{(1)})-h_n^{(2)}(\gamma)}{(\tilde a(k/n))^2}+(\hat\gamma^{(1)}-\gamma)\frac{8\gamma^2+24\gamma+12}{(\gamma+1)^2(2\gamma+1)^2}\right)=o_p(1). \qquad (4.24)$$

From (4.13) and (4.14), we have that as $n\to\infty$,

$$\frac{h_n^{(1)}(\gamma)}{\tilde a(k/n)}\xrightarrow{P}\frac{1}{\sqrt{2\gamma+1}(\gamma+1)},$$

and

$$\frac{h_n^{(2)}(\gamma)}{(\tilde a(k/n))^2}\xrightarrow{P}\frac{2}{(2\gamma+1)(\gamma+1)}.$$

Considering with (4.23) and (4.24), we also have that as $n\to\infty$,

$$\begin{aligned}
\frac{h_n^{(1)}(\hat\gamma^{(1)})}{\tilde a(k/n)} &\xrightarrow{P} \frac{1}{\sqrt{2\gamma+1}(\gamma+1)},\\
\frac{h_n^{(2)}(\hat\gamma^{(1)})}{(\tilde a(k/n))^2} &\xrightarrow{P} \frac{2}{(2\gamma+1)(\gamma+1)}.
\end{aligned}$$

Finally, by using (4.23), (4.24) and the four equations above, we can calculate that

$$\begin{aligned}
&k^{1/2}\left(\frac{(h_n^{(1)}(\hat\gamma^{(1)}))^2}{h_n^{(2)}(\hat\gamma^{(1)})}-\frac{(h_n^{(1)}(\gamma))^2}{h_n^{(2)}(\gamma)}\right)\\
=\quad&k^{1/2}\left(\frac{h_n^{(1)}(\hat\gamma^{(1)})-h_n^{(1)}(\gamma)}{\tilde a(k/n)}\right)\left(\frac{h_n^{(1)}(\hat\gamma^{(1)})+h_n^{(1)}(\gamma)}{\tilde a(k/n)}\right)\frac{1}{h_n^{(2)}(\hat\gamma^{(1)})/(\tilde a(k/n))^2}
\end{aligned}$$

$$
\begin{aligned}
&+k^{1/2}\left(\frac{h_n^{(2)}(\hat\gamma^{(1)})-h_n^{(2)}(\gamma)}{(\tilde a(k/n))^2}\right)\cdot(-1)\frac{\left(h_n^{(1)}(\gamma)/\tilde a(k/n)\right)^2}{\left(h_n^{(2)}(\gamma)/(\tilde a(k/n))^2\right)\left(h_n^{(2)}(\hat\gamma^{(1)})/(\tilde a(k/n))^2\right)}\\
=\;& k^{1/2}(\hat\gamma^{(1)}-\gamma)\left(-\frac{2\gamma^2+6\gamma+3}{(\gamma+1)^2(2\gamma+1)^{3/2}}\cdot 2\frac{\sqrt{2\gamma+1}}{2}+\frac{8\gamma^2+24\gamma+12}{(\gamma+1)^2(2\gamma+1)^2}\frac{2\gamma+1}{4}\right)+o_p(1)\\
=\;& o_p(1).
\end{aligned}
$$

The lemma has been proved. □

Lemma 4.2.2 shows that

$$
\sqrt{k}\left(\frac{1}{2(\varphi(\gamma)+1)}-\frac{1}{2(\hat\gamma_{STEP}+1)}\right)=o_p(1),
$$

which implies (4.20) as a direct consequence. Hence we complete the proof of Theorem 4.2.2. □

Remark 4.2.1 *From the definition of the 2-step estimator, it is clear that, if the first step estimator is shift and scale invariant, the final estimator should be the same. So we can choose the Pickands' estimator $\hat\gamma_P$ mentioned in Section 4.1 as the first step estimator. Although the Pickands' estimator itself does not perform very well in most of the cases, after the 2-step procedure, it will be close enough to the maximum likelihood estimator.*

Remark 4.2.2 *Obviously, we can also use the final 2-step estimator as the first step estimator, and iterate the same procedure once more. It results in a 3-step estimator. If the first step estimator is shift and scale invariant, so is the final 3-step one. Simulations in Section 4.3 will show that the 3-step estimator is even more accurate.*

Remark 4.2.3 *The weighted moments $WM_n^{(j)}$ $(j=1,2)$ can be represented in another way as*

$$
\begin{aligned}
WM_n^{(1)} &= \frac{1}{\hat\gamma^{(1)}+1}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{\hat\gamma^{(1)}+1}(X_{n,n-i+1}-X_{n,n-i}),\\
WM_n^{(2)} &= \frac{1}{2\hat\gamma^{(1)}+1}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{2\hat\gamma^{(1)}+1}(X_{n,n-i+1}-X_{n,n-i})(X_{n,n-i+1}+X_{n,n-i}-2X_{n,n-k}).
\end{aligned}
$$

4.3 Simulations

Simulations have been done for 3 cases: γ positive, negative and $\gamma=0$. We also try to simulate for both large and small sample size.

For large sample size simulation, a sample with sample size 10,000 from a certain distribution is generated. In case $\gamma>0$, we choose the Cauchy distribution which has a positive extreme value index $\gamma = 1$ and a second order index $\rho = -2$. In case $\gamma = 0$, we choose the standard normal distribution. Both the extreme value index γ and the second order index ρ are equal to 0. In case $\gamma < 0$, we choose the Reversed Burr distribution. Such a distribution function is given as

$$F(x) = 1 - \left(\frac{4}{4 + x^{-2}}\right)^2, \qquad x < 0$$

It belongs to the domain of attraction of the extreme value distribution with extreme value index $\gamma = -1/4$ and $\rho = -1/2$.

We choose the Pickands estimator as the first step estimator, and calculate the 2-step estimator and the maximum likelihood estimator. The 3-step estimators described in Remark 4.2.2 are also demonstrated in the figures. In order to study the sensitivity of the first step estimator, we also use the moment estimator as the first step estimator for the same simulated samples. For γ positive, they are presented separately in Figure 4.1 and 4.2. For $\gamma = 0$ and γ negative, the results are shown in Figure 4.3-4.6.

From these figures we observe that, the 2-step estimator is close enough to the maximum likelihood estimator. Hence, it can be a good substitute of the maximum likelihood estimator with explicit formula. Furthermore, the 3-step estimator is even closer, i.e. it will be better to iterate the procedure for more steps. With the moment estimator as the first step estimator, the performance of the 2-step is improved. Hence, it will be helpful to choose an accurate first step estimator, even if it is not location invariant.

Secondly, we turn to small sample size. We generate 500 samples with sample size 1,000 each, calculate the maximum likelihood estimator and 2-step estimator in each sample, and take the average of the estimators among the samples. We also calculate the mean squared error (MSE) for both the maximum likelihood estimator and the 2-step estimator. Denote the calculated estimator as $\hat{\gamma}_i$ for sample i, where $1 \leq i \leq 500$. Then the mean squared error is defined as follows

$$MSE = \frac{1}{500} \sum_{i=1}^{500} (\hat{\gamma}_i - \gamma)^2,$$

where γ is the known extreme value index.

Because the Pickands estimator does not perform very well for small sample size, we use the moment estimator as the first step estimator. In these simulations, the 3-step estimator is ignored. For γ positive, we change to the Pareto distribution with $\gamma = 1/2$, i.e. the distribution function is $F(x) = 1 - 1/x^2$. In this case $\rho = -\infty$. Together with the large sample size simulation study, all these simulation studies cover the entire range of all possible ρ, i.e. $\rho \in [-\infty, 0]$.

The averaged estimations for γ positive are shown in Figure 4.7 with its corresponding MSE pictures in Figure 4.8. Figure 4.9-4.12 present the results for $\gamma = 0$ and γ negative.

From the multi-sample simulations, we again observe that the 2-step estimator is close enough to the maximum likelihood estimator, while the MSE is in a comparable level.

Furthermore, we also make simulations for even smaller sample size, for example, 100. The results are no longer comparable with the maximum likelihood estimator. From the simulation study, we recommend the 2-step estimator for relatively larger sample size, for instance, at least 1,000.

4.4 Conclusion

In the literature of estimating the extreme value index, a variety of estimators are proposed. A good estimator should have the following properties:
1) performing a smaller estimation error;
2) satisfying theoretical properties such as shift and scale invariance;
3) easy to calculate.

Most of the explicit estimators do not satisfy the shift invariance property or perform a relatively worse estimation, while the maximum likelihood estimator is shift and scale invariant and provides a reasonably nice performance. However, it is not explicitly given.

In this chapter, we propose an explicit 2-step estimator which is close enough to the maximum likelihood estimator. Therefore it has the same asymptotic behavior. Furthermore, by a suitable choice of the first step estimator, it is shift and scale invariant. By iteration, we can get 3-step or even more step estimators which performs better according to an extensive simulation study.

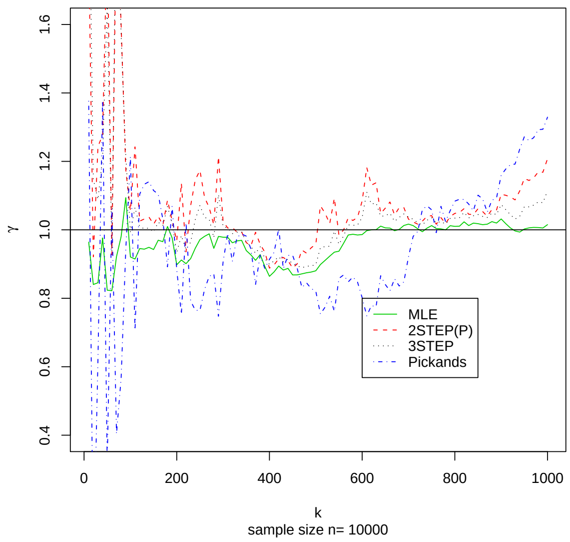

Figure 4.1: Large sample: Cauchy 1
First step: the Pickands' estimator
$(\gamma = 1, \rho = -2)$

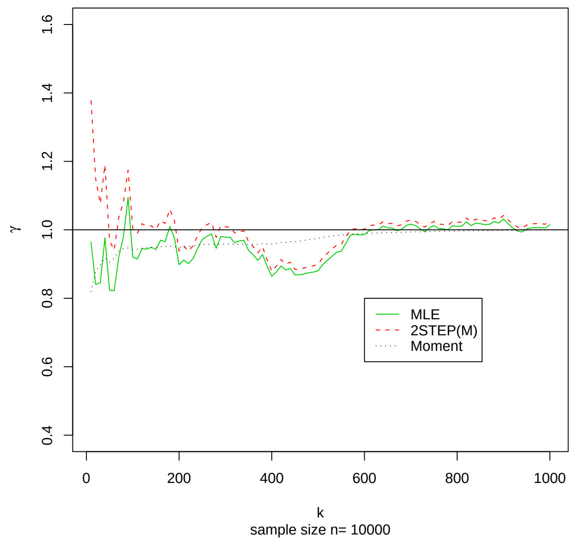

Figure 4.2: Large sample: Cauchy 2
First step: the moment estimator
$(\gamma = 1, \rho = -2)$

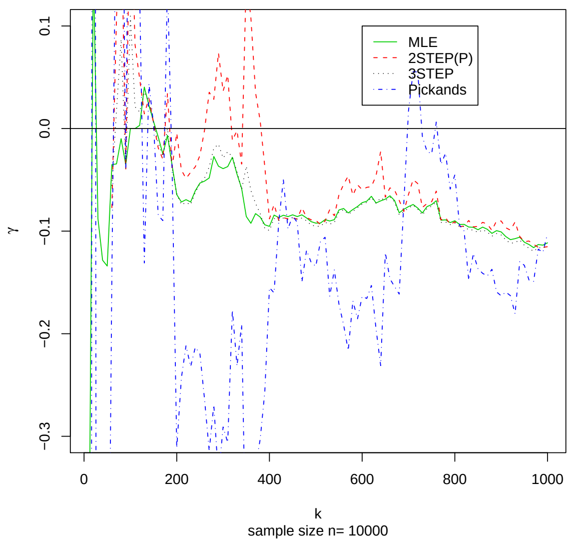

Figure 4.3: Large sample: normal 1
First step: the Pickands' estimator
$(\gamma = 0, \rho = 0)$

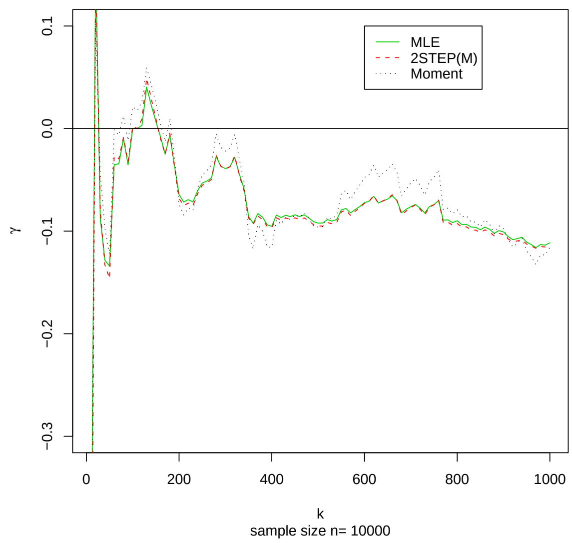

Figure 4.4: Large sample: normal 2
First step: the moment estimator
$(\gamma = 0, \rho = 0)$

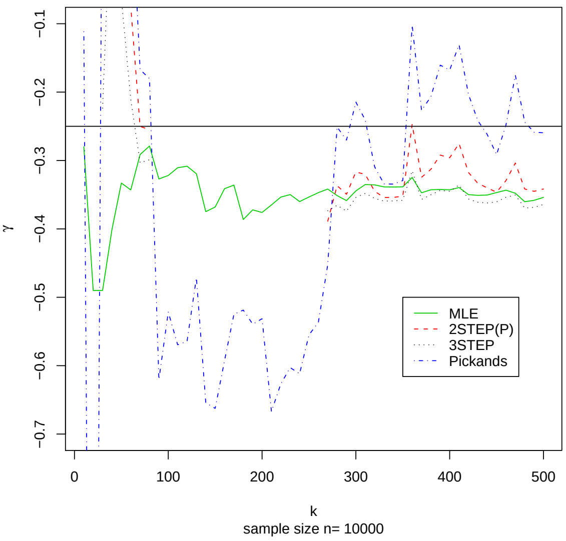

Figure 4.5: Large sample: R-Burr 1
First step: the Pickands' estimator
($\gamma = -1/4, \rho = -1/2$)

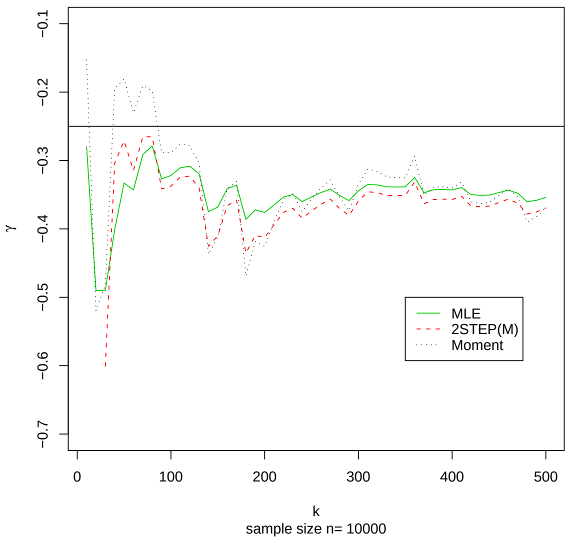

Figure 4.6: Large sample: R-Burr 2
First step: the moment estimator
($\gamma = -1/4, \rho = -1/2$)

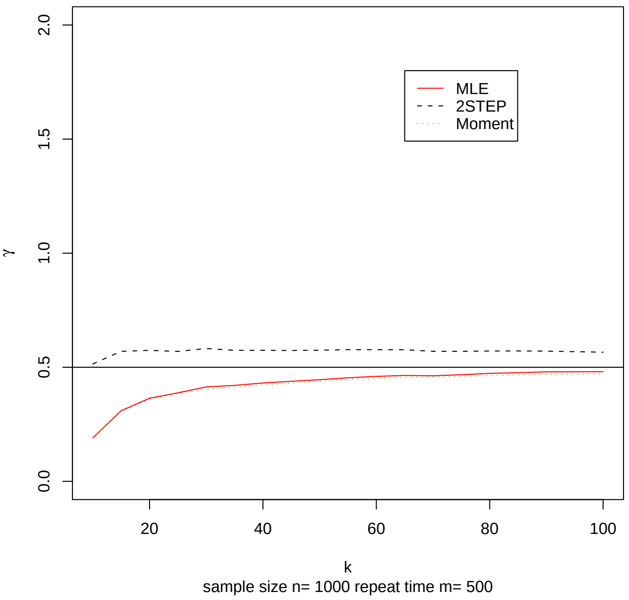

Figure 4.7: Small sample: Pareto
($\gamma = 1/2, \rho = -\infty$)

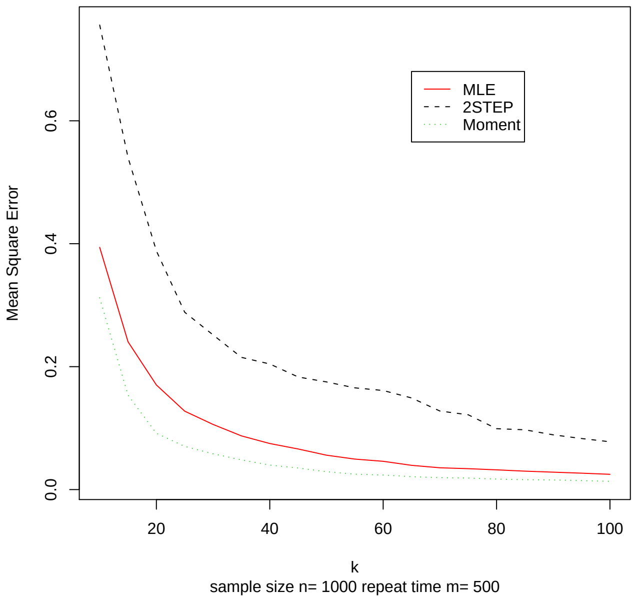

Figure 4.8: Small sample MSE: Pareto
($\gamma = 1/2, \rho = -\infty$)

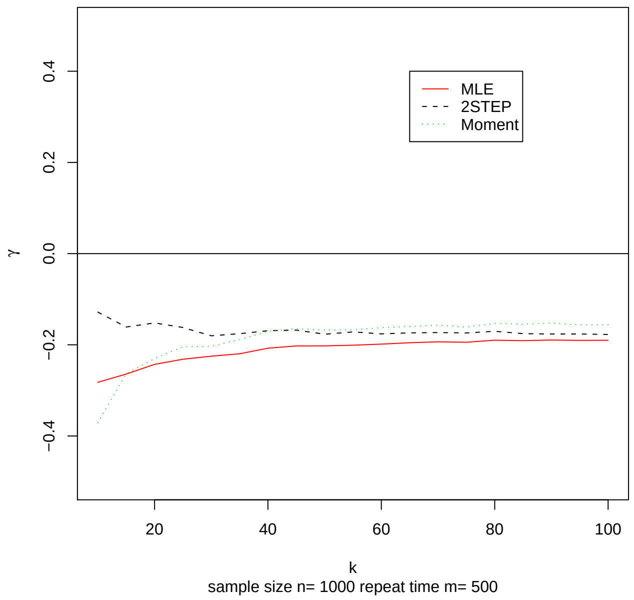

Figure 4.9: Small sample: normal
$(\gamma = 0, \rho = 0)$

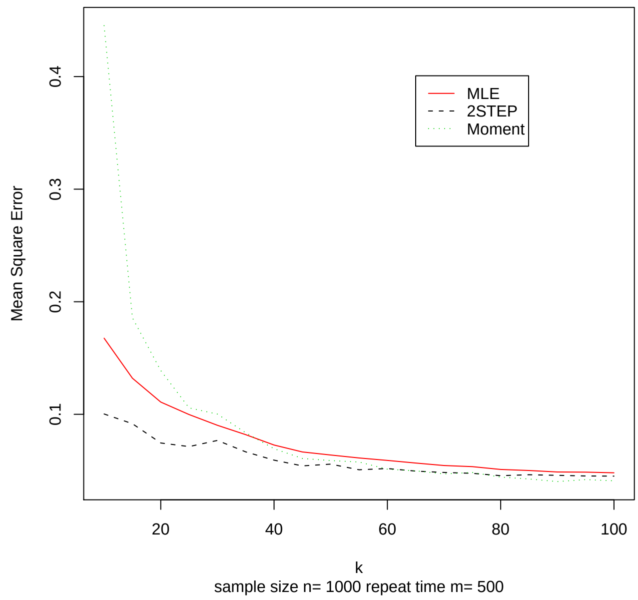

Figure 4.10: Small sample MSE: normal
$(\gamma = 0, \rho = 0)$

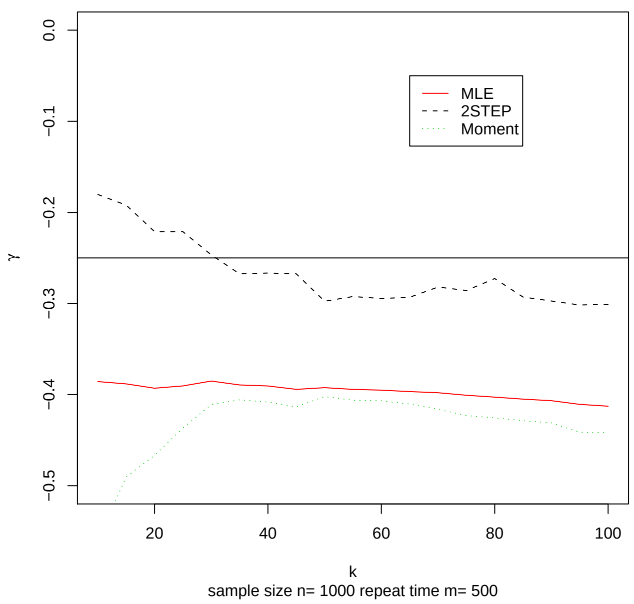

Figure 4.11: Small sample: R-Burr
$(\gamma = -1/4, \rho = -1/2)$

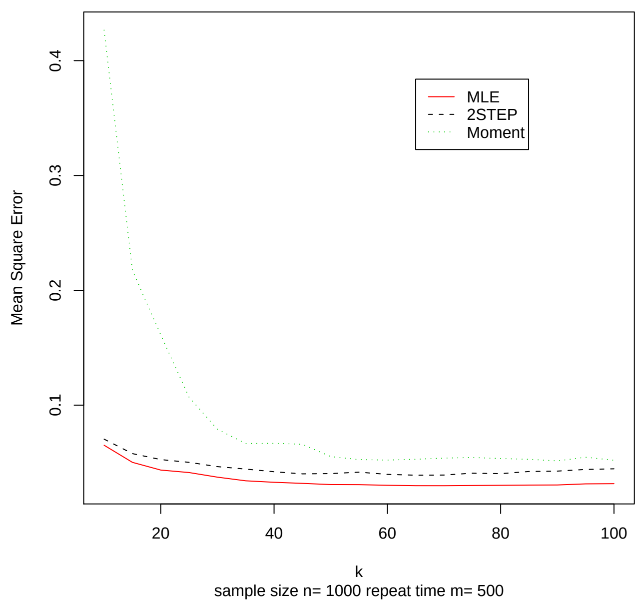

Figure 4.12: Small sample MSE: R-Burr
$(\gamma = -1/4, \rho = -1/2)$

Part II

Portfolio Optimization

Chapter 5

Portfolio Selection with Secondary Risk Indicators of Heavy Tailed Distributions

5.1 Introduction

In the financial industry, risk managers use portfolios to diversify away the risk of individual securities. Risk indicators based on different criteria lead to different methods for constructing the optimal portfolio. For example, initiated by Markowitz (1952), the classical mean-variance approach takes the variance as a risk indicator. Therefore, the principal component analysis is the main instrument in the portfolio selection procedure. Usually, the optimal portfolio under the mean-variance approach selects non-zero proportions for most of the individual securities.

In contrast to the classical approach, Roy (1952) and Arzac and Bawa (1977) introduced the safety-first criterion which uses the downside risk as risk indicator in portfolio selection. Value at Risk (VaR) becomes the risk indicator under such kind of criterion. With the assumption of Gaussian distributed returns, Gouriéroux *et al.* (2000) investigated the sensitivity of VaR. However, it is well documented that to model the financial returns by the Gaussian distribution is not a realistic assumption. In Fama and Miller (1972, Chap 6, Sec V), the symmetric stable distribution is employed to model the returns. The stable distribution is a rather narrow class and difficult to apply for empirical analysis because it has no explicit distribution or density function. Recent development in Extreme Value Theory (EVT) creates the possibility to model the non-Gaussian returns by the heavy tailed distributions, see, e.g., Jansen and de Vries (1991).

In the stable or heavy tailed framework, VaR is always considered as the risk indictor for the portfolio selection problem. In Fama and Miller (1972)'s symmetric stable framework, the characteristic exponent α of a stable distribution is considered to be equal among all assets, while the scale parameter of a stable distribution is introduced as an alternative to the variance in the Gaussian framework. Fama and Miller (1972) considered the portfolio optimization problem by minimizing the scale parameter of the constructed portfolio. Thus the scale parameter of a stable distribution which plays the role as the risk indicator. The VaR concept had not yet been developed at that time. But, since the VaR of a symmetric stable distribution is determined by the characteristic exponent α and the scale parameter, the scale minimization method is in fact a VaR approach.

Alongside the portfolio optimization problem, the portfolio diversification effect is always considered simultaneously. We still take the result in Fama and Miller (1972, Chap 6, Sec V) as an example. By modeling the returns as the symmetric stable distributions, they concluded that the diversification effects can only be realized when $\alpha > 1$. When $\alpha = 1$, in general, diversification has no effect. When $\alpha < 1$, increased diversification usually causes the risk indicator to increase which means a negative diversification effect.

The result in the symmetric stable framework can be extended to heavy tailed setup without much difficulty due to the fact that the tail of a heavy tailed distribution is asymptotically Pareto distributed. The tail index of a heavy tailed distribution plays the same role as the characteristic exponent α of a symmetric stable distribution, while a functional scale takes the place of the scale parameter in the symmetric stable framework. Based on the heavy tailed setup, the VaR can be approximately calculated thanks to the explicit Pareto distribution.

With heavy tailed assumption, Jansen *et al.* (2000) started the empirical exercises on portfolio selection under VaR approach, followed by Jansen (2001) and Susmel (2001). Since the datasets in these papers are the market indices of either markets across different countries or different asset markets such as equities and bonds, it is reasonable to assume that the individual returns are independent. However, for a general portfolio selection problem–in particular, constructing an optimal portfolio based on individual stocks trading on a specific stock market–the independence assumption is not realistic.

In a series of papers by Hyung and de Vries (2002, 2005, 2007), the portfolio optimization problem is studied by assuming specific dependence structure as well as the diversification effects. In these papers, the well-known single factor model, Capital Asset Pricing Model (CAPM), is used to model the dependence structure among the stock returns. Based on such kind of dependence structure, Hyung and de Vries (2002) confirmed Fama and Miller (1972)'s conclusion in the heavy tailed framework: the diversification effects depend on whether the tail index is higher than 1. From their empirical study, they also find that *"it is important to account for differences in scales when computing the diversification effects of portfolio investment. Per contrast, differences in tail shape were not large and did not seem to matter much for the diversification effect."*

Since the CAPM model only creates a relatively simple dependence structure, modeling the complex dependence among the stock returns, in particular, the dependence within the downside tails, is still an open question. The development in multivariate EVT helps to build more general models. Even a non-parametric approach is possible, see, e.g. Huang (1992). For empirical applications, Hartmann *et al.* (2004) applied multivariate EVT to test the tail dependence among the stock markets of the G-5 countries as well as the bond markets without any parametric assumption on the dependence structure. A portfolio selection application based on multivariate EVT is shown in Poon *et al.* (2004). However, the dependence structure is not non-parametrically modeled. On the contrary, several parametric models are employed.

In all of the portfolio selection exercises mentioned above, the optimal portfolio is only based on 2 or 3 financial return series, because the proportion of each security should be numerically calculated, i.e., the proportion of each security is controlled as a varying parameter from 0 to 1, the optimal portfolio is selected by comparing VaR when the parameters are varying. This method is difficult to extend, when there are a large amount of securities to select.

In this chapter, we start by linking VaR to the scale parameter and the tail index of a heavy tailed distribution. To consider the VaR, the scale parameter is recognized as the *secondary risk indicator* alongside the *primary risk indicator (PRI)*: the tail index. As suggested in Hyung and de Vries (2002), for the stock returns in the same market, their tail indices do not differ a lot from each other. Hence, minimizing the secondary risk indicator, the scale parameter, is the way to construct optimal portfolio. Then, by assuming that the stock returns belong to the domain of attraction of a multivariate extreme value distribution, we calculate the scale parameter of any specified portfolio. This result creates the possibility to solve the portfolio optimization problem without assuming parametric dependence structure, and even in the case when numerous stocks are involved. Since the dependence structure is non-parametrically estimated, this result is more objective than the CAPM approach. Meanwhile, the portfolio diversification effects are also discussed in this general dependence structure framework. Results similar to Fama and Miller (1972, Chap 6, Sec V) are confirmed within our framework.

In contrast to the studies in literature, we also discuss the theoretical case when the tail indices are all not higher than 1. In this case, diversification can not reduce the risk. Hence, the individual stock that has the lowest secondary risk indicator will be automatically the optimal choice. However, if the secondary risk indicators are also at the same level, (for instance, the lowest scale is shared among several stocks) the individual VaR must be the same as well. Then VaR is no longer a suitable criterion for choosing the optimal stock. In this case, we introduce the *probability of dominance* as a third risk indicator to find the optimal choice. Intuitively, the one who has the lower probability of dominance is less connected to the systemic risk, therefore, is a relatively less risky choice.

We apply our selection procedure to construct an optimal portfolio from 15 chosen stocks listed on the S&P 100 index in March, 2001. We compare the optimal portfolio selected by our procedure to those selected by the classical mean-variance approach. Our portfolio is better in terms of lower downside risk in extremal situation.

This chapter is organized as follows. Section 5.2 introduces the scale parameter as the secondary risk indicator. Section 5.3 provides the calculation of the scale parameter for any specified portfolio. Section 5.4 discusses the diversification effects and the portfolio selection in general. When the primary and secondary risk indicators are all equal across securities and the PRI is not higher than 1, the probability of dominance is introduced as the third risk indicator in choosing the optimal security as in Section 5.5. Section 5.6 is an empirical exercise. The conclusion and further extension are discussed in Section 5.7.

5.2 The secondary risk indicator

We start our study by linking the VaR to the scale parameter in the one dimensional case. Let X denote the downside return of a specific asset, i.e. if the return is R, we take $X = -R$. Suppose X follows a heavy tailed distribution, i.e. X has a regularly varying tail. Denote F as the distribution function of X, we have

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

where α is the tail index. It implies that $t^\alpha(1-F(t))$ is a slowly varying function. Considering a narrow case that

$$\lim_{t\to\infty} t^\alpha(1-F(t)) = A,$$

where $A > 0$ is a finite number, we call A *the scale of* X, denoted as

$$\sigma(X) = A.$$

Notice that the distribution function of R has the following representation

$$P(R \le -x) = P(X \ge x) = Ax^{-\alpha}[1 + o(1)],$$

as $x \to \infty$. This is the same setup as in Hyung and de Vries (2002, 2005).

Because of the asymptotic relation, the tail index and the scale are recognized as risk criteria. Given the tail index α and the scale A, it is easy to see that $VaR(\delta)$ with tail probability δ can be approximately calculated from

$$\delta = P(X > VaR(\delta)) \approx AVaR(\delta)^{-\alpha}.$$

Hence,

$$VaR(\delta) \approx (A/\delta)^{1/\alpha}.$$

By estimating the VaR, α and A with suitable estimators, the above relation turns to be an exact equation. For example, consider the Hill-type estimators for the tail index, the scale and the quantile in literature[1] (the first two are in Hill (1975), for the quantile, see Weissman (1978)),

$$\begin{aligned} \hat{\alpha}_H &= \left\{ \frac{1}{k} \sum_{i=1}^{k} \log X_{n,n-i+1} - \log X_{n,n-k} \right\}^{-1}, \\ \hat{A} &= \frac{k}{n} X_{n,n-k}^{\hat{\alpha}_H}, \\ Va\hat{R}(\delta) &= X_{n,n-k} \left\{ \frac{k}{n\delta} \right\}^{1/\hat{\alpha}_H}, \end{aligned}$$

where k is the number of upper order statistics used in estimation, which is chosen in advance, see, e.g. an extra criterion on choosing k in Danielsson *et al.* (2000). The estimators satisfy

$$Va\hat{R}(\delta) = (\hat{A}/\delta)^{1/\hat{\alpha}_H}. \tag{5.1}$$

Relation (5.1) shows that the approximate relation to calculate VaR turns to be an exact equation when using estimators. Although (5.1) is only a mathematical reformulation of the existing estimators, in practice, it is also useful. Noticing that the choice of k–the number of order statistics used in estimation–is always a considerable issue in statistical extremes, (5.1) actually has no k in the formula. It means that, although potentially we need to choose k in estimating the tail index and the scale, the VaR estimation is a direct calculation from the estimations. This argument is especially useful in multi-dimensional case as we shall come back to this point later.

Relation (5.1) shows that, when considering the safety-first criterion–VaR, alongside the PRI–tail index, the scale parameter plays a role as the *secondary risk indicator (SRI)*. As we discussed in Section 5.1, Hyung and de Vries (2002) found that the PRIs of stock returns trading on the same market do not differ from each other. Therefore, to study the scale of a heavy-tailed distributed random variable is useful. We start from some basic properties of the scale $\sigma(\cdot)$.

The function $\sigma(\cdot)$ has the following properties:

1) $\sigma(cX) = c^{\alpha}\sigma(X)$, for all $c > 0$;

2) $\sigma(X + Y) = \sigma(X) + \sigma(Y)$, if X and Y are independent and share the same tail index.

[1] Notice that the definition of VaR is exactly the same as the quantile of a certain distribution function.

The second property follows from the Feller theorem, see Feller (1971, section VIII, 8). This is the starting point of calculating the portfolio scale. However, it is based on the assumption of independence. We are going to extend this to a general dependent case.

5.3 The scale of a portfolio

Let $X = (X_1, \cdots, X_d)^T$ denote the downside returns of d individual stocks. We consider the downside risk as in Section 5.2. To model the tail dependence, we suppose $X = (X_1, \cdots, X_d)^T$ belongs to the domain of attraction of a d-dimensional extreme value distribution, i.e. for i.i.d. copies of X, $X^{(1)}, \cdots, X^{(n)}, \cdots$, there exist a distribution function G and two sequences of constants $\left\{a_i^{(n)} > 0\right\}_{n=1}^{\infty}$ and $\left\{b_i^{(n)}\right\}_{n=1}^{\infty}$ for $1 \leq i \leq d$, such that

$$\lim_{n\to\infty} P\left(\frac{\bigvee_{j=1}^n X_1^{(j)} - b_1^{(n)}}{a_1^{(n)}} \leq x_1, \cdots, \frac{\bigvee_{j=1}^n X_d^{(j)} - b_d^{(n)}}{a_d^{(n)}} \leq x_d\right) = G(x_1, \cdots, x_d) \quad (5.2)$$

holds for any continuity points $(x_1, \cdots, x_d)$ of G. In this case, each marginal of the random vector X follows the one-dimensional extreme value condition, while the dependence structure within the marginals follows a wider scope. Suppose the marginal tail indices are $\alpha_1, \alpha_2, \cdots, \alpha_d > 0$, and the marginal scales are all finite. Without loss of generality, we assume that $\alpha_1 = \min_{1\leq i\leq d} \alpha_i$. Firstly, we assume that α_1 is the unique minimum tail index. Then we get the following proposition. The proof is postponed to the appendix.

Proposition 5.3.1 *For any positive numbers $c_1, \cdots, c_d$, $\sum_{i=1}^d c_i X_i$ has a regularly varying tail with tail index α_1 and the scale is $c_1^{\alpha_1}\sigma(X_1)$.*

From Proposition 5.3.1, we can see that, when constructing a portfolio based on d individual stock returns satisfying our multivariate EVT assumption, the tail index of the portfolio is dominated by the minimum marginal tail index. Moreover, the scale is also dominated by the one who has the minimum marginal tail index. Therefore, to minimize the risk, we should always pick up the stock with the maximum tail index.

Now the only difficulty remains is for the situation in which the maximum tail index is not unique. In this situation, we need to consider these marginals which share the same tail indices as the maximum. They still belong to the domain of attraction of a finite-dimensional extreme value distribution, only the dimension might be reduced. To simplify but without loss of generality, we assume $\alpha_1 = \cdots = \alpha_d = \alpha$ in the rest of this section. Actually, this is what Hyung and de Vries (2002) suggested for the stocks trading on a specific market.

Since the discussion later is relatively complicated and long, we divide it into 3 subsections. In Subsection 5.3.1, we review the multivariate EVT, and calculate the scale of the portfolio based on the simple max-stable distribution. In Subsection 5.3.2, we extend the theory from the simple max-stable distribution to the general case: the domain of attraction.

5.3.1 Special case: simple max-stable distribution

According to multivariate EVT, by a suitable choice of the sequences of constants in (5.2), we can have a limit distribution function G which has marginal distribution functions: $F(x) = \exp\{-1/x\}, x > 0$. F is the standard Fréchet distribution function. This kind of limit distribution function G is the so-called *simple max-stable distribution* function. The simple max-stable distribution G belongs to the domain of attraction of itself. Hence, we can take G as a special case to model the stock returns. Suppose $U = (U_1, \cdots, U_d)^T$ follows the simple max-stable distribution G. We first study the scale of a portfolio based on U.

From the standard Fréchet distribution, we have that all the marginal tail indices are 1, and all the marginal scales are 1. There are several ways to exhibit the dependence structure of U. Here we use one which serves our purpose. The dependence structure of U depends on a probability measure H, see de Haan and Ferreira (2006, Chap 6). Let H be any probability measure on

$$W = \{w = (w_1, \cdots, w_d) : w_1 + \cdots + w_d = 1, w_i \geq 0, i = 1, 2, \cdots, d\},$$

such that

$$\int_W w_i H(dw) = \frac{1}{d}, \quad \text{for } 1 \leq i \leq d.$$

Any qualified H leads to a simple max-stable distribution function G as follows,

$$G(x_1, \cdots, x_d) = \exp\left(-d \int_W \left(\frac{w_1}{x_1} \vee \cdots \vee \frac{w_d}{x_d}\right) H(dw)\right).$$

Conversely, any simple max-stable distribution G has the above representation with a suitable H. H is called the *spectral measure on* W.

The following proposition shows how to calculate the scale of a specific portfolio based on U.

Proposition 5.3.2 *Suppose $U = (U_1, \cdots, U_d)^T$ follows a simple max-stable distribution, for any positive constants $c_1, \cdots, c_d$, $\sum_{i=1}^d c_i U_i$ has tail index 1, and the scale is calculated as*

$$\sigma(\sum_{i=1}^d c_i U_i) = \sum_{i=1}^d c_i. \tag{5.3}$$

Proof of Proposition 5.3 with discrete H

To prove this proposition we need to study the probability measure H which exhibits the dependence structure. In order to simplify the discussion, we only prove the case when H is a discrete measure. A proof for general H is given in Appendix 5.A. Suppose the discrete probability measure H has the following representation

$$H = \sum_{j=1}^{\infty} p_j \delta_{a_j},$$

where $\left\{a_j = (a_{1j}, \cdots, a_{dj})^T\right\}_{j=1}^{\infty}$ are points on W, i.e. $\sum_{i=1}^{d} a_{ij} = 1$ for $j = 1, 2, \cdots$, and $\sum_{j=1}^{\infty} p_j = 1$. According to the requirements for H we have, for any $1 \leq i \leq d$,

$$\sum_{j=1}^{\infty} p_j a_{ij} = \int_W w_i H(dw) = 1/d.$$

Now, we turn to decompose the random vector U by constructing a $d-$dimensional random vector of simple structure that has the same distribution function as U. Suppose $V_1, V_2, \cdots$ are i.i.d. standard Fréchet distribution. Define, for $1 \leq i \leq d$,

$$U_i = d \bigvee_{j=1}^{\infty} p_j a_{ij} V_j.$$

It is not very difficult to verify that in this way $U = (U_1, \cdots, U_d)^T$ has the distribution function G. Hence, we have a different view on the dependence structure of U in terms of the independent random variables $V_1, V_2, \cdots$.

From the construction of U, we have that

$$\begin{aligned}
\sum_{i=1}^{d} c_i U_i &= \sum_{i=1}^{d} c_i d \bigvee_{j=1}^{\infty} p_j a_{ij} V_j \\
&\leq \sum_{i=1}^{d} c_i d \sum_{j=1}^{\infty} p_j a_{ij} V_j \\
&= \sum_{j=1}^{\infty} (\sum_{i=1}^{d} c_i d p_j a_{ij}) V_j
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i=1}^{d} c_i U_i &= \sum_{i=1}^{d} c_i d \bigvee_{j=1}^{\infty} p_j a_{ij} V_j \\
&\geq \bigvee_{j=1}^{\infty} (\sum_{i=1}^{d} c_i d p_j a_{ij}) V_j.
\end{aligned}$$

Since V_j are i.i.d standard Fréchet distributed, $\sigma(V_j) = 1$, from e.g. Embrechts *et al.* (1997, pages 40 & 50), we get that, both $\sum_{j=1}^{\infty}(\sum_{i=1}^{d} c_i p_j a_{ij})V_j$ and $\bigvee_{j=1}^{\infty}(\sum_{i=1}^{d} c_i p_j a_{ij})V_j$ have tail index 1 and share the same scale as

$$\sum_{j=1}^{\infty}(\sum_{i=1}^{d} c_i d p_j a_{ij}) = \sum_{i=1}^{d} c_i.$$

Hence, $\sum_{i=1}^{d} c_i U_i$ must have tail index 1, and its scale must be the same as that of the two bounds, i.e.

$$\sigma(\sum_{i=1}^{d} c_i U_i) = \sum_{i=1}^{d} c_i. \square$$

Proposition 5.3.2 solves the scale problem for the simple max-stable distributions. From the proposition, we have the following remark.

Remark 5.3.1 *The scale of the portfolio based on a simple max-stable distribution does not depend on the dependence structure, but only on the sum of the portfolio weights.*

Up to now, we have shown that for the simple max-stable distribution, we can easily calculate the scale of any specific portfolio by simply adding up the weights. We will try to extend this from the simple max-stable case to the domain of attraction.

5.3.2 From the limit to the domain of attraction

In this subsection, we go back from U to X. Suppose $X = (X_1, \cdots, X_d)^T$ belongs to the domain of attraction of a d-dimensional extreme value distribution, the marginal tail indices are all equal to α, and the marginal scales are denoted as $\sigma(X_i) = \sigma_i$ for $1 \leq i \leq d$. Our purpose is to calculate the scale of a specified portfolio $\sum_{i=1}^{d} c_i X_i$ for any positive numbers $c_1, \cdots, c_d$.

In this case, the tail index α also plays a role in the calculation of the portfolio scale. The next result is the main theorem in this chapter. The proof is still in Appendix 5.A.

Theorem 5.3.1 *Suppose $X = (X_1, \cdots, X_d)^T$ belongs to the domain of attraction of a d-dimensional extreme value distribution with all marginal tail indices equal to α and all marginal scales finite. Let H be the corresponding spectral measure on W. Denote $\sigma(X_i) = \sigma_i$, for $1 \leq i \leq d$. Then, for any positive constants $c_1, \cdots, c_d$, the linear combination $\sum_{i=1}^{d} c_i X_i$ must have tail index α, and the scale is calculated as*

$$\sigma(\sum_{i=1}^{d} c_i X_i) = d \int_W \left(\sum_{i=1}^{d} c_i (\sigma_i w_i)^{1/\alpha} \right)^{\alpha} H(dw).$$

Remark 5.3.2 *When $\alpha = 1$, by considering the property of H, the result of the theorem is simplified as*

$$\sigma(\sum_{i=1}^{d} c_i X_i) = \sum_{i=1}^{d} c_i \sigma_i.$$

Hence, similar to Remark 5.3.1, we remark that the scale of the portfolio based on a random vector that belongs to the domain of attraction with marginal tail indices 1 does not depend on the dependence structure. It equals the sum of the weighted marginal scales.

With a general dependence structure assumption, Theorem 5.3.1 shows how to calculate the SRI–scale for a specified portfolio, when the PRI are all the same across the stocks under consideration. This creates the possibility to perform a portfolio selection procedure.

5.4 Diversification effects and portfolio selection

5.4.1 Diversification effects

We study the diversification effects by comparing the SRIs of the portfolios under different diversified levels. As in Fama and Miller (1972), to study the diversification effects, it is assumed that the individual stocks have the same characteristics, i.e. $\alpha_i = \alpha$ and $\sigma_i = \sigma$ for $1 \leq i \leq d$. Meanwhile, the diversified portfolio is constructed by assigning equal weights to the considered stocks. We study whether the SRI is increasing or decreasing as the number of stocks considered in the portfolio d varies.

In case $\alpha = 1$, from Remark 5.3.2, the scale of a portfolio does not depend on the dependence structure. Hence, the result is straightforward,

$$\sigma(\sum_{i=1}^{d} \frac{1}{d} X_i) = d \int_W \left(\sum_{i=1}^{d} \frac{1}{d} \sigma_i w_i \right) H(dw) = \sigma.$$

Therefore, one can never construct a portfolio which has smaller secondary risk indicator than the minimum among the individuals. Meanwhile, as d changes, the diversified portfolio will have the same SRI. It confirms Fama and Miller (1972)'s conclusion that the diversification has no effect when $\alpha = 1$.

The empirical studies in literature always observe $\alpha > 1$ for financial returns. In this case, considering that x^α is a convex function, we have, from the Jensen inequality,

$$\begin{aligned}\sigma(\sum_{i=1}^{d} \frac{1}{d} X_i) &= d\sigma \int_W \left(\sum_{i=1}^{d} \frac{1}{d} w_i^{1/\alpha} \right)^\alpha H(dw) \\ &\leq d\sigma \int_W \sum_{i=1}^{d} \frac{w_i}{d} H(dw)\end{aligned}$$

$$= \sigma. \tag{5.4}$$

Therefore, the diversified portfolio will have a smaller risk than individual stocks considered in this portfolio.

The diversification effects depend on the dependence structure. Consider the completely independent case, where H has positive measure on only d points:

$$(1, 0, \cdots, 0)^T, (0, 1, \cdots, 0)^T, \cdots, (0, 0, \cdots, 1)^T$$

with probability $1/d$ assigned to each point. The scale of $\bar{X}$ is then calculated as $\sigma/(d^{\alpha-1})$. It is an decreasing function as d increases. This agrees Fama and Miller (1972)'s conclusion that the more diversified, the less risky when $\alpha > 1$.

However, the inequality (5.4) turns to be an equality when H concentrates all its measure on a single point $(1/d, \cdots, 1/d)^T$. In multivariate EVT, that is the case in which we have completely tail dependence. In this case, the diversification has no effects. This extends Fama and Miller (1972)'s discussion.

The case $\alpha < 1$ is rarely observed in reality, however it is still theoretically interesting. In this case, considering that x^α is a concave function, by the Jensen inequality, we have for any $\sum_{i=1}^d c_i = 1$,

$$\begin{aligned}
\sigma(\sum_{i=1}^d c_i X_i) &= d \left(\sum_{i=1}^d c_i \sigma^{1/\alpha} \right)^\alpha \int_W \left(\sum_{i=1}^d \frac{c_i}{\sum_{i=1}^d c_i} w_i^{1/\alpha} \right)^\alpha H(dw) \\
&\geq d\sigma \int_W \frac{\sum_{i=1}^d c_i w_i}{\sum_{i=1}^d c_i} H(dw) \\
&= \sigma.
\end{aligned}$$

Hence, one can never construct a better portfolio than the individual securities. In particular, when $c_i = 1/d$, that confirms Fama and Miller (1972)'s conclusion, but now we have included the dependent case.

Summarizing the above discussion, we have the following three statements:

1) when $\alpha = 1$, the diversification has no effect regardless the dependence structure;
2) when $\alpha < 1$, the diversification has negative effects, i.e. leading to a higher risk;
3) when $\alpha > 1$, the diversification is in general leading to a lower risk. However, in the completely tail dependent case, the diversification has no effect. Hence, the diversification effect depends on the dependence structure.

Therefore, only for the case $\alpha > 1$, the portfolio optimization problem has a non-marginal solution. Otherwise, when $\alpha \leq 1$, the optimal portfolio is the individual stock which has the lowest SRI. An extra problem occurs when the lowest SRI is shared by several stocks. That will be discussed in Section 5.5.

5.4.2 Portfolio selection in case $\alpha > 1$

In case $\alpha > 1$, it is possible to achieve a smaller SRI, even smaller than the smallest SRI among the marginals, by constructing a diversified portfolio. In order to do so, we are going to minimize $\sigma(\sum_{i=1}^{d} c_i X_i)$ with the constraint that $\sum_{i=1}^{d} c_i = 1$.

In practice, the marginal scales σ_i can be individually estimated, while the dependence structure H can be non-parametrically estimated, see de Haan and Ferreira (2006). Hence, the scale of a specified portfolio can be calculated based on those estimates. In statistical application, we again take the advantage of introducing the scale as the secondary risk indicator in the following sense. The number of order statistics used in estimation, k, is a considerable issue not only in the marginal estimation but also in the estimation of the spectral measure H on W. However, our formula in Theorem 5.3.1 shows that the scale of a portfolio only depends on the portfolio scheme (the weight on each stock), the marginal scales and the probability measure H, but does not directly depend on the choice of k. So we can actually use different k in estimating different marginal scales, even a different k in estimating H. But it does not influence the calculation of the portfolio scale.

Finally, after the marginal scales and H are estimated, to find the optimal portfolio is a well-known convex optimization problem. Theoretically, there exists a unique solution. Empirically, this can be solved by the Newton method, see, e.g. Dennis and Schnabel (1996) and Boyd and Vandenberghe (2004). We will give an empirical example in Section 5.6.

5.5 Case $\alpha \leq 1$: probability of dominance

In the case $\alpha \leq 1$, as we discussed in Section 5.4.1, to minimize the risk by considering the SRI, the best choice is not to diversify but put all eggs in the same basket, i.e. to choose the individual stock with the lowest marginal scale. Even though this situation rarely happens in reality, it is better to complete the whole discussion because of the theoretical interest. Now, we should still consider a special case when the lowest scale is shared by several marginals. Without loss of generality, we assume that not only the PRIs-tail indices, but also the SRIs-scales, are equal, i.e. $\alpha_i = \alpha \leq 1$ and $\sigma_i = \sigma$, for all $1 \leq i \leq d$.

In this case, the marginal VaRs are all equal. Hence the individual stocks do not differ if we take VaR as the risk measure. It means that each stock has the same probability to fall into crisis, i.e. X_i is higher than a certain threshold. However, if the market is in a crisis situation, the performance of each stock can still differ. Obviously, the probability that the $i-$th stock performs the worst, more specifically, $X_i = \bigvee_{j=1}^{d} X_j$, can be different for different i. Since we are interested in the tail, we consider this probability given that the market is in a crisis situation, more specifically, $\bigvee_{j=1}^{d} X_j$ is higher than a high threshold. Intuitively, each individual stock may have a different relation to the systemic risk. Therefore, we introduce the following new measurement to compare the stocks. Denote the limiting conditional probability

$$p_i := \lim_{t \to \infty} P(X_i = \bigvee_{j=1}^{d} X_j | \bigvee_{j=1}^{d} X_j > t),$$

where $1 \leq i \leq d$. p_i is called the *probability of dominance (POD)* of the $i-$th stock. Intuitively, the POD indicates the probability to be the "worst" one when the system is in a "worse" case. The following theorem shows how to calculate the POD. For the proof, see Appendix 5.A.

Theorem 5.5.1 *Suppose $X = (X_1, \cdots, X_d)^T$ belongs to the domain of attraction of a d-dimensional extreme value distribution, the marginal tail indices are all equal and the marginal scales are all finite and equal. Let H be the spectral measure on W. For a given i, $1 \leq i \leq d$, the POD of the $i-th$ marginal is calculated as*

$$p_i = \frac{\int_{w_i = \max_{1 \leq j \leq d} w_j} w_i H(dw)}{\int_W \max_{1 \leq j \leq d} w_j H(dw)}.$$

Remark 5.5.1 *Similar results as in Theorem 5.5.1 can be found in de Haan (1984b) and Resnick and Roy (1990). However, in these papers, the results are not expressed in terms of the H measure.*

Like the SRI, the POD of each marginal also can be statistically estimated from Theorem 5.5.1. Therefore, the one who has the minimum POD will be an optimal choice in the sense that, it is less likely to be the "worst" choice in a "worse" situation. In other words, the one who has the minimum POD has the minimum connection with the systemic risk. If the minimum POD is shared within several stocks, it means that they even have the same connection with the systemic risk. Hence, there is no difference within choosing any one of them.

5.6 Empirical application

We apply our portfolio selection procedure to the dataset employed by Hyung and de Vries (2002). The dataset consists of the daily returns (close-to-close data), including cash dividends, for 15 companies listed on the S&P 100 index in March of 2001. The 15 stocks are arbitrarily chosen. The daily price series for each stock run from January 2, 1980 through March 6, 2001. Based on the daily prices, the daily (logarithmic) returns are derived, which gives a sample size (n) of 5,525 for each stock. The dataset covers more than 20 years of data including the 1987 Crash. Table 5.1 presents the selected stocks and their descriptive statistics.

Table 5.1: Selected stocks and descriptive statistics

Series	Name	Mean	Std	Min
1	ALCOA	0.0554	1.94	-27.57
2	AT & T	0.0428	1.73	-23.89
3	BLACK & DECKER	0.0204	2.24	-21.73
4	CAMPBELL SOUP	0.0630	1.75	-14.11
5	DISNEY (WALT)	0.0663	1.95	-34.38
6	ENTERGY	0.0453	1.63	-19.97
7	GEN.DYNAMICS	0.0576	1.81	-15.42
8	HEINZ HJ	0.0658	1.59	-10.01
9	JOHNSON & JOHNSON	0.0692	1.63	-20.27
10	MERCK	0.0756	1.58	-13.98
11	PEPSICO	0.0739	1.77	-15.39
12	RALSTON PURINA	0.0702	1.63	-11.59
13	SEARS ROEBUCK	0.0488	1.96	-29.20
14	UNITED TECHNOLOGIES	0.0611	1.68	-17.06
15	XEROX	0.0102	2.19	-29.75

The numbers in this table are percentages.

We start with estimating the left tails of these stock returns. To estimate the marginal tail indices and the marginal scales, we use the Hill-type estimators introduced in Section 5.2. Different from Hyung and de Vries (2002), we do not use the bootstrapping method as in Danielsson *et al.* (2000) to choose the number of upper order statistics for the estimation, k. Instead, we follow the method in de Haan and de Ronde (1998). For each stock, we plot the estimators of the tail index and the scale against k, then choose a particular k around which the two estimates exhibit constant levels simultaneously. The estimated tail indices $\hat{\alpha}$ and their corresponding standard deviation are presented in Table 5.2 as well as the estimators of the scales $\hat{A}$. The results are slightly different from Hyung and de Vries (2002) due to the different method of choosing k. However, it still agrees with the observation that the tail indices are very similar while the scales are quite different.

To construct the optimal portfolio, we should first select the stocks with the maximum tail indices. By ranking the estimated tail indices among the stocks, the series 10 provides the maximum, 3.871. However, the second highest tail index is provided by series 14 as 3.759. Considering their standard deviation, 0.325 and 0.310 respectively, by a hypothesis test, they do not significantly differ from each other. Following this idea, we make hy-

Table 5.2: Estimation for individual stocks

Series	k	$\hat{\alpha}$	std	$\hat{A}$
1	161	3.559	0.281	2.320
2	214	2.846	0.195	0.656
3	131	3.133	0.274	2.581
4	89	3.509	0.372	1.954
5	379	2.670	0.137	0.677
6	263	2.473	0.153	0.329
7	163	3.189	0.250	1.358
8	110	3.373	0.322	1.065
9	261	3.363	0.208	1.023
10	142	3.871	0.325	2.045
11	158	3.195	0.254	1.234
12	161	3.148	0.248	0.790
13	135	3.335	0.287	1.961
14	147	3.759	0.310	2.211
15	291	2.379	0.140	0.620

The column k shows the number of upper order statistics used in estimation. The column std gives the estimated standard deviation of the corresponding tail index.

pothesis tests for each pair of stocks to see whether they are significantly different. Under a significant level 0.2, the hypothesis that the top 4 series, series 10, 14, 1 and 4, are in the same level is not rejected. However, the tail index of the series 10 is significant different from the 5th highest tail index provided by series 8, 3.373. Notice that we use a rather high significant level, because the null hypothesis is that the two indices are equal while we are afraid of involving potential smaller tail indices. Hence, our optimal portfolio will be constructed from only 4 stocks: 10, 14, 1 and 4. We take their tail indices as the average of their estimations, 3.674.

The next step is to estimate the dependence structure H within those 4 series. We follow the estimation procedure in de Haan and Ferreira (2006, Section 7.3). Denote these 4 return series as R_i^t, where $i = 1, 2, 3, 4$, $t = 1, 2, \cdots, n$. Here n is the sample size 5,525. Since we focus on the downside risk, we define $X_i^t = -R_i^t$. By ranking $\{X_i^t\}_{1 \le t \le n}$, we get 4 rank sequences $\{Y_i^t\}_{1 \le t \le n}$ as $Y_i^t = rank(X_i^t)$, for $i = 1, 2, 3, 4$. By choosing a proper m, we find those observation points s such that

$$s \in S := \left\{ t : \bigvee_{i=1}^{4} Y_i^t \ge n + 1 - m \right\}.$$

For any $s \in S$, we connect the point

$$((n+1-Y_1^s, n+1-Y_2^s, n+1-Y_3^s, n+1-Y_4^s)) \in \mathbb{R}_+^4$$

and the origin by a straight line, and find the intersection with the plane

$$W_4 := \left\{(x_1, x_2, x_3, x_4) \in \mathbb{R}_+^4 : \quad x_1 + x_2 + x_3 + x_4 = 1\right\}.$$

We get the points

$$\frac{1}{4n+4-\sum_{i=1}^4 Y_i^s}\left((n+1-Y_1^s, n+1-Y_2^s, n+1-Y_3^s, n+1-Y_4^s)\right),$$

for all $s \in S$. By assigning equal weights to those points on W_4, we get the estimation of the spectral measure H on W_4. The parameter m plays a similar role as the number of upper order statistics k used in the marginal estimation. To choose m, we use the tail dependence measure

$$\kappa := \left(\int_{W_4} \max_{1\le i\le 4} w_i H(dw)\right)^{-1}$$

defined in de Haan and Ferreira (2006, Section 7.4). Notice that $1 \le \kappa \le 4$. $\kappa = 1$ implies completely tail independence and $\kappa = 4$ implies completely tail dependence. An intermediate κ shows how dependent they are. With each m, κ can be estimated from the estimated H. Hence, we plot the estimation of the tail dependence measure κ against m. Similar to the choice of k for the marginals, we choose a particular m around which the estimated κ is at a constant level. Finally, with $m = 91$, we get the estimated spectral measure H on W_4 which consists of 456 points.

With the estimation of the marginal tail indices, scales and the spectral measure H on W_4, we can use the formula in Theorem 5.3.1 to calculate the scale of any portfolio with specified weights. Therefore, it is possible to minimize the scale of the portfolio thanks to the optimization toolbox in Matlab. The optimal weights are shown in Table 5.3 as Portfolio 1.

To compare with our optimal portfolio, we also construct other portfolios using the classical mean-variance approach. The weights are calculated by using the portfolio allocation package in Matlab. One portfolio based on stocks 10, 14, 1 and 4 is constructed, while the other one is based on all 15 stocks. Their weights are shown in Table 5.3 as Portfolio 2 and 3.

From the different weights, we calculate the 5,525 daily returns of those 3 portfolios. In order to compare the downside risk of those portfolios, we estimates their VaRs. We present the estimations of the left quantiles with probabilities 5%, 1%, 0.1% and 0.01%. Since we consider daily data, these probabilities more or less represent the extremal events

Table 5.3: Weights of portfolios

Series	Portfolio 1	Portfolio 2	Portfolio 3
1	0.1981	0.1163	0.0332
2			0.0046
3			
4	0.3099	0.2186	0.0467
5			0.0657
6			0.0845
7			0.0836
8			0.1397
9			0.0514
10	0.2633	0.4396	0.1782
11			0.0930
12			0.1559
13			
14	0.2287	0.2255	0.0636
15			

Portfolio 1 is the optimal portfolio following our portfolio selection procedure. Portfolio 2 is the optimal portfolio based on stocks 10, 14, 1 and 4 by mean-variance approach. Portfolio 3 is the optimal portfolio based on all 15 stocks by mean-variance approach.

happened once per month, half year, 4 years and 40 years. Notice that, since we have 5,525 observations, the non-parametric estimations of the left quantiles with probabilities 5%, 1% and 0.1% are simply the 276th, 55th and 5th lower order statistics of the portfolio returns. Meanwhile, with one-dimensional EVT, the quantiles with small tail probabilities are also possible to be estimated by the Hill-type estimator introduced in Section 5.2. In particular, the left quantile with probability 0.01% can only be estimated in this way. We present both the non-parametric estimation and the EVT estimation in Table 5.4 and Table 5.5 respectively.

Table 5.4: Downside VaR: non-parametric estimation

	5%	1%	0.1%
Portfolio 1	-1.74	-2.91	-6.24
Portfolio 2	-1.70	-2.93	-6.50
Portfolio 3	-1.46	-2.43	-6.10

The estimations for the left quantiles with probability 5%, 1% and 0.1% are simply the 276th 55th and 5th lower order statistics.

From these tables, we observe that regardless the choice of tail probability and es-

Table 5.5: Downside VaR: EVT estimation

	5%	1%	0.1%	0.01%
Portfolio 1	-1.72	-2.88	-6.00	-12.53
Portfolio 2	-1.74	-2.93	-6.18	-13.05
Portfolio 3	-1.37	-2.72	-7.23	-19.24

The estimations are from the Hill-type quantile estimators.

timation method, the VaR of our optimal portfolio, Portfolio 1, always stays in a safer level than Portfolio 2, which is the optimal portfolio based on the 4 chosen stocks by the mean-variance approach. The difference is enhanced when considering a smaller tail probability. Hence our portfolio is safer than Portfolio 2 in extremal situation.

Comparing to the optimal portfolio based on all 15 stocks by the mean-variance approach, Portfolio 3, our portfolio does not perform better in the non-parametric estimation. However, from the EVT estimation, Portfolio 1 is better when the tail probability is 0.1% and 0.01%. Hence, our portfolio is better protected in extremal situation.

An extra comparison shows that the differences between Portfolio 1 and 3 are larger than the differences between Portfolio 1 and 2, when the tail probability is 0.1% and 0.01% in Table 5.5. This phenomenon agrees with the result in Proposition 5.3.1 in the following sense. Portfolio 3 assign positive weights to most of the 15 stocks. It must have a lower tail index than Portfolio 1 and 2 which are only based on the 4 stocks sharing the maximum tail index. Hence, the risk in extremal situation is eventually higher than the other two portfolios when the corresponding tail probability is small enough.

5.7 Conclusion and further extension

This chapter studies the portfolio selection in a multivariate EVT framework. By employing VaR as the risk criterion, we first link the VaR to the scale parameter of a heavy tailed distribution. Therefore, in the VaR optimization problem, alongside the primary risk indicator–tail index, the scale parameter can be recognized as the secondary risk indicator.

To study the portfolio selection and diversification effect problems, we uses multivariate EVT to model the tail dependence structure among the stocks non-parametrically. Compared to the CAPM model or other parametric dependence structures, this approach is much wider. Under the assumption that the joint stock returns belong to the domain of attraction of a multivariate extreme value distribution, we propose the following procedure to select the optimal portfolio.
1) By estimating the marginal tail indices (PRI), the optimal portfolio should be constructed from those have the maximum PRI.

2) Within these candidates, we estimate their marginal scale parameters (SRI) and dependence structure given by the spectral measure H on W. Notice that although for marginal estimations and the estimation of H, the number of upper order statistics in use, k, is always a considerable issue, the calculation of the portfolio scale is straightforward from the estimates.
3) If those equivalent tail indices are higher than 1, Theorem 5.3.1 shows how to calculate the SRI of a specific portfolio. Therefore, the portfolio selection problem can be numerically solved by minimizing the portfolio SRI.
4a) If those equivalent tail indices are not higher than 1, the optimal choice is to stay in the stock with the minimum SRI.
4b) If there are more than one stock sharing the same minimum SRI, Theorem 5.5.1 shows how to calculate the probability of dominance (POD) for each marginal. The optimal choice is to pick up the one who has the minimum POD. If the PODs are even equal among several stocks, there is no difference in choosing any of them, because they really have no difference not only in terms of the marginal risk, but also in terms of how they are connected with the systemic risk.

Our study leads to a similar discussion on the diversification effect as in Fama and Miller (1972) under more general conditions. For $\alpha = 1$, there is no diversification effect. For $\alpha < 1$, the diversification causes a negative effect that is the increase of the risk. For $\alpha > 1$, our conclusion depends on the dependence structure. Usually, the diversification leads to a decrease of the risk. However, if the individual securities are completely tail dependent, the diversification has no effect.

An empirical application shows that to construct an optimal portfolio with our procedure will indeed result in a portfolio which is more safe than the classical mean-variance approach in extremal situation.

In Section 5.2, we introduced the scale parameter as the SRI by restricting the tail distribution to a narrow case, i.e.

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

In the standard EVT setup, we face much wider situation: $t^{\alpha}(1-F(t))$ is a slowly varying function. In this case the scale may not be well defined. However, we can still define a relative scale for multivariate case. We present the definition and the result while the proofs are omitted.

Assume now that X belongs to the domain of attraction of a d-dimensional extreme value distribution, i.e. (5.2) holds. Similar to proposition 5.3.1, it is not difficult to prove that when the tail indices α_i are not equal, without defining SRI, the optimal portfolio is the individual stock who has the maximum tail index (PRI).

Suppose for all $1 \leq i \leq d$, $\alpha_i = \alpha$. In the narrow case where we can define the scale parameter, we have $(a_i^{(n)})^{\alpha_i}/n \to \sigma_i$ as $n \to \infty$, where $a_i^{(n)}$ are the normalization constants series in (5.2). In the general case, this limit does not necessarily exist. Hence, the scale can not be well-defined. Instead, there exist the scale functions $a_i(t)$, for $1 \leq i \leq d$, such that $a_i^{(n)} = a_i(n)$ for $1 \leq i \leq d$ and $a_i(t)$ is a regularly varying function with index $1/\alpha_i$.

Suppose all $a_i(t)$ are comparable, i.e. for any $1 \leq i, j \leq d$, $a_i(t)/a_j(t)$ converges to a finite positive number as $t \to \infty$. We can still define a relative scale as

$$\tilde{\sigma}_i = \lim_{t\to\infty} \frac{a_i(t)}{\sum_{j=1}^d a_j(t)},$$

for each $1 \leq i \leq d$. This relative scale can be recognized as a *pseudo secondary risk indicator (PSRI)*, because it can be proved that, when the marginal tail indices are all equal, i.e. $\alpha_i = \alpha$,

$$\tilde{\sigma}_i = \lim_{\delta\to 0} \frac{VaR_{X_i}(\delta)}{\sum_{j=1}^d VaR_{X_j}(\delta)}.$$

Therefore, the PSRI is a relative scale to compare the VaRs of the individuals.

For the portfolio selection and diversification effect problems, we define a relative scale for a portfolio similar to the PSRI above. For any portfolio, $P = \sum_{i=1}^d c_i X_i$, it can be proved that,

$$\tilde{\sigma}_P := \lim_{\delta\to 0} \frac{VaR_P(\delta)}{\sum_{j=1}^d VaR_{X_i}(\delta)}$$

is always a finite positive number. With this definition, Proposition 5.3.1 and Theorem 5.3.1 still holds by changing all σ_i into $\tilde{\sigma}_i$ and the scale of portfolio into $\tilde{\sigma}_P$. Therefore this is still a relative scale to compare the VaRs of the portfolios. The portfolio selection procedure in this chapter still applies by minimizing the PSRI. Notice that, the formula of POD does not include the marginal scales. Hence, it is not difficult to verify that Theorem 5.5.1 still holds under the case that all PRIs are equal and not higher than 1, and all PSRIs are also equal.

To summarize, although our approach starts from the definition of scale which requires some extra condition, it can be relaxed to a much wider case where the marginal scale functions are comparable. In this case, a relative scale plays a role as a PSRI, which leads to a similar procedure in portfolio selection.

5.A Appendix A

Proof of Proposition 5.3.1

Since X belongs to the domain of attraction of a $d-$dimensional extreme value distribution with positive marginal tail indices α_i, $1 \le i \le d$, X is a regularly varying random vector with tail index $\min_{1\le i\le d} \alpha_i$. For the definition of regularly variation for a random vector, see, e.g. Basrak *et al.* (2002). From Theorem 1.1 in Basrak *et al.* (2002), the tail index of a linear combination of a regularly varying random vector is dominated by the minimum marginal tail index. Hence, $\sum_{i=1}^d c_i X_i$ must have tail index α_1. Now, we only need to prove for the scale part.

From the domination of the minimum tail index theorem mentioned above, $\sum_{i=2}^d c_i X_i$ must have the tail index $\min_{2\le i\le d} \alpha_i > \alpha_1$. For any given $0 < \delta < 1$,

$$P(\sum_{i=1}^d c_i X_i > t) \le P(c_1 X_1 > (1-\delta)t) + P(\sum_{i=2}^d c_i X_i > \delta t).$$

Therefore,

$$\limsup_{t\to\infty} t^{\alpha_1} P(\sum_{i=1}^d c_i X_i > t) \le \frac{c_1^{\alpha_1}\sigma(X_1)}{(1-\delta)^{\alpha_1}}.$$

By taking $\delta \to 0$, we have

$$\limsup_{t\to\infty} t^{\alpha_1} P(\sum_{i=1}^d c_i X_i > t) \le c_1^{\alpha_1}\sigma(X_1).$$

Meanwhile,

$$\liminf_{t\to\infty} t^{\alpha_1} P(\sum_{i=1}^d c_i X_i > t) \ge \liminf_{t\to\infty} t^{\alpha_1} P(c_1 X_1 > t) = c_1^{\alpha_1}\sigma(X_1).$$

Combining the lower and upper boundaries, the proposition is proved. □

Proof of Proposition 5.3.2 with a general H

We start from approximating H by a series of finite discrete measures. Suppose $\{H^{(m)}\}_{m=1}^{\infty}$ is a series of probability measures on W such that
1) $H^{(m)}$ is a finite(no more than m) points discrete probability measure on W.
2) $H^{(m)} \xrightarrow{v} H$ as $m \to \infty$.
Here $\xrightarrow{v}$ means the vague convergence. The existence of this kind of measure series can be proved by a real construction. Subdividing W into m $H-$measurable sets, and concentrating the measures in each set on its gravity center, we get a suitable series $H^{(m)}$.

By a similar proof as Proposition 5.3.2 with a discrete H in Section 5.3.2, we can get similar result for $H^{(m)}$, though the marginal expectations of $H^{(m)}$ are not necessarily $1/d$.

Suppose

$$H^{(m)} = \sum_{j=1}^{m} p_j \delta_{a_j},$$

where $\left\{a_j = (a_{1j}, \cdots, a_{dj})^T\right\}_{j=1}^m$ are m points on W, i.e. $\sum_{i=1}^d a_{ij} = 1$, and $\sum_{j=1}^m p_j = 1$. Denote, for any $1 \leq i \leq d$,

$$\sigma_i^{(m)} = \int_W w_i H^{(m)}(dw) = \sum_{j=1}^{m} p_j a_{ij}.$$

As $m \to \infty$, $H^{(m)} \xrightarrow{v} H$ implies that $\sigma_i^{(m)} \to 1/d$.

Similarly, we construct a $d-$dimensional random vector using the i.i.d. standard Fréchet distributed random variable $V_1, V_2, \cdots$. Define, for $1 \leq i \leq d$,

$$U_i^{(m)} = d \bigvee_{j=1}^{m} p_j a_{ij} V_j.$$

Then, it is not difficult to verify that $U^{(m)} = (U_1^{(m)}, \cdots, U_d^{(m)})^T$ has the distribution function

$$G^{(m)}(x_1, \cdots, x_d) = \exp\left(-d \int_W \left(\frac{w_1}{x_1} \vee \cdots \vee \frac{w_d}{x_d}\right) H^{(m)}(dw)\right).$$

Because $H^{(m)} \xrightarrow{v} H$ as $m \to \infty$, for the random vector, we have $U^{(m)} \xrightarrow{d} U$.

Similar to the proof for discrete H, we can prove that

$$\sigma(\sum_{i=1}^{d} c_i U_i^{(m)}) = \sum_{i=1}^{d} c_i d \sigma_i^{(m)}.$$

From the fact that $\sigma_i^{(m)} \to 1/d$ and $U^{(m)} \xrightarrow{d} U$ as $m \to \infty$, by taking $m \to \infty$ we proved the proposition for the general case of H. □

For the rest of the appendix we need to introduce an alternative way to study the dependence structure of a simple max-stable distribution: the exponent measure. For detail, see e.g. de Haan and Ferreira (2006).

Denote $\mathbb{R}_+ = [0, \infty)$. Suppose ν is a measure defined for all Borel sets $A \subset \mathbb{R}_+^d$ with

$$\inf_{(x_1, \cdots, x_d) \in A} \max(x_1, \cdots, x_d) > 0 \tag{5.5}$$

such that

1) Homogeneity: for any Borel set A satisfying (5.5) and $\nu(\partial A) = 0$, and any $a > 0$,

$$\nu(aA) = a^{-1} \nu(A).$$

2) Marginal condition: for any $1 \leq i \leq d$ and any $x > 0$,

$$\nu \left\{ (x_1, \cdots, x_d) \in \mathbb{R}_+^d \ : \ x_i > x \right\} = 1/x.$$

By denoting

$$A_{x_1,\cdots,x_d} = \{(s_1, \cdots, s_d) \ : \ \exists 1 \leq i \leq d \ \text{ s.t. } \ s_i > x_i\},$$

we can get a simple max-stable distribution G as $G(x_1, \cdots, x_d) = \exp(-\nu(A_{x_1,\cdots,x_d}))$. Conversely, any simple max-stable distribution G has such a representation with a suitable ν. ν is called the *exponent measure*.

These two measures H and ν can be transformed from one to the other in the following way. A point $x = (x_1, \cdots, x_d) \in \mathbb{R}_+^d / \{(0, \cdots, 0)\}$ can be mapped to $(r, w) \in (0, \infty) \times W$ by $r = \sum_{1 \leq i \leq d} x_i$ and $w = x/(\sum_{i=1}^d x_i)$. It is a one-to-one mapping. Denote this mapping as π. For any Borel set A satisfying (5.5), $\pi(A)$ must be a Borel set in $(0, \infty) \times W$. In particular, for any $(x_1, \cdots, x_d)^T \in \mathbb{R}_+^d / \left\{(0, \cdots, 0)^T\right\}$,

$$\begin{aligned} \pi(A_{x_1,\cdots,x_d}) &= \{(r, w) \ : \ \exists 1 \leq i \leq d \ \text{ s.t. } rw_i > x_i\} \\ &= \left\{ (r, w) \ : \ r > \frac{x_1}{w_1} \wedge \cdots \wedge \frac{x_d}{w_d} \right\}. \end{aligned}$$

Hence, we have

$$\begin{aligned} \nu(A_{x_1,\cdots,x_d}) &= -\log G(x_1, \cdots, x_d) \\ &= d \int_W \frac{w_1}{x_1} \vee \cdots \vee \frac{w_d}{x_d} H(dw) \\ &= d \int_W \left(\int_{\frac{x_1}{w_1} \wedge \cdots \wedge \frac{x_d}{w_d}}^{\infty} \frac{1}{r^2} dr \right) H(dw) \\ &= d \int_{(r,w) \in \pi(A_{x_1,\cdots,x_d})} \frac{1}{r^2} dr H(dw). \end{aligned}$$

This relation holds for any set $A_{x_1,\cdots,x_d}$, therefore, it also holds for any other Borel set A satisfying (5.5), i.e.

$$\nu(A) = d \int_{(r,w) \in \pi(A)} \frac{1}{r^2} dr H(dw). \tag{5.6}$$

From Proposition 5.3.2, we can get the following corollary.

Corollary 5.A.1 *For any $x > 0$ and $\tilde{A}_x := \left\{ (x_1, \cdots, x_d) \in \mathbb{R}_+^d \ : \ \sum_{i=1}^d c_i x_i > x \right\}$,*

$$\nu(\tilde{A}_x) = \frac{\sum_{i=1}^d c_i}{x}.$$

Hence the measure ν on the set $\tilde{A}_x$ does not depend on how the weights are distributed but only on their sum.

Proof of Corollary 5.A.1

Considering that G is in the domain of attraction of itself, i.e.

$$\lim_{n\to\infty} n\left(1 - P\left(\frac{U_1}{n} \le x_1, \cdots, \frac{U_d}{n} \le x_d\right)\right) = -\log G(x_1, \cdots, x_d) = \nu(A_{x_1,\cdots,x_d})$$

holds for any $(x_1, \cdots, x_d) \in \mathbb{R}_+^d/(0, \cdots, 0)$, a similar relation must hold for other Borel sets A satisfying (5.5). In particular, for the set $\tilde{A}_x := \left\{(x_1, \cdots, x_d) \in \mathbb{R}_+^d \ : \ \sum_{i=1}^d c_i x_i > x\right\}$ where $x > 0$, we have

$$\lim_{n\to\infty} nP\left(\left(\frac{U_1}{n}, \cdots, \frac{U_d}{n}\right) \in \tilde{A}_x\right) = \nu(\tilde{A}_x),$$

Notice that

$$\lim_{n\to\infty} nP\left(\left(\frac{U_1}{n}, \cdots, \frac{U_d}{n}\right) \in \tilde{A}_x\right) = \lim_{n\to\infty} nP\left(\frac{\sum_{i=1}^d c_i U_i}{n} > x\right) = \frac{\sigma(\sum_{i=1}^d c_i U_i)}{x}.$$

Hence, from Proposition 5.3.2, the corollary is proved.

An alternative proof of corollary 5.A.1 is carried out by straightforward calculation from (5.6) as follows.

$$\begin{aligned}
\nu(\tilde{A}_x) &= d\int_{\pi(\tilde{A}_x)} \frac{1}{r^2} dr H(dw) \\
&= d\int_{r > \frac{x}{\sum_{i=1}^d c_i w_i}} \frac{1}{r^2} dr H(dw) \\
&= d\int_W \frac{\sum_{i=1}^d c_i w_i}{x} H(dw) \\
&= \frac{\sum_{i=1}^d c_i}{x}. \square
\end{aligned}$$

Proof of Theorem 5.3.1

We prove Theorem 5.3.1 by using the exponent measure ν introduced above.

Suppose X satisfies (5.2) with all marginal tail indices α. Define $Y = (Y_1, \cdots, Y_d)^T$ by $Y_i = X_i^\alpha$, then for each marginal Y_i, the tail index is 1, and the scale is the same as σ_i. Denote $Z_i = Y_i/\sigma_i$, we have that $Z = (Z_1, \cdots, Z_d)^T$, Y and X share the same simple max-stable distribution G, where Z has unit scales. Because the tail index of $\sum_{i=1}^d c_i X_i$ is alpha, the scale of $\sum_{i=1}^d c_i X_i$ can be calculated as

$$\begin{aligned}
\sigma(\sum_{i=1}^d c_i X_i) &= \lim_{t\to\infty} t^\alpha P\left(\sum_{i=1}^d c_i X_i > t\right) \\
&= \lim_{t\to\infty} t^\alpha P\left(\sum_{i=1}^d c_i (\sigma_i Z_i)^{1/\alpha} > t\right)
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} nP\left(\sum_{i=1}^{d} c_i(\sigma_i Z_i)^{1/\alpha} > n^{1/\alpha}\right) \\
&= \lim_{n\to\infty} nP\left(\sum_{i=1}^{d} c_i\left(\sigma_i \frac{Z_i}{n}\right)^{1/\alpha} > 1\right) \\
&= \lim_{n\to\infty} nP\left(\left(\frac{Z_1}{n}, \frac{Z_2}{n}, \cdots, \frac{Z_d}{n}\right) \in A^*\right),
\end{aligned}
$$

where $A^* = \left\{(x_1, \cdots, x_d) \ : \ \sum_{i=1}^{d} c_i(\sigma_i x_i)^{1/\alpha} > 1\right\}$.

Because Z is in the domain of attraction of G and Z has tail indices 1 and unit scales, we have, similar to the proof of Corollary 5.A.1,

$$
\lim_{n\to\infty} nP\left(\frac{Z_1}{n} \le x_1, \cdots, \frac{Z_d}{n} \le x_d\right) = -\log G(x_1, \cdots, x_d) = \nu(A_{x_1,\cdots,x_d})
$$

holds for any $(x_1, \cdots, x_d) \in \mathbb{R}_+^d/(0, \cdots, 0)$. Hence it holds for any Borel set A satisfying (5.5), in particular, A^*. Therefore, the calculation of the portfolio scale can be continued as

$$
\begin{aligned}
\sigma\left(\sum_{i=1}^{d} c_i X_i\right) &= \nu(A^*) \\
&= d\int_{(r,w)\in\pi(A^*)} \frac{1}{r^2} dr H(dw) \\
&= d\int_{\sum_{i=1}^{d} c_i(\sigma_i r w_i)^{1/\alpha} > 1} \frac{1}{r^2} dr H(dw) \\
&= d\int_{r > \left(\sum_{i=1}^{d} c_i(\sigma_i w_i)^{1/\alpha}\right)^{-\alpha}} \frac{1}{r^2} dr H(dw) \\
&= d\int_{W} \left(\sum_{i=1}^{d} c_i(\sigma_i w_i)^{1/\alpha}\right)^{\alpha} H(dw).
\end{aligned}
$$

The proof of Theorem 5.3.1 is complete. □.

Proof of Theorem 5.5.1

To calculate the POD, we start with the same notation $Z_i = X_i^{\alpha}/\sigma$ as in the proof of Theorem 5.3.1. Then $Z = (Z_1, \cdots, Z_d)^T$ and X share the same simple max-stable distribution G, where Z has standardized tail indices and scales. Hence,

$$
\begin{aligned}
p_i &= \lim_{t\to\infty} P\left(X_i = \bigvee_{i=1}^{d} X_i \,\Big|\, \bigvee_{i=1}^{d} X_i > t\right) \\
&= \lim_{t\to\infty} P\left(Z_i = \bigvee_{i=1}^{d} Z_i \,\Big|\, \bigvee_{i=1}^{d} Z_i > t\right)
\end{aligned}
$$

$$= \lim_{t\to\infty} \frac{P(Z \in A_i \bigcap B_t)}{P(Z \in B_t)},$$

where $A_i =: \left\{(x_1, \cdots, x_d) \in \mathbb{R}_+^d \ : \ x_i = \bigvee_{s=1}^d x_s\right\}$ and $B_t := \left\{(x_1, \cdots, x_d) \in \mathbb{R}_+^d \ : \ \bigvee_{s=1}^d x_s > t\right\}$. Suppose ν is the corresponding exponent measure. Since Z is in the domain of attraction of G, from the proof of Theorem 5.3.1, we have that, for any Borel set $A \subset \mathbb{R}_+^d$ satisfying (5.5),

$$\lim_{t\to\infty} tP(Z \in tA) = \nu(A).$$

Notice that $B_t = tB_1$, both B_1 and $A_i \bigcap B_1$ satisfy (5.5). We conclude that

$$\lim_{t\to\infty} tP(Z \in A_i \bigcap B_t) = \nu(A_i \bigcap B_1),$$

and

$$\lim_{t\to\infty} tP(Z \in B_t) = \nu(B_1),$$

Finally, we get

$$p_i = \nu(A_i \bigcap B_1)/\nu(B_1).$$

To continue with the calculation, we use (5.6) to transform the ν measure to the H measure.

$$\begin{aligned}
\nu(A_i \bigcap B_1) &= d \int_{\pi(A_i \bigcap B_1)} \frac{1}{r^2} dr H(dw) \\
&= d \int_{rw_i > 1, w_i = \max_{1\le i\le d} w_i} \frac{1}{r^2} dr H(dw) \\
&= d \int_{w_i = \max_{1\le i\le d} w_i} w_i H(dw).
\end{aligned}$$

Similarly,

$$\nu(B_1) = d \int_W \max_{1\le i\le d} w_i H(dw).$$

Combining these two, the proof is complete. □

Part III

Extremal Rainfall

Chapter 6

On Spatial Extremes: with Application to a Rainfall Problem

6.1 Introduction

Extreme rainfall statistics are frequently used when a damaging flood has occurred to answer questions about the rarity of the event. Engineers often need extreme rainfall statistics for the design of structures for flood protection. A typical question is e.g. what is the amount of rain in a given area on one day that is exceeded once in 100 years? Or, more mathematically, what is the 100-year quantile of the total rainfall in the area on one day? In this chapter this question is investigated for a low-lying flat area in the northwest of the Netherlands. The area is shown in Figure 6.1. Because it roughly covers the province of North Holland, it will shortly be indicated as North Holland.

There are 32 rainfall stations in the area for which daily data were available for the 30-year period 1971-2000. Only the fall season, i.e. the months September, October and November, is considered. In this season the likelihood of flooding and its impact are relatively large. Because of the restriction to the fall season it is reasonable to assume stationarity in time. Stationarity in space, except for location and scale, is also assumed.

Since we have to extrapolate from a 30-year to a 100-year period, our problem is an extreme value problem. There is also a clear spatial aspect.

Engineers often make use of areal reduction factors (ARFs) to convert quantiles for point rainfall to the corresponding quantiles of areal rainfall. ARFs have been derived empirically by estimating the areal rainfall as a function of point rainfall measurements (e.g., NERC (1975); Bell (1976)) or by statistical modelling (e.g., Bacchi and Ranzi (1996); Sivapalan and Blöschl (1998); Veneziano and Langousis (2005)). The latter requires assumptions on distributions, spatial correlation and/or scaling behavior. The resulting ARF for the 100-year quantile is generally very uncertain.

Some attempts have been made to estimate ARFs from weather radar data (Allen

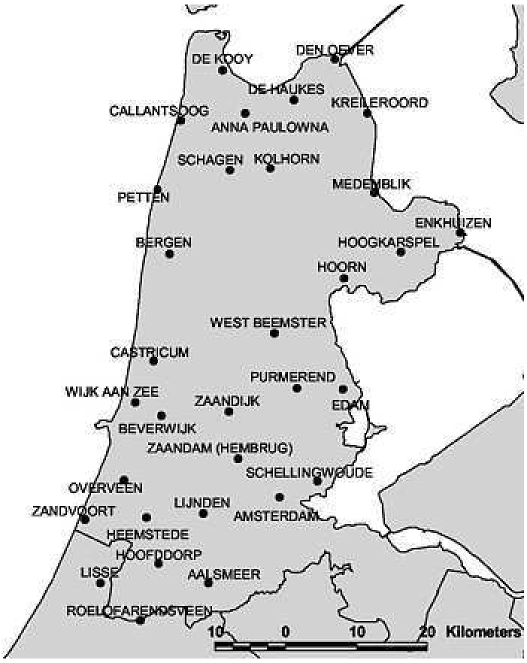

Figure 6.1: The study area: North Holland

and Allen and DeGaetano (2005); Stewart (1989)). One difficulty is that the raw rainfall intensities from the radar reflectivities need to be adjusted for systematic deviations from the values observed at the rainfall stations. Another difficulty is that archived radar data cover a relatively short time period (in the Netherlands only 10 years).

Regional Climate Model (RCM) simulations driven by weather reanalysis data are a potential source for areal aggregated rainfall. A reanalysis is an estimate of the state of the atmosphere based on observations and a numerical weather forecast. The RCM is necessary to increase the spatial resolution. Various 40-year simulations have been performed recently with spatial resolutions of 50 km $\times$ 50 km and 25 km $\times$ 25 km, in particular within the framework of the EU-funded project ENSEMBLES (www.ensembles-eu.org). In addition to the limited length and rather coarse resolution for our application, there are systematic differences between simulated and observed rainfall. For the KNMI RCM driven by ERA40 reanalysis data, Leander and Buishand (2007) report differences up to 20% in seasonal average rainfall for the river Meuse basin, situated south of the Netherlands. For North Holland the differences can even be larger because it is much smaller than the Meuse basin and it is surrounded by water.

Statisticians have used max-stable processes to obtain the quantiles of the distribution of spatially aggregated rainfall. Coles and Tawn, in a series of papers (Coles (1993); Coles and Tawn (1996)) have developed methods to deal with spatial extremes based on the spectral representation (de Haan (1984b); see also Schlather (2002)). Here we follow a different approach based on random fields. Apart from the parameters that characterize the upper tail of the marginal distributions our model has a parameter that determines spatial dependence. This parameter is estimated from the tails of the empirical two-dimensional marginal distributions of daily rainfall in North Holland. The 100-year quantile of the total rainfall over the area is found by simulating synthetic daily rainfall fields using the estimated model.

In order to motivate our solution we first explain some relevant aspects of extreme value theory, in $\mathbb{R}^1$, $\mathbb{R}^d$ $(d > 1)$ and $C[0,1]$ (Section 6.2). In Section 6.3, we specify the stochastic process used in the simulation. This process is used only to simulate "extreme" rainfall. For non-extreme rainfall we sample from the available data. In Section 6.4 we explain how we combine the two to get a simulated day of rainfall. The estimation of the dependence parameter is dealt with in Section 6.5. Section 6.6 discusses the outcome of the simulation and the answer to our problem. Section 6.7 summarizes our main conclusions.

6.2 Extreme value background

We now explain the background of our approach by reviewing some aspects of extreme value theory and the related theory of excursions over a high threshold. This will be done first in the one-dimensional case (Section 6.2.1), then the finite-dimensional case (Section 6.2.2) and finally the case of continuous stochastic processes (Section 6.2.3). The results in the various cases are similar but of increasing complexity. That is why we start with the one-dimensional case which is well-known (Gnedenko (1943) and Pickands III (1975) respectively).

6.2.1 One-dimensional space

Suppose that the distribution function F is in the domain of attraction of an extreme value distribution, i.e., if $X_1, X_2, \cdots$ are i.i.d. with distribution function F, there are a positive function a and a function b, such that

$$\lim_{n\to\infty} P\left(\max_{1\le i\le n} \frac{X_i - b(n)}{a(n)} \le x\right) = G(x)$$

a non-degenerate distribution function. We denote this by $F \in \mathcal{D}$. Then a and b can be chosen such that

$$G(x) = G_\gamma(x) = \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\}$$

for all x with $1 + \gamma x > 0$. Then we also say $F \in \mathcal{D}(G_\gamma)$.

Let X be a random variable with distribution function F. Then there exists a positive function a_0 and real shape parameter γ (the *extreme value index*), such that for all x with

$1 + \gamma x > 0$,

$$\lim_{t \uparrow x^*} P\left(\frac{X-t}{a_0(t)} > x | X > t\right) = (1+\gamma x)^{-1/\gamma} =: 1 - Q_\gamma(x).$$

Here $x^* := \sup\{x : F(x) < 1\}$. This means that the larger observations in a sample follow approximately the probability distribution Q_γ - the generalized Pareto distribution (GPD, c.f. Bajari and Hortaçsu (2003); Pickands III (1975)). Note that $1-Q_\gamma(x) = -\log G_\gamma(x)$.

Let R be a random variable with distribution function Q_γ. Then,

$$P\left(\frac{R-t}{1+\gamma t} > x | R > t\right) = P(R > x)$$

for x and t such that $1+\gamma t > 0$ and $1+\gamma x > 0$. We call this property *excursion stability.*

Suppose that we have observed a sample $X_1, X_2, \cdots, X_n$ from F. Since the approximate distribution of the large values is completely specified, it is possible to use it as a basis to simulate more "large observations", even larger than those in the sample. Thus, by resampling the non-extreme part of the sample and simulating extreme observations from the GPD distribution one can produce more and more "observations", even extreme ones. Using partly simulation and partly resampling is the main idea behind what we intend to do. Hence we sample from

$$\bar{F}(x) = \begin{cases} F(x) & \text{if } x < t, \\ 1 - (1-F(t))Q_\gamma\left(\frac{x-t}{a_0(t)}\right) & \text{if } x \geq t. \end{cases} \tag{6.1}$$

We can implement this by letting t be one of the upper order statistics and using estimators for F, γ and a_0.

The extra "extreme" observations are sampled from the tail model and they are independent of the "non-extreme" observations. This is justified by the so-called "découpage de Lévy" stating roughly that cutting up a sequence of i.i.d. random variables in two subsequences according to whether their values are in a set B or in its complement B^c, results in two independent i.i.d. sequences. The result has been formulated and proved carefully in Resnick (1987), pages 212 and 215. A similar argument applies in higher dimensional space (Section 6.2.2 and 6.2.3).

6.2.2 Finite-dimensional space

Let us now consider the finite-dimensional case, or rather the two-dimensional case for simplicity. Let (X, Y) be a random vector with distribution function F. Suppose $F \in \mathcal{D}$, i.e. if $(X_1, Y_1), (X_2, Y_2), \cdots$ are i.i.d. with distribution function F, there are positive functions a and c and functions b and d, such that

$$\lim_{n\to\infty} P\left(\max_{1\leq i\leq n} \frac{X_i - b(n)}{a(n)} \leq x, \max_{1\leq i\leq n} \frac{Y_i - d(n)}{c(n)} \leq y\right) = G(x, y),$$

a distribution function with non-degenerate marginals. If this is the case, we say $F \in \mathcal{D}(G)$ and G is a (multivariate) extreme value distribution. Then, as in the one-dimensional case, there exists a related 2-dimensional GPD distribution Q_H, obtained for example as follows:

$$\begin{aligned}&\lim_{t\to\infty} P\left(\frac{X-b(t)}{a(t)} > \frac{x^{\gamma_1}-1}{\gamma_1} \quad \text{or} \quad \frac{Y-d(t)}{c(t)} > \frac{y^{\gamma_2}-1}{\gamma_2} | X > b(t) \quad \text{or} \quad Y > d(t)\right)\\ =&2\int_0^1 \max\left(\frac{s}{x}, \frac{1-s}{y}\right) H(ds) =: 1 - Q_H(x,y),\end{aligned}$$

for $(x, y) \in D_H = \left\{(x,y) : 2\int_0^1 \max\left(\frac{s}{x}, \frac{1-s}{y}\right) H(ds) \leq 1\right\} \supset \{(x,y) : x, y \geq 2\}$, where γ_1 and γ_2 are the marginal extreme value indices, and H is a probability distribution function on $[0, 1]$ with mean $1/2$. Any distribution H with mean $1/2$ may occur (see e.g., de Haan and Ferreira (2006, Chapter 6)). H characterizes the dependence in the tail. It is different from the traditional dependence measure: correlation coefficient which measures the dependence at moderate levels. If H concentrates all its measure on point $1/2$, (X, Y) is completely tail dependent. If H is a discrete measure on only two points: 0 and 1 with weight $1/2$ each, (X, Y) is completely tail independent. Similar to the one-dimensional case we have

$$-\log G\left(\frac{x^{\gamma_1}-1}{\gamma_1}, \frac{y^{\gamma_2}-1}{\gamma_2}\right) = 1 - Q_H(x,y).$$

Q_H is a probability distribution function on D_H with the properties:

1. Standard one-dimensional GPD marginals: $Q_H(x, \infty) = Q_H(\infty, x) = 1 - 1/x$, for $x \geq 1$.

2. Homogeneity: $1 - Q_H(tx, ty) = t^{-1}(1 - Q_H(x, y))$ for $t > 1$ and $(x, y) \in D_H$, in particular $Q_H \in \mathcal{D}$:

$$Q_H^n(nx, ny) = (1-(1-Q_H(x,y))/n)^n \to \exp\{-(1-Q_H(x,y))\} = G\left(\frac{x^{\gamma_1}-1}{\gamma_1}, \frac{y^{\gamma_2}-1}{\gamma_2}\right).$$

Sometimes the function $1 - Q_H(1/x, 1/y)$ is called the *asymptotic dependence function* of F. It determines the tail dependence between the two components without specifying the marginal distributions.

3. Excursion stability: If (R, S) is a random vector with distribution function Q_H, then with $c := 1 - Q_H(1, 1)$, we have for $x, y \in D_H$, $t > c$

$$P\left(R > \frac{tx}{c} \quad \text{or} \quad S > \frac{ty}{c} | R > t \quad \text{or} \quad S > t\right) = P(R > x \quad \text{or} \quad S > y).$$

We remark that a random vector with an arbitrary extreme value distribution can be constructed as follows. Let $E_1, E_2, \cdots$ be i.i.d. standard exponential random variables. Let V be a random variable with distribution function H and consider i.i.d. copies

$V_1, V_2, \cdots$ of V. Let the sequences $\{E_i\}$ and $\{V_i\}$ be independent. Then the random vector

$$\left(\max_{i\geq 1} 2V_i/(E_1+E_2+\cdots+E_i), \max_{i\geq 1} 2(1-V_i)/(E_1+E_2+\cdots+E_i)\right)$$

has an extreme value distribution with marginal distribution functions $\exp(-1/x), x > 0$.

We want to follow the line of reasoning from the one-dimensional situation and propose to use Q_H to simulate more "large observations", to be combined with resampling from the available sample. However, simulation from a multivariate distribution is more complicated than in the one-dimensional case. It is more convenient if we can find a random vector that is easy to simulate and that has the same distribution. Consider the random vector $(2YV, 2Y(1-V))$ with Y and V independent, Y has distribution function $1-1/x$, $x \geq 1$ and V has distribution function H. It is easy to check that the distribution function $Q_H^0(x, y)$ of $(2YV, 2Y(1-V))$ coincides with $Q_H(x, y)$ for $x, y \geq 2$. The fact that the distribution function is not exactly the same is not a problem: we are dealing with an asymptotic property and the important thing is that Q_H^0 has the asymptotic dependence function $1 - Q_H(1/x, 1/y)$, i.e.

$$\lim_{t\to\infty} P\left(2YV > tx \;\text{ or }\; 2Y(1-V) > ty | 2YV > t \;\text{ or }\; 2Y(1-V) > t\right) = 1 - Q_H(x, y) \quad (6.2)$$

for $x, y > 1$. In fact any distribution function in the domain of attraction of G would do since the asymptotic dependence structure is the same as for the limiting extreme value distribution.

Now the random vector $(2YV, 2Y(1-V))$ is useful but not flexible enough: the set of conditions $V \in [0, 1]$ and $EV = 1/2$ is rather restrictive. Hence let us consider the random vector (YA_1, YA_2) with Y and the vector (A_1, A_2) independent, Y as before and A_1 and A_2 positive with $EA_1 = EA_2 = 1$. The distribution function Q^* of (YA_1, YA_2) satisfies the following properties.

1^*. $1 - Q^*(x, \infty) = E\min\left(1, \frac{A_1}{x}\right)$ for $x > 0$, hence $\lim_{t\to\infty} t(1 - Q^*(tx, \infty)) = 1/x$, similarly for $Q^*(\infty, x)$;

2^*. $\lim_{t\to\infty} t(1 - Q^*(tx, ty)) = E\max\left(\frac{A_1}{x}, \frac{A_2}{y}\right)$ for $x, y > 0$, i.e. $Q^* \in \mathcal{D}$;

3^*.

$$\lim_{t\to\infty} P(YA_1 > tx/c \;\text{ or }\; YA_2 > ty/c | YA_1 > t \;\text{ or }\; YA_2 > t) = E\max\left(\frac{A_1}{x}, \frac{A_2}{y}\right)$$

for $x, y > 0$ with $c := E\max(A_1, A_2)$.

We can easily simulate from Q^*, but this distribution satisfies only approximately (not exactly) the three properties 1, 2 and 3. Because of property 2^* (meaning that the distribution function of (YA_1, YA_2) has the same asymptotic dependence function as the distribution function of $\max\left(0, 1 - E\max\left(\frac{A_1}{x}, \frac{A_2}{y}\right)\right)$, c.f. (6.2)), we can still use Q^* for simulation albeit with caution.

6.2.3 Extremes of continuous stochastic processes

What do we mean by extremes in $C[0,1]$, the space of continuous functions defined on the unit interval? The setup is as follows. Let $\{X(s)\}_{s\in[0,1]}$ be a stochastic process in $C[0,1]$. Consider independent copies $X_1, X_2, \cdots$ of the process X. Compose for each n a continuous stochastic process

$$\left\{\max_{1\le i\le n} X_i(s)\right\}_{s\in[0,1]}.$$

Suppose that for some positive functions $a_s(n)$ and real functions $b_s(n)$, the sequence of processes

$$\left\{\max_{1\le i\le n} \frac{X_i(s)-b_s(n)}{a_s(n)}\right\}_{s\in[0,1]}$$

converges in $C[0,1]$. If this is the case, we say $X \in \mathcal{D}$. Let us call the limiting process $\{U(s)\}_{s\in[0,1]}$. Then we say $X \in \mathcal{D}(U)$. The following proposition is useful for our purposes (de Haan and Lin (2001)).

Proposition 6.2.1 *$X \in \mathcal{D}$ if and only if the following two statements hold:*
1. (uniform convergence of the marginal distributions) There exists a continuous function $\gamma(s)$ such that, for $x > 0$

$$\lim_{n\to\infty} P\left(\max_{1\le i\le n} \frac{X_i(s)-b_s(n)}{a_s(n)} \le \frac{x^{\gamma(s)}-1}{\gamma(s)}\right) = \exp\left(-\frac{1}{x}\right),$$

uniformly for $s \in [0,1]$.
2. (convergence of the standardized process) With $F_s(x) := P(X(s) \le x)$ for $s \in [0,1]$,

$$\left\{\max_{1\le i\le n} \frac{1}{n(1-F_s(X_i(s)))}\right\} \xrightarrow{d} \{\eta(s)\} \quad \textit{(say)}$$

in $C[0,1]$. Note that all one-dimensional marginal distributions of the process $1/(1-F_s(X_i(s)))$ are equal to $1-1/x$, $x \ge 1$.

The process η satisfies: if $\eta_1, \eta_2, \cdots$ are i.i.d. copies of η, then

$$\frac{1}{n}\max_{1\le i\le n} \eta_i \stackrel{d}{=} \eta,$$

i.e. the process is *simple max-stable.* (The word "simple" indicates that all marginal distributions are standard Fréchet distributions, $\exp(-1/x), x > 0$.) Moreover, we have that

$$\{U(s)\} \stackrel{d}{=} \left\{\frac{(\eta(s))^{\gamma(s)}-1}{\gamma(s)}\right\}.$$

As a consequence of this proposition, we can study the "simple" process η first and go back to U later, in a straightforward way.

Two characterizations of simple max-stable processes are known. One of them can serve our purposes. It is given in the following proposition. The other characterization is discussed at the end of this subsection.

Proposition 6.2.2 *(Schlather (2002), de Haan and Ferreira (2006) Corollary 9.4.5) Every simple max-stable process η in $C[0,1]$ can be generated in the following way. Let $E_1, E_2, \cdots$ be i.i.d. standard exponential random variables. Further consider i.i.d. positive stochastic processes $V, V_1, V_2, \cdots$ in $C[0,1]$ with $EV(s) = 1$ for all $s \in [0,1]$ and $E \sup_{0 \le s \le 1} V(s) < \infty$. Let the sequences $\{E_i\}$ and $\{V_i\}$ be independent. Then*

$$\eta \stackrel{d}{=} \max_{i \ge 1} V_i/(E_1 + E_2 + \cdots + E_i).$$

Conversely, each process with this representation is simple max-stable. One can take the stochastic process V such that

$$\sup_{0 \le s \le 1} V(s) = c \quad a.s.$$

with c some positive non-random constant.

Now recall the "generalized Pareto" results in one- and finite-dimensional extremes, that allowed us to simulate from the tail of the distribution. What is the situation in this spatial setup?

One way to proceed is as in the finite-dimensional case. Let Y be a random variable with distribution function $1 - 1/x$, $x \ge 1$ (i.e. one-dimensional GPD). Let V be a positive stochastic process in $C[0,1]$ that satisfies the conditions of Proposition 6.2.2: $EV(s) = 1$ for $s \in [0,1]$ and $\sup_{0 \le s \le 1} V(s) = c$, a non-random constant. Let Y and V be independent. Consider *the GPD process*

$$\{\xi(s)\}_{s \in [0,1]} := \{YV(s)\}_{s \in [0,1]}\,.$$

The process ξ is in $C[0,1]$ and satisfies

1. Standard GPD tail: $P(YV(s) > x) = 1/x$ for $x > c$;
2. Homogeneity;
3. Excursion stability: The distribution of $\{cYV(s)/t\}$ given $\sup_{0 \le s \le 1} YV(s) > t$ is the same as that of $\{YV(s)\}$ for $t > c$.

The validity of the three properties requires the condition that $\sup_{0 \le s \le 1} V(s) = c$, a non-random constant. If we only know $E \sup_{0 \le s \le 1} V(s) < \infty$, the properties do not hold as they stand, but we still have an asymptotic version of them as in the finite-dimensional case. In particular, $\xi \in \mathcal{D}(\eta)$.

We remark that the stochastic process $\{YV(s)\}$ is in the domain of attraction of the process $\{\eta(s)\}$, hence the asymptotic dependence structure of the two processes is the same (c.f. Section 6.2.2) and either of the processes can be used for simulating extreme events. This is also true for the process $\{YV(s)\}$ with the weaker side condition $E\sup_{0\le s\le 1} V(s) < \infty$. Hence there are three candidate processes for simulating extremal rainfall.

We finish this section with two remarks.

In all of the above we can replace $[0, 1]$ by any compact subset of an Euclidean space, i.e. we can deal with spatial extremes.

An alternative approach to areal rainfall is presented in Coles (1993) and Coles and Tawn (1996), as mentioned before. Rather than the representation of Proposition 6.2.2, their approach is based on an alternative, more analytical, representation involving spectral functions originating from Drees and Huang (1998) and an unpublished manuscript by Smith (1990). They sketch how to calculate (rather than simulate) a quantile of the areal rainfall. The model developed by Coles (1993) consists of a multivariate extreme-value distribution that describes the extremes at a subset of the rainfall stations and deterministic "storm profile functions" to obtain the amount of rain in the remaining points of the area. A consequence of this approach is that the model depends on the positions of the rainfall stations. Coles and Tawn (1996) applied this model to calculate quantiles of extreme daily areal rainfall in the winter season for a region in south-west England. It is assumed that heavy rainfall always takes place throughout the region. Schlather (2002) advocated the use of the representation of Proposition 6.2.2 to simulate extreme widespread rainfall, like winter rainfall in south-west England.

6.3 Stochastic process for simulating "extreme" rainfall

The starting point for the simulation of the rainfall process is Proposition 6.2.2, the representation of simple max-stable processes and its counterpart, the excursion stable process $\{YV(s)\}$. Conceptually, as explained in Section 6.2, the excursion stable process is the right one to use.

However, non-parametric estimation of the characteristics of the process V is presently beyond our reach. Hence we choose to work with a tractable parametric model for V. Unfortunately, the condition $\sup_{0\le s\le 1} V(s) = c$, that makes the process $\{YV(s)\}$ excursion stable, is very stringent and we could not find a reasonable parametric model for such a process. Hence, it seems better to stay with the model $\{YV(s)\}$ but replace the condition $\sup_{0\le s\le 1} V(s) = c$ by $E\sup_{0\le s\le 1} V(s) < \infty$ as allowed by Proposition 6.2.2. Then the excursion stability is still approximately true, i.e. the process has the same asymptotic dependence structure. But we meet another problem. In order to tie the simulated process to the observed non-extreme rainfall, it is imperative that the marginal distributions of the simulated process has a GPD tail (c.f. relation (6.6) below). As explained in Section 6.2, this is not correct for $\{YV(s)\}$ with $E\sup_{0\le s\le 1} V(s) < \infty$, worse, the marginal distribution is quite untractable, hence a transformation to repair this problem seems difficult to find.

Only the third possibility remains: to choose the simple max-stable process from Proposition 6.2.2 for the simulation. Then the asymptotic dependence structure of the process is the same as that of the corresponding GPD-type process $\{YV(s)\}$ (they are in the same domain of attraction) and the marginal distributions are all the same hence they can easily be transformed to the distribution function $1-1/x$, $x \ge 1$. The transformation on marginal distributions will not change the asymptotic dependence structure.

This is what we do in the simulation. For V we choose the so-called exponential martingale (c.f. Øksendal (1992), exercise 4.10). Also we have to extend the process to a process with a two-dimensional index set. We choose the model

$$\eta(s_1,s_2) := \max_{i\ge 1} \frac{\exp\{W_{1i}(\beta s_1) + W_{2i}(\beta s_2) - \beta(|s_1|+|s_2|)/2\}}{E_1 + E_2 + \cdots + E_i} \tag{6.3}$$

for $(s_1,s_2) \in \mathbb{R}^2$ (or rather the area under study, North Holland). Here $\{E_i\}$ is an i.i.d. sequence of standard exponential distributed random variables. The processes $W_{11}, W_{21}, W_{12}, W_{22}, W_{13}, W_{23}, \cdots$ are independent copies of double-sided Brownian motions W defined as follows. Take two independent Brownian motions B_1 and B_2. Then

$$W(s) := \begin{cases} B_1(s), & s \ge 0; \\ B_2(-s), & s < 0. \end{cases} \tag{6.4}$$

The positive constant β reflects the amount of spatial dependence at high levels of rainfall: "β small" means strong dependence and "β large" means weak dependence. The model assumes that the dependence between extreme rainfall at two locations depends only on the distance between the locations as we shall see later on.

The process η satisfies the requirements of Proposition 6.2.2:

$$E\exp\{W_1(\beta s_1) + W_2(\beta s_2) - \beta(|s_1|+|s_2|)/2\} = 1 \quad \text{for} \quad (s_1,s_2) \in \mathbb{R}^2,$$

and

$$E \sup_{\substack{a_1\le s_1\le b_1 \\ a_2\le s_2\le b_2}} \exp\{W_1(\beta s_1) + W_2(\beta s_2) - \beta(|s_1|+|s_2|)/2\} < \infty \quad \text{for all } a_1<b_1, a_2<b_2 \text{ real.}$$

By Proposition 6.2.2, the one-dimensional marginal distributions of (6.3) are all $e^{-1/x}, x > 0$. The two-dimensional marginal distributions are calculated in de Haan and Zhou (2008).

They are invariant under a shift. The same holds for the higher-dimensional marginal distributions (the proof is in de Haan and Zhou (2008)). Hence the process is shift stationary as it should be for our application.

The choice of this particular process is mainly one of convenience: the process is not too crude and it allows easy simulation and estimation of the dependence parameter.

For the simulation of our process we need to simulate (6.3), the maximum of infinitely many terms. However, since the denominators form an increasing sequence, one can approximate the process η by taking the maximum of only finitely many terms. In fact it turns out that even 4 terms are sufficient to get a reasonable result.

We have now a simple max-stable process that can be simulated rather well. But - taking into account our discussion of the finite-dimensional case - in fact we need a process that has generalized Pareto marginals, not the standard Fréchet extreme value distribution as marginals. Hence we use the process η from (6.3) but transform the marginal distributions to the generalized Pareto distribution $1 - 1/x$, $x \geq 1$:

$$\xi(s_1, s_2) := \frac{1}{1 - \exp\left\{-\frac{1}{\eta(s_1,s_2)}\right\}} \tag{6.5}$$

for (s_1, s_2) in the area.

The last step is a further transformation of the marginal distribution that adapts the process to the local shape (γ), scale (a) and shift (b) parameters. These parameters can be estimated from each station separately, using the local sample. However, the resulting estimates may not be accurate enough, due to the small sample size (there is a large number of days with no rain). To increase precision, it is often assumed in the hydrological and climatological literature that the shape parameter γ is constant over the region of interest (e.g., NERC (1975); AlilaAlila (1999); Gellens (2002); Fowler and Kilsby (2003)). A reliable estimate of γ is then obtained using all extreme values (usually in the literature this concerns the seasonal or annual maxima) in the region. This can be done by combining the extremes into a single record (the so-called station-year method), by averaging a local estimate of γ or a skewness statistic over the region of interest, or by maximizing a log likelihood with a common γ and local scale and location parameters. For annual maximum daily rainfall in the Netherlands, Buishand (1991) compared the maximum likelihood approach with the averaging of a local estimate of γ. Almost the same results were obtained.

Here we use the average of the local estimates of γ. We found the value $\hat{\gamma} = 0.1082$. This value is comparable with the estimates of the shape parameter found for daily maximum rainfall in the winter half-year (October-March) in the Netherlands (Buishand (1983)) and Belgium (Gellens (2002)). Of course our model allows γ to vary over the area.

The final transformation results into the process

$$X(s_1, s_2) := \hat{a}_{(s_1,s_2)}(n/k) \left(\frac{\xi(s_1, s_2)^{\hat{\gamma}_{n,k}} - 1}{\hat{\gamma}_{n,k}} \right) + \hat{b}_{(s_1,s_2)}(n/k). \tag{6.6}$$

Note that

$$\begin{aligned}
&P\left(\frac{X(s_1, s_2) - b_{(s_1,s_2)}(n/k)}{a_{(s_1,s_2)}(n/k)} > x \right) \\
=&P\left(\frac{\xi(s_1, s_2)^{\gamma} - 1}{\gamma} > x \right) \\
=&P\left(\eta(s_1, s_2) > \frac{1}{-\log(1 - (1 + \gamma x)^{-1/\gamma})} \right) \\
=&(1 + \gamma x)^{-1/\gamma},
\end{aligned}$$

hence, all marginal distributions are GPD.

The estimation for γ, a and b (c.f. Proposition 6.2.1)at any location is based on the "extreme" part of the local sample, i.e. the upper k order statistics of that sample. In the asymptotic theory, when the sample size n is going to infinity, k will go to infinity: $k = k(n) \to \infty$; but of lower order than n: $k(n)/n \to 0, n \to \infty$.

The estimation of the shift $b_{(s_1,s_2)}(n/k)$ is particularly simple: $\hat{b}_{(s_1,s_2)}(n/k)$ is the k-th largest order statistics of the local sample. There are various estimators of γ and $a_{(s_1,s_2)}(n/k)$ that converge at speed $k^{-1/2}$. In the present application we use the so-called moment estimator for γ (see e.g. de Haan and Ferreira (2006, Section 3.9)) and the accompanying estimator for $a_{(s_1,s_2)}(n/k)$ (see e.g. de Haan and Ferreira (2006, Section 4.2)).

As explained before, to obtain a global shape parameter, we take the average of the local estimates of γ among all the stations. However, we keep the local estimates of the scale and shift at each station.

To choose the number of upper order statistics k used for the estimation of the shape parameter, we plot the average estimate of this parameter across monitoring stations against k. This average is constant when k is around 125. Similar plots for each individual station confirmed that to choose $k = 125$ is also reasonable for most stations. Therefore, we keep it also for estimating the scale a and the shift b throughout the area. The sample size n is 2730.

With these estimations, the process (6.6) provides the simulated (extreme) rainfall in the area.

6.4 Simulating a day of rainfall

On an arbitrary day, there will be "extreme" rainfall in part of the area and "non-extreme" rainfall (or no rainfall at all) in the rest of the area.

We achieve this in the simulation as follows: on the one hand, we simulate the process (6.6) for the whole area; on the other hand, we choose at random a day out of the 30*(30+31+30)=2730 days of observed rainfall and we connect the two as follows:

For each station we check whether the observed rainfall on the chosen day is larger than the shift parameter $\hat{b}_{(s_1,s_2)}(n/k)$ for that station. If so, we use (6.6) (i.e., the simulated process) to get the rainfall at that station. If not, we just use the observed rainfall for the chosen day at that station.

How do we extend this to obtain the rainfall in the entire area?

First we connect the monitoring stations with each other, so as to cover the area with Triangles. The division is presented in Figure 6.2 as the solid lines. The numbers refer to the local 100-year quantile for each station, see Section 6.6. We write Triangles since later on we shall also deal with smaller triangles, also we write Vertex and Edge for a vertex and edge of a Triangle. Any Triangle can be extreme or non-extreme.

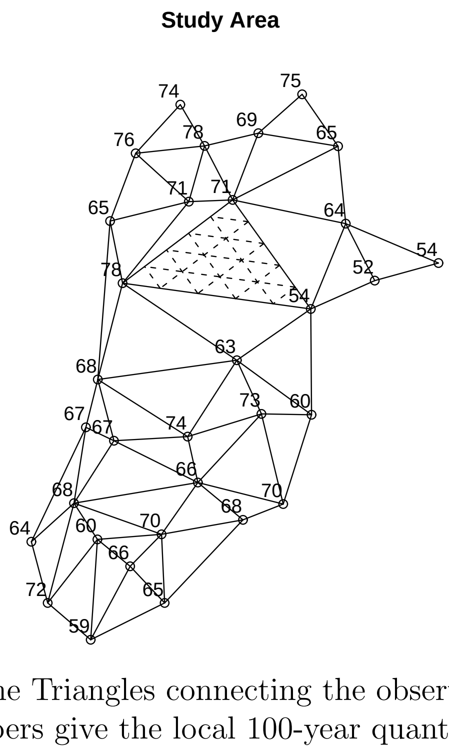

Figure 6.2: The Triangles connecting the observation stations
(The numbers give the local 100-year quantile in mm)

1. Non-extreme: this is the case if all Vertices of the Triangle are non-extreme. The rainfall in such a Triangle is just a linear function whose value at the Vertices are the observed values.

2. Extreme: all other cases. In that case the rainfall is mainly determined by the process (6.6) where the functions $a_{(s_1,s_2)}(n,k)$ and $b_{(s_1,s_2)}(n,k)$ on the Triangle are chosen as linear functions whose values at the Vertices are the values obtained by local estimation. Note that this mix of extreme and non-extreme simulation is similar to (6.1).

More specifically we proceed as follows:

2.a) Subdivide each Edge into d intervals of equal length. Connect the separating points on the Edges with each other using lines parallel to the Edges as indicated by the dashed lines in Figure 6.2. This results into d^2 triangles inside a Triangle. We used $d = 5$ in the simulation.

2.b) Next we determine the rainfall process in each vertex (i.e. vertex of a triangle). For Vertices we already determined the process. For the vertices, there are two cases.

2.b.1 On an Edge connecting two non-extreme Vertices in an extreme Triangle, the rainfall is chosen to be the linear function whose values at the Vertices are the observed values. This determines the rainfall for all vertices on such an Edge. The process (6.6) plays no role.

2.b.2 The rainfall for every other vertex in an extreme Triangle is determined by the process (6.6).

2.c) In order to carry out the numerical integration we simplify the rainfall process on each triangle in an extreme Triangle. The rainfall in each triangle is given as a linear function whose value at the vertices is the one obtained in part **2.b**.

This is the way we obtained a day of rainfall. An example of a simulation for Oct 11, 1997, is given in Figure 6.3. Rainfall for this day is extreme over large parts of the north and the middle of the study area. The left figure is based on the real data for Oct 11, 1997 and the right one presents a simulation. The latter produces very extreme rainfall over a small region in the middle of the study area with a steep gradient to the south. Note that the process is continuous and that it is easy to integrate numerically.

We remark that on 2299 out of the 2730 days of observation, none of the Vertices (stations) is extreme, so that no simulation is necessary. On the other hand, there are 44 days on which all Triangles are extreme, so that the whole area is simulated.

The choice for Triangles with monitoring stations as Vertices is one of convenience: triangles fit together easily to produce a continuous process and are relatively simple to handle.

6.5 Estimation of the dependence parameter

One problem remains: we do not know β, the global dependence parameter in (6.3). It has to be estimated. This can be done along the lines indicated in de Haan and Pereira

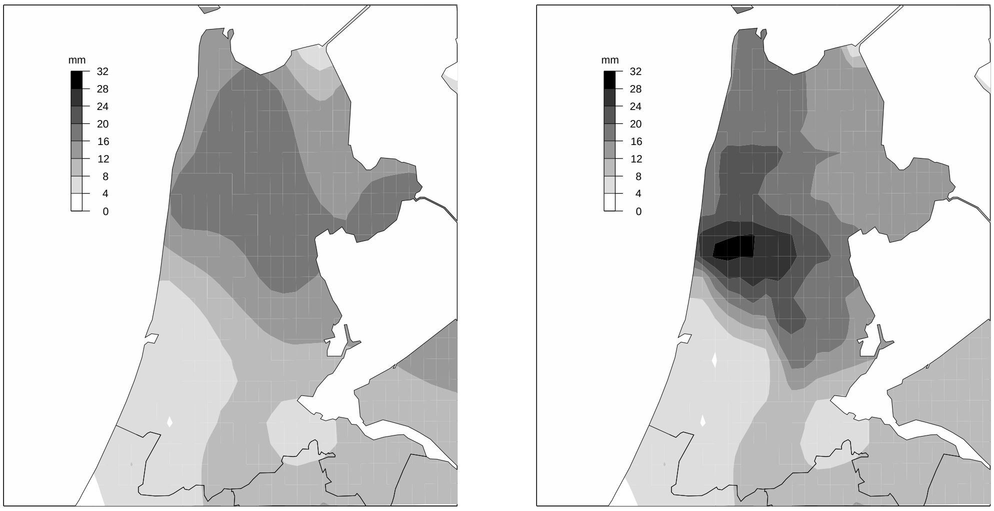

Figure 6.3: Observed (left) and simulated (right) rainfall for Oct 11, 1997

(2006).

We need to calculate the two-dimensional marginal distributions of the process η (defined in (6.3)) at locations (u_1, u_2) and (v_1, v_2), say. This is done in de Haan and Zhou (2008). The result is as follows: for x, y real with $h := |u_1 - v_1| + |u_2 - v_2|$,

$$\begin{aligned} &P(\eta(u_1, u_2) \le e^x, \eta(v_1, v_2) \le e^y) \\ =&\exp\left\{ -\left(e^{-x}\Phi\left(\frac{\sqrt{\beta h}}{2} + \frac{y-x}{\sqrt{\beta h}}\right) + e^{-y}\Phi\left(\frac{\sqrt{\beta h}}{2} + \frac{x-y}{\sqrt{\beta h}}\right)\right)\right\}, \end{aligned} \tag{6.7}$$

where Φ is the standard normal distribution function. Taking $x = y = 0$, we find

$$P(\eta(u_1, u_2) \le 1, \eta(v_1, v_2) \le 1) = \exp\left\{-2\Phi\left(\frac{\sqrt{\beta h}}{2}\right)\right\},$$

and consequently

$$\beta = \frac{4}{h}\left(\Phi^{\leftarrow}\left(-\frac{1}{2}\log P(\eta(u_1, u_2) \le 1, \eta(v_1, v_2) \le 1)\right)\right)^2.$$

Hence we can estimate β if we know how to estimate

$$L_{(u_1,u_2),(v_1,v_2)}(1, 1) := -\log P(\eta(u_1, u_2) \le 1, \eta(v_1, v_2) \le 1).$$

This is a problem of two-dimensional extreme value theory that has been solved by Huang and Mason (see Huang (1992) and Drees and Huang (1998)).

Let the continuous process X be in $\mathcal{D}$ (c.f. beginning of Section 6.2.3). Let $X_1, X_2, \cdots$ be i.i.d. copies of X. Write $\{X_{i,n}(s_1, s_2)\}_{i=1}^n$ for the order statistics at location (s_1, s_2). Then the estimator

$$\hat{L}^{(k)}_{(u_1,u_2),(v_1,v_2)}(1, 1) := \frac{1}{k}\sum_{j=1}^{n} 1_{\left\{X_j(u_1,u_2) \ge X_{n-k+1,n}(u_1,u_2) \quad \text{or} \quad X_j(v_1,v_2) \ge X_{n-k+1,n}(v_1,v_2)\right\}}$$

is consistent provided $k = k(n) \to \infty$, $k(n)/n \to 0$, $n \to \infty$. It is asymptotically normal under certain mild extra conditions.

Now label the monitoring stations with the numbers $1, 2, \cdots, N (N = 32)$ and define for $p < q \leq N$,

$$\hat{\beta}_{p,q} = \frac{4}{h} \left(\Phi^{\leftarrow} \left(\frac{1}{2} \hat{L}^{(k(p,q))}_{(u_1,u_2),(v_1,v_2)}(1,1) \right) \right)^2,$$

where (u_1, u_2) and (v_1, v_2) are the coordinates of station p and q respectively, $k(p,q)$ is the number of higher order statistics used in the estimation. Our estimator for β is

$$\hat{\beta} := \frac{2}{N(N-1)} \sum_{q=2}^{N} \sum_{p=1}^{q-1} \hat{\beta}_{p,q}$$

(consistent and asymptotically normal).

We found that $\hat{\beta} = 0.04277$.

Note that the estimators $\hat{\gamma}$, $\hat{a}$ and $\hat{b}$ come from one-dimensional extreme value theory, the estimator $\hat{\beta}$ comes from finite-dimensional extreme value theory and the process η comes from extreme value theory in $C[0,1]$.

6.6 Application

Our purpose is to study extremes of the total rainfall in North Holland. In particular we want to determine how severe the areal rainfall is that occurs once in 100 years. To be precise, it is once in 100*(30+31+30)=9100 days. In other words, we are studying the 1-1/9100 quantile of the daily total rainfall in the area. This quantile will be briefly indicated as the 100-year quantile.

Before presenting the simulation result, we would like to introduce some statistics and results for separate stations. Take Station West Beemster as an example (it is located in the middle of the area, and considered as the origin point when simulating the dependence process). The largest observed rainfall in the 30 years is 68.2 mm. By fitting the GPD with shape parameter $\hat{\gamma} = 0.1082$ to the observed extreme daily rainfall amounts at West Beemster, we can estimate the 1-1/9100 quantile for this station. The point estimator is 63.0 mm.

The 1-1/9100 quantiles for the other monitoring stations were obtained in the same way. Figure 6.2 gives the result for each station. We get that the average 1-1/9100 quantile among all the stations is 67.0 mm.

The simulation procedure in Section 6.4 has been repeated 91,000 times. This results in a sample of 91,000 days of rainfall in North Holland. For each day we calculate the total rainfall as the numerical integral of the rainfall process on the area. We take the 10th largest order statistic of this sample, i.e. we determine the 1-1/9100 sample quantile of the integrated rainfall. Dividing by the total area, 2010 km^2, we get the average rainfall in the area. We replicate this procedure 60 times. A histogram of the 60 simulated quantiles is given in Figure 6.4.

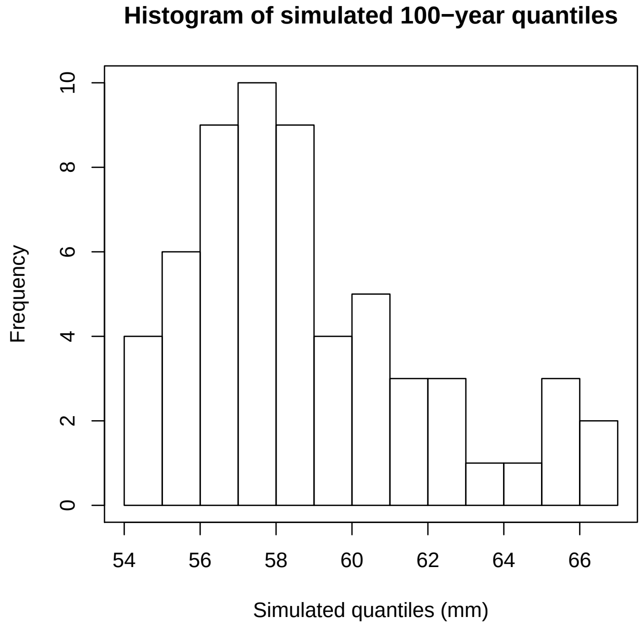

Figure 6.4: Histogram of simulated 100-year quantiles

Table 6.1: Statistics of simulated 100-Year quantiles of area-average rainfall

Mean (mm)	58.8	Sample Std (mm)	3.16
Min	54.4	Max	66.7

Some statistics of the 60 simulated quantiles are given in Table 6.6. From the table, the sample mean of the simulated quantiles is 58.8 mm, and the sample standard deviation is 3.16 mm. Hence the standard deviation of the sample mean is 0.41 mm.

When estimating the dependence parameter β in Section 6.5, we take the average of 496 estimates from all pairs of the monitoring stations, 0.04277. This is what we used in the above simulation. In order to study the sensitivity of the dependence parameter β, we take the 25% and 75% quantiles of the 496 estimates, 0.0339 and 0.0496, as β in the simulated model to repeat the above analysis. From 10 simulated 100-year quantiles for each new β, we get sample means of 58.4 mm and 60.0 mm respectively. Hence, the result does not change much.

A setback of the exponential martingale model is the dependence on the coordinate axes. In order to see how important the choice of the axes is, we have repeated the analysis after a 45 degree rotation of the axes and the result does not change a lot (from 10 simulated 100-year quantiles, we get sample mean 58.2 mm, and sample standard deviation 3.1 mm). After rotation one of the axes is more or less the prevailing wind direction.

It is interesting to compare the estimated 100-year quantile with the value obtained by fitting the GPD with $\hat{\gamma} = 0.1082$ to the extremes of the average daily rainfall of the 32 stations in the area. The use of the same shape parameter as that for the extremes at the individual stations can be justified from multivariate extreme value theory (Coles and Tawn (1994)). The resulting value of 57.8 mm for the 100-year quantile is slightly smaller than the average of our simulations.

The quantile for the area-average rainfall is smaller than the average of the corresponding quantile for the individual measurement stations. The areal reduction factor ARF is the ratio of these two quantities, $ARF = 58.8/66.9 = 0.88$. It is remarkable that from the graph in the UK Flood Studies Report (see NERC (1975)), a similar value of ARF is found for an area of 2010 km^2. The latter refers to annual maximum rainfall rather than seasonal maximum rainfall. The ARF from the GPD fit to the extreme average daily rainfall equals $57.8/66.9 = 0.86$.

6.7 Conclusion

The theory of extremes of continuous processes was used to estimate the 100-year quantile of the daily area-average rainfall over North Holland. The estimation of this quantile was done by simulating the daily process.

Regions with large rainfall were generated using a specific max-stable spatial process. It was argued that direct simulation from the excursion process is not feasible.

The estimated 100-year quantile for the areal average rainfall turns out to be 12% lower than the average 100-year quantile of the 32 measurement stations.

The model used in this chapter is convenient: it is really infinite-dimensional (does not depend on a finite number of random variables only); the two-dimensional distributions can be calculated and it can be simulated easily. A disadvantage is that the process is not invariant with respect to rotation of the coordinate axes (see end of Section 6.6). This will be the subject of future research.

Chapter 7

On extreme value analysis of a spatial process

7.1 Introduction

Problems of spatial statistics connected with high values of the spatial process need to be dealt with using extreme value theory (EVT), since the dependence between locations at high levels may differ from the dependence at moderate levels.

A case in point is the estimation of high quantiles of the total rainfall in a certain area. Engineers often need extreme rainfall statistics for the design of structures for flood protection. A typical question is e.g. what is the amount of rain in a given area on one day that is exceeded once in 100 years? Or, more mathematically, what is the 100-year quantile of the total rainfall in the area on one day? Buishand *et al.* (2008) investigated this question for a low-lying flat area in the northwest of the Netherlands. The observed rainfall data is only available on a few fixed monitoring stations. In order to study the high quantiles of the total rainfall, it is necessary to model the extreme rainfall process with dependence.

Considering the dependence structure, Cooley and Naveau (2007) used a Bayesian hierarchical model: locally the extreme rainfall is modeled by a one-dimensional EVT distribution and the parameters of this distribution follow some spatial dependence model.

A different way of introducing dependence is via a max-stable process. The mathematical setting of a max-stable process is as follows. Consider independent replications of a stochastic process with continuous sample paths

$$\{X_n(t)\}_{t\in\mathbb{R}},$$

$n = 1, 2, \cdots$. Suppose that the process is in the domain of attraction of a max-stable process, that is, there are sequences of continuous functions $a_n > 0$ and b_n such that as

$n \to \infty$

$$\left\{ \frac{\max_{1\le i\le n} X_i(t) - b_n(t)}{a_n(t)} \right\}_{t\in\mathbb{R}} \xrightarrow{w} \{\tilde{\eta}(t)\}_{t\in\mathbb{R}} \tag{7.1}$$

in $C-$space. Necessary and sufficient conditions have been given by de Haan and Lin (2001). The limit process $\{\tilde{\eta}(t)\}$ is a max-stable process. Without loss of generality we can assume that the marginal distribution of $\tilde{\eta}$ can be written as

$$\exp\left\{-(1+\gamma(t)x)^{-1/\gamma(t)}\right\}$$

for all x with $1+\gamma(t)x > 0$ where the function γ is continuous.

Buishand *et al.* (2008) simulated extreme rainfall from a max-stable process. Combining simulations of extreme rainfall with resampling from the non-extreme observations, an overview on the total rainfall can be generated. This is a novel solution for problems connected to both spatial statistics and extreme value analysis.

A major difficulty in the above methodology is to find a reasonable model for the max-stable process. With a suitable standardization, we can restrict ourselves to discussing the standardized process, called simple max-stable,

$$\{\eta(t)\} := \left\{(1+\gamma(t)\tilde{\eta}(t))_+^{1/\gamma(t)}\right\},$$

whose marginal distribution functions are all standard Fréchet: $\exp(-1/x), x > 0$.

For application, it would be nice to have a stationary simple max-stable process. There are two different representations of stationary simple max-stable processes in literature. One is from Theorem 9.6.10 in de Haan and Ferreira (2006), as follows.

A mapping Φ from L_1^+(the non-negative integrable functions on $\mathbb{R}$) to L_1^+ is called a *piston* if for $h \in L_1^+$

$$\Phi(h(t)) = r(t)h(H(t))$$

with H a one-to-one measurable mapping from $\mathbb{R}$ to $\mathbb{R}$ and r a positive measurable function, such that for every $h \in L_1^+$

$$\int_{\mathbb{R}} \Phi(h(t))dt = \int_{\mathbb{R}} h(t)dt.$$

Let $\{(Z_i, T_i)\}_{i=1}^{\infty}$ be a realization of a Poisson point process on $(0, +\infty] \times \mathbb{R}$ with mean measure $(dr/r^2) \times d\lambda$ (λ is the Lebesgue measure). If the stochastic process $\{\eta(s)\}_{s\in\mathbb{R}}$ is simple max-stable, strictly stationary and continuous a.s., then there exist a function $h \in L_1^+$ with $\int_{\mathbb{R}} h(t)dt = 1$ and a continuous group of pistons $\{\Phi_s\}_{s\in\mathbb{R}}$ (continuous, i.e. $\Phi_{s_n}(h(t)) \to \Phi_s(h(t))$ as $s_n \to s$ for almost all $t \in \mathbb{R}$) with

$$\int_{\mathbb{R}} \sup_{s\in I} \Phi_s(h(t))dt < \infty$$

for each compact interval I, such that

$$\{\eta(s)\}_{s\in\mathbb{R}} \stackrel{d}{=} \left\{\max_{i\geq 1} Z_i \Phi_s(h(T_i))\right\}_{s\in\mathbb{R}}. \tag{7.2}$$

Conversely every stochastic process of the form exhibited at the right-hand side of (7.2) with the stated conditions, is simple max-stable, strictly stationary and a.s. continuous.

A special case is obtained by setting $\Phi_s(h(t)) := h(t+s)$, where h is a continuous probability density. It leads to a few strictly stationary, continuous, simple max-stable process as discussed in de Haan and Pereira (2006). For example, one of them is the normal density h as

$$h(t) := \frac{\beta}{\sqrt{2\pi}} \exp\left\{-\frac{\beta^2 t^2}{2}\right\}$$

with β a positive constant. The parameter β indicates the amount of dependence in the process: small values of β indicates high dependence. By estimating β using the rainfall observation at the available locations, we get the dependence structure. It is shown in de Haan and Pereira (2006) that for estimating β, it is sufficient to calculate explicitly the two-dimensional marginal distributions of the max-stable process.

An alternative representation of max-stable process is as follows, see Corollary 9.4.5 in de Haan and Ferreira (2006).

All simple max-stable process in $C^+(\mathbb{R})$(the positive continuous functions on $\mathbb{R}$) can be generated in the following way. Consider a Poisson point process on $(0, +\infty]$ with mean measure dr/r^2. Let $\{Z_i\}_{i=1}^{\infty}$ be a realization of this point process. Further consider i.i.d. stochastic processes $V, V_1, V_2, \cdots$ in $C^+(\mathbb{R})$ with $EV(s) = 1$ for all $s \in \mathbb{R}$ and $E\sup_{s\in I} V(s) < \infty$ for all compact interval I. Let the point process and the sequence $V, V_1, V_2, \cdots$ be independent. Then

$$\{\eta(s)\}_{s\in\mathbb{R}} \stackrel{d}{=} \left\{\max_{i\geq 1} Z_i V_i(s)\right\}_{s\in\mathbb{R}} \tag{7.3}$$

is a simple max-stable process. Conversely each simple max-stable process has such a representation.

We use this result in a two-dimensional context and propose the following model

$$\eta(s_1, s_2) := \max_{i\geq 1} Z_i \exp\{W_{1i}(\beta s_1) + W_{2i}(\beta s_2) - \beta(|s_1| + |s_2|)/2\} \tag{7.4}$$

for $(s_1, s_2) \in \mathbb{R}^2$. The processes $W_{11}, W_{21}, W_{12}, W_{22}, W_{13}, W_{23}, \cdots$ are independent copies of double-sided Brownian motions W defined as follows. Take two independent Brownian motions B_1 and B_2. Then

$$W(s) := \begin{cases} B_1(s), & s \geq 0; \\ B_2(-s), & s < 0. \end{cases} \tag{7.5}$$

The positive constant β reflects the amount of spatial dependence at high levels of rainfall: "β small" means strong dependence and "β large" means weak dependence. From this model, we shall prove that the dependence between extreme rainfall at two locations depends only on the distance between the locations.

The process η satisfies the requirements as follows:

$$E \exp\{W_1(\beta s_1) + W_2(\beta s_2) - \beta(|s_1| + |s_2|)/2\} = 1 \quad \text{for} \quad (s_1, s_2) \in \mathbb{R}^2,$$

and

$$E \sup_{\substack{a_1 \le s_1 \le b_1 \\ a_2 \le s_2 \le b_2}} \exp\{W_1(\beta s_1) + W_2(\beta s_2) - \beta(|s_1| + |s_2|)/2\} < \infty \quad \text{for all } a_1 < b_1, a_2 < b_2 \text{ real.}$$

Meanwhile, the one-dimensional marginal distribution functions of (7.4) are all the same as the standard Fréchet distribution function, $e^{-1/x}, x > 0$.

Similar to de Haan and Pereira (2006), in order to use this model in studying the areal rainfall, we have to prove that the process η is shift stationary and we have to calculate the two-dimensional marginal distributions.

Since the two-dimensional process η is a combination of two one-dimensional processes, for the stationarity it is sufficient to prove the same for the one-dimensional version, i.e. that the process

$$\eta'(s_1) := \max_{i \ge 1} Z_i \exp\{W_{1i}(\beta s_1) - \beta |s_1|/2\} \tag{7.6}$$

is stationary. This follows from the fact that the process η' can be obtained as the limit of the pointwise maximum of i.i.d. Ornstein-Uhlenbeck processes (see e.g. Example 9.8.2 in de Haan and Ferreira (2006)). The stationarity follows from the stationarity of the Ornstein-Uhlenbeck process.

It remains to calculate the two-dimensional marginal distributions. This is done in Section 7.2.

7.2 The two-dimensional marginal distribution of η

The two-dimensional marginal distribution of η' in (7.6) is calculated in de Haan and Ferreira (2006, Section 9.8). We state it as the following proposition.

Proposition 7.2.1 *Suppose $\{\eta'(s)\}_{s \in \mathbb{R}}$ is defined as in (7.6). Then for $x, y \in \mathbb{R}$ and $s_1, s_2 \in \mathbb{R}$,*

$$\begin{aligned} &- \log P(\eta'(s_1) \le e^x, \eta'(s_2) \le e^y) \\ =& e^{-x} \Phi\left(\frac{\sqrt{|s_1 - s_2|}}{2} + \frac{-x + y}{\sqrt{|s_1 - s_2|}} \right) + e^{-y} \Phi\left(\frac{\sqrt{|s_1 - s_2|}}{2} + \frac{x - y}{\sqrt{|s_1 - s_2|}} \right). \end{aligned}$$

This is useful in similar calculation for the two-dimensional process η. Besides Proposition 7.2.1, we need the following Lemma.

Lemma 7.2.1 *Suppose N is normally distributed with mean 0, variance u, then with non-random constants $a > 0$ and b,*

$$Ee^{N-u/2}\Phi(aN+b) = \Phi\left(\frac{au+b}{\sqrt{a^2u+1}}\right). \tag{7.7}$$

Proof

Suppose N_1 is standard normally distributed, and independent of N, then we have

$$Ee^{N-u/2}1_{N_1\le aN+b} = E_N E(e^{N-u/2}1_{N_1\le aN+b}|N) = Ee^{N-u/2}\Phi(aN+b),$$

which is the left side of (7.7). By Fubini's Theorem, it can be recalculated in the following way

$$\begin{aligned}
&Ee^{N-u/2}1_{N_1\le aN+b}\\
=&E_{N_1}E(e^{N-u/2}1_{N_1\le aN+b}|N_1)\\
=&E_{N_1}\int_{\frac{N_1-b}{a}}^{\infty} e^{t-u/2}\frac{1}{\sqrt{2\pi u}}e^{-\frac{t^2}{2u}}\,dt\\
=&E_{N_1}\int_{\frac{N_1-b}{a}}^{\infty} \frac{1}{\sqrt{2\pi u}}e^{-\frac{(t-u)^2}{2u}}\,dt\\
=&E_{N_1}\left(1-\Phi\left(\frac{N_1-b}{a\sqrt{u}}-\sqrt{u}\right)\right).
\end{aligned}$$

By a similar trick - introducing a standard normal variable N_2 independent of N_1, the calculation can be finished to prove the lemma.

$$\begin{aligned}
&E_{N_1}\left(1-\Phi\left(\frac{N_1-b}{a\sqrt{u}}-\sqrt{u}\right)\right)\\
=&E_{N_1}E(1_{N_2\ge\frac{N_1-b}{a\sqrt{u}}-\sqrt{u}}|N_1)\\
=&E_{N_1,N_2}1_{N_2\ge\frac{N_1-b}{a\sqrt{u}}-\sqrt{u}}\\
=&P(N_2\ge\frac{N_1-b}{a\sqrt{u}}-\sqrt{u})\\
=&\Phi\left(\frac{au+b}{\sqrt{a^2u+1}}\right).
\end{aligned}$$

We remark that the last calculation is similar to that of Lemma 2.1 in Gupta *et al.* (2004). □

The lemma can be used to derive the two-dimensional marginal distributions as follows. As in the proof of Proposition 7.2.1 (see de Haan and Ferreira (2006, Section 9.8)), we have

$$-\log P(\eta(u_1,u_2)\le e^x, \eta(v_1,v_2)\le e^y)$$

$$
\begin{aligned}
&= E\max\left(e^{W_1(\beta u_1)+W_2(\beta u_2)-(|\beta u_1|+|\beta u_2|)/2-x}, e^{W_1(\beta v_1)+W_2(\beta v_2)-(|\beta v_1|+|\beta v_2|)/2-y}\right)\\
&= E_{W_1}E\left(\max\left(e^{W_1(\beta u_1)+W_2(\beta u_2)-(\beta|u_1|+\beta|u_2|)/2-x}, e^{W_1(\beta v_1)+W_2(\beta v_2)-(\beta|v_1|+\beta|v_2|)/2-y}\right)|W_1\right)\\
&= Ee^{-x+W_1(\beta u_1)-\beta|u_1|/2}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right)\\
&+ Ee^{-y+W_1(\beta v_1)-\beta|v_1|/2}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{x-y+W_1(\beta v_1)-W_1(\beta u_1)-\beta|v_1|/2+\beta|u_1|/2}{\sqrt{\beta|u_2-v_2|}}\right).
\end{aligned}
\tag{7.8}
$$

Now we can calculate the two parts in (7.8) separately. Without loosing generality, we only focus on the first part.

Case 1: $0 \le u_1 \le v_1$

In this case $e^{-x+W_1(\beta u_1)-\beta|u_1|/2}$ is independent of the other part. Hence,

$$
\begin{aligned}
&Ee^{-x+W_1(\beta u_1)-\beta|u_1|/2}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right)\\
&= e^{-x}E\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x-(W_1(\beta v_1)-W_1(\beta u_1)-\beta(v_1-u_1)/2)}{\sqrt{\beta|u_2-v_2|}}\right)\\
&= e^{-x}P\left(N\le\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x-(W_1(\beta v_1)-W_1(\beta u_1)-\beta(v_1-u_1)/2)}{\sqrt{\beta|u_2-v_2|}}\right)\\
&= e^{-x}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|+\beta(v_1-u_1)}}{2}+\frac{y-x}{\sqrt{\beta|u_2-v_2|+\beta(v_1-u_1)}}\right).
\end{aligned}
$$

Case 2: $0 \le v_1 < u_1$

Note that $Ee^{W_1(\beta v_1)-\beta v_1/2} = 1$ and $W_1(\beta v_1)$ is independent of $W_1(\beta u_1)-W_1(\beta v_1)$, we have

$$
\begin{aligned}
&Ee^{-x+W_1(\beta u_1)-\beta|u_1|/2}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right)\\
&= e^{-x}Ee^{W_1(\beta u_1)-W_1(\beta v_1)-\beta(u_1-v_1)/2}\\
&\quad\cdot\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right).
\end{aligned}
$$

Since $W_1(\beta u_1) - W_1(\beta v_1)$ is normally distributed with mean 0, variance $\beta(u_1 - v_1)$, we can apply Lemma 7.2.1 with the constants $a = 1/\sqrt{\beta|u_2-v_2|}$, $u = \beta(u_1 - v_1)$ and

$$
b = \frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x-\beta u_1/2+\beta v_1/2}{\sqrt{\beta|u_2-v_2|}}.
$$

The final result is

$$
Ee^{-x+W_1(\beta u_1)-\beta|u_1|/2}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right)
$$

$$=e^{-x}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|+\beta(u_1-v_1)}}{2}+\frac{y-x}{\sqrt{\beta|u_2-v_2|+\beta(u_1-v_1)}}\right).$$

Case 3: $v_1 < u_1 < 0$ and $u_1 \leq v_1 < 0$

These two cases are similar to Case 1 and 2 respectively. The final results are all the same as follows.

$$Ee^{-x+W_1(\beta u_1)-\beta|u_1|/2}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right)$$
$$=e^{-x}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|+\beta|u_1-v_1|}}{2}+\frac{y-x}{\sqrt{\beta|u_2-v_2|+\beta|u_1-v_1|}}\right).$$

Case 4: u_1 and v_1 have different signs.

In this case $W_1(\beta u_1)$ and $W_1(\beta v_1)$ are independent, we can calculate the expectation with respect to $W_1(\beta v_1)$ first, then with respect to $W_1(\beta u_1)$.

$$Ee^{-x+W_1(\beta u_1)-\beta|u_1|/2}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right)$$
$$=e^{-x}Ee^{W_1(\beta u_1)-\beta|u_1|/2}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|+\beta|v_1|}}{2}+\frac{y-x+W_1(\beta u_1)-\beta|u_1|/2}{\sqrt{\beta|u_2-v_2|+\beta|v_1|}}\right).$$

Now we can again apply Lemma 7.2.1 with the constants $a = 1/\sqrt{\beta|u_2-v_2|+\beta|v_1|}$, $u = \beta|u_1|$ and

$$b=\frac{\sqrt{\beta|u_2-v_2|+\beta|v_1|}}{2}+\frac{y-x-\beta|u_1|/2}{\sqrt{\beta|u_2-v_2|+\beta|v_1|}}$$

to get that

$$Ee^{-x+W_1(\beta u_1)-\beta|u_1|/2}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right)$$
$$=e^{-x}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|+\beta(|u_1|+|v_1|)}}{2}+\frac{y-x}{\sqrt{\beta|u_2-v_2|+\beta(|u_1|+|v_1|)}}\right).$$

Notice that due to the different signs of u_1 and v_1, $|u_1-v_1|=|u_1|+|v_1|$.

By defining $h=|u_1-v_1|+|u_2-v_2|$, all these cases can be combined together as

$$Ee^{-x+W_1(\beta u_1)-\beta|u_1|/2}\Phi\left(\frac{\sqrt{\beta|u_2-v_2|}}{2}+\frac{y-x+W_1(\beta u_1)-W_1(\beta v_1)-\beta|u_1|/2+\beta|v_1|/2}{\sqrt{\beta|u_2-v_2|}}\right)$$
$$=e^{-x}\Phi\left(\frac{\sqrt{\beta h}}{2}+\frac{y-x}{\sqrt{\beta h}}\right).$$

Symmetrically, the second part of (7.8) can be simplified as

$$e^{-y}\Phi\left(\frac{\sqrt{\beta h}}{2}+\frac{x-y}{\sqrt{\beta h}}\right).$$

Combining these two parts, we get the following theorem about the two-dimensional marginal distribution of η.

Theorem 7.2.1 *Suppose the simple max-stable process η is defined in (7.4). Given any two coordinates (u_1, u_2) and (v_1, v_2) on $\mathbb{R}^2$, denote the distance between them as $h := |u_1 - v_1| + |u_2 - v_2|$. Then the two-dimensional distribution function of $(\eta(u_1, u_2), \eta(v_1, v_2))$ is*

$$\begin{aligned}&P(\eta(u_1,u_2) \le e^x, \eta(v_1,v_2) \le e^y)\\ &= \exp\left\{-\left(e^{-x}\Phi\left(\frac{\sqrt{\beta h}}{2} + \frac{y-x}{\sqrt{\beta h}}\right) + e^{-y}\Phi\left(\frac{\sqrt{\beta h}}{2} + \frac{x-y}{\sqrt{\beta h}}\right)\right)\right\}, \end{aligned} \tag{7.9}$$

where Φ is the standard normal distribution function and $x, y \in \mathbb{R}$.

Note that the two-dimensional marginal distribution depends on only h. It agrees with the shift stationarity discussed in Section 7.1.

Similar to de Haan and Pereira (2006), Theorem 7.2.1 is useful in estimating β. By taking $x = y = 0$, we get that

$$P(\eta(u_1, u_2) \le 1, \eta(v_1, v_2) \le 1) = \exp\left\{-2\Phi\left(\frac{\sqrt{\beta h}}{2}\right)\right\}.$$

Consequently, we have that

$$\beta = \frac{4}{h}\left(\Phi^{\leftarrow}\left(-\frac{1}{2}\log P(\eta(u_1,u_2) \le 1, \eta(v_1,v_2) \le 1)\right)\right)^2.$$

Hence we can estimate β if we know how to estimate

$$L_{(u_1,u_2),(v_1,v_2)}(1,1) := -\log P(\eta(u_1,u_2) \le 1, \eta(v_1,v_2) \le 1).$$

In fact, this problem has been solved by Huang and Mason (see Huang (1992) and Drees and Huang (1998)). Suppose we have i.i.d. observations of η as $\eta_1, \eta_2, \cdots$. Write $\{\eta_{i,n}(s_1, s_2)\}_{i=1}^n$ for the order statistics at location (s_1, s_2). Then the estimator

$$\hat{L}^{(k)}_{(u_1,u_2),(v_1,v_2)}(1,1) := \frac{1}{k}\sum_{j=1}^n 1_{\{\eta_j(u_1,u_2) \ge \eta_{n-k+1,n}(u_1,u_2) \quad \text{or} \quad \eta_j(v_1,v_2) \ge \eta_{n-k+1,n}(v_1,v_2)\}}$$

is consistent provided by $k = k(n) \to \infty$, $k(n)/n \to 0$ as $n \to \infty$. It is asymptotically normal under certain mild extra conditions.

Hence, from the two-dimensional marginal distribution, we can estimate β when we have observations at two specific locations. An application of this method is in Buishand *et al.* (2008), Section 5.

Part IV

Internet Auction

Chapter 8

The Extent of Internet Auction Markets

8.1 Introduction

Internet auctions (IA) have rapidly become a highly popular mechanism of exchange. The popularity of IA sites is partly due to the enlargement of the market for rare items. Since there are only a few buyers and sellers for such goods, creating an easily accessible nation-wide or global market potentially improves the match between supply and demand. Nevertheless, also many standard commodities, such as notebooks, are sold through online auction sites. It seems intuitive that a seller is better off the larger is the extent of the market. Indeed, Bulow and Klemperer (1996) showed that under the hypothesis of the Independent Private Values Paradigm (IPVP), the seller is better off by enlarging the market.[1] Similarly, from the buyer's perspective a larger choice of items and sellers is often the better.

One of the perhaps surprising facts of IA markets is the low number of *active bidders*, notwithstanding the popularity of the mechanism. Consider the well known Ockenfels and Roth (2006) study of laptop and antiques auctions on eBay and Amazon. This study covers 120 eBay laptop auctions with 740 active bidders in total and 120 Amazon laptop auctions with a total of 595 active bidders. This implies an average number of 6.17 active bidders per auction on ebay and 4.96 active bidders per auction on Amazon. The Bajari and Hortaçsu (2003) study of eBay coin auctions finds 3.08 active bidders on average, with a standard deviation of 2.51 and a maximum of 14 bidders. As another example, the study of the online English auction with fixed ending time in Korea by Park and Bradlow (2005) reports an average number of 5.80 bidders and 8.4 bids per auction.

Most IA sites have two ways of bidding. First, there is the possibility to bid manu-

[1] But in multi-unit second price auctions entry can lead to lower revenues, see Ausubel and Milgrom (2002).

ally, which is the standard English auction mechanism. Second, the possibility of proxy bidding, in which a machine bids on behalf of the buyer, adds the Vickrey feature to the auction. In multiple day English auctions not every manual bidder is continuously active. Some intervening bids might therefore not materialize, whereas these would be placed under proxy bidding. Nevertheless, whatever the exact number of the active bidders is, it is a rather low number. The *potential bidders* are those bidders who check the auction website with or without placing a bid. Potential bidders are in principle interested in buying the item, but may not be willing to pay the going price. Thus the number of potential bidders is much larger than the number of actual bidders. We will try to explain the low average number of active bidders relative to the potential extent of the market.

Consider the number of potential bidders n as the extent of the market. Under the independent private value paradigm (IPVP) we show that, under a mild additional assumption, the valuations of the active bidders forms a sequence of records and second records. By using the probability theory of records, we then prove that if the number n of potential bidders is large, the number of active bidders is approximately equal to $2\log n$. This explains the relative inactivity, since $(2\log n)/n \to 0$ as $n \to \infty$. With this relationship in hand one can address questions such as by how much the extent of the market has increased through the creation of IA.

The large amount of data generated by IA is conducive to empirical auction research. A key difficulty in the analysis of IA data, though, is the fact that the number of potential bidders is unknown. Many standard auction models require the number of bidders as a known parameter for identification, see Krishna (2002). Paarsch (1992) is a good example of an empirical study in which the number of bidders is known. Several empirical studies combine the data from multiple auctions to study the IA without knowing the number of potential bidders for a specific auction. However, this requires an extra assumption regarding the distribution of the number of bidders. Laffont *et al.* (1995) use the multiple-auctions approach by assuming that the numbers of bidders is the same across all auctions under consideration. Bajari and Hortaçsu (2003) and McAfee and McMillan (1987) both analyze structural econometric models for multiple auctions under the assumption that the number of bidders follows a specific stochastic distribution.

Consider the identification problem in auction theory as in Athey and Haile (2002). Song (2004) discusses this issue for IA and shows that under the IPVP, it is impossible to identify the distribution of the bidder's valuations from the empirical distribution of bids, without knowing the number of potential bidders, if only the second largest order statistic is observed. The parent distribution, though, is identified without knowing the number of potential bidders, when at least two order statistics are observed, i.e. the second and third largest order statistics. The structural analysis of auctions requires that one can identify the distribution of valuations from the data under the conjectured equilibrium bid strategy. The Vickrey mechanism weakly dominant strategies are proxy bids which permit such identification. But since most IA are a combination of proxy bids and manual bids, an extra assumption is required for identification. If it can be assumed that manual bidders respond immediately once they are overbid, this guarantees that their final bid is either the winning bid, or that it equals their valuation of the item, see Song (2004). Then both the second and the third largest order statistics can be observed.

While most empirical papers in fact proceed by pooling data across different auctions, tacitly assuming homogeneity in the distribution of valuations and numbers of potential bidders, we follow Caserta and de Vries (2005) and study IA on a per auction basis. But to be able to do this, we need that the number of potential bidders is relatively large. The $2 \log n$ rule would lend itself easily to empirical scrutiny if the numbers of active and potential bidders were known. Suppose that the number of page views can be used as an indicator of potential interest. Fortunately, some smaller sites do report this number. We will use this information to measure n and to test for the $2 \log n$ rule via regression analysis.

The large IA sites such as eBay and Yahoo! do not record the number of page views and hence provide no direct information on the number of potential bidders. But information on the bid arrival time is often publicly recorded. We show that this information can be used to test the $2 \log n$ rule indirectly, still using the per auction approach, under one extra maintained assumption. If potential bidders arrive according to a Poisson process, with a preference for auctions with short remainder time, then the asymptotic property of the inter arrival times implies the $2 \log n$ rule. We also provide indirect evidence for the $2 \log n$ rule using these bid arrival times.

Both the regression evidence and the indirect evidence via the bidding time for the $2 \log n$ rule can be interpreted as a weak test of the IPVP. Alternative settings like common values and interdependent signals paradigms do not necessarily imply the specific record sequence which results under the IPVP. Thus insofar the $2 \log n$ rule is not rejected, the evidence weakly supports the IPVP hypothesis. Notice that most tests of the IPVP contain several other maintained hypotheses, such as homogeneity across different auctions and the distribution of the valuations. Here we do not need to maintain these other hypotheses, except for the assumption regarding the response of the manual bidders in case they are overbid. But the test is still weak as it only looks at one implication of the auctions under IPVP and the alternative paradigms do not necessarily destroy the $2 \log n$ rule, so that the alternative hypothesis is not well specified. In Bajari and Hortaçsu (2003) the IPVP is tested versus the common value pardigm by means of the winner's curse, which is more severe the larger the number of bidders. For rare coin auctions the common value paradigm is the natural null hypothesis as is pointed out by Bajari and Hortaçsu. In this chapter we focus on laptops which a priory better fits the IPVP.

This chapter is organized as follows. In Section 8.2, we review certain aspects of IA. Identification strategies are discussed in Section 8.3. Section 8.4 discusses bids as a record sequence. Our main theorem in Section 8.5 gives the asymptotic distribution of the index of the record sequence. Section 8.6 discusses the empirical evidence of the $2 \log n$ rule via the number of pageviews. Simulations and empirical evidence for the $2 \log n$ rule via the timing of the bids are provided in Section 8.7. Section 8.8 concludes. Proofs are relegated to the Appendix 8.A. An example clarifying some notation in this chapter is given in Appendix 8.B.

8.2 The online auction

In comparison to offline auctions, the IA exhibit three main differences: the bidding system, the termination rule and the reminder system. We discuss these features in the following three subsections.

8.2.1 The bidding system

IA sites usually permit a choice between alternative bid procedures. Most common is the choice between *manual* and *proxy bidding*. For a manual bid, the bidder just enters an amount higher than the currently prevailing price. Manual bidding is like the first price open ascending bid in an English auction. The system immediately places the bid at the amount which is entered. A proxy bidder (secretely) communicates the maximum amount he is willing to bid to the server of the auction site, after which the machine takes over the bidding for this bidder. The proxy bidding procedure captures the second price sealed bid mechanism studied by Vickrey (1962). Thus, if the newly entered manual bid is below the maximum willingness to pay of one previous proxy bidders, the system will raise the price to the minimum increment above the newly entered manual bid. Otherwise the manual bid becomes the currently prevailing price. Similarly, a new proxy bidder may find that he is immediately outbid by another proxy bidder. When two proxy bids are placed, the current price will automatically jump to the lower of the two maximum willingness to bid submissions plus the smallest possible increment.

8.2.2 The termination rules

There exist broadly two alternative termination rules. Either the auction ends after a pre-announced fixed lapse of time, or there is variable termination time. On eBay, the auctions have a fixed ending time. The winner is the highest bidder at the time of the close. Per contrast, Amazon type auctions use an auto-extension termination rule. Before the auction starts an initial ending time is announced. If no bidding takes place during the last ten minutes, the auction stops at the initial pre-announced time. But if there are some bids in the last ten minutes, the ending time is automatically extended by another ten minutes. This rule is also applied to the new extension period. Thus the auction will only end once no bidding activity has occurred within the last ten minutes before the previous ending time, otherwise, the ending time is extended automatically. On the Yahoo! auction site the sellers can choose which of the two termination rules they adopt.

The influence of the different termination systems on the bidding strategy is considerable. The fixed ending time on eBay invites strategic last minute bidding, called sniping. This avoids competition, since other bidders may be unable to respond due to a lack of time, see Ockenfels and Roth (2006). In the Amazon-style auction sniping is not observed.

8.2.3 The reminder system

Most IA sites have an email based reminder system to inform active bidders about the fact that they are outbid. This is particularly relevant for those active bidders who use the manual system. Since many of these auctions run for several days or even longer, those agents may rely on such automated signalling before becoming active again. With the reminder system, the IA for the manual bidder in effect becomes a continuous open English auction. But note that the English auction under the IPVP is strategically equivalent to a second price auction. Most IA sites go through great pains in trying to explain this equivalence, so that bidders are stimulated to use the proxy bidding. But given the strategic incentives for sniping with the fixed ending time, manual bidding is nevertheless often observed. Moreover, for items like collectibles for which the IPVP may not be appropriate, manual bidding may again be preferable for strategic reasons.

8.3 Maintained hypothesis

Before we can come to the main theme of the chapter, we first have to ascertain that the bid sequence is revealing with respect to the valuations. To analyze this issue for IA, first consider a hypothetical IA on which only the proxy bidding mechanism is available and to which the Amazon termination rule applies. Under the IPVP the weakly dominant strategy is to bid one's valuation. If bidders act in this way, each active bidder's valuation will be observed as his last bid, except for the winner.

Next, consider the hybrid IA with the possibility of manual bidding added to the proxy bidding mechanism. To ensure that the third largest and lower valuations are observed as actual bids in the case with manual bidding, we use an assumption introduced by Song (2004) for identification of the third largest valuation. Suppose that when the third highest valuation bidder is overbid, he is immediately notified by the reminder system if he is bidding manually. This enables the manual bidder to respond directly. We will assume that the manual bidder indeed immediately counters with a higher bid as soon as he is overbid and as long as his valuation is above the current price. This assumption ensures that his valuation is observed as the second highest bid once the auction has terminated.

In this chapter the analysis is in principle concluded on a per auction basis, i.e. we do not necessarily pool across different auctions by making homogeneity assumptions. To enable this analysis, we therefore like to use the bids from bidders other than the top three, to increase the information content. For this reason, we will assume that all the (active) manual bidders immediately respond to a counter bid:

Assumption 8.3.1 *Each active (manual) bidder immediately returns to the IA and increases his bid as soon as he is overbid and his valuation is above the prevailing price.*

We already noted that Assumption 8.3.1 is automatically satisfied when there are only proxy bidders present. For the manual bidders, as we discussed in Subsection 8.2.2, they may have an incentive to bid later when there is a fixed ending time. However, with the Amazon style auction termination rule, the active bidders (both proxy bidder and manual bidder) have no incentive to wait. In this case, the assumption provides a lower bound to the number of active bidders. Without an immediate response, some other manual bids might intervene, whereas these would not be placed in the case Assumption 8.3.1 applies.

Given the Assumption 8.3.1, the currently prevailing price must be equal to the second highest valuation among all the potential bidders up to that moment. Therefore the current price faced by a new potential bidder must be the second highest valuation among all the potential bidders who were actively bidding earlier on. Hence, in order to motivate a new potential bidder to bid, his valuation must be higher than the current second highest valuation. This conclusion is summarized in the first proposition.

Proposition 8.3.1 *Consider an IA with a hybrid system of manual and proxy bids. Suppose that Assumption 8.3.1 within the IPVP setting applies, then each active bidder's valuation is the highest or second-highest among all the valuations of the potential bidders who were active before.*

8.4 Bids as a specific record sequence

Proposition 8.3.1 implies that the bids can be viewed as records of the valuations of the potential bidders. The sequence of bids constitutes a quite particular record sequence, which we can characterize. There exists a well developed theory of records in probability theory, see the book by Resnick (1987). This theory will be used to derive the novel $2\log n$ rule. We first introduce the concept *record and 2-record sequence.* Let $i = 1, 2, \cdots, n$ denote the order in which the n potential bidders check the auction site. Suppose the valuation of all potential bidders are i.i.d. random variables $X_1, X_2, \cdots, X_n$ with distribution function $F(x)$. Define the rank sequence $\{R_i\}_{i=1}^n$ as

$$R_i := \sum_{k=1}^{i} 1_{\{X_k \geq X_i\}}, \tag{8.1}$$

where R_i is the rank of the valuation of the $i-$th potential bidder among the valuations of all the potential bidders who checked the auction before agent i. The valuation X_i is called a *record* if $R_i = 1$. Similarly, it is a *2-record* if $R_i = 2$. Denote the indices of the records and 2-records as $\{J(j)\}_{j=1}^m$. This index sequence is given by

$$J(1) = 1,\ J(2) = 2 \tag{8.2}$$

$$J(j+1) = \min\{i > J(j) : R_i \leq 2\},\ j = 2, 3, \cdots, m-1, \tag{8.3}$$

where m is the number such that $R_i > 2$ for all $i > J(m)$, i.e. m is the number of active bidders. Then, the active bidders' valuations constitute the record and 2-record sequence $\left\{X_{J(j)}\right\}_{j=1}^m$. An example to clarify the notation is presented in Appendix 8.B.

Under Assumption 8.3.1, from Proposition 8.3.1 we get the following corollary.

Corollary 8.4.1 *Under Assumption 8.3.1, the active bidders' valuations are* $\left\{X_{J(j)}\right\}_{j=1}^m$. *If an active bidder is not the winner, his valuation is observed as his last bid.*

8.5 Main theorem

Corollary 8.4.1 implies that the record and 2-record sequence is $\left\{X_{J(j)}\right\}_{j=1}^m$. Our main theorem studies the property of the index sequence $\{J(j)\}$. The proof is relegated to Appendix 8.A.

Theorem 8.5.1 *As the number of potential bidders* $n \to \infty$*, the number of active bidders* $m \to \infty$ *as well. Given that* $k \to \infty$*, the sequence* $\{\log J(k+j) - \log J(k+j-1)\}_{j=1}^\infty$ *is asymptotically an i.i.d sequence with exponential distribution and mean* $1/2$.

Remark 8.5.1 *In fact, based on the record and 2-record sequence, one can prove–in a way analogous to the proof of Corollary 4.5 of Resnick (1987) for the record sequence–that the point process with points*

$$\left\{\frac{1}{2}\log J(j) - \frac{1}{2}\log n\right\}_{j=1}^{\infty}$$

converges to a homogeneous Poisson point process. However, our proof of Theorem 8.5.1 is not based on point processes, and follows a simpler and novel approach.

Intuitively, Theorem 8.5.1 states that the differences of the sequence $\log J(j)$ are asymptotically i.i.d. and have an exponential distribution with mean 1/2. This indicates that $\log J(m)$ is approximately $m/2$ for sufficiently large m. Note that m active bidders will be observed until there are $J(m)$ potential bidders; conversely, one could say that if there are n potential bidders, the number of active bidders will be approximately $2\log n$.

We make this result precise by studying the asymptotic behavior of $J(m)$. To this end we first need to introduce two more sequences of random variables:

$$\xi_i = 1_{\{R_i \leq 2\}}, \quad \text{and} \quad N(i) = \sum_{k=1}^{i} \xi_k. \tag{8.4}$$

The number $\{\xi_i\}$ indicates whether the $i-$th potential bidder is an active bidder or not. The sequence $\{N(i)\}$ gives the number of active bidders among the first i potential bidders. An example of these two other sequences is also shown in Appendix 8.B.

The following two lemmas study the asymptotic normality of both the $N(n)$ and $J(m)$ sequences. The proofs are given in Appendix 8.A.

Lemma 8.5.1 *With the notation $N(n)$ defined in (8.4), the sequence*

$$\frac{N(n) - 2\log n}{\sqrt{2\log n}}$$

is asymptotically standard normal, as $n \to \infty$.

Lemma 8.5.2 *The sequence*

$$\frac{2\log J(m) - m}{\sqrt{m}}$$

is also asymptotically standard normal, as $m \to \infty$.

These lemmas imply the following convergence (in probability) results. As $n \to \infty$ and/or $m \to \infty$, we have that

$$\frac{N(n)}{2\log n} \xrightarrow{P} 1 \quad \text{and} \quad \frac{2\log J(m)}{m} \xrightarrow{P} 1. \tag{8.5}$$

Notice that $N(n)$ is the number of active bidders, while n is the number of potential bidders. Conversely, $J(m)$ is the number of potential bidders when we do observe m active bidders. Thus (8.5) gives the asymptotic relationship between the two sequences. This yields the $2\log n$ rule.

Rule 8.5.1 ($2\log n$) *If the number of potential bidders n is large, the number of active bidders is approximately equal to $2\log n$.*

The $2\log n$ rule relates the extent of the IA as measured by n to the market activity as measured by the number of active bidders $N(n)$. In the empirical sections we will test for this rule.

The $2\log n$-rule has implications for the time sequence of the bids in Amazon type auctions. Define the time at which a record or second record occurs as the *entering time*. Obviously, the entering time is related to the index sequence $\{J(j)\}_{j=1}^{m}$. Note that most auction sites do report the timing of the bids, but do not report the number of page views as an indicator of the number of potential bidders. Therefore a direct test of the $2\log n$ rule is not possible with publicly available data on such kind of auction sites. An indirect test may be feasible, however, by using the entering time sequence of the records. To see this, we model the arrival process of the potential bidders. The simplest model is the homogenous Poisson arrival process. Under Poisson arrivals, the appearance of potential bidders is random from the viewpoint of the seller and is independent from the time the auction has been running. Since we consider Amazon type auctions, there is no strategic reason for late bidding, but bidders may nevertheless display a preference for auctions which are close to the end of their run times. Most auction sites offer the possibility to easily rank order the relevant auctions on the remaining time to the announced deadline. Suppose agents actively use this feature for selecting auctions which are soon to close. Everything else equal, this preference arises from the cost of having to wait until the end of the auction. Therefore, we assume that the Poisson arrival rate λ increases as time progresses according to the following function

$$\lambda(t) = \lambda_0 e^{\theta t}, \tag{8.6}$$

where θ is the time preference factor, and where the beginning of the auction is at $t = 0$. Reversing time, θ can also be seen as the discount factor under continuous discounting. Since the Amazon type auction can be extended, we do not specify the end time. The time preference turns the homogenous Poisson arrival process into a non-homogeneous Poisson process with the instantaneous arrival rate $\lambda(t)$, see e.g., Klein and Roberts (1984).

For this non-homogeneous Poisson arrival process, we derive the asymptotic form of the entering time process of the record and 2-record sequence. The result is presented in the following theorem. The proof of the theorem is again relegated to Appendix 8.A.

Theorem 8.5.2 *Suppose the potential bidders* $1, 2, \cdots, n$ *arrive at times* $T(1), T(2), \cdots, T(n)$. *Let* $\{T(i)\}_{i=1}^{\infty}$ *be the arrival times of a non-homogeneous Poisson process where the rate of occurrence function is given as in (8.6). Then the arrival times of the active bidders is* $\{T(J(j))\}_{j=1}^{m}$. *For* $l \to \infty$, *the sequence* $\{T(J(l+j)) - T(J(l+j-1))\}_{j=1}^{\infty}$ *is asymptotically an i.i.d sequence with exponentially distributed innovations that have mean* $1/(2\theta)$.

The result of this theorem permits an indirect test of the $2 \log n$ Rule 8.5.1. Instead of directly using the number of potential bidders, Theorem 8.5.2 implies that one might as wel use the entering time sequence of the records. This is possible if one is willing to buy the extra maintained assumption of the specific non-homogeneous Poisson process due to the time preference function in (8.6).

8.6 Initial empirical evidence from pageviews

We provide some evidence for the $2 \log n$ rule on the basis of some aggregate statistics from other studies and a Dutch auction site that reports the number of page views.

8.6.1 Aggregate evidence on the $2 \log n$ rule

From the relationship between the number of active bidders and the number of potential bidders, (8.5), given a certain number of potential bidders, we can estimate the corresponding number of active bidders by the $2 \log n$ rule. But how large is the set of potential bidders? Clearly, the number of potential bidders is bounded, at least, by the size of the world population. Actually, a realistic number of potential bidders is likely to be much smaller, even though the internet has expanded the extent of the market. In Table 8.1 we provide some simple calculations for the number of potential bidders. The table shows that our rule implies that the number of active bidders should hardly ever exceed 50. Actually, a realistic upper bound is perhaps around 14, which is corresponds to around 1,000 potential bidders. By comparing these numbers, we can see that, although the number of active bidders is relatively small, it does not mean that the extent of the IA market is small. In other words, the $2 \log n$-rule illustrates that the extent of market is considerable, and has possibly greatly benefited from the fact that these items can now be sold through the internet, facilitating national and even international reach.

Number of potential bidders (n)	6 billion	1 million	10,000	5,000	1,000	500	100
Estimates of active bidders ($2 \log n$)	49.64	27.63	18.42	17.03	13.81	12.43	9.21
Standard deviation ($\sqrt{2 \log n}$)	7.05	5.26	4.29	4.13	3.72	3.53	3.03

Table 8.1: Estimates of the number of active bidders

This argument is supported by data from Yahoo! auctions. We collected the number of active bidders from the closed laptop auctions on Yahoo! with the Amazon auction style. We only consider the auctions which have more than 30 bids, including multiple bids from the same active bidders. The purpose of this requirement is to select the auctions which have the highest number of active bidders. There were 64 laptop auctions which matched for our selection in the period July 2006 - Dec. 2006. The histogram of the number of active bidders is shown in Figure 8.1. The sample mean is 5.80, with standard deviation 2.66. Actually, the maximum is 13, which is below the guesstimate upper bound of 14. These data show that modeling the bidding process as a record and 2-record sequence,

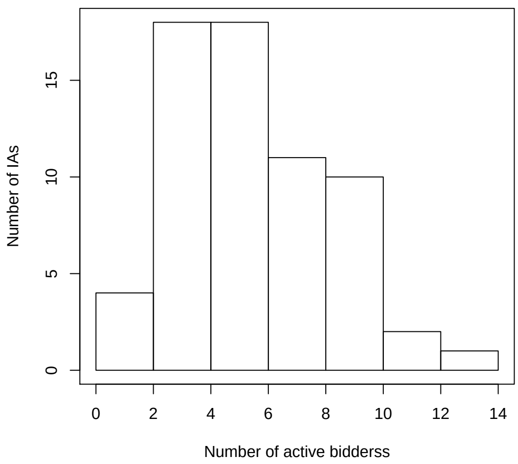

Figure 8.1: Histogram of the number of active bidders

is a possible explanation for why there are mostly few active bidders participating in a specific IA.

8.6.2 Regression evidence on the $2 \log n$ rule

On the large IA websites of eBay, Amazon and Yahoo!, the number of potential bidders is not reported. This hampers a direct evaluation of the $2 \log n$ rule. Fortunately, on some websites a system is in place that reports the number of people who checked a specific webpage. Potential bidders are identified uniquely by their IP addresses. In this way we are able to record the number of potential bidders. For example, the eBay owned Dutch advertising website www.marktplaats.nl (denoted as *Marktplaats* in the rest of this chapter) reports such kind of data. This creates a possibility to verify the model by (8.5).

Marktplaats is an advertising site with the option to bid or to negotiate. The auction is of the Amazon variety without a termination rule. Moreover, bids are non-binding as the bidder and seller have to finalize their agreement through an email contact. The website has the additional Buy It Now (BIN) feature that gives the bidders the possibility to contact the seller directly via email to bargain over the price. In contrast to several other sites with the BIN feature, it does not come with a posted price. Instead, the BIN feature is of the bargaining variety in which the buyer can make an email offer that can be accepted or countered by the seller. One of the attractions of an auction is that it reduces the transactions cost in the sense that it cuts out direct negotiations between the seller and the potential buyers. Thus a percentage of buyers can be expected to prefer not to enter into the costly process of bargaining, while some others, including the seller, may be tempted into bargaining.

Assume that there is a fixed percentage p of pageviewer who want to proceed by placing a bid. It means that, if there are n potential bidders, there will be only np who consider to place a bid on website, while the others choose the outside option to directly approach the seller. The number of active bidders therefore reads $N(np)$. From Lemma 8.5.1, we have that

$$\frac{N(np)}{\sqrt{\log n}} = \frac{2\log np + \sqrt{2\log np}\cdot\varepsilon_n}{\sqrt{\log n}} = \beta_1\sqrt{\log n} + \frac{\beta_0}{\sqrt{\log n}} + \varepsilon'_n, \tag{8.7}$$

where $\beta_1 = 2$, $\beta_0 = 2\log p$ and is a negative number, ε_n is asymptotically standard normal, while ε'_n is just asymptotically normal distributed with variance 2. The website provides data for $N(np)$ and n. So we can use linear regression to verify whether (8.7) holds or not. Moreover, the coefficient β_0 provides information on percentage of agents who choose the outside option. Note that we have divided both sides of (8.7) with $\sqrt{\log n}$. This ensures the homoscedasticity of the error terms, which would not follow from a direct specification of the BIN modified $2\log n$ rule.

We choose the specific category of laptops, since for these items the IPVP is a priori the natural setting. We used the advertisements in the period Jan.2006 - Mar.2006. We only look at auctions with at least 5 active bidders. This requirement was imposed to ensure that there are enough potential bidders, since (8.7) is an approximate relationship that applies only if the number of potential bidders is sufficiently large. In particular, considering that only a small proportion of potential bidders prefer to bid online due to the BIN feature, the requirement is useful to ensure a sufficiently large number of potential bidders. In total, 32 auctions qualified. The statistics of the number of potential bidders and the number of active bidders are shown in Table 8.2.

Variable	Mean	Std Dev	Min	Max	Median
The number of potential bidders	603.7	326.4	164	1331	552
The number of active bidders	6.9	2.3	5	15	6

Table 8.2: Statistics of the data

Linear regression analysis is employed to evaluate (8.7). The result is shown in Table 8.3. According to (8.7), we know that the theoretical value of the coefficients should be $\beta_0 = 2\log p \leq 0$ and $\beta_1 = 2$. We test two hypotheses on the two coefficients:

$$H_{0,0} : \beta_0 > 0 \text{ versus } H_{0,1} : \beta_0 \leq 0,$$

and

$$H_{1,0} : \beta_1 = 2 \text{ versus } H_{1,1} : \beta_1 \neq 2.$$

The p-values of these two tests are shown in the last column in Table 8.3. So one can reject $H_{0,0}$ at the 10% significant level 0.1. But one can not reject the hypothesis $H_{1,0}$ at any conventional level of significance. We conclude that the empirical data seem to fit the theoretical model quite well.

Parameter	Estimate	Std. Error	t Stat.	p-value for $H_{i,0}$
β_0	-5.43	3.80	-1.43	0.082
β_1	1.98	0.61	3.24**	0.970
R-Squared: 0.153	Adjusted R-squared: 0.125			

**: Significant level 5%.

Table 8.3: Empirical test on (8.7)

An additional result is that we can estimate p as $e^{\hat{\beta}_0/2} = 0.066$. It means that 6.6% potential bidders choose not to directly negotiate via email, but instead prefer to bid first. Thus, only a small percentage of the bidders in the end prefer the auction over the bargaining. Since the auction part of Marktplaats does not offer the proxy bid feature, active bidding can be very time consuming and moreover it reveals information to competitors. Nevertheless, the estimated p is not zero since there are also costs associated with bargaining, i.e. bargaining with the seller can also be time intensive. Apparently the tradeoff between BIN and the auction is in favor of the bargaining process. But we do not want to pursue this issue further given the $2\log n$ rule being the topic of this chapter and the fact that this is one of the advertising sites.

8.7 Empirical evidence from the timing of the bids

Under the assumption of Poisson arrivals, we can also test Theorem 8.5.1 indirectly via the timing of the bids. In Subsection 8.7.1 we run a small simulation study to illustrate the methodology. An empirical study on real data from a large IA site is given in Subsection 8.7.2.

8.7.1 Simulation

The simulation exercise starts by simulating the valuation of the potential bidders. We record the index sequence of the active bidders to illustrate Theorem 8.5.1. To this end we generate 10,000 i.i.d. random variables. We simulate from the uniform distribution, though we notice that choice of the distribution does not affect the record sequence. Then, by the definition of record and 2-record sequence, we find the corresponding index sequence $\{J(j)\}$. By taking the difference between logarithm of $\{J(j)\}$ sequence, we may verify whether the innovations are asymptotically i.i.d. exponentially distributed with mean 1/2. Because this result holds only as $l \to \infty$, see Theorem 8.5.1, we only use the upper 2/3 of the differences. To verify the exponential distribution feature, we employ the QQ-plot device. The plot is drawn for the empirical quantile of $\{\log J(l+j) - \log J(l+j-1)\}$ against the exponential distribution with mean 1/2, see Figure 8.2.

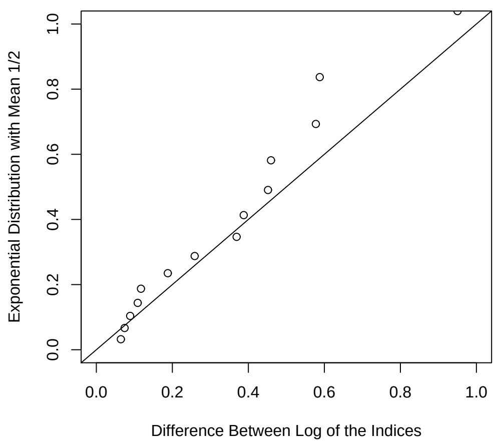

Figure 8.2: QQ-plot on indicies

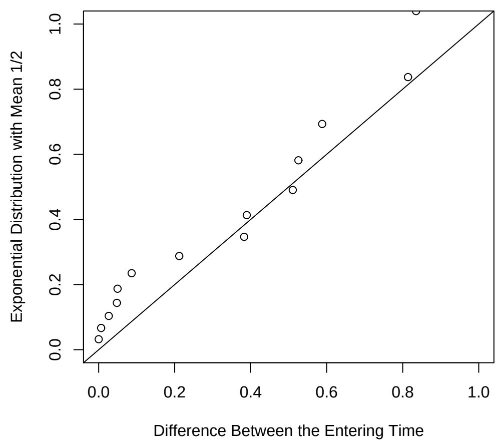

Figure 8.3: QQ-plot on entering times

Figure 8.2 shows that the exponential distribution fits well. This is also confirmed by the Kolmogorov-Smirnov (KS) test: the test statistics is 0.1333 with a corresponding p-value 0.9998. The simulation illustrates that the QQ-plot method can be used to verify whether an index sequence is really the index of a record and 2-record sequence.

To be able to apply this methodology on real data, we can not use the index sequence as it is not observable, but instead we can use the entering time sequence of potential bidders, as we discussed in Theorem 8.5.2. Hence, besides the valuation sequence, we also simulate the arrival process of the potential bidders. Following the discussion in connection with Theorem 8.5.2, we simulate the non-homogeneous Poisson arrival process. In this simulation we combine the $\{J(j)\}$ sequence simulated above with the independently simulated non-homogeneous Poisson arrival process $\{T(i)\}$ with $\theta = 1$.[2] Then we can use

[2] For simulating non-homogeneous Poisson process with a log linear rate function, see Lewis and Shedler

the simulated entering time sequence $\{T(J(j))\}$ to replace the index sequence $\{J(j)\}$. By taking only the upper 2/3 of the differences, we can again make a QQ-plot, see Figure 8.3. The KS test statistics is 0.2667 with a p-value 0.6781. The KS test again indicates a good fit, but of course due to the extra noise, the quality of the fit is somewhat lower. The extra noise due to the introduction of the non-homogeneous Poisson arrivals can also be seen by comparing Figure 8.3 with Figure 8.2. A cross plot of these two simulations is also made to check whether the distributions of these two simulated samples are about equal, see Figure 8.4. The KS test for equality yields a p-value of 0.3885. Since the difference is not significant, it seems that the methodology of using QQ-plot to verify the specific record model is reasonable, even under the assumption of stochastic arrival times.

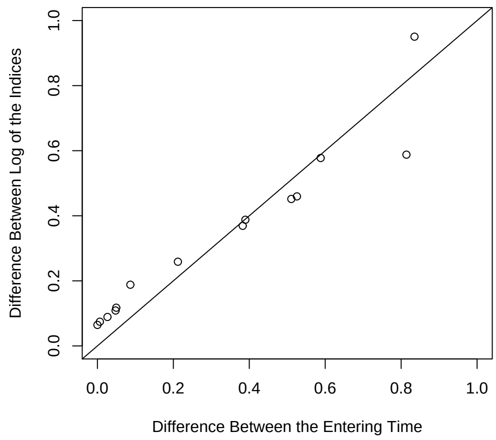

Figure 8.4: QQ-plot between two simulated samples

8.7.2 Empirical application

The methodology of Subsection 8.7.1 is now applied to real data in order to verify whether the bidding process follows a record and 2-record sequence, in other words, whether Theorem 8.5.2 has a bite.

We choose auctions from the Yahoo! IA site. As discussed in Subsection 8.2.2, on Yahoo! there are two kinds of auctions. The Amazon type auction has no fixed termination time. In this kind of auction people have no incentive to delay their bid for strategic reasons (as in the eBay type auction with a hard close). This fact corresponds with the Assumption 8.3.1. So, we choose an Amazon type auction for our investigation.

First, a single auction on a Dell Inspiron 1500 laptop is studied. The auction started March 26, 2007 and lasted for ten days. There were 9 active bidders in this auction. Ranking the 9 active bidders' entering time, we take only the upper 2/3 as the relevant

(1976).

sample since our result only holds asymptotically. Hence, there are only 6 entering times under consideration. By taking the differences between these 6 entering times, with the assumption that the potential bidders come to the auction according to a non-homogeneous Poisson process, the 5 differences should be asymptotically i.i.d. exponentially distributed. The estimated time preference parameter turned out to be $\hat{\theta} = 0.007$, which implies a half-life of 99 minutes.[3] By normalizing the mean to 1, we drew a QQ-plot with respect to the exponential distribution with unit mean, see Figure 8.5.

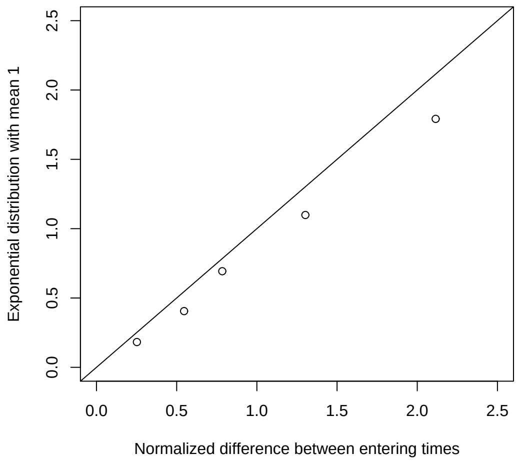

Figure 8.5: QQ-plot for a single Yahoo! Auction

Due to the low number of observations, Figure 8.5 does not clearly show whether the fit is good or bad. The KS test, however, still gives a p-value 0.918. Hence we can not reject the null hypothesis that the differences follow the exponential distribution. Thus the entering times in this auction seem to be in accordance with Theorem 8.5.2.

Since our model is based on IPVP and Assumption 8.3.1, the above test can be viewed as a weak test of IPVP. Notice that the IPVP tests in the literature mostly employ data from multiple auctions. Therefore, assumptions such as homogeneity across auctions are always required. In comparison with the literature, the advantage of our single-auction test is that it is only based on Assumption 8.3.1 and the maintained hypothesis of Poisson arrivals; this does not require homogeneity across auctions. But we can nevertheless try to pool the data from different auctions to increase efficiency. If we do so, we will preserve the per auction approach philosophy by allowing θ to differ across auctions.

We collected a number of other Amazon type laptop auctions on Yahoo!. All the auctions between October 2006 and April 2007 with at least 4 active bidders and at least 25 bids were collected. The restrictions are again mainly to ensure there are enough

[3]By solving the equation $e^{-\hat{\theta}t} = 0.5$, we get $t = 99$. In other words, each time 99 minutes have passed, the incentive for people to check this auction doubled.

potential bidders since our theorems only reveal the asymptotic properties. Ten auctions qualified.[4] For each auction, we use the same procedure as for the single auction discussed above. The 10 estimated time preference parameters have a mean 0.0079 (implied half-life is one hour and a half), with standard deviation 0.008. The maximum and minimum estimated time preference parameters are 0.0256 and 0.0003, with half-lives of half an hour and one and half day, respectively. These estimates imply that agents have a quite variable and sometimes high preference for auctions that are near closing time. After the standardization with θ, we combine all the normalized differences together and construct a QQ-plot with respect to exponential distribution with mean 1. The QQ-plot is given in Figure 8.6. The KS test has p-value 0.378. Thus the null hypothesis of exponential

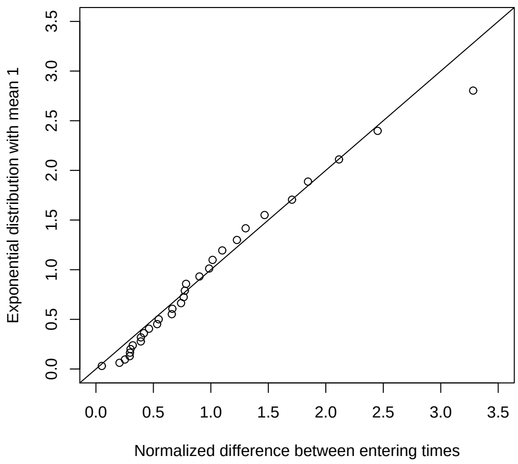

Figure 8.6: QQ-plot for combined auction data

distribution is again not rejected. Therefore, also in lager samples, the $2 \log n$ rule can not be rejected either.

8.8 Conclusion

Internet auctions attract numerous agents, but only a few become active bidders. The number of potential bidders is the extent of the market. This number, however, is unknown to the internet auctioneer. In this chapter, we study the connection between the number of potential bidders and the number of active bidders, in order to explain the low number of active bidders.

Our study started from the bidding process of the Amazon type IA. These IAs are a hybrid of the English auction and the second price auction, but without the strategic last

[4] There were 3 auctions with 4 active bidders, 2 auctions with 5 active bidders and the other 5 auctions had 6,7,8,9 and 11 active bidders respectively. An eleventh auction qualified, but this auction posted multiple identical items and was therefore not taken into account.

minute bidding. Under the assumption that all active bidders are notified immediately when overbid and respond directly, the active bidders' valuations can be modeled as the record and 2-record sequence of the potential bidders' valuation sequence.

We proved that the logarithmic difference of the index of the record and 2-record sequence is asymptotically i.i.d. exponentially distributed with mean 1/2. The number of potential bidders are thus connected with the number of active bidders through the $2\log n$ rule. Data from a small Dutch site were used to test for this relationship empirically.

On large IA websites such as eBay, Amazon and Yahoo!, the number of potential bidders is not reported. If, however, the potential bidders come to the auction according to a non-homogeneous Poisson process, then we can test for the $2\log n$ rule. Under the Poisson arrival process the publicly available entering time sequence of the active bidders can substitute for the logarithm of the index of the record and 2-record sequence. The empirical study using Yahoo! data supports the theoretical model, but showed quite some variation in the discount factor. Future research whereby this variation is explained by product characteristics seems of interest.

The $2\log n$ rule explains why there are always few active bidders in an IA. Since the number of active bidders is logarithmically related to the number of bidders potentially showing interest for a particular IA, the extent of the market can be much larger than is revealed by direct observation of the active bidders. Table 8.1 shows that, with 10,000 potential bidders, 10 active bidders are within the 95% confidence band, but 25 active bidders are also compatible. Thus the extent of the market may at times be quite different given the activity levels in terms of the number of active bidders. In our sample, we never observed more than 11 active bidders. If 11 active bidders are considered as the upper bound of the confidence band, such a real activity is still compatible with as few as 21 potential bidders but also as many as 21,249 potential bidders. Thus the $2\log n$ rule explains the low observed bidding activity, but does not necessarily imply a very large number of potential bidders in a particular auction.

To see the economic relevance of the $2\log n$ rule, consider a case of a laptop IA where there are 500 potentially interested agents. Suppose the distribution of valuations is uniform on $[0, 500]$.[5] By Table 8.1, this implies that the expected number of active bidders is 12. If the seller were to calculate the potential gain from the auction by only considering the number of active bidders, he would arrive at $11/13 \cdot 500 = 423.08$$. This would be a considerable underestimate of the true gains. From the 500 potential bidders, the expected revenue is in fact $499/501 \cdot 500 = 498.00$$. This is 75$ higher than the back of an envelope calculated guesstimate on basis of the directly observed market extent.

[5]In one of the ten laptop auctions studied in the previous section, the BIN price was posted at 500$, while the first bid was at 1$. The uniform distribution is commonly used in auction theory

8.A Appendix A

Proof of Lemma 8.5.1

From the independence of the $\{X_i\}_{i=1}^n$, we get that $\{R_i\}_{i=1}^n$ is a sequence of independent random variables, see Resnick (1987), Proposition 4.3(i). Thus, the $\{\xi_i\}_{i=1}^n$ are also independent. Since $N(n)$ is the partial sum sequence of $\{\xi_i\}_{i=1}^n$, by using the central limit theorem for independent bounded random variables, we get the asymptotic normality immediately.

The asymptotic mean and variance are calculated from the distribution of the $\{\xi_i\}_{i=1}^n$. From Proposition 4.3(i) in Resnick (1987), we have

$$P(R_i = s) = 1/i,\ s = 1, 2, \cdots, i,$$

so that

$$P(\xi_i = 1) = \frac{2}{i} \text{ and } P(\xi_i = 0) = 1 - \frac{2}{i}. \tag{8.8}$$

This implies

$$EN(n) = 1 + \sum_{i=2}^{n} \frac{2}{i} = 2\log n + o(\sqrt{\log n}),$$

and

$$Var(N(n)) = \sum_{i=2}^{n} \frac{2}{i} - \sum_{i=2}^{n} \frac{4}{i^2} \sim 2\log n.$$

This proves the lemma. □

Proof of Lemma 8.5.2

For fixed positive number x, denote

$$n_0(m, x) = [\exp(\frac{xm^{1/2} + m}{2})].$$

Then,

$$\frac{m - 2\log n_0}{\sqrt{2\log n_0}} \to -x \quad as\ m \to \infty$$

Since $J(m)$ is an integer, we have

$$P(\frac{2\log J(m) - m}{\sqrt{m}} \leq x) = P(J(m) \leq n_0(m, x)).$$

Notice that the two events $\{J(m) \leq n_0\}$ and $\{N(n_0) \geq m\}$ are actually the same. Therefore,

$$P(\frac{2\log J(m) - m}{\sqrt{m}} \leq x) \quad = \quad P(J(m) \leq n_0)$$

$$\begin{aligned} &= P(N(n_0) \geq m) \\ &= P(\frac{N(n_0) - 2\log n_0}{\sqrt{2\log n_0}} \geq \frac{m - 2\log n_0}{\sqrt{2\log n_0}}) \\ &\to 1 - \Phi(-x) = \Phi(x) \quad as\ m \to \infty. \end{aligned}$$

Here we use Lemma 8.5.1 in the last step. This proves Lemma 8.5.2. □

Proof of Theorem 8.5.1

As a consequence of Lemma 8.5.1, when $n \to \infty$, $m = N(n) \to \infty$. Similarly, from Lemma 8.5.2, $J(m) \xrightarrow{P} \infty$ when $m \to \infty$.

As we discussed in the proof of Lemma 8.5.1, the random variables $\{\xi_i\}_{i=1}^n$ are independent, and the distribution is given in (8.8). So we have

$$\begin{aligned} & P(J(j+1) > s | J(j) = s_n, J(j-1) = s_{n-1}, \cdots, J(1) = s_1) \\ = \ & P(\xi_{s_n+1} = 0, \cdots, \xi_{s-1} = 0, \xi_s = 0) \\ = \ & \left(1 - \frac{2}{s_n+1}\right)\left(1 - \frac{2}{s_n+2}\right)\cdots\left(1 - \frac{2}{s-1}\right)\left(1 - \frac{2}{s}\right) \\ = \ & \frac{s_n - 1}{s_n+1} \cdot \frac{s_n}{s_n+2} \cdot \frac{s_n+1}{s_n+3} \cdots \frac{s-3}{s-1} \cdot \frac{s-2}{s} \\ = \ & \frac{s_n(s_n-1)}{s(s-1)} \end{aligned} \tag{8.9}$$

which implies that $\{J(j)\}_{j=1}^n$ is a Markov process. From (8.9), subsequently,

$$\begin{aligned} & P(\log J(j+1) - \log J(j) > x | J(j) = s_n, \cdots, J(1) = s_1) \\ = \ & P(J(j+1) > e^x s_n | J(j) = s_n, J(j-1) = s_{n-1}, \cdots, J(1) = s_1) \\ = \ & \frac{s_n(s_n-1)}{[e^x s_n]([e^x s_n] - 1)} \\ \to \ & e^{-2x}, \ \ when\ s_n \to \infty. \end{aligned} \tag{8.10}$$

When $l \to \infty$, combining $J(l) \xrightarrow{P} \infty$ with (8.10), the theorem is proved. □

Proof of Theorem 8.5.2

Let $M(t)$ be the number of potential bidders arriving at the auction site before time t. According to the property of a non-homogeneous Poisson process, $M(t)$ follows a Poisson distribution with mean

$$\mu(t) = \int_0^t \lambda(s)\, ds = \int_0^t \lambda_0 e^{\theta s}\, ds = \lambda_0 \frac{e^{\theta t} - 1}{\theta}.$$

Hence, when $t \to \infty$, $\mu(t) \to \infty$ and

$$\frac{\mu(t)}{e^{\theta t}} \to \frac{\lambda_0}{\theta} =: c.$$

Consider the family of random variables $\{M(t)/\mu(t)\}$. Notice that as $t \to \infty$,

$$Var(M(t)/\mu(t)) = Var(M(t))/(\mu(t))^2 = 1/\mu(t) \to 0.$$

So we get that $M(t)/\mu(t) \to 1$ in probability. Therefore, as $t \to \infty$

$$\frac{M(t)}{e^{\theta t}} \xrightarrow{P} c,$$

which implies that $\log M(t) - \theta t \to \log c$ in probability.

Replace t with $T(J(l+j))$ and let $l \to \infty$. Considering the fact that $M(T(J(l+j))) = J(l+j)$, we get that

$$\log J(l+j) - \theta T(J(l+j)) \xrightarrow{P} \log c.$$

Replacing j with $j-1$ in this relation, we also have that

$$\log J(l+j-1) - \theta T(J(l+j-1)) \xrightarrow{P} \log c.$$

Combining these two results implies,

$$(\log J(l+j) - \log J(l+j-1)) - \theta(T(J(l+j)) - T(J(l+j-1))) \xrightarrow{P} 0.$$

By the conclusion of Theorem 8.5.1, the Theorem 8.5.2 follows. $\square$

8.B Appendix B

Here we present an example to understand the notation in this chapter. Suppose we have the index sequence $i = 1, 2, \cdots 6$ with the valuation sequence $X_1 = 30, X_2 = 10, X_3 = 60, X_4 = 20, X_5 = 50, X_6 = 40$. In Table 8.4, we present the rank sequence R_i, the indicator sequence ξ_i, and the number of active bidder sequence $N(i)$. In this example, the corresponding index sequence of the record and 2-record $J(j)$ is given as $J(1) = 1, J(2) = 2, J(3) = 3, J(4) = 5$. In this case, $m = 4$. Moreover, suppose all bids are proxy bids and that 1 is the minimum bid increment.

i	1	2	3	4	5	6
X_i	30	10	60	20	50	40
R_i	1	2	1	3	2	3
ξ_i	1	1	1	0	1	0
$N(i)$	1	2	3	3	4	4
j	1	2	3	–	4	–
$J(j)$	1	2	3	–	5	–
Price	1	11	31	–	51	–

Table 8.4: Example for notations

Chapter 9

The Expected Payoff to Internet Auctions

9.1 Introduction

Internet auctions (IA) provide an easily accessible platform for trade. This has increased the extent of the market for many items to nationwide or even global markets. IA improve the matching between the supply and demand side. Bulow and Klemperer (1996) showed that under the hypothesis of the Independent Private Values Paradigm (IPVP), the seller is better off if the market is larger. By the same fact, buyers are also better off, since it becomes more likely that the agent with highest consumer surplus is matched with the seller.

Auction theory derives the optimal bid functions for specific auction mechanisms, such as the Dutch (descending price) or the English (ascending price) auctions, and given a specific demand function. The demand function is modeled as a distribution of valuations of the object to be auctioned. Both the seller and the buyer have an interest in knowing the final price that might materialize to answer such questions as: Is it worthwhile to put the item up for sale? and, is it worth my time to bid? Sellers may want extrapolate from a single auction to predict total revenues from repeat sales. Competition authorities have an interest in the price that is to be expected in order to determine whether the bidding process was fair (e.g. was not hampered by a bidding ring). The analysis may also lead to a prediction of the final price, given that the bids observed half way through the auction (IA typically run for multiple days). Fortunately, auction theory under IPVP makes quite a robust prediction about the expected price given the number of buyers and the specific distribution of valuations.

This prediction is as follows. For any standard auction, the Revenue Equivalence Principle (REP) holds, which means that under IPVP, the expected revenue of the seller does not depend on the auction mechanism, see e.g. Krishna (2002).[1] The expected revenue for a standard mechanism is equal to the expectation of the second highest order statistic of the valuations. For the second-price sealed bid auction, this is easily shown to be the case. In a second-price sealed bid auction, the winner pays the second highest bid. Since for this auction agents have an incentive to exactly bid their valuations, the claim follows. Auction theory shows that this revenue result holds for all standard auctions.

Most IAs have two mechanisms for placing a bid, i.e. the manual bid and the proxy bid. This induces a hybrid of an English auction and a Vickrey auction. In an English auction, bidders publicly compete with each other by placing ascending bids. The Vickrey mechanism uses sealed bids, i.e. bidders do not see the bids of competitors. In a Vickrey auction, the winner pays the amount bid by the second highest bidder. Since both of these two mechanisms are standard, the IA is also standard in the sense of revenue equivalence. Compared to studies of classical auction mechanisms, the empirical analysis of an IA is severely hampered due to the unobserved number of potential bidders. In an IA, besides the *active bidders* who indeed place a bid on the websites, there are also a large number of *potential bidders* who only check the website with or without placing a bid. The number of potential bidders plays a role similar to the number of bidders in classical auction mechanisms, i.e. the number of bidders sitting in the auction hall. However, for IA, it is hard to observe this number. Firstly, most of the large IA sites do not provide the number of page views. Secondly, even the number of page views would not be a clean estimate of the number of potential bidders, since a potential bidder may check the website multiple times.

Lacking the knowledge of the number of potential bidders severely restricts the statistical analysis on the expected revenue for the seller. Therefore, a few papers recently focus on analyzing data under some additional assumptions regarding the number of potential bidders. Bajari and Hortaçsu (2003) and Paarsch (1992) do not require knowledge of the number of potential bidders, but they assume that the observed bidders are the only potential bidders willing to pay the reserve price. This assumption appears implausible for IA. For instance, de Haan *et al.* (2008b) argued that the actual extent of the IA market, i.e. the number of potential bidders, is far beyond the observed number of active bidders. Alternative approaches are based on modeling the number of potential bidders. For example, McAfee and McMillan (1987) analyzed the case when the number of bidders is stochastic. Another example is Laffont *et al.* (1995), who assumed that the unknown number of potential bidders is the same across all auctions under consideration.

Still different is Song (2004) who considered the nonparametric estimation of the dis-

[1] In classical auction theory, an auction is called *standard* if the rule dictates that the person who bids the highest amount is awarded the object.

tribution of bidder's valuation without having any information on the number of potential bidders. Song (2004) argued that without knowing the number of potential bidders, the distribution of the bidder's valuation is not identified if only the payoff, i.e. the second highest order statistic, is observed. But if one can observe the bid history, then using the two top order statistics identifies the parent distribution, even if the number of potential bidders is unknown.

Under the IPVP, the potential bidder's valuations are assumed to be identically and independently distributed random variables. Because the payoff of an IA is the second largest valuation among all the potential bidders[2], when the number of potential bidders is sufficiently large, the payoff only depends on the tail of the distribution of the bidders' valuations. Thus there is only a need to model the tail of the distribution. Semi-parametric Extreme Value Theory (EVT) provides an approximation to the tail of the distribution. Caserta and de Vries (2005) applied the EVT approach to investigate the expected payoff. However, the number of potential bidders is a major difficulty for their analysis as they assume that the number of actual bids equals the number of potential bidders.

Existing econometric analysis of auctions, c.f. Paarsch (1992), often proceeds on the basis that the number of bidders is known and that the different auctions are homogeneous, possibly controlled for covariates. This allows for the pooling of the data from different auctions in order to estimate the demand curve (distribution of valuations) and to test for the IPVP. In the current analysis, we do not necessarily want to make this maintained homogeneity assumption, as we want to investigate the expected price of a particular, possibly unique, auction. For this purpose the EVT approach appears appropriate.

Similar to Caserta and de Vries (2005), we also model the tail of the distribution of the bidder's valuation as in the EVT setup. Unfortunately, the message of the paper is somewhat bleak. We show that while for distributions of valuations in the max-domain of attraction with positive extreme value index, the logarithm of the expected payoff can be estimated after application of a correction factor, the expected payoff cannot be estimated consistently. A somewhat similar result is obtained for the negative case. Only for a subset of distributions in the max-domain of attraction with zero extreme value index does a consistent estimator exist with a certain speed of convergence under a suitable second order condition.

This chapter is organized as follows. In Section 9.2, the record and 2-record model is revisited. Section 9.3 demonstrates the EVT approach with positive, negative and zero

[2]In fact, the final payoff should be the second largest valuation plus a minimum increment because the winner has to overbid the second largest valuation. We assume that the minimum increment is negligible compared to the value.

extreme value index. For the zero case, a subclass model of the domain is introduced. Section 9.4 concludes this chapter.

9.2 The bidding activities in Internet auction

IAs have some features that differentiate these from the standard auction mechanisms, but are otherwise just the internet version of known auctions. The differences pertain to the bidding systems and the termination rules. The Internet facilitates the use of two bidding systems simultaneously. Most of the IA sites allow for *manual* and *proxy bidding*. Manual bidding is similar to the first price open ascending bid in an English auction, while the proxy bidding procedure captures the second price sealed bid mechanism studied by Vickrey (1962). Proxy bidding proceeds by providing the server of the IA with the maximum value a person would be willing to pay. The machine then takes over and keeps on overbidding on behalf of the proxy bidder as long as the other bids are below this maximum. Regarding the termination rules, there are also two alternatives. One type of IA ends after a pre-announced fixed lapse of time, while the other type has a variable auto-extended termination time. Typical examples are the eBay auctions and the Amazon auctions. The eBay auctions have a fixed ending time. The Amazon type auctions use the auto-extension termination rule.[3] At the beginning of the Amazon type auction, an initial ending time is announced. If no bidding takes place during the last ten minutes, the auction stops at the announced ending time. But if there are some bids in the last ten minutes, the ending time is automatically extended by another ten minutes. This rule is also applied to the new extension period. On the Yahoo! auction site the sellers can choose between these two termination rules.

The analysis of auctions can be divided into two classes. Either it is assumed that the bidders' valuations are independent from each other, or they are dependent. The former case is usually referred to as the independent private values paradigm (IPVP). Valuations are considered to be draws from some given distribution. This is the paradigm that we consider in this chapter as well. Standard commodities are well modeled on the IPVP assumption. Rare items, collectibles and works of art are usually considered to be in the other class. At the extreme end of the other class is the common value case. Under the IPVP and the Amazon type termination rule, de Haan *et al.* (2008b) argued that the active bidders come to the IA as a record and 2-record arrival process, while their valuations form the record and 2-record sequence of the valuations among all potential

[3] Although www.amazon.com has terminated their auction platform, since they used the feature of auto-extension termination rule, we still call auctions with such kind of setup the Amazon type auction. On the Yahoo! platform, this feature is still in use.

bidders.

To explain this model, let $i = 1, 2, \cdots, n$ denote the order in which the n potential bidders arrive at the auction site. IPVP assumes that the valuation of all potential bidders are i.i.d. random variables $X_1, X_2, \cdots, X_n$ with distribution function $F(x)$. Define the rank sequence $\{R_i\}_{i=1}^n$ as

$$R_i := \sum_{k=1}^{i} 1_{\{X_k \geq X_i\}}. \tag{9.1}$$

Intuitively, R_i is the rank of the valuation of the $i-$th potential bidder among the valuations of all the potential bidders who checked the auction earlier (before i). The valuation X_i is called a *record* if $R_i = 1$. Similarly, for $k = 2, 3, \cdots$, it is a *k-record* if $R_i = k$, see Resnick (1987). Denote the indices of the records and 2-records as $\{J(j)\}_{j=1}^m$. This index sequence is given by

$$J(1) = 1,\ J(2) = 2 \tag{9.2}$$

$$J(j+1) = \min\{i > J(j) : R_i \leq 2\},\ j = 2, 3, \cdots, m-1, \tag{9.3}$$

where m is the number such that $R_i > 2$ for all $i > J(m)$.

With the maintained hypothesis that *"each active (manual) bidder immediately returns to the IA and increases his bid as soon as he is overbid and his valuation is above the prevailing price."*,[4] the active bidders must have the indices $\{J(j)\}_{j=1}^m$ in the potential bidders sequence. So m is the number of active bidders. Then, the active bidders' valuations are obviously the record and 2-record sequence $\left\{X_{J(j)}\right\}_{j=1}^m$. Actually, the first $m-1$ active bidders' valuations can be observed as their last bids. The winner's valuation $X_{J(m)}$ is obviously unobservable, just as in the English auction. For the Amazon type auction, since there is no motivation for bidders to postpone their bids for strategic reasons, it can be assumed that the bids reflect the first $m-1$ records and 2-records. de Haan *et al.* (2008b) tested this model by employing Yahoo! IA data.

9.3 EVT approaches

Our purpose is to compare the observed payoff and its expectation for a specific IA. With n potential bidders, and m active bidders, there are two ways to represent the payoff following the record and 2-record model. One way is to consider $M_{n-1:n}$ as the second largest order statistics of $X_1, X_2, \cdots, X_n$. The other way is to view the payoff as the $X_{J(m-1)}$, where $\{J(j)\}_{j=1}^m$ is the record and 2-record index sequence as defined in the previous section.

[4]Note that this assumption is automatically satisfied when there are only proxy bidders present.

Since the payoff is determined by the largest order statistics, it is reasonable to make assumptions only on the right tail of the valuation distribution $F(x)$. Caserta and de Vries (2005) suggested to use the EVT approach, and assume that the distribution of the bidder's valuation belongs to the max-domain of attraction of an extreme value distribution. This setup is as follows.

Suppose the bidder's valuations are i.i.d random variables $X_1, X_2, \cdots, X_n, \cdots$ with common distribution function F. Denote $M_n = \max\{X_1, \cdots, X_n\}$. We say that F belongs to the max-domain of attraction, if there exist a non-degenerate distribution function G, a positive sequence $\{a_n\}_{n=1}^{\infty}$ and a real sequence $\{b_n\}_{n=1}^{\infty}$, such that

$$\lim_{n\to\infty} P\left\{\frac{M_n - b_n}{a_n} \le x\right\} = G(x)$$

for all continuity points of G. Denote this domain of attraction feature as $F \in \mathcal{D}(G)$.

The necessary and sufficient condition for a distribution function to belong to the max-domain of attraction is the extreme value condition , see e.g. de Haan (1984a).

Proposition 9.3.1 *Let* $U := \left(\frac{1}{1-F}\right)^{\leftarrow}$ *be the generalized inverse function of* $1/(1-F)$. *Then* $F \in \mathcal{D}(G)$ *if and only if there exists a function* $a(t) > 0$ *such that*

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, \tag{9.4}$$

for some $\gamma \in \mathbb{R}$ *and all* $x > 0$.

Here γ is called the *extreme value index*. Under the extreme value condition, the following proposition is proved in Caserta and de Vries (2005).

Proposition 9.3.2 *Let* $X_1, \cdots, X_n, \cdots$ *be i.i.d. sequence with common distribution function F belonging to domain of attraction, i.e. (9.4) holds for some* $\gamma < 0$. *Then, we have that*

$$\lim_{n\to\infty} \frac{EM_{n-1:n} - U(n)}{a(n)} = -\Gamma(2-\gamma) \tag{9.5}$$

From Proposition 9.3.2, a possible estimator of $EM_{n-1:n}$ is $\hat{U}(n) - \hat{a}(n)\Gamma(2-\hat{\gamma})$, where $\hat{U}(n)$, $\hat{a}(n)$,and $\hat{\gamma}$ are proper estimators for the location, scale and shape parameter in the EVT model. The literature offers several alternative estimations for these parameters. We note here that (9.5) holds for $0 \le \gamma < 2$ as well.

Since in our model, only the record and 2-record sequence is observed, it is necessary to have proper estimators based on only those observations. When γ is positive, Berred (1992) derived an estimator for γ based on the record sequence, which can be generalized to our case of the record and 2-record sequence.

The main difficulty in this approach is that the number of potential bidders n is in fact unknown. This lack of information inhibits the estimation of $U(n)$ and $a(n)$. In Caserta and de Vries (2005) n is assumed to be equal to the number of bids. Thus multiple bids from the same bidder are considered as coming from different potential bidders. Although this estimate of n is larger than the observed m, it is a rather inaccurate estimation. In de Haan *et al.* (2008b), it is shown that $m \sim 2\log n$ as $n \to \infty$. All in all, to consider the expected payoff as the expectation of the second largest valuation seems to be an approach of limited value due to the unknown number of potential bidders.

That leaves us with the the second representation to model the payoff based on the record theory. We first study the record and 2-record sequence via its point process representation. Let $\{L_k(n)\}_{n=1}^{\infty}$ be the indices of the $k-$record sequence, that is

$$L_k(1) = 1,\ L_k(n+1) = \min\{j > L(n) : R_j = k\},\ n = 1, 2, \cdots .$$

Define the $k-$record point process N_k by

$$N_k := \sum_{n=1}^{\infty} \varepsilon_{X_{L_k(n)}},$$

for $k = 1, 2, \cdots$. Then according to Proposition 4.30 in Resnick (1987), the point processes $\{N_k\}_{k=1}^{\infty}$ are i.i.d. random elements.

It is clear that the combination of L_1 and L_2 sequences constitutes the record and 2-record sequence J. According to Proposition 4.1(ii) in Resnick (1987), N_1 and N_2 are homogeneous Poisson processes on $(0, +\infty)$. Then, the point process of the records and 2-records must be

$$N = N_1 + N_2$$

which is the sum of two independent homogeneous Poisson process on $(0, +\infty)$. So N is a Poisson process with intensity measure 2μ, where μ is the Lebesgue measure on $(0, +\infty)$. In other words, we have the following Lemma.

Lemma 9.3.1 *Suppose $E_1, E_2, \cdots$ are i.i.d standard exponentially distributed random variables and $\{J_E(m)\}$ is the index sequence of the records and 2-records of $\{E_n\}$. Then*

$$\left\{E_{J_E(m)}\right\}_{m=2}^{\infty} \stackrel{d}{=} \{\Gamma_n\}_{m=2}^{\infty}$$

where $\Gamma_m = \sum_{i=1}^{m} E_i'$ is the partial sum of the sequence $\{E_i'\}_{i=1}^{\infty}$ which is an i.i.d sequence with exponential distribution and mean $1/2$.

By defining $Q := (-\log(1-F))^{\leftarrow}$, the i.i.d sequence $\{X_n\}$ can be represented as $\{Q(E_n)\}$, where $E_1, E_2, \cdots$ are i.i.d standard exponentially distributed random variables. Hence, a direct implication of Lemma 9.3.1 is as follows.

Corollary 9.3.1 *The record and 2-record sequence can be represented as*

$$\left\{X_{J(m)}\right\}_{m=2}^{\infty} \overset{d}{=} \left\{Q(\Gamma_m)\right\}_{m=2}^{\infty}$$

where $\Gamma_m = \sum_{i=1}^{m} E_i'$ *is the partial sum of the sequence* $\{E_i'\}_{i=1}^{\infty}$ *which is an i.i.d sequence with exponential distribution and mean* $1/2$*.*

By definition, the Q and U functions are connected by $Q(t) = U(e^t)$. Therefore, the extreme value condition in (9.4) can be rewritten in terms of the Q function as

$$\lim_{t\to\infty} \frac{Q(t+x) - Q(t)}{a(e^t)} = \frac{e^{\gamma x} - 1}{\gamma}, \tag{9.6}$$

where γ is the extreme value index. We discuss in the three separate subsections the cases $\gamma > 0$, $\gamma < 0$ and $\gamma = 0$.

9.3.1 Positive case: $\gamma > 0$

When (9.4) holds with $\gamma > 0$, we have that $\lim_{t\to\infty} U(t) = \infty$ and U is a regularly varying function at infinity, i.e.

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\gamma}.$$

According to Proposition B.1.9 in de Haan and Ferreira (2006), $\log U(t)/\log t \to \gamma$. Hence for any $\delta > 0$, there exists $t(\delta) > 0$ such that for any $t > t(\delta)$,

$$t^{\gamma-\delta} < U(t) < t^{\gamma+\delta}.$$

Correspondingly, for all $t > t_1(\delta) := \log t(\delta)$, we have that

$$e^{(\gamma-\delta)t} < Q(t) < e^{(\gamma+\delta)t}. \tag{9.7}$$

Therefore, for $\delta < \gamma$, we have that

$$\begin{aligned} Q(\Gamma_m) &> e^{(\gamma-\delta)\Gamma_m} 1_{\Gamma_m > t_1(\delta)} + Q(\Gamma_m) 1_{\Gamma_m \le t_1(\delta)} \\ &> e^{(\gamma-\delta)\Gamma_m} (1 - 1_{\Gamma_m \le t_1(\delta)}) \\ &> e^{(\gamma-\delta)\Gamma_m} - e^{(\gamma-\delta)t_1(\delta)}. \end{aligned} \tag{9.8}$$

and

$$\begin{aligned} Q(\Gamma_m) &< e^{(\gamma+\delta)\Gamma_m} 1_{\Gamma_m > t_1(\delta)} + Q(\Gamma_m) 1_{\Gamma_m \le t_1(\delta)} \\ &< e^{(\gamma+\delta)\Gamma_m} + Q(t_1(\delta)). \end{aligned} \tag{9.9}$$

The expectation of $e^{\lambda\Gamma_m}$ for $\lambda \in \mathbb{R}$ is calculated in the following lemma.

Lemma 9.3.2 *For $\lambda \geq 2$ and any fixed integer m, $Ee^{\lambda\Gamma_m} = +\infty$. For $\lambda < 2$ and any fixed integer m*

$$Ee^{\lambda\Gamma_m} = \left(\frac{2}{2-\lambda}\right)^m.$$

When $\gamma > 2$, there exists a $\delta > 0$ such that $\gamma - \delta > 2$. By taking expectations at the two sides of (9.8), we get that

$$EQ(\Gamma_m) \geq Ee^{(\gamma-\delta)\Gamma_m} - e^{(\gamma-\delta)t_1(\delta)} = +\infty.$$

Therefore, we conclude that if $\gamma > 2$, for any finite level m, $EQ(\Gamma_m)$ does not exist. In other words, the current observed price necessarily underestimates the expected payoff. However, $\gamma > 2$ is not very realistic for most items offered on the IA platforms. For most, if not all, items the payoff does have an expectation. In case of a finite expected payoff, i.e. $\gamma < 2$, we have following theorem.

Theorem 9.3.1 *Suppose (9.6) holds for $0 < \gamma < 2$. Then $EQ(\Gamma_m)$ is finite for any fixed m. We have the following two limit relations*

$$\lim_{m\to\infty} \frac{EQ(\Gamma_m)}{Q(\Gamma_m)} = +\infty, \tag{9.10}$$

and

$$\lim_{m\to\infty} \frac{\log EQ(\Gamma_m)}{\log Q(\Gamma_m)} = c, \tag{9.11}$$

where $c := \frac{\log\left(\frac{2}{2-\gamma}\right)}{\gamma/2}$ is a constant larger than 1.

Proof of Theorem 9.3.1

Since Q is a monotone function, we only need to prove that $EQ(\Gamma_m)$ is finite for large m. Choose $\delta < \min(\gamma, 2-\gamma)$. By taking expectations at the two sides of (9.9), we get that

$$EQ(\Gamma_m) \leq Ee^{(\gamma+\delta)\Gamma_m} + Q(t_1(\delta)) < +\infty.$$

Hence $EQ(\Gamma_m)$ is finite. Similarly, we have the lower bound of $EQ(\Gamma_m)$ as

$$EQ(\Gamma_m) \geq Ee^{(\gamma-\delta)\Gamma_m} - e^{(\gamma-\delta)t_1(\delta)} = \left(\frac{2}{2-\gamma+\delta}\right)^m - e^{(\gamma-\delta)t_1(\delta)}.$$

Together with (9.9), we find that

$$\frac{EQ(\Gamma_m)}{Q(\Gamma_m)} \geq \frac{\left(\frac{2}{2-\gamma+\delta}\right)^m - e^{(\gamma-\delta)t_1(\delta)}}{e^{(\gamma+\delta)\Gamma_m} + Q(t_1(\delta))} = \frac{1 - \left(\frac{2}{2-\gamma+\delta}\right)^{-m} e^{(\gamma-\delta)t_1(\delta)}}{e^{(\gamma+\delta)\Gamma_m - m\log\left(\frac{2}{2-\gamma+\delta}\right)} + \left(\frac{2}{2-\gamma+\delta}\right)^{-m} Q(t_1(\delta))}. \tag{9.12}$$

Denote $c(\delta) := 2\frac{\log\left(\frac{2}{2-(\gamma-\delta)}\right)}{\gamma+\delta}$. Since $\frac{2}{2-(\gamma-\delta)} > 1$, (9.12) is continued as

$$\frac{EQ(\Gamma_m)}{Q(\Gamma_m)} \geq \frac{1-o(1)}{e^{(\gamma+\delta)(\Gamma_m - mc(\delta)/2)} + o(1)}. \tag{9.13}$$

Note that as $\delta \to 0$, $c(\delta) \to c$. From the inequality that $\log \frac{1}{1-x} > x$ for all $x < 1$, we get that $c > 1$. Hence, we can choose δ small enough such that $c(\delta) > 1$. From central limit theorem, we have that $\frac{\Gamma_m - m/2}{\sqrt{m}/2}$ is asymptotically standard normally distributed. Thus, for any $c(\delta) > 1$, $\Gamma_m - mc(\delta)/2 \xrightarrow{P} -\infty$ as $m \to \infty$. Therefore, as $m \to \infty$, the right side of (9.13) goes to $+\infty$ which completes the proof of (9.10).

From the boundaries of $EQ(\Gamma_m)$, we have that

$$\left(\frac{2}{2-\gamma+\delta}\right)^m - e^{(\gamma-\delta)t_1(\delta)} \leq EQ(\Gamma_m) \leq \left(\frac{2}{2-\gamma-\delta}\right)^m + Q(t_1(\delta)).$$

Hence, by taking logarithms and asking $m \to \infty$, we get

$$\log\left(\frac{2}{2-\gamma+\delta}\right) \leq \liminf_{m\to\infty} \frac{\log EQ(\Gamma_m)}{m} \leq \limsup_{m\to\infty} \frac{\log EQ(\Gamma_m)}{m} \leq \log\left(\frac{2}{2-\gamma-\delta}\right).$$

By taking $\delta \to 0$, it follows that

$$\lim_{m\to\infty} \frac{\log EQ(\Gamma_m)}{m} = \log\left(\frac{2}{2-\gamma}\right). \tag{9.14}$$

Since $\lim_{t\to\infty} \log Q(t)/t \to \gamma$ and $\Gamma_m \xrightarrow{P} \infty$ as $m \to \infty$, we get that

$$\lim_{m\to\infty} \log Q(\Gamma_m)/\Gamma_m = \gamma.$$

From the Law of Large Numbers, we have that $\Gamma_m/(m/2) \xrightarrow{P} 1$ as $m \to \infty$. Thus

$$\lim_{m\to\infty} \log Q(\Gamma_m)/m = \gamma/2.$$

Together with (9.14), this complete the proof of (9.11). □

Theorem 9.3.1 implies that, for $0 < \gamma < 2$, although the expected payoff is bounded, the observed payoff always underestimates its expectation if there are numerous active bidders. The following remark gives the essential reason for the underestimation.

Remark 9.3.1 *Under the EVT model with $0 < \gamma < 2$, the comparison between $EQ(\Gamma_m)$ and $Q(\Gamma_m)$ is essentially a comparison between $Ee^{\gamma\Gamma_m}$ and $e^{\gamma\Gamma_m}$. From Jensen's inequality, we have that*

$$Ee^{\gamma\Gamma_m} \geq e^{\gamma E\Gamma_m}.$$

From the Law of Large Numbers, $\Gamma_m \sim E\Gamma_m$ as $m \to \infty$. Thus, we intuitively see why $Q(\Gamma_m)$ underestimates $EQ(\Gamma_m)$.

A question whether it is possible to correct the underestimation. Theorem 9.3.1 shows that the logarithm of the expected payoff can be approximated by the logarithm of the observed payoff multiplied with an adjustment factor c that is always higher than 1. Notice that c is a function of the extreme value index γ which can be consistently estimated as a function of the observed record sequence, see Berred (1992). By estimating c, a consistent estimator for the logarithm of the expected payoff can be constructed. However, a consistent estimate at the log-level does not provide a consistent estimator for the expected payoff itself because both of the expected payoff and the observed payoff go to infinity as the number of active bidders m go to infinity. The situation is similar to the $2 \log n$ rule in de Haan *et al.* (2008b). Given the number of potential bidders n, the number of active bidder m is consistently estimated as $2 \log n$. However, the $2 \log n$ rule does not provide a consistent estimator for the number of potential bidders given the number of active bidders. Therefore, within the framework of Theorem 9.3.1, it is not possible to have a consistent estimator of the expected payoff based on the observed record and 2-record sequence.

9.3.2 Negative case: $\gamma < 0$

In case $\gamma < 0$, the distribution function of the bidders' valuations F has a right endpoint, i.e. $Q(\infty) := \lim_{x \to +\infty} Q(x) < \infty$. Hence, the bidders' valuations are never above $Q(\infty)$.[5] In such a case, the expected payoff is always finite. Caserta and de Vries (2005) argued that this is a realistic model for most items sold through IA. For example, the new price is often a realistic upper bound of a second-hand consumer item sold through IA.

Since $\Gamma_m \xrightarrow{P} +\infty$ as $m \to \infty$, we get that $Q(\Gamma_m) \to Q(\infty)$. The following theorem studies the asymptotic difference between $Q(\Gamma_m)$ and $EQ(\Gamma_m)$.

Theorem 9.3.2 *Suppose (9.6) holds for $\gamma < 0$. Then $EQ(\Gamma_m) \to Q(\infty)$ as $m \to \infty$. Thus*

$$\lim_{m \to \infty} \frac{EQ(\Gamma_m)}{Q(\Gamma_m)} = 1.$$

Furthermore, we have the following two limit relations,

$$\lim_{m \to \infty} \frac{Q(\infty) - EQ(\Gamma_m)}{Q(\infty) - Q(\Gamma_m)} = +\infty, \tag{9.15}$$

and

$$\lim_{m \to \infty} \frac{\log(Q(\infty) - EQ(\Gamma_m))}{\log(Q(\infty) - Q(\Gamma_m))} = c, \tag{9.16}$$

where c is defined as in Theorem 9.3.1. Notice that for negative γ we have $c < 1$.

[5] We remark that the uniform distribution, which is a commonly used distribution in auction theory, belongs to this case with $\gamma = -1$.

We start by proving the following useful lemma.

Lemma 9.3.3 *For any fixed constants $T > 0$, $\varepsilon > 0$,*

$$\lim_{m\to\infty} e^{\varepsilon m} P(\Gamma_m \le T) = 0$$

Proof of Lemma 9.3.3

Notice that $2\Gamma_m$ follows a Gamma distribution with shape parameter m, i.e. the density function of $2\Gamma_m$ is $f(x) := \frac{x^{m-1}e^{-x}}{\Gamma(m)} 1_{x>0}$. We have that

$$\begin{aligned}
0 < e^{\varepsilon m} P(\Gamma_m \le T) &= e^{\varepsilon m} \int_0^{2T} \frac{x^{m-1}e^{-x}}{\Gamma(m)} dx \\
&= \int_0^{2T} \frac{(e^{\varepsilon}x)^{m-1}e^{-x}}{\Gamma(m)} d(e^{\varepsilon}x) \\
&= \int_0^{2T} e^{(e^{\varepsilon}-1)x} \frac{(e^{\varepsilon}x)^{m-1}e^{-(e^{\varepsilon}x)}}{\Gamma(m)} d(e^{\varepsilon}x) \\
&\le e^{(e^{\varepsilon}-1)2T} \int_0^{2Te^{\varepsilon}} \frac{x^{m-1}e^{-x}}{\Gamma(m)} dx \\
&= e^{(e^{\varepsilon}-1)2T} P(\Gamma'_m \le 2Te^{\varepsilon})
\end{aligned}$$

where Γ'_m is a Gamma distributed random variable with the same density function f. Because $\Gamma'_m \to +\infty$ as $m \to \infty$, the lemma is proved. □

Proof of Theorem 9.3.2

When $\gamma < 0$, $1/(Q(\infty) - Q(t))$ is a regularly varying function at $+\infty$ with index $-\gamma$. Similar to the inequality (9.7) in the positive case, the following inequality holds. For any $\delta > 0$, there exists $t_2(\delta)$ such that for any $t > t_2(\delta)$

$$e^{(\gamma-\delta)t} < Q(\infty) - Q(t) < e^{(\gamma+\delta)t}. \tag{9.17}$$

Therefore, we have that

$$\begin{aligned}
Q(\infty) - Q(\Gamma_m) &> e^{(\gamma-\delta)\Gamma_m} 1_{\Gamma_m > t_2(\delta)} + (Q(\infty) - Q(\Gamma_m)) 1_{\Gamma_m \le t_1(\delta)} \\
&> e^{(\gamma-\delta)\Gamma_m} (1 - 1_{\Gamma_m \le t_2(\delta)}) \\
&> e^{(\gamma-\delta)\Gamma_m} - 1_{\Gamma_m \le t_2(\delta)}.
\end{aligned} \tag{9.18}$$

and

$$\begin{aligned}
Q(\infty) - Q(\Gamma_m) &< e^{(\gamma+\delta)\Gamma_m} 1_{\Gamma_m > t_2(\delta)} + (Q(\infty) - Q(\Gamma_m)) 1_{\Gamma_m \le t_2(\delta)} \\
&< e^{(\gamma+\delta)\Gamma_m} + Q(\infty) 1_{\Gamma_m \le t_2(\delta)}.
\end{aligned} \tag{9.19}$$

By taking expectations at the two sides of (9.19), we get that

$$0 < Q(\infty) - EQ(\Gamma_m) \le \left(\frac{2}{2-\gamma-\delta}\right)^m + Q(\infty) P(\Gamma_m \le t_2(\delta)).$$

By taking $\delta < -\gamma$, we get that $\frac{2}{2-\gamma-\delta} < 1$. Since $P(\Gamma_m \leq t_2(\delta)) \to 0$ as $m \to \infty$, we get that $\lim_{m\to\infty} EQ(\Gamma_m) = Q(\infty)$.

We turn to compare $Q(\infty) - EQ(\Gamma_m)$ with $Q(\infty) - Q(\Gamma_m)$. By taking expectations on the two sides of (9.18), we get that

$$Q(\infty) - EQ(\Gamma_m) \geq \left(\frac{2}{2-\gamma+\delta}\right)^m - P(\Gamma_m \leq t_2(\delta)).$$

Notice that $\Gamma_m \xrightarrow{P} +\infty$ as $m \to \infty$. The inequality (9.19) implies that eventually

$$Q(\infty) - Q(\Gamma_m) < e^{(\gamma+\delta)\Gamma_m}.$$

Hence

$$\frac{Q(\infty) - EQ(\Gamma_m)}{Q(\infty) - Q(\Gamma_m)} \geq \frac{\left(\frac{2}{2-\gamma+\delta}\right)^m - P(\Gamma_m \leq t_2(\delta))}{e^{(\gamma+\delta)\Gamma_m}}. \tag{9.20}$$

Lemma 9.3.3 shows that $P(\Gamma_m \leq t_2(\delta))$ goes to 0 at a higher speed than any exponential speed. Thus, (9.20) is continued as

$$\frac{Q(\infty) - EQ(\Gamma_m)}{Q(\infty) - Q(\Gamma_m)} \geq \frac{1 - o(1)}{e^{(\gamma+\delta)(\Gamma_m - d(\delta)m/2)}}, \tag{9.21}$$

where $d = 2\frac{\log\left(\frac{2}{2-\gamma+\delta}\right)}{\gamma+\delta}$. Notice that as $\delta \to 0$, $d(\delta) \to c$ and $c < 1$ holds for $\gamma < 0$. Hence, we can choose δ small enough such that $d(\delta) < 1$. Then, $\Gamma_m - d(\delta)m/2 \xrightarrow{P} +\infty$ as $m \to \infty$. Therefore, as $m \to \infty$, the right side of (9.21) goes to $+\infty$ which completes the proof of the theorem. The proof of (9.16) is similar to that of (9.11). □

Theorem 9.3.2 studies the case $\gamma < 0$ and tells a story just opposite to the positive case. For $\gamma < 0$, the observed payoff might be considered as a consistent estimator of its expectation because both the observed and expected payoff converge to the right endpoint of the bidder's valuation. Nevertheless, the distance between the observed payoff and the right endpoint is eventually smaller than the distance between the expected payoff and the right endpoint. Hence, if there are sufficiently many active bidders, the observed payoff always overestimates its expectation. A consistent estimator on the log-level of the difference is given by multiplication with an adjustment factor c that is always lower than 1. However, similar to the positive case, a consistent estimate at the log-level would not provide a consistent estimator of the difference itself. Therefore, it is not possible to correct the overestimation.

To sum up, Theorem 9.3.1 and Theorem 9.3.2 show that if the extreme value index of bidder's valuation distribution is not 0, the observed payoff is never a satisfactory estimator for the expected payoff.

9.3.3 Zero case: $\gamma = 0$

In the previous two Subsections, we found that the observed payoff underestimates or overestimates its expectation when γ is positive or negative respectively. The remaining case is $\gamma = 0$, i.e. when F belongs to the Gumbel domain. In this section, we show that at least for a subclass of the Gumbel domain, the observed payoff is a reasonable estimator for its expectation.

Model specification

As we discussed before, in order to avoid the problem of the unknown number of potential bidders, the model should be based on the Q function and the payoff should be taken as the last observation in the record and 2-record sequence.

We introduce a refinement and assume that the Q function itself is regularly varying or generalized regularly varying. Furthermore, in order to study the asymptotic properties, we assume that a second-order condition holds.

We start from the regularly varying model. Suppose Q function itself is regularly varying with index $\lambda > 0$. We also assume that it is second-order regularly varying with second-order index $\rho \le 0$, i.e.

$$\frac{\frac{Q(tx)}{Q(t)} - x^\lambda}{A(t)} \to H(x) := x^\lambda \frac{x^\rho - 1}{\rho} \tag{9.22}$$

as $t \to \infty$, for some suitable function $A(t) \in RV_\rho$ and all $x > 0$. We call this *Regularly Varying Q-function (RVQ)* model.

Remark 9.3.2 *The RVQ model with $\rho < -1$ is a subclass of the Gumbel domain.*

Proof of Remark 9.3.2

From (9.22), we have the following inequality (See de Haan and Ferreira (2006, Appendix B)). Given any $\varepsilon > 0$, there is a $t_0(\varepsilon)$ such that for all $tx > t_0(\varepsilon)$,

$$\left| \frac{\frac{Q(tx)}{Q(t)} - x^\lambda}{A(t)} - H(x) \right| \le \varepsilon x^{\lambda+\rho+\varepsilon}. \tag{9.23}$$

Therefore,

$$\left| \frac{\frac{Q(t+x)}{Q(t)} - (\frac{t+x}{t})^\lambda}{A(t)} - H(\frac{t+x}{t}) \right| \le \varepsilon (\frac{t+x}{t})^{\lambda+\rho+\varepsilon},$$

for all $t > t_0(\varepsilon)$ and positive x. When $\rho < -1$ and $A(t) \in RV_\rho$, we get $tA(t) \to 0$ as $t \to \infty$. It leads to the fact that

$$t\left\{ \frac{Q(t+x) - Q(t)}{Q(t)} - ((1 + x/t)^\lambda - 1) \right\} \to 0.$$

Since

$$t((1+x/t)^{\lambda}-1)\to\lambda x,$$

we finally get that

$$\frac{Q(t+x)-Q(t)}{\lambda Q(t)/\sqrt{t}}\to x,$$

i.e. the corresponding U function satisfies condition (9.4) with $\gamma=0$. □

Thus, the RVQ model is only a special case in the Gumbel domain ($\gamma=0$), and therefore is more narrow. Fortunately, quite some well-known parametric distributions belong to this model. For example, both the normal distribution and the exponential distribution satisfy this model. (For the normal distribution, $\lambda=1/2$, $\rho=-\infty$. For the exponential distribution, $\lambda=1$, $\rho=-\infty$.)

In the RVQ model, the original distribution function can not have a finite right endpoint. In order to include distributions with finite right endpoint, we extend the model as follows. Suppose the Q function is second-order generalized regularly varying with first-order index $\lambda\in\mathbb{R}$ and second-order index $\rho\leq 0$, i.e.

$$\frac{\frac{Q(tx)-Q(t)}{a(t)}-\frac{x^{\lambda}-1}{\lambda}}{A(t)}\to H(x) \tag{9.24}$$

as $t\to\infty$, for all $x>0$ and some suitable function $a(t)\in RV_{\lambda}$, $A(t)\in RV_{\rho}$ and $H(x)$. We call this *Generalized Regularly Varying Q-function (GRVQ)* model.

In the GRVQ model, if $\lambda>0$, it can be simplified to the RVQ model. When λ is negative, $Q(\infty)<\infty$. Hence, in this case, the original distribution function F must have a finite right endpoint.[6] We note though that a random variable K with distribution function

$$1-e^{-((a-x)/b)^{1/\lambda}}$$

for all $x\leq a$, where $b>0,\lambda<0$ satisfies the requirements of the GRVQ model (λ is the regularly varying index in (9.24) and $\rho=-\infty$). Notice that $-K$ follows the inverse Weibull distribution.

Statistical inference

We turn to study the expected payoff. The following lemma gives the asymptotic properties of the observed and expected payoff.

[6] The commonly used distribution functions for the valuations in auction theory are the uniform distributions and other distributions with a finite right endpoint. Notice that the uniform distribution belongs to the case $\gamma=-1$ which has been discussed in the previous section.

Lemma 9.3.4 *Suppose the RVQ model holds with $\rho < -1/2$, then*

$$\sqrt{m}\left(\frac{X_{J(m)}}{Q(m/2)} - 1\right) \xrightarrow{d} \lambda W; \tag{9.25}$$

$$\sqrt{m}\left(\frac{EX_{J(m)}}{Q(m/2)} - \frac{\Gamma(\lambda+m)}{n^{\lambda}\Gamma(m)}\right) \rightarrow 0. \tag{9.26}$$

where W is a standard normal distributed random variable and Γ is the Gamma function.

Proof of Lemma 9.3.4

Applying (9.23), we get that, eventually,

$$\left|\frac{\frac{Q(\Gamma_m)}{Q(m/2)} - (\frac{\Gamma_m}{m/2})^{\lambda}}{A(m/2)} - H(\frac{\Gamma_m}{m/2})\right| \leq \varepsilon(\frac{\Gamma_m}{m/2})^{\lambda+\rho\pm\varepsilon}. \tag{9.27}$$

The symbol $\pm$ means taking the suitable sign according to whether $\frac{\Gamma_m}{m/2}$ is higher or lower than 1. Since $\sqrt{m}A(m/2) \rightarrow 0$, $\frac{\Gamma_m}{m/2} \xrightarrow{P} 1$ and $\lim_{x\rightarrow 1} H(x) \pm \varepsilon x^{\lambda+\rho\pm\varepsilon}$ exists, we find that

$$\sqrt{m}\left(\frac{Q(\Gamma_m)}{Q(m/2)} - (\frac{\Gamma_m}{m/2})^{\lambda}\right) \xrightarrow{P} 0$$

From Central Limit Theory, we have that

$$\sqrt{m}\left(\frac{\Gamma_m}{m/2} - 1\right) \xrightarrow{d} W,$$

where W is a standard normal distributed random variable. According to Cramèr's delta method,

$$\sqrt{m}\left((\frac{\Gamma_m}{m/2})^{\lambda} - 1\right) \xrightarrow{d} \lambda W.$$

Thus, (9.25) is a direct consequence.

Taking expectation of the two sides of inequality (9.27) on the set $\{\Gamma_m > t_0(\varepsilon)\}$, because the absolute value function is a convex function, we get that

$$\left|\frac{\frac{EQ(\Gamma_m)1_{\{\Gamma_m > t_0(\varepsilon)\}}}{Q(m/2)} - E(\frac{\Gamma_m}{m/2})^{\lambda}1_{\{\Gamma_m > t_0(\varepsilon)\}}}{A(m/2)} - EH(\frac{\Gamma_m}{m/2})1_{\{\Gamma_m > t_0(\varepsilon)\}}\right|$$
$$\leq \varepsilon E(\frac{\Gamma_m}{m/2})^{\lambda+\rho\pm\varepsilon}1_{\{\Gamma_m > t_0(\varepsilon)\}}.$$

Similar to above discussion, we can conclude that

$$\sqrt{m}\left(\frac{EQ(\Gamma_m)1_{\{\Gamma_m > t_0(\varepsilon)\}}}{Q(m/2)} - E(\frac{\Gamma_m}{m/2})^{\lambda}1_{\{\Gamma_m > t_0(\varepsilon)\}}\right) \rightarrow 0. \tag{9.28}$$

Since $\sqrt{m}P(\Gamma_m \leq x) \rightarrow 0$ as $m \rightarrow \infty$ for all $x > 0$, we have that

$$\sqrt{m}\left(\frac{EQ(\Gamma_m)1_{\{\Gamma_m \leq t_0(\varepsilon)\}}}{Q(m/2)}\right) \leq \sqrt{m}\left(\frac{Q(t_0(\varepsilon))P(\{\Gamma_m \leq t_0(\varepsilon)\})}{Q(m/2)}\right) \rightarrow 0,$$

and

$$\sqrt{m}E(\frac{\Gamma_m}{m/2})^\lambda 1_{\{\Gamma_m \le t_0(\varepsilon)\}} \le \sqrt{m}E(\frac{t_0(\varepsilon)}{m/2})^\lambda P(\{\Gamma_m \le t_0(\varepsilon))\} \to 0.$$

It follows that (9.28) can be rewritten as

$$\sqrt{m}\left(\frac{EQ(\Gamma_m)}{Q(m/2)} - E(\frac{\Gamma_m}{m/2})^\lambda\right) \to 0 \tag{9.29}$$

Since $2\Gamma_m$ follows the Gamma distribution with shape parameter m, the expectation of $(2\Gamma_m)^\lambda$ is $\frac{\Gamma(\lambda+m)}{\Gamma(\lambda)}$ (see e.g., Papoulis (1984), pp. 103-104). Then

$$E(\frac{\Gamma_m}{m/2})^\lambda = \frac{\Gamma(\lambda+m)}{n^\lambda \Gamma(m)}.$$

Thus, (9.26) is proved. □

Now, we prove that

$$\lim_{m\to\infty} \sqrt{m}\left(\frac{\Gamma(\lambda+m)}{n^\lambda\Gamma(m)} - 1\right) = 0. \tag{9.30}$$

Since $\Gamma(x) = e^{-x}x^{x-1/2}\sqrt{2\pi}(1 + 1/12xo(1)))$ as $x \to \infty$, we have that

$$\begin{aligned}
\frac{\Gamma(\lambda+m)}{n^\lambda\Gamma(m)} &= e^{-\lambda}(1+\lambda/m)^{m-1/2}(1+\lambda/m)^\lambda \frac{1+\frac{\lambda+m}{12}o(1)}{1+\frac{m}{12}o(1)} \\
&= e^{(m-1/2)\log(1+\lambda/m)-\lambda}(1+\lambda^2/m+o(1/m))(1+o(1/m)) \\
&= e^{-\frac{\lambda-\lambda^2}{2m}+o(1/m)}(1+\lambda^2/m+o(1/m))(1+o(1/m)) \\
&= 1 + O(1/m).
\end{aligned}$$

Thus (9.30) is proved.

We can estimate the expected payoff as follows. Suppose we have m active bidders in an IA. We observe the bidders' valuations $X_{J(1)}, \cdots, X_{J(m-1)}$ as their final bids, except for the winner. The payoff will be $X_{J(m-1)}$. We estimate the expected payoff $EX_{J(m-1)}$ by this observation. The asymptotic property of this estimator is given by the following theorem.

Theorem 9.3.3 *Suppose the RVQ model holds for $\rho < -\frac{1}{2}$. Then*

$$\sqrt{m-1}\left(\frac{X_{J(m-1)}}{EX_{J(m-1)}} - 1\right) \xrightarrow{d} \lambda W \tag{9.31}$$

as $m \to \infty$, where W is a standard normally distributed random variable.

Proof of Theorem 9.3.3

The theorem is proved by combining (9.25), (9.26) and (9.30). □

Theorem 9.3.3 shows that for the RVQ model the observed payoff is an accurate estimator for the expected payoff and the corresponding asymptotic normality holds under a second order condition.

Starting from the GRVQ model, similar results can be obtained as in the previous subsection. Here we only present the conclusion, the proof is omitted. The proof for the GRVQ model is essentially the same as the proof for the RVQ model.

Theorem 9.3.4 *Suppose the GRVQ model holds with $\rho < -\frac{1}{2}$. Then as $m \to \infty$,*

$$\sqrt{m-1}\frac{X_{J(m-1)} - EX_{J(m-1)}}{a((m-1)/2)} \xrightarrow{d} W,$$

where W is a standard normally distributed random variable. In particular, when $\lambda > 0$, we have that (9.31) holds. When $\lambda < 0$, we have that $Q(\infty) < \infty$, and as $m \to \infty$,

$$\sqrt{m-1}\left(\frac{Q(\infty) - X_{J(m-1)}}{Q(\infty) - EX_{J(m-1)}} - 1\right) \xrightarrow{d} \lambda W.$$

9.4 Conclusion

Internet auctions are a hybrid of the standard second-price Vickrey auction and the first-price English auction, for which the expected payoff is equal to the expectation of the second highest valuation among all the potential bidders. The expected payoff acts as a benchmark of the reasonableness of the price that is paid for the purchased item. Since the number of potential bidders is not observable, the expected value is difficult to estimate accurately. We approached this problem by considering the bids as a record and 2-record sequence of the potential bidder's valuation. The observed payoff is thus one of the records and 2-records.

In this chapter, we use the EVT models to model the tail distribution of the bidder's valuation and study the expected payoff. We first argue that assuming that the extreme value index γ is higher than 2 is not a realistic model because in that case the expected payoff is unbounded. For $0 < \gamma < 2$, we show that the observed payoff underestimates the expected payoff. At the log-level, an adjusted estimator exists for the expected payoff based on the logarithm of the observed payoff. We show that this is not possible at the level of the expected payoff. Hence, the consistency is an issue. One may argue that $\gamma > 0$ is not a realistic setup for the distribution of the bidder's valuation, because such a distribution function has no finite right endpoint which does not reflect the reality of IA.

For $\gamma < 0$ the distribution function of the bidder's valuation has a finite endpoint, which is a more realistic setup for IA. Both the expected and observed payoff converge to the right endpoint as the number of active bidders m goes to infinity. However, the distances to the endpoint converge to 0 at different speeds. The distance between the observed payoff and the right endpoint goes to 0 faster. Therefore, the observed payoff always overestimates the expected payoff. In this case, though the observed final price consistently estimates the expected payoff, the overestimation cannot be redressed consistently.

For $\gamma = 0$, i.e. the distribution function of the bidder's valuation belongs to the Gumbel domain, the observed payoff can be a consistent estimator for the expected revenue. We introduced a subclass of the Gumbel domain as the model of bidder's valuation distribution. Within this subclass and under a second order condition, the observed payoff consistently converges to the expected payoff and the corresponding asymptotic normality holds.

All in all, in an IA the observed payoff is the final price of the deal, while the expected payoff is what the seller should get from holding such an IA. Our study shows that by assuming that the tail of the bidder's valuation distribution belongs to the domain of attraction of an extreme value distribution, the final price does not always reflect what the seller deserves.

Even though our message is not so positive, there remain some interesting issues. For example, our analysis focused on general EVT setup on the distribution of bidder's valuation. In case one is willing to make an explicit parametric assumption regarding the distribution of valuations, consistent estimation of the expected payoff may be possible outside the limited class that we could handle. Further research on estimating the extreme value index from the observed record and 2-record sequence is also of interest and may help to identify the situation for a specific IA.

Summary

Rare events such as natural disasters or financial crises often have a large impact on our lives. To model and analyze such events is important, particularly for risk control. However, it is not an easy job due to the scarce observations on such kind of events. In this thesis, we study Extreme Value Statistics and show that it is a proper instrument for modeling rare events.

A major issue in one-dimensional EVT is to estimate the extreme value index. In the setup of EVT, the tail of a distribution function is approximated by an explicit parametric model. This creates the possibility to apply the maximum likelihood procedure to estimate the parameters. However, since the model is an approximation rather than a real parametric approach, the regular theory on maximum likelihood does not apply. Therefore, it is necessary to develop the theory of the maximum likelihood estimator for the extreme value index from the original setup: the extreme value condition. Part I deals with this problem in a thorough way and helps to complete the literature on this aspect.

In finance, one may consider to construct a diversified portfolio in order to diversify away the individual risks. However, this is not always the case when considering large losses as risks. In particular, the dependence among the large losses across different securities might be complicated. Part II applies multivariate EVT to study this problem and provide an applicable portfolio selection procedure.

Besides evaluating risks in financial investment, it is also important to control risks of extremal weather situation. As an example, Part III evaluates the level of "once-in-100-year" total rainfall in the province North Holland (The Netherlands). The major difficulty of this problem is that one should consider the dependence of the rainfall amount across the concerning area while observations are only on a few fixed monitoring stations. Notice that the dependence of extremal rainfall could be quite different from the moderate level. This has to be done via a infinite-dimensional EVT approach.

After demonstrating different applications of EVT in different fields, an interesting question arises: does the EVT instrument always work in modeling tails? Part IV provides a somewhat different example. We study an extreme-value-type problem in Internet auctions. In an Internet auction, only few bidders on auction websites are observed,

but there may be a multitude of potentially interested bidders. We first explain the discrepancy by using record theory. Secondly, when studying the expected revenue of an Internet auction, since the final payoff is one of those high bids, it seems to be again a ”tail problem” which might be dealt with EVT. However, it turns out that the EVT setup does not lead to a reasonable estimator for the expected payoff in most of the cases.

All in all, besides contributing to the theoretical literature of Extreme Value Statistics, the thesis is devoted to show different applications of Extreme Value Statistics in different fields. Those applications exhibit the strong potential of Extreme Value Statistics in modeling and analyzing rare events.

Nederlandse samenvatting (Summary in Dutch)

Zeldzame gebeurtenissen zoals natuurrampen hebben gewoonlijk grote gevolgen voor ons leven. Het is belangrijk om zulke gebeurtenissen te modelleren en te analyseren, in het bijzonder met het oog op risicobeheersing. Dit is echter geen eenvoudig werk vanwege het geringe aantal waarnemingen van zulke gebeurtenissen in het verleden. In dit proefschrift bestuderen we extreme waarden statistiek en we tonen aan dat dit het geigende gereedschap is om zeldzame gebeurtenissen te modelleren.

Een belangrijk probleem in eendimensionale EVT (Extreme Value Theory) is hoe de extreme waarden index te schatten. De EVT theorie zegt ons dat de start van kansverdeling (het deel dat gelieerd is aan extreme waarden) goed benaderd kan worden door een expliciet parametrisch model. Dit opent de mogelijkheid om de methode van grootste aannemelijkheid (maximum likelihood) te gebruiken om de parameter te schatten, dat wil zeggen, de extreme waarden index. Jammer genoeg kunnen de bekende resultaten over de grootste aannemelijkheidschatter niet direct gebruikt worden omdat het model een benadering is, niet de werkelijkheid. Daarom is het belangrijk een theorie van grootste aannemelijkheid te ontwikkelen voor deze specifieke situatie uitgaande van extreme waarden voorwaarde waaraan de oorspronkelijke verdeling voldoet. Deel 1 van dit proefschrift behandelt dit probleem op een grondige manier; bekende resultaten vinden hier een noodzakelijke aanvulling.

In de financiering gebruikt men een gediversifieerde portfolio om de individuele risico's te neutraliseren. Dit heeft echter niet altijd het beoogde effect als men grote verliezen bij risico's beschouwt. In deel 2 van dit proefschrift wordt meerdimensionale EVT theorie toegepast om dit probleem te bestuderen en om een toepasbare procedure voor portfolioselectie te ontwikkelen.

We bestuderen niet alleen risico's bij financiële investeringen. Het is ook belangrijk risico's bij extreme weersomstandigheden in de hand te houden. In deel 3 van dit proefschrift wordt als voorbeeld berekend welke niveau van regenval in Noord-Holland (dat wil zeggen het niveau van de totale regenval in die provincie) eens per eeuw voorkomt.

Dus we berekenen hoe zwam de meest ernstige regenval is die we eens per eeuw kunnen verwachten. Het grootste probleem hierbij is dat we rekening moeten houden met de afhankelijkheid in hoeveelheid regenval tussen alle plaatsen in het gebied, terwijl er alleen maar waarnemingen gedaan zijn op een beperkt aantal in het gebied. Merk hierbij op dat de afhankelijkheid bij extreme regenval nogal verschillend kan zijn van de afhankelijkheid die heerst bij matige regenval. De aanpak gebruikt oneindig-dimensionale EVT.

Na op deze manier verschillende toepassingen bestudeerd te hebben, komt een interessante vraag op: werkt het EVT gereedschap altijd voor het modelleren van de staart van de verdeling? Deel 4 van dit proefschrift geeft een enigszins verschillend voorbeeld. We nemen een EVT-achtig probleem onder de loep bij veilingen op internet. In een dergelijke veiling zijn er meestal maar erg weinig bieders. We hebben aangetoond dat het aantal potentiële bieders die belangstelling hebben en de website bezoeken zeer veel hoger moet zijn dan het aantal werkelijke bieders. Vervolgens is gekeken naar de verwachte prijs in een internet veiling. Omdat deze in feite een van de hoge biedingen is, lijkt het dat we hier weer een EVT probleem hebben. Toch blijkt het dat een rechtstreekse toepassing van EVT geen redelijke schatter geeft. Een enigszins andere aanpak is dus noodzakelijk.

Dit proefschrift geeft dus zowel een bijdrage aan de theoretische literatuur over EVT als een selectie van toepassingen in verschillende gebieden. Deze toepassingen laten zien dat EVT methoden van grote waarde zijn voor het modelleren en analyseren van zeldzame gebeurtenissen.

Bibliography

Alila, Y. (1999), A hierarchical approach for the regionalization of precipitation annual maxima in Canada, *J. Geophys. Res.*, **104(D24)**, 31,645–31,655.

Allen, R. and A. DeGaetano (2005), Considerations for the use of radar-derived precipitation estimates in determining return intervals for extreme areal precipitation amounts, *J. Hydrol.*, **315**, 203–219.

Arzac, E. and V. Bawa (1977), Portfolio choice and equilibrium in capital markets with safety-first investors, *J. Finan. Econ.*, **4**, 277–288.

Athey, S. and P. A. Haile (2002), Identification of standard auction models, *Econometrica*, **70**, 2107–2140.

Ausubel, L. M. and P. R. Milgrom (2002), Ascending Auctions with Package Bidding, *Front. Theor. Econ.*, **1**, 1019–1019.

Bacchi, B. and R. Ranzi (1996), On the derivation of the areal reduction factor of storms, *Atmos. Res.*, **42**, 123–135.

Bajari, P. and A. Hortaçsu (2003), Winner's curse, reserve Prices and endogenous entry: empirical insights from eBay auction, *RAND J. Econ.*, **34**, 329–355.

Balkema, A. and L. de Haan (1974), Residual life time at great age, *Ann. Probab.*, **2**, 792–804.

Basrak, B., R. A. Davis, and T. Mikosch (2002), A characterization of multivariate regular variation, *Ann. Appl. Probab.*, **12**, 908–920.

Beirlant, J., P. Vynckier, and J. Teugels (1996), Tail index estimation, Pareto quantile plots, and regression diagnostics, *J. Amer. Statist. Assoc.*, **91**, 1659–1667.

Bell, F. (1976), The areal reduction factor in rainfall frequency estimation, Tech. Rep. 35, Cent. for Ecol. and Hydrol., Wallingford, U.K.

Berred, M. (1992), On record values and the exponent of a distribution with regularly varying upper tail, *J. Appl. Probab.*, **29**, 575–586.

Boyd, S. and L. Vandenberghe (2004), *Convex Optimization*, Cambridge University Press.

Buishand, T. (1983), Extremely high rainfall amounts and the theory of extreme values (in Dutch), *Cultuurtechnisch Tijdschrift*, **23**, 9–20.

Buishand, T. (1991), Extreme rainfall estimation by combining data from several sites, *Hydrol. Sci. J.*, **36**, 345–365.

Buishand, T., L. de Haan, and C. Zhou (2008), On spatial extremes: with application to a rainfall problem, *Ann. Appl. Statist.*, **2**, 624–642.

Bulow, J. I. and P. D. Klemperer (1996), Auctions vs. negotiations, *Amer. Econ. Rev.*, **86**, 180–194.

Caserta, S. and C. G. de Vries (2005), Auctions with numerous bidders, *Tinbergen Institute Discussion Paper*, **05-031/2**.

Coles, S. (1993), Regional modelling of extreme storms via max-stable processes, *J. R. Statist. Soc. B*, **55**, 797–816.

Coles, S. and J. Tawn (1994), Statistical methods for multivariate extremes: an application to structural design (with discussion), *Appl. Statist.*, **43**, 1–48.

Coles, S. and J. Tawn (1996), Modelling Extremes of the Areal Rainfall Process, *J. R. Statist. Soc. B*, **58**, 329–347.

Cooley, D., D. Nychka and P. Naveau (2007), Bayesian spatial modeling of extreme precipitation return levels, *J. Amer. Statist. Assoc.*, **102**, 824–840.

Danielsson, J., L. de Haan, L. Peng, and C. G. de Vries (2000), Using a bootstrap method to choose the sample fraction in tail index esitmation, *J. Multivariate Anal.*, **76**, 226–248.

de Haan, L. (1984a), *Slow variation and characterization of domains of attraction.* In Statistical Extremes and Application (J. Tiago de Oliveira, ed), Reidel Publishing, pp. 31–38.

de Haan, L. (1984b), A spectral representation for max-stable processes, *Ann. Probab.*, **12**, 1194–1204.

de Haan, L. and J. de Ronde (1998), Sea and wind: multivariate extremes at work, *Extremes*, **1**, 7–45.

de Haan, L., C. G. de Vries, and C. Zhou (2008a), The expected payoff to Internet auctions, *Unpublished manuscript*.

de Haan, L., C. G. de Vries, and C. Zhou (2008b), The extent of Internet auction markets, *Tinbergen Institute Discussion Paper*, **2008-041/2**.

de Haan, L. and A. Ferreira (2006), *Extreme Value Theory: An Introduction*, Springer.

de Haan, L. and T. Lin (2001), On convergence towards an extreme value distribution in C[0,1], *Ann. Prob.*, **29**, 467–483.

de Haan, L. and L. Peng (1998), Comparison of tail index estimators, *Stat. Neerl.*, **52**, 60–70.

de Haan, L. and T. Pereira (2006), Spatial extremes: models for the stationary case, *Ann. Statist.*, **34**, 146–168.

de Haan, L. and U. Stadtmüller (1996), Generalized regular variation of second order, *J. Austral. Math. Soc. Ser. A*, **61**, 381–395.

de Haan, L. and C. Zhou (2008), On extreme value analysis of a spatial process, *Revstat*, **6**, 71–81.

Deheuvels, P., E. Haeusler, and D. Mason (1988), Almost sure convergence of the Hill estimator, *Math. Proc. Camb. Philos. Soc.*, **104**, 371–381.

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

Dennis, J. E. and R. B. Schnabel (1996), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.*, Society for Industrial and Applied Mathematics.

Drees, H. (1998), On smooth statistical tail functionals, *Scand. J. Statist.*, **25**, 187–210.

Drees, H., A. Ferreira, and L. de Haan (2004), On maximum likelihood estimation of the extreme value index, *Ann. Appl. Probab.*, **14**, 1179–1201.

Drees, H. and X. Huang (1998), Best attainable rates of convergence for estimators of the stable tail dependence function, *J. Multivariate Anal.*, **64**, 25–47.

Einmahl, J. and D. Mason (1988), Strong limit theorems for weighted quantile processes, *Ann. Probab.*, **16**, 1623–1643.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997), *Modelling extremal events: for insurance and finance*, Springer.

Falk, M. (1995), Some best parameeter estimates for distributions with finite endpoint, *Statistics*, **27**, 115–125.

Fama, E. and M. Miller (1972), *The Theory of Finance*, Holt, Rinehart and Winston, New York.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, vol. 2, John Wiley & Sons, New York.

Fowler, H. and C. Kilsby (2003), A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000, *Int. J. Climatol.*, **23**, 1313–1334.

Gellens, D. (2002), Combining regional approach and data extension procedure for assessing GEV distribution of extreme precipitation in Belgium, *J. Hydrol.*, **268**, 113–126.

Geluk, J. and L. de Haan (1987), *Regular variation, extensions and Tauberian theorems*, CWI Tract 40, Center for Mathematics and Computer Science, Amsterdam, Netherlands.

Gnedenko, B. (1943), Sur la distribution limite du terme maximum d'une série aléatoire, *Ann. Math.*, **44**, 423–453.

Gouriéroux, C., J. Laurent, and O. Scaillet (2000), Sensitivity analysis of Values at Risk, *J. Empirical Finance*, **7**, 225–245.

Grimshaw, S. (1993), Computing maximum likelihood estimates for the generalized Pareto distribution, *Technometrics*, **35**, 185–191.

Gupta, A., G. González-Farías, and J. Domínguez-Molina (2004), A multivariate skew normal distribution, *J. Multivariate Anal.*, **89**, 181–190.

Hartmann, P., S. Straetmans, and C. G. de Vries (2004), Asset market linkages in crisis periods, *Rev. Econ. Statist.*, **86**, 313–326.

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

Hosking, J. and J. Wallis (1987), Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, **29**, 339–349.

Huang, X. (1992), *Statistics of bivariate extreme values*, Ph.D. thesis, Tinbergen Institute.

Hyung, N. and C. G. de Vries (2002), Portfolio diversification effects and regular variation in financial data, *Allgemeines Statistisches Arch.*, **86**, 69–82.

Hyung, N. and C. G. de Vries (2005), Portfolio diversification effects of downside risk, *J. Financial Econometrics*, **3**, 107–125.

Hyung, N. and C. G. de Vries (2007), Portfolio selection with heavy tails, *J. Empirical Finance*, **14**, 383–400.

Jansen, D. (2001), Limited downside risk in portfolio selection among U.S. and Pacific basin equities, *Int. Econ. J.*, **15**, 1–22.

Jansen, D. and C. G. de Vries (1991), On the frequency of large stock returns: putting booms and busts into perspective, *Rev. Econ. Statist.*, **73**, 18–24.

Jansen, D., K. Koedijk, and C. G. de Vries (2000), Portfolio selection with limited downside risk, *J. Empirical Finance*, **7**, 247–269.

Klein, R. W. and S. D. Roberts (1984), A time-varying Poisson arrival process generator, *Simulation*, **43**, 193–195.

Krishna, V. (2002), *Auction Theory*, Academic Press.

Laffont, J., H. Ossard, and Q. Vuong (1995), Econometrics of first price auctions, *Econometrica*, **63**, 953–980.

Leander, R. and T. Buishand (2007), Resampling of regional climate model output for the simulation of extreme river flows, *J. Hydrol.*, **332**, 487–496.

Lewis, P. A. W. and G. S. Shedler (1976), Simulation of nonhomogeneous Poisson processes with log linear rate function, *Biometrika*, **63**, 501–505.

Markowitz, H. (1952), Portfolio selection, *J. Finance*, **7**, 77–91.

McAfee, R. P. and J. McMillan (1987), Auctions with a stochastic number of bidders, *J. Econ. Theory*, **43**, 1–19.

NERC (1975), Flood Studies Report, Vol. II Meteorological Studies, *Cent. for Ecol. and Hydrol.*, Wallingford , U.K.

Ockenfels, A. and A. E. Roth (2006), Late and multiple bidding in second price Internet auctions: theory and evidence concerning different rules for ending an auction, *Games Econ. Behav.*, **55**, 297–320.

Øksendal, B. (1992), *Stochastic differential equations*, third edn., Springer.

Paarsch, H. J. (1992), Deciding between the common and private value paradigms in empirical models of auctions, *J. Econometrics*, **51**, 191–215.

Papoulis, A. (1984), *Probability, Random Variables, and Stochastic Processes, 2nd ed.*, New York: McGraw-Hill.

Park, Y. H. and E. T. Bradlow (2005), An integrated model for bidding behavior in Internet auctions: whether, who, when, and how much, *J. Marketing Res.*, **42**, 470–482.

Pickands III, J. (1975), Statistical inference using extreme order statistics, *Ann. Statist.*, **3**, 119–131.

Poon, S., M. Rockinger, and J. Tawn (2004), Extreme value dependence in financial markets: diagnostics, models, and financial implications, *Rev. Finan. Stud.*, **17**, 581–610.

Resnick, S. and R. Roy (1990), Multivariate extremal processes, leader processes and dynamic choice models, *Adv. Appl. Probab.*, **22**, 309–331.

Resnick, S. I. (1987), *Extreme Values, Regular Variation, and Point Processes*, Springer-Verlag.

Roy, A. (1952), Safety first and the holding of assets, *Econometrica*, **20**, 431–449.

Schlather, M. (2002), Models for stationary max-stable random fields, *Extremes*, **5**, 33–44.

Shorack, G. and J. Wellner (1986), *Empirical processes with applications to statistics*, John Wiley & Sons, Inc.

Sivapalan, M. and G. Blöschl (1998), Transformation of point to areal rainfall: Intensity-duration-frequency curves, *J. Hydrol.*, **204**, 150–167.

Smith, R. (1987), Estimating tails of probability distributions, *Ann. Statist.*, **15**, 1174–1207.

Smith, R. (1990), Max-stable processes and spatial extremes, *Unpublished manuscript*.

Song, U. (2004), Nonparametric estimation of an eBay auction model with an unknown number of bidders, *Working Paper, see http://faculty.arts.ubc.ca/usong/eBay.pdf*.

Stewart, E. (1989), Areal reduction factors for design storm construction: Joint use of raingauge and radar data, in *New Directions for Surface Water Modeling (Proceedings of the Baltimore Symposium, May 1989)*, IAHS Publ. no. 181, International Association of Hydrological Sciences (IAHS).

Susmel, R. (2001), Extreme observations and diversification in Latin American emerging equity markets, *J. Int. Money Finance*, **20**, 971–986.

Veneziano, D. and A. Langousis (2005), The areal reduction factor: A multifractal analysis, *Water Resour. Res.*, **41**, W07008.

Vickrey, W. (1962), Auctions and bidding games, *Recent Advances in Game Theory*, **29**, 15–27.

Weissman, I. (1978), Estimators of parameters and large quantiles based on the k largest observations, *J. Amer. Statist. Assoc.*, **73**, 812–815.

Zhou, C. (2007), Extending maximum likelihood estimation of the extreme value index, *Unpublished manuscript*.

Zhou, C. (2008a), A 2-step estimator of the extreme value index, *Extremes*, **11**, 281–302.

Zhou, C. (2008b), Existence and consistency of the maximum likelihood estimator for the extreme value index, *J. Multivariate Anal.*, To appear.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

382. Z. ŠAŠOVOVÁ, *Liking and disliking: The dynamic effects of social networks during a large-scale information system implementation.*
383. P. RODENBURG, *The construction of instruments for measuring unemployment.*
384. M.J. VAN DER LEIJ, *The economics of networks: Theory and empirics.*
385. R. VAN DER NOLL, *Essays on internet and information economics.*
386. V. PANCHENKO; *Nonparametric methods in economics and finance: dependence, causality and prediction.*
387. C.A.S.P. SÁ, *Higher education choice in The Netherlands: The economics of where to go.*
388. J. DELFGAAUW, *Wonderful and woeful work: Incentives, selection, turnover, and workers' motivation.*
389. G. DEBREZION, *Railway impacts on real estate prices.*
390. A.V. HARDIYANTO, *Time series studies on Indonesian rupiah/USD rate 1995 – 2005.*
391. M.I.S.H. MUNANDAR, *Essays on economic integration.*
392. K.G. BERDEN, *On technology, uncertainty and economic growth.*
393. G. VAN DE KUILEN, *The economic measurement of psychological risk attitudes.*
394. E.A. MOOI, *Inter-organizational cooperation, conflict, and change.*
395. A. LLENA NOZAL, *On the dynamics of health, work and socioeconomic status.*
396. P.D.E. DINDO, *Bounded rationality and heterogeneity in economic dynamic models.*
397. D.F. SCHRAGER, *Essays on asset liability modeling.*
398. R. HUANG, *Three essays on the effects of banking regulations.*
399. C.M. VAN MOURIK, *Globalisation and the role of financial accounting information in Japan.*
400. S.M.S.N. MAXIMIANO, *Essays in organizational economics.*
401. W. JANSSENS, *Social capital and cooperation: An impact evaluation of a women's empowerment programme in rural India.*
402. J. VAN DER SLUIS, *Successful entrepreneurship and human capital.*
403. S. DOMINGUEZ MARTINEZ, *Decision making with asymmetric information.*
404. H. SUNARTO, *Understanding the role of bank relationships, relationship marketing, and organizational learning in the performance of people's credit bank.*
405. M.Â. DOS REIS PORTELA, *Four essays on education, growth and labour economics.*
406. S.S. FICCO, *Essays on imperfect information-processing in economics.*

407. P.J.P.M. VERSIJP, *Advances in the use of stochastic dominance in asset pricing.*
408. M.R. WILDENBEEST, *Consumer search and oligopolistic pricing: A theoretical and empirical inquiry.*
409. E. GUSTAFSSON-WRIGHT, *Baring the threads: Social capital, vulnerability and the well-being of children in Guatemala.*
410. S. YERGOU-WORKU, *Marriage markets and fertility in South Africa with comparisons to Britain and Sweden.*
411. J.F. SLIJKERMAN, *Financial stability in the EU.*
412. W.A. VAN DEN BERG, *Private equity acquisitions.*
413. Y. CHENG, *Selected topics on nonparametric conditional quantiles and risk theory.*
414. M. DE POOTER, *Modeling and forecasting stock return volatility and the term structure of interest rates.*
415. F. RAVAZZOLO, *Forecasting financial time series using model averaging.*
416. M.J.E. KABKI, *Transnationalism, local development and social security: the functioning of support networks in rural Ghana.*
417. M. POPLAWSKI RIBEIRO, *Fiscal policy under rules and restrictions.*
418. S.W. BISSESSUR, *Earnings, quality and earnings management: the role of accounting accruals.*
419. L. RATNOVSKI, *A Random Walk Down the Lombard Street: Essays on Banking.*
420. R.P. NICOLAI, *Maintenance models for systems subject to measurable deterioration.*
421. R.K. ANDADARI, *Local clusters in global value chains, a case study of wood furniture clusters in Central Java (Indonesia).*
422. V.KARTSEVA, *Designing Controls for Network Organizations: A Value-Based Approach*
423. J. ARTS, Essays *on New Product Adoption and Diffusion*
424. A. BABUS, *Essays on Networks: Theory and Applications*
425. M. VAN DER VOORT, *Modelling Credit Derivatives*
426. G. GARITA, *Financial Market Liberalization and Economic Growth*
427. E.BEKKERS, Essays *on Firm Heterogeneity and Quality in International Trade*
428. H.LEAHU, *Measure-Valued Differentiation for Finite Products of Measures: Theory and Applications*
429. G. BALTUSSEN, *New Insights into Behavioral Finance*
430. W. VERMEULEN, *Essays on Housing Supply, Land Use Regulation and Regional Labour Markets*
431. I.S. BUHAI, *Essays on Labour Markets: Worker-Firm Dynamics, Occupational Segregation and Workplace Conditions*