# MODELLING AND INFERENCE FOR EXTREMAL EVENTS: METHODS AND TECHNIQUES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Yuelin Sun

August 2018

MODELLING AND INFERENCE FOR EXTREMAL EVENTS: METHODS AND TECHNIQUES

Yuelin Sun, Ph.D.

Cornell University 2018

Extreme events are frequently observed in nature and in human activities; they tend to have severe and often negative impact. For this reason they are well-studied, and the underlying body of work is usually referred to as extreme value theory. The theory often deals with the behavior in the tail of probability distributions or data sets. A key notation is that of heavy-tailed probability distributions. Univariate heavy-tailed distributions exhibit interesting mathematical properties practical for modelling purposes. However, many types of univariate heavy-tailed distributions do not have natural multivariate extensions. Another area of interest in extreme value theory is that of the clustering of extremes in stationary sequences. Inference of cluster sizes tends to be difficult, partly due to the scarcity of data. Clustering also introduces heavy serial dependence in extremal observations, which in turn influences statistical analysis. This thesis seeks to address the aforementioned problems and difficulties. Key contributions include: a multivariate model for a particular class of heavy-tailed distributions, the subexponential distributions, that allows for the approximation of ruin probabilities; a multilevel approach to extremal inference that partially addresses the issue of data scarcity and that improves the variance of extremal estimators; and an algorithmic method to reduce the impact of serial dependence in extremal inference.

## BIOGRAPHICAL SKETCH

Julian Sun received his undergraduate degree from University of Waterloo in Mathematics and began his Ph.D. studies in the School of Operations Research and Information Engineering at Cornell University in 2013. Upon completing his doctoral studies, he will be working as a Senior Operations Research Analyst at Wayfair in Boston.

This thesis is dedicated to my family and friends.

**ACKNOWLEDGEMENTS**

First and foremost, I would like to thank my Ph.D. advisor Gennady Samorodnitsky. His deep mathematical knowledge, astute intuitions, and passion for research guided me throughout my time as a Ph.D. student. His mentorship is the main reason that I am able to work through many obstacles in my research. I feel very fortunate to have Gena as my Ph.D. advisor.

I would also like to thank Sidney Resnick for many long conversations often pertaining to my research. I am grateful for his many forms of guidance, understanding, and encouragement.

My gratitude goes out to my parents for their love and unwavering support, and the many sacrifices they made for me and my education.

I will give my last thanks to all the friends I made along the way, for making the time fly by.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

CHAPTER 1

# INTRODUCTION

Extreme events arise frequently in nature. Excessive rainfall, floods and earthquakes are all examples of extreme events. They also occur in human activities in the form of, for example, large files over network traffic, catastrophe insurance losses and stock market crashes. Due to their severe and often negative impact, extreme events are heavily studied, and there have been major efforts to model and predict such events.

Closely associated with concept of extreme events is that of heavy-tailed probability distributions. There is no universal definition for heavy-tailed distributions. Some authors use the term to refer to distributions that do not have finite first or second moments, while others may be referring to distributions whose (right) tails are not exponentially bounded. Broadly speaking, in the univariate case, heavy-tailed distributions refer to probability distributions in which the probability of a large value is relatively big. Heavy-tailed distributions, along with heavy-tailed phenomena, are extensively studied in [47].

There are many classes of distributions that are considered heavy-tailed. In one dimension, examples of such classes include regularly varying distributions, subexponential distributions, long-tailed distributions, dominated varying tail distributions, etc. Each of these classes of distributions exhibits many unique and interesting properties, making them powerful instruments for studying and modelling extreme events. For example, regularly varying distributions are characterized as probability distributions whose (right) tail decays like a "power law". Examples of such distributions include Pareto, Cauchy, $t$ and $F$ distributions. Financial log returns have been observed to exhibit reg-

ularly varying tail behaviour. Correspondingly, under some mild conditions, many stochastic processes used to model financial returns, such as the ARCH and GARCH processes, have regularly varying marginal distributions.

However, despite the fact that univariate heavy-tailed distributions play a prominent role in modelling extreme events, often times one cannot consider events as a single random variable. A financial portfolio usually consists of a large number of dependent assets, and should be modelled as a multidimensional vector. In this instance, the concept of multivariate heavy-tailed distributions come into play. Classes such as regularly varying distributions have natural multivariate extensions with interesting mathematical characterizations and properties; see e.g. [3]. Those properties can again be used for practical purposes such as in [33].

Unfortunately, many types of univariate heavy-tailed distributions do not have natural multivariate extensions, neither do some of their more useful properties. A central property that makes univariate heavy-tailed distributions attractive for modelling purposes is that of the single large jump principle. In broad strokes, the property states that the most likely way for the sum of some random variables to be large is if one of the summands is large. This property is very useful in the study of risk theory. Subexponential distributions are a large class of heavy-tailed distributions that exhibit the single large jump principle. There has been considerable interest in the study of subexponential distributions; see e.g. [46], [24], [19]. However, there is no natural multivarite extension of the subexponential distribution, or of the single large jump principle.

One of the difficulties of transitioning from a univariate to a multivariate notion of heavy-tailed distributions is that of the "tail". In one dimension, it is

very intuitive and clear what is the "tail" of a distribution; in the multivariate case, it is less obvious what is meant by the notion of a "tail". In the univariate sense, heavy-tailed distributions can be thought of as probability distributions whose tail decays slowly. In order to extend univariate definitions of a heavy-tailed distribution to a multivariate sense, it may be beneficial to introduce a rigorous definition of a multivariate "tail". We tackle this problem in Chapter 2. We will introduce a class of multivariate subexponential distributions that exhibit the single large jump principle, and showcase an application in ruin theory.

The study of extreme events, however, is not only restricted to heavy-tailed distributions. Extreme value theory is an academic discipline that studies the behavior of extreme events. It is a blend of mathematics, probability theory, and statistics. Many aspects of extreme value theory are extensively studied in [21]. As extreme events usually only concern the tail behavior of probability distributions and underlying data sets, many approaches and techniques in extreme value theory differ from those of classical probabilistic and statistical analysis.

As a result, there are many difficulties when devising statistical methods for extreme events. One of the most notable and challenging problems in extreme value theory is the scarcity of data. By nature extreme events occur very rarely, and therefore the size of the available "extreme" data set is usually very small. This problem is compounded as one often needs to make predictions about occurrences outside the range of the available data set.

With the need to make statistical inference using small data sets, the problem of using data efficiently arises. There are many approaches to tackling this problem, and in this thesis we will introduce the following techniques.

Since statistical inference in extreme value theory is often only concerned with the tail of underlying data sets, data towards the mean, or "center", of the data set tend to be ignored. This practice is predicated on the belief that the largest observations in a data set contain the best and most accurate information about the tail of the underlying distribution. While this may be true, ignoring smaller observations may not be the best practice. For example, by only using data above some high threshold, inference results usually come with a high variance. Techniques to lower the variance or overall mean squared error for specific estimators have been explored, such as in [52] and [11]. We devise a more general approach. In Chapter 3 we explore the idea of including lower levels of observations - with the caveat that larger observations receive more weight - for inference. In particular, we consider stationary sequences that exhibit serial dependence. Those sequences often manifest clusters of large values. One prominent statistic related to the size of such clusters is the *extremal index*. We apply our approach to an estimator of the extremal index.

When a data set comes from an underlying stochastic process, then, as mentioned, the data will often exhibit serial dependence. Large observations tend to come in the form of extremal clusters. Extremal clustering behavior in stationary sequences has been well-studied; see e.g. [32], [14]. Observations that come from the same "cluster" can be highly dependent with each other, and can often "pollute" extremal inference results, in the form of large biases in estimators. Therefore a declustering method is often helpful to ensure that at most one observation from each "cluster" is used for extremal inference. General declustering methods tend to be difficult, and generally resort to a blocks method or runs method; see e.g. [30] and [56]. Those methods tend to work for very specific estimators, and are sometimes difficult to generalize. We again propose

a more general approach to declustering. Since observations from the same "cluster" appear close to each other, we propose a distance based declustering method in Chapter 4. We will establish asymptotic consistency when using this declustering method in the inference for the index of regular variation.

As a word on notation, in the rest of this thesis we use the standard notation $\overline{F} = 1 - F$ to denote the tail of a distribution $F$.

CHAPTER 2

# MULTIVARIATE SUBEXPONENTIAL DISTRIBUTIONS AND THEIR APPLICATIONS

## 2.1 Introduction

Subexponential distributions are commonly viewed as the most general class of heavy tailed distributions. The notion of subexponentiality was introduced by [8] for distributions supported by $[0, \infty)$. If $F$ is a distribution function on the nonnegative real numbers, and $X_1$, $X_2$ are i.i.d. random variables with the law $F$, then $F$ is subexponential if

$$\lim_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X_1 > x)} = 2\,. \tag{2.1}$$

This notion was later extended to distributions supported by the entire real line $(-\infty, \infty)$; see e.g. [60]

The best known subclass of subexponential distributions is that of regularly varying distributions, which have power-like tails; regularly varying distributions along with the basic information on one-dimensional subexponential distributions will be reviewed in Section 2.3.

Subexponential distributions exhibit many properties that are desirable for modelling purposes. The definition (2.1) of subexponential distributions implies that the sum of two i.i.d. random variables with a subexponential distribution is large only when one of these random variables is large. The same turns out to be true for the sum of an arbitrary finite number of terms and, in many cases, for the sum of a random number of terms. Theoretically, this leads to the

"single large jump" structure of large deviations for random walks with subexponentially distributed steps; see e. g. [23]. In practice, this has turned out to be particularly important in applications to ruin probabilities. In ruin theory the situation where the claim sizes (often assumed to be independent with identical distribution) have a subexponential distribution is usually referred to as the non-Cramér case. The "single large jump" property of subexponential distributions leads to a well known form of the asymptotic behaviour of the ruin probability, and to a particular structure of the surplus path leading to the ruin; see e.g. [21] and [1].

It is desirable to have a notion of a multivariate subexponential distribution. The task is of a clear theoretical interest, and it is of an obvious interest in applications. A typical insurance company, for instance, has multiple insurance portfolios, with dependent claims, so it would be useful if one could build a model in which claims could be said to have a multivariate subexponential distribution. Such models and accompanying theory exist in one dimension, when claims are treated as univariate subexponential random variables; see e.g. [1].

There exists a well developed notion of a multivariate distribution with regularly varying tails; see e.g. [47]. In comparison, a notion of a multivariate subexponential distribution has not been developed to nearly the same extent. To the best of our knowledge, a notion of multivariate subexponentiality has been introduced twice, in [9] and in [43]. Both of these papers define a class (or classes) of multivariate distributions that extend the one-dimensional notion of a subexponential distribution in a natural way. They show that their notions of multivariate subexponentiality possess multidimensional analogs of important properties of one-dimensional subexponential distributions. Nonetheless, these

notions have not become as widely used as that of, say, a multivariate distribution with regularly varying tails.

With ruin probability applications in mind, we introduce yet another notion of multivariate subexponential distribution.

This chapter is organized as follows. In Section 2.2 we outline a univariate ruin problem and the challenges of extending the accompanying results to a multivariate case. In Section 2.3 we review the basic properties of one-dimensional subexponential distributions, in order to have a benchmark for the properties we would like a multivariate subexponential distribution to have. In Section 2.4 we discuss the definitions of multivariate subexponentiality of [9] and in [43]. Our notion of multivariate subexponential distributions is introduced in Section 2.5. Some applications of that notion to multivariate ruin problems are discussed in Section 2.6.

## 2.2 The Ruin Problem

In this section we introduce a univariate version of the ruin problem. We will present an asymptotic result regarding the univariate problem, and outline a few differences and challenges faced when extending such a result to a multivariate case.

The classical one-dimensional (Cramér-Lundberg) ruin problem can be described as follows. Suppose that an insurance company has an initial capital $u > 0$. The company receives a stream of premium income at a constant rate $c > 0$ per unit of time. The company has to pay claims that arrive according

to a rate $\lambda$ Poisson process. The claim sizes are assumed to be i.i.d. with a finite mean $\mu$ and independent of the arrival process. If $U(t)$ is the capital of the company at time $t \geq 0$, then the ruin probability is defined as the probability the company runs out of money at some point. This probability is, clearly, a function of the initial capital $u$, and it is often denoted by

$$\psi(u) = P\big(U(t) < 0 \text{ for some } t \geq 0\big).$$

The *positive safety loading*, or the *net profit condition*,

$$\rho := \frac{c}{\lambda\mu} - 1 > 0$$

says that, on average, the company receives more in premium income than it spends in claim payments. If the net profit condition fails, then an eventual ruin is certain. If the net profit condition holds, then the ruin probability is a number in $(0, 1)$, and its behaviour for large values of the initial capital $u$ strongly depends on the properties of the distribution $F$ of the claim sizes. Let

$$F_I(x) = \frac{1}{\mu}\int_0^x \overline{F}(y)\,dy,\ x \geq 0$$

be the integrated tail distribution. Using the "single large jump" principle as mentioned earlier, one can show that if $F_I$ is a subexponential distribution, then

$$\psi(u) \sim \rho^{-1} F_I(u) \text{ as } u \to \infty; \tag{2.2}$$

see Theorem 1.3.6 in [21].

As mentioned in Section 2.1, a typical insurance company would have multiple lines of business with dependent claims, it would therefore be beneficial to be able to model claims as multivariate random vectors, and extend the above result to a multivariate case. There are, however, challenges to making such an

extension, as there are many important distinctions between $\mathbb{R}$ and $\mathbb{R}^d$ that are relevant to this problem.

The first challenge is the definition of ruin, namely, that of "running out of money". In the univariate case, this has a very simple definition of reserves falling below zero. In the multivariate case, however, there could be several alternative definitions. If the insurance company's different lines of business cannot transfer reserves from one to another, then one can say that the company runs out of money as soon as the reserves for one line of business falls below zero. Alternatively, if different lines of business can freely transfer reserves amongst each other, then the company is in ruin when the total reserves amount falls below zero. The company can also have other policies or regulatory constraints, leading to various scenarios for ruin. In order to extend the result (2.2), one must be able to extend the definition of ruin robustly to accommodate a wide range of policies and regulations, but also simply enough to achieve theoretical results.

A second - related, but more subtle - challenge is that of the geometry in higher dimensional space. Specifically and more intuitively, we refer to the concept of "directions" in $\mathbb{R}^d$. In the univariate case for an insurance company, there are only two such directions: the position of the reserves can increase when receiving premium, or decrease when paying out a claim. In $\mathbb{R}^d$, however, there are infinitely many such directions that the position of reserves can take. It would be beneficial to have a similar concept of "increasing" and "decreasing" of reserves positions in $\mathbb{R}^d$ as one does in $\mathbb{R}$.

We will address those challenges in Section 2.5, when we introduce our notion of multivariate subexponentiality.

## 2.3 A review of one-dimensional subexponentiality

Many properties of subexponential distributions are critical to the result (2.2). In this section we review the basic properties of one-dimensional subexponential distributions. We denote the class of such distributions (and random variables with such distributions) by $\mathscr{S}$. Unless stated explicitly, we do not assume that a random variable with a subexponential distribution $F$ is nonnegative; such a random variable (or its distribution) is called subexponential if the nonnegative random variable $X_+ = \max(X, 0)$ is subexponential. Most of the not otherwise attributed facts stated below can be found in [20].

We first review the implications of the membership in the class $\mathscr{S}$. If a distribution $F \in \mathscr{S}$, then $F$ is *long-tailed*: for any $y \in \mathbb{R}$,

$$\lim_{x\to\infty} \frac{\overline{F}(x+y)}{\overline{F}(x)} = 1 \tag{2.3}$$

(implicitly assuming that $\overline{F}(x) > 0$ for all $x$.) The class of all long-tailed distributions is denoted by $\mathscr{L}$. The class of subexponential distributions $\mathscr{S}$ is in fact a proper subset of $\mathscr{L}$; see e.g. [19]. The class $\mathscr{L}$ of long-tailed distributions is well-studied. Notably it is closed under convolutions, while the class $\mathscr{S}$ of subexponential distributions is not, see [36].

A distribution $F$ has a regularly varying right tail if there is $\alpha \geq 0$ such that for every $b > 0$,

$$\lim_{x\to\infty} \frac{\overline{F}(bx)}{\overline{F}(x)} = b^{-\alpha}\,. \tag{2.4}$$

The parameter $\alpha$ is called the exponent of regular variation. The class of distributions with a regularly varying right tail is denoted by $\mathscr{R}$ (or $\mathscr{R}(\alpha)$ if we wish to emphasize the exponent of regular variation.) The class of regularly varying distributions is a subclass of subexponential distributions. Namely $\mathscr{R} \subset \mathscr{S}$.

In light of the definition (2.4), on can view $\mathscr{R}$ as the class of distributions with "power-like" right tails, in which case all distributions with "power-like" right tails are subexponential. This statement, however, should be treated carefully; other classes of distributions can be referred to as having "power-like" right tails, and not all of them form subclasses of $\mathscr{S}$. Indeed, consider the class $\mathscr{D}$ of distributions with *dominated varying tails*, defined by the property

$$\liminf_{x\to\infty} \frac{\overline{F}(2x)}{\overline{F}(x)} > 0\,. \tag{2.5}$$

One could view a distribution $F \in \mathscr{D}$ as having a "power-like" right tail. However, $\mathscr{D} \not\subset \mathscr{S}$. We note, on the other hand, that it is still true that $\mathscr{D} \cap \mathscr{L} \subset \mathscr{S}$; see [24]. Namely, if a distribution has a dominated varying tail, and is also long-tailed, then it is subexponential.

Many distributions that do not have "power-like" right tails are subexponential as well. Examples include the log-normal distribution, as well as the Weibull distribution with the shape parameter smaller than 1; see e.g. [46].

Let $X_1, X_2, \ldots$ be i.i.d. random variables with a subexponential distribution. The defining property (2.1) extends, automatically, to any finite number of terms, i.e.

$$\lim_{x\to\infty} \frac{P(X_1 + \ldots + X_n > x)}{P(X_1 > x)} = n \text{ for any } n \geq 1. \tag{2.6}$$

Moreover, the number of terms can also be random. Let $N$ be a random variable independent of the i.i.d. sequence $X_1, X_2, \ldots$ and taking values in the set of nonnegative integers. If

$$E\tau^N < \infty \text{ for some } \tau > 1, \tag{2.7}$$

then

$$\lim_{x\to\infty} \frac{P(X_1 + \ldots + X_N > x)}{P(X_1 > x)} = EN\,. \tag{2.8}$$

## 2.4 Existing definitions of multivariate subexponentiality

Before proceeding, let us clarify that we will use $F$ interchangeably to denote a distribution or a distribution function. Further, in this chapter we will denote $\overline{F} = 1 - F$ for a distribution function $F$ on $\mathbb{R}^d$.

The first known definition of multivariate subexponential distributions was introduced by [9]. They consider distributions supported by the entire $d$-dimensional space $\mathbb{R}^d$ (and not only by the nonnegative orthant). That paper defines both multivariate subexponential distributions, and multivariate *exponential distributions*. In our discussion here we only consider the subexponential case. The definition is tied to a function $\mathbf{b}(t) = (b_1(t), \ldots, b_d(t))$ such that $b_i(t) \to \infty$ as $t \to \infty$ for $i = 1, \ldots, d$.

One starts with defining the class of long-tailed distributions in $\mathbb{R}^d$, i.e. a multivariate analog of the class $\mathscr{L}$ in (2.3). Let $E = [-\infty, \infty]^d \setminus \{-\boldsymbol{\infty}\}$, and let $\nu$ be a finite measure on $E$ concentrated on the purely infinite points, i.e. on $\{-\infty, \infty\}^d \setminus \{-\boldsymbol{\infty}\}$, and such that $\nu(\mathbf{x} \in E : x_i = \infty) > 0$ for each $i = 1, \ldots, d$. Then a probability distribution $F$ is said to belong to the class $\mathscr{L}(\nu; \mathbf{b})$ if, as $t \to \infty$,

$$tF\big(\mathbf{b}(t) + \cdot\big) \xrightarrow{v} \nu \tag{2.9}$$

vaguely in $E$ (see [50] for a thorough treatment of vague convergence of measures.) The class of subexponential distributions (with respect to the same function $\mathbf{b}$ and the same measure $\nu$) is defined to be that subset $\mathscr{S}(\nu; \mathbf{b})$ of distributions $F$ in $\mathscr{L}(\nu; \mathbf{b})$ for which

$$tF * F\big(\mathbf{b}(t) + \cdot\big) \xrightarrow{v} 2\nu$$

vaguely in $E$, where $*$ denotes the convolution of distribution functions.

Corollary 2.4 in [9] shows that $F \in \mathscr{S}(\nu; \mathbf{b})$ if and only if $F \in \mathscr{L}(\nu; \mathbf{b})$ and the marginal distribution $F_i$ of $F$ is in the one-dimensional subexponential class $\mathscr{S}$ for each $i = 1, \ldots, d$.

It is shown in [9] that the distributions in $\mathscr{S}(\nu, \mathbf{b})$ possess the natural multivariate extensions of the properties of the one-dimensional subexponential distributions mentioned in Section 2.3. For example, if $F \in \mathscr{S}(\nu, \mathbf{b})$, then for any $n \geq 1$, $F^{*n} \in \mathscr{S}(n\nu, \mathbf{b})$. More generally, if $N$ is a random variable satisfying (2.7), and $H = \sum_{n=0}^{\infty} P(N = n) F^{*n}$, then $H \in \mathscr{S}(EN\nu, \mathbf{b})$.

The distributions in $\mathscr{S}(\nu, \mathbf{b})$ also possess the right relation with the distributions with multivariate regularly varying tails. Denote the nonnegative quadrant $\mathbb{R}_+^d = [0, \infty)^d$. A distribution $F$ supported by $\mathbb{R}_+^d$ is said to have regularly varying tails if there is a Radon measure $\mu$ on $[0, \infty]^d \setminus \{\mathbf{0}\}$ concentrated on finite points, and a function $\mathbf{b}$ as above such that, as $t \to \infty$,

$$tF\big(\mathbf{b}(t)\cdot\big) \xrightarrow{v} \mu \tag{2.10}$$

vaguely in $[0, \infty]^d \setminus \{\mathbf{0}\}$; see [47]. Note that (2.10) allows for different scaling in different directions, hence also different marginal exponents of regular variation. This situation is sometimes referred to as *non-standard regular variation*. If we denote by $\mathscr{R}(\mu, \mathbf{b})$ the class of distributions with regularly varying tails satisfying (2.10), then, as shown in [9], $\mathscr{R}(\mu, \mathbf{b}) \subset \mathscr{S}(\nu, \mathbf{b})$ for some $\nu$.

As mentioned above, this definition of multivariate subexponentiality requires, beyond marginal subexponentiality for all components, only the joint long tail property (2.9). This property, together with the nature of the limiting measure, makes this notion somewhat inconvenient in applications, because it is not easy to see how to use it on sets in $\mathbb{R}^d$ that are not "asymptotically rectangular".

Another observation worth making is that in probability theory, many well established multivariate extensions of important one-dimensional notions have a "stability property" with respect to projections on one-dimensional subspaces (i.e., with respect to taking linear combinations of the components.) Specifically, if the distribution of a random vector $(X^{(1)}, \ldots, X^{(d)})$ has, say, a property $\mathscr{G}_d$ (the subscript $d$ specifying the dimension in which the property holds), then ideally the distribution of any (non-degenerate) linear combination $\sum_1^d a_i X^{(i)}$ should have the property $\mathscr{G}_1$. This is true, for instance, for multivariate regular variation, multivariate Gaussianity, stability and infinite divisibility. Unfortunately, the definition of multivariate subexponentiality by $\mathscr{S}(\nu, \mathbf{b})$ does not have this feature, as the following example shows.

**Example 2.4.1.** Consider a 2-dimensional random vector $(X, Y)$ with nonnegative coordinates such that $P(X + Y = 2^n) = 2^{-(n+1)}$ for $n \geq 0$, with the mass distributed uniformly on the simplex $\{(x, y) : x, y \geq 0,\, x + y = 2^n\}$ for each $n \geq 0$. It is elementary to check that $X, Y \in \mathscr{L} \cap \mathscr{D} \subset \mathscr{S}$. Furthermore, for $2^n \leq x \leq 2^{n+1}$, $n = 0, 1, 2, \ldots$ we can compute

$$P(X > x) = P(Y > x) = 2^{-(n+1)} - \frac{x}{3} 2^{-(2n+1)} = 2P(X > x, Y > x)\,.$$

If we define a function $b$ by $tP(X > b(t)) = 1$ for $t \geq 2$, then it immediately follows that $(X, Y) \in \mathscr{L}(\nu; \mathbf{b})$ with $\mathbf{b}(t) = (b(t), b(t))$ and

$$\nu = \frac{1}{2}\delta_{(-\infty,\infty)} + \frac{1}{2}\delta_{(\infty,-\infty)} + \frac{1}{2}\delta_{(\infty,\infty)}\,,$$

and the result of [9] tells us that $(X, Y) \in \mathscr{S}(\nu; \mathbf{b})$. It is clear, however, that

$$\liminf_{x\to\infty} \frac{P(X + Y > x + 1)}{P(X + Y > x)} = \frac{1}{2}\,,$$

so $X+Y$ does not even have a long-tailed, let alone subexponential, distribution.

The second existing definition of multivariate subexponentiality we are aware of is due to [43]. Once again, this definition concerns rectangular regions on the space $\mathbb{R}^d$. The paper presents 3 versions of the definition. The versions are similar, and we concentrate only on one of them. Let $F$ be a probability distribution on $\mathbb{R}^d$ with $F(\mathbf{0}+) = 0$. Then one says that $F \in S(\mathbb{R}^d)$ if for all $\mathbf{x} \in (0,\infty]^d$ with $\min(x_i) < \infty$,

$$\lim_{t\to\infty} \frac{\overline{F^{*2}}(t\mathbf{x})}{\overline{F}(t\mathbf{x})} = 2. \tag{2.11}$$

This definition, like the definition of [9], has the following property: a distribution $F \in S(\mathbb{R}^d)$ if and only if each marginal distribution $F_i$ of $F$ is a one-dimensional subexponential distribution, and a multivariate long-tail property holds. In the present case the long-tail property is

$$\lim_{t\to\infty} \frac{\overline{F}(t\mathbf{x}-\mathbf{a})}{\overline{F}(t\mathbf{x})} = 1 \tag{2.12}$$

for each $\mathbf{x} \in (0,\infty]^d$ with $\min(x_i) < \infty$ and each $\mathbf{a} \in [0,\infty)^d$. This follows from Theorem 7 and Corollary 11 in [43].

The following statement shows that, in fact, the definition (2.11) of multivariate subexponentiality requires *only* marginal subexponentiality of each coordinate.

**Proposition 2.4.2.** *Let $F$ be a probability distribution supported by the positive quadrant in $\mathbb{R}^d$. Then $F \in S(\mathbb{R}^d)$ if and only if all marginal distributions $F_i$ of $F$ are subexponential in one dimension.*

*Proof.* By choosing $\mathbf{x}$ with only one finite coordinate, we immediately see that if $F \in S(\mathbb{R}^d)$, then $F_i \in \mathscr{S}$ for each $i = 1,\dots,d$.

In the other direction, we know by the results of [43], that only the long-tail property (2.12) is needed, in addition to the marginal subexponentiality, to establish that $F \in S(\mathbb{R}^d)$. Therefore, it is enough to check that the long-tail property (2.12) follows from the marginal subexponentiality. In fact, we will show that, if each $F_i$ is long-tailed, i.e. satisfies (2.3), $i = 1, \ldots, d$, then (2.12) holds as well.

Let $\epsilon > 0$. Fix $\mathbf{x} = (x_1, \ldots, x_d) \in (0, \infty)^d$ (allowing some of the components of $\mathbf{x}$ be infinite only leads to a reduction in the dimension), and $\mathbf{a} = (a_1, \ldots, a_d) \in [0, \infty)^d$.

Since $F_i \in \mathscr{L}$, $i = 1, \ldots, d$, for sufficiently large $t$ we have

$$0 \leq \overline{F_i}(tx_i - a_i) - \overline{F_i}(tx_i) < \epsilon \overline{F_i}(tx_i)$$

for $i = 1, \ldots, d$. Further, it is clear that

$$0 \leq \overline{F}(t\mathbf{x} - \mathbf{a}) - \overline{F}(t\mathbf{x}) \leq \sum_{i=1}^{d} \left( \overline{F_i}(tx_i - a_i) - \overline{F_i}(tx_i) \right).$$

Hence for sufficiently large $t$,

$$\begin{aligned} 0 \leq \frac{\overline{F}(t\mathbf{x} - \mathbf{a}) - \overline{F}(t\mathbf{x})}{\overline{F}(t\mathbf{x})} &\leq \sum_{i=1}^{d} \frac{\left( \overline{F_i}(tx_i - a_i) - \overline{F_i}(tx_i) \right)}{\overline{F}(t\mathbf{x})} \\ &\leq \sum_{i=1}^{d} \frac{\left( \overline{F_i}(tx_i - a_i) - \overline{F_i}(tx_i) \right)}{\overline{F_i}(tx_i)} \\ &< d\epsilon. \end{aligned}$$

Letting $\epsilon \to 0$ gives the desired result. □

**Remark 2.4.3.** It is worth noting that the above statement and Corollary 11 in [43] show that for any probability distribution $F$ supported by the positive

quadrant in $\mathbb{R}^d$, such that the marginal distribution $F_i$ of $F$ is subexponential for every $i = 1, \ldots, d$, we have, for all $\mathbf{a} \in [0, \infty)^d$, $\mathbf{x} \in (0, \infty)^d$ and $n \geq 1$,

$$\lim_{t \to \infty} \frac{\overline{F^{*n}}(t\mathbf{x} - \mathbf{a})}{\overline{F}(t\mathbf{x})} = n. \tag{2.13}$$

**Remark 2.4.4.** Proposition 2.4.2 was apparently a by-product of Proposition 11 in [2], when the authors were investigating the difference between $\overline{F^{*n}}(\mathbf{x})$ and $n\overline{F}(\mathbf{x})$.

**Remark 2.4.5.** As an extension to long-tailed and subexponential distributions in one dimension, for $\gamma \geq 0$, one says that $F \in L_\gamma$ if $\lim_{x \to \infty} \overline{F}(x - y)/\overline{F}(x) = e^{\gamma y}$ for all $y \in \mathbb{R}$; and $F \in S_\gamma$ if $F \in L_\gamma$ and $\lim_{x \to \infty} \overline{F^{*2}}(x)/\overline{F}(x) = D < \infty$. In [44], the definition of $S(\mathbb{R}^d)$ was extended in a similar fashion. One says that $\overline{F} \in L(\mathbb{R}^d, \nu)$ if for some real function $\nu$,

$$\lim_{t \to \infty} \frac{\overline{F}(t\mathbf{x} - \mathbf{a})}{\overline{F}(t\mathbf{x})} = \nu(\mathbf{x}, \mathbf{a}) \tag{2.14}$$

for each $\mathbf{x} \in (0, \infty]^d$ with $\min(x_i) < \infty$ and each $\mathbf{a} \in [0, \infty)^d$. Further, $F \in S(\mathbb{R}^d, \alpha)$ if some real function $\alpha$ and for all $\mathbf{x} \in (0, \infty]^d$ with $\min(x_i) < \infty$,

$$\lim_{t \to \infty} \frac{\overline{F^{*2}}(t\mathbf{x})}{\overline{F}(t\mathbf{x})} = \alpha(\mathbf{x}). \tag{2.15}$$

The authors went on to show in Theorem 3 of the paper, that as long as for each $i = 1, \ldots, d$, the marginal distribution $F_i$ satisfies $F_i \in S_{\gamma_i}$ for some $\gamma_i \geq 0$, and that $\overline{F} \in L(\mathbb{R}^d, \nu)$ for some real function $\nu$, then $F \in S(\mathbb{R}^d, \alpha)$ for some $\alpha$. However, with a proof similar to that of Proposition 2.4.2, one can show that as long as $F_i \in S_{\gamma_i}$ for some $\gamma_i \geq 0$ for each $i = 1, \ldots, d$, then $\overline{F} \in L(\mathbb{R}^d, \nu)$ for some real function $\nu$. So only the first condition is required.

Given Proposition 2.4.2, using (2.11) as a definition of multivariate subexponentiality is, therefore, equivalent to merely requiring one-dimensional subexponentiality for each marginal distribution. Such requirement, in particular,

cannot guarantee one-dimensional subexponentiality of the linear combinations, as we have seen in Example 2.4.1. In fact, it was shown in [36] that even the sum of independent random variables with subexponential distributions does not need to have a subexponential distribution.

## 2.5 Multivariate Subexponential Distributions

In this section we introduce a new notion of a multivariate subexponential distribution. We approach the task with the multivariate ruin problem in mind. We start with a family $\mathcal{R}$ of open sets in $\mathbb{R}^d$. Recall that a subset $A$ of $\mathbb{R}^d$ is increasing if $\mathbf{x} \in A$ and $\mathbf{a} \in [0,\infty)^d$ imply $\mathbf{x}+\mathbf{a} \in A$. Let

$$\mathcal{R} = \{A \subset \mathbb{R}^d : \ A \text{ open, increasing, } \ A^c \text{ convex, } \ \mathbf{0} \notin \overline{A}\}. \tag{2.16}$$

**Remark 2.5.1.** Note that $\mathcal{R}$ is a cone with respect to the multiplication by positive scalars. That is, if $A \in \mathcal{R}$, then $uA \in \mathcal{R}$ for any $u > 0$. Further, half-spaces of the form

$$H = \{\mathbf{x} : \ a_1x_1 + \cdots + a_dx_d > b\}, \ b > 0, \ a_1, \ldots, a_d \geq 0 \text{ with } a_1 + \ldots + a_d = 1 \tag{2.17}$$

are members of $\mathcal{R}$.

**Remark 2.5.2.** We can write a set $A \in \mathcal{R}$ (in a non-unique way) as $A = \mathbf{b}+G$ ($\mathbf{x} \in A$ if and only if $\mathbf{x} = \mathbf{b} + \mathbf{y}$ for some $\mathbf{y} \in G$), with $\mathbf{b} \in (0,\infty)^d$ and $\mathbf{0} \in \partial G$ (with $\partial G$ being the boundary of $G$). It is clear that the set $G$ is then also increasing. We will adopt this notation in some of the proofs to follow.

To see a connection with the multivariate ruin problem, imagine that for a fixed set $A \in \mathcal{R}$ we view $A$ as the "ruin set" in the sense that if, at any time, the

excess of claim amounts over the premia falls in $A$, then the insurance company is ruined. Note that, in the one-dimensional situation, all sets in $\mathcal{R}$ are of the form $A = (u, \infty)$ with $u > 0$, so the ruin corresponds to the excess of claim amounts over the premia being over the initial capital $u$. The different shapes of sets in $\mathcal{R}$ can be viewed as allowing different interactions between multiple lines of business. For example, choosing $A$ of the form

$$A = \{\mathbf{x} : \, x_i > u_i \text{ for some } i = 1, \ldots, d\}, \;\; u_1, \ldots, u_d > 0$$

corresponds to completely separate lines of business, where a ruin of one line of business causes the ruin of the company. On the other hand, using as $A$ a half-space of the form (2.17) corresponds to the situation where there is a single overall initial capital $b$ and the proportion of $a_i$ in a shortfall in the $i$th line of business is charged to the overall capital $b$. This addresses the first issue raised in Section 2.2. The connections to the ruin problem are discussed more thoroughly in Section 2.6.

Before we introduce our notion of multivariate subexponentiality, we collect, in the following lemma, certain facts about the family $\mathcal{R}$. Note that part (d) is a general property of convex sets.

**Lemma 2.5.3.** *Let $A \in \mathcal{R}$.*

*(a) If $G = A - \mathbf{b}$ for some $\mathbf{b} \in \partial A$, then $G^c \supset (-\infty, 0]^d$.*

*(b) If $u_1 > u_2 > 0$ then $u_1 A \subset u_2 A$.*

*(c) There is a set of vectors $I_A \subset \mathbb{R}^d$ such that*

$$A = \left\{\mathbf{x} \in \mathbb{R}^d : \, \mathbf{p}^T \mathbf{x} > 1 \text{ for some } \mathbf{p} \in I_A\right\}.$$

*(d) Let $C$ be any convex set, and $u_1, u_2 > 0$, then $u_1 C + u_2 C = (u_1 + u_2) C$.*

*Proof.* (a) Since $G^c$ is closed, it contains the origin. Since $G$ is increasing, $G^c$ contains the entire quadrant $(-\infty, 0]^d$.

(b) This is an immediate consequence of the fact that $A^c$ is convex and $\mathbf{0} \in A^c$.

(c) Let $\mathbf{x}_0 \in \partial A$. Since $A^c$ is convex, the supporting hyperplane theorem (see e.g. Corollary 11.6.2 in [53]) tells us that there exists a (not necessarily unique) nonzero vector $\mathbf{p}_{\mathbf{x}_0}$ such that $\mathbf{p}_{\mathbf{x}_0}^T \mathbf{x} \leq \mathbf{p}_{\mathbf{x}_0}^T \mathbf{x}_0$ for all $\mathbf{x} \in A^c$. Since $\mathbf{0} \in A^c$, we must have $\mathbf{p}_{\mathbf{x}_0}^T \mathbf{x}_0 \geq 0$. Since $A$ is increasing, the case $\mathbf{p}_{\mathbf{x}_0}^T \mathbf{x}_0 = 0$ is impossible, so $\mathbf{p}_{\mathbf{x}_0}^T \mathbf{x}_0 > 0$.

We scale each $\mathbf{p}_{\mathbf{x}_0}$ so that $\mathbf{p}_{\mathbf{x}_0}^T \mathbf{x}_0 = 1$. Let $I_A$ be the set of all such $\mathbf{p}_{\mathbf{x}_0}$ for all $\mathbf{x}_0 \in \partial A$. Since a closed convex set equals the intersection of the half-spaces bounded by its supporting hyperplanes (see e.g. Corollary 11.5.1 in [53]), the collection $I_A$ has the required properties.

(d) Let $\mathbf{x} \in u_1 C$ and $\mathbf{y} \in u_2 C$, then $\frac{\mathbf{x}}{u_1}, \frac{\mathbf{y}}{u_2} \in C$, and by convexity $\frac{u_1}{u_1+u_2}\frac{\mathbf{x}}{u_1} + \frac{u_2}{u_1+u_2}\frac{\mathbf{y}}{u_2} = \frac{1}{u_1+u_2}(\mathbf{x}+\mathbf{y}) \in C$, so $\mathbf{x}+\mathbf{y} \in (u_1+u_2)C$, implying that $u_1 C + u_2 C \subset (u_1+u_2)C$. The other direction is obvious. □

**Remark 2.5.4.** It is clear that, once we have chosen a collection $I_A$ for some $A \in \mathcal{R}$, for any $u > 0$ we can use $I_A/u$ as $I_{uA}$.

We are now ready to define multivariate subexponentiality. Let $F$ be a probability distribution on $\mathbb{R}^d$ supported by $[0, \infty)^d$. For a fixed $A \in \mathcal{R}$ it follows from part (b) of Lemma 2.5.3 that the function $F_A$ on $[0, \infty)$ defined by

$$F_A(t) = 1 - F(tA),\ t \geq 0\,,$$

is a probability distribution function on $[0, \infty)$.

**Definition 2.5.5.** For any $A \in \mathcal{R}$, we say that $F \in \mathscr{S}_A$ if $F_A \in \mathscr{S}$, and we write $\mathscr{S}_{\mathcal{R}} := \cap_{A\in\mathcal{R}}\mathscr{S}_A$.

We view the class $\mathscr{S}_{\mathcal{R}}$ as the class of subexponential distributions. However, for some applications we can use a larger class, such as $\mathscr{S}_A$ for a fixed $A \in \mathcal{R}$, or the intersection of such classes over a subset of $\mathcal{R}$.

We now introduce some useful properties of the class $\mathscr{S}_{\mathcal{R}}$. By Remark 2.5.1, if $\mathbf{X}$ is a random vector in $\mathbb{R}^d$ whose distribution is in $\mathscr{S}_{\mathcal{R}}$, then all non-degenerate linear combinations of the components of $\mathbf{X}$ with nonnegative coefficients have one-dimensional subexponential distributions. More generally, we have the following stability property. We say that a linear transformation $T : \mathbb{R}^d \to \mathbb{R}^k$ is increasing if $T\mathbf{x} \in [0,\infty)^k$ for any $\mathbf{x} \in [0,\infty)^d$.

**Proposition 2.5.6.** *Let $T : \mathbb{R}^d \to \mathbb{R}^k$ be an increasing linear transformation. If $\mathbf{X}$ is a random vector in $\mathbb{R}^d$ whose distribution is in $\mathscr{S}_{\mathcal{R}}$ (in $\mathbb{R}^d$), then the same is true (in $\mathbb{R}^k$) for the distribution of the random vector $T\mathbf{X}$.*

*Proof.* It suffices to show that for any $A \in \mathcal{R}$ in $\mathbb{R}^k$, the set $T^{-1}A$ is in $\mathcal{R}$ in $\mathbb{R}^d$.

We check that $T^{-1}A$ satisfies each of the conditions in (2.16). That $T^{-1}A$ is open follows from that fact that $T$ is continuous and $A$ is open. To show that $T^{-1}A$ is increasing, let $\mathbf{x} \in T^{-1}A$, and $\mathbf{a} \geq \mathbf{0}$. In this case $T\mathbf{x} = \mathbf{y}$ for some $\mathbf{y} \in A$, and $T\mathbf{a} = \mathbf{b}$ for some $\mathbf{b} \geq \mathbf{0}$. Hence by linearity and the fact that $A$ is increasing, it follows that $T(\mathbf{x}+\mathbf{a}) = \mathbf{y}+\mathbf{b} \in A$. As for convexity of $(T^{-1}A)^c$, notice that $(T^{-1}A)^c = T^{-1}A^c$. For any $\mathbf{x},\mathbf{y} \in (T^{-1}A)^c$, then, $T\mathbf{x}, T\mathbf{y} \in A^c$. For $\lambda \in (0,1)$, $T(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) = \lambda T\mathbf{x}+(1-\lambda)T\mathbf{y} \in A^c$, and convexity of $(T^{-1}A)^c = T^{-1}A^c$ follows. Lastly, notice that $\overline{T^{-1}A} = T^{-1}\overline{A}$, so $\mathbf{0} \in \overline{T^{-1}A}$ leads to the contradiction that $\mathbf{0} \in \overline{A}$. □

**Remark 2.5.7.** Note that Proposition 2.5.6 also holds if instead of the entirety of $\mathscr{S}_{\mathcal{R}}$, we restrict the distributions to those in $\mathscr{S}_{\mathcal{R}_H}$, where $\mathcal{R}_H$ consists only of half-spaces of the form (2.17), and $\mathscr{S}_{\mathcal{R}_H} := \cap_{A\in\mathcal{R}_H}\mathscr{S}_A$. This follows easily from the above proof and the fact that $T^{-1}H$ is still a half-space for any half-space $H$.

The following lemma is introduces certain useful equivalences.

**Lemma 2.5.8.** *Let $A \in \mathcal{R}$, and let $\mathbf{X}$ be a random vector with distribution $F$. Let $I_A$ be as defined in the proof of Lemma 2.5.3. Define $Y_A = \sup\{u : \mathbf{X} \in uA\}$. Then the following equivalences hold.*

$$\mathbf{X} \in uA \Leftrightarrow Y_A > u \Leftrightarrow \sup_{\mathbf{p}\in I_A} \mathbf{p}^T\mathbf{X} > u. \tag{2.18}$$

*Proof.* This follows directly from that fact that for any $u > 0$, we can write the event $\{\mathbf{X} \in uA\}$ as $\left\{\sup_{\mathbf{p}\in I_A} \mathbf{p}^T\mathbf{X} > u\right\}$; see Lemma 2.5.3 and Remark 2.5.4. □

In light of Lemma 2.5.8, one can view $Y_A$ as a projection of $\mathbf{X}$ onto $\mathbb{R}$, which helps address the second issue raised in Section 2.2.

The next lemma is useful and follows quite naturally from Lemma 2.5.8.

**Lemma 2.5.9.** *For any $A \in \mathcal{R}$ and $n \geq 1$,*

$$\overline{(F_A)^{*n}}(t) \geq F^{*n}(tA). \tag{2.19}$$

*Proof.* Let $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ be independent random vectors with distribution $F$. Let $Y_1, \ldots, Y_n$ be one-dimensional random variables defined by

$$Y_i = \sup\{u : \mathbf{X}^{(i)} \in uA\} = \sup_{\mathbf{p}\in I_A} \mathbf{p}^T\mathbf{X}^{(i)},\ i = 1, \ldots, d\,.$$

By Lemma 2.5.8, $P(Y_i > t) = \overline{F_A}(t)$. Hence it follows that

$$\begin{aligned} F^{*n}(tA) &= P(\mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)} \in tA) \\ &= P\big(\sup_{\mathbf{p}\in I_A} \mathbf{p}^T(\mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)}) > t\big) \\ &\leq P\big(\sup_{\mathbf{p}\in I_A} \mathbf{p}^T\mathbf{X}^{(1)} + \cdots + \sup_{\mathbf{p}\in I_A} \mathbf{p}^T\mathbf{X}^{(n)} > t\big) \\ &= P(Y_1 + \ldots + Y_n > t) \\ &= \overline{(F_A)^{*n}}(t)\,, \end{aligned}$$

as required. □

In spite of this result, the two probabilities in (2.19) are in fact asymptotically equivalent.

**Corollary 2.5.10.** *$A \in \mathcal{R}$. Let $\mathbf{X}, \mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ be independent random vectors with distribution $F$. If $F \in \mathscr{S}_A$ for some $A \in \mathcal{R}$, then for all $n \geq 1$,*

$$\lim_{u\to\infty} \frac{P(\mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)} \in uA)}{P(\mathbf{X} \in uA)} = n. \tag{2.20}$$

*Proof.* It follows from Lemma 2.5.9 that only an asymptotic lower bound needs to be established. However, since $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ are all nonnegative, and $A$ is an increasing set, it must be that if $\mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)} \in uA^c$, then each $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)} \in uA^c$. Therefore,

$$\begin{aligned} P(\mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)} \in uA^c) &\leq P(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)} \in uA^c) \\ &= P(\mathbf{X} \in uA^c)^n. \end{aligned}$$

It follows that

$$\liminf_{u\to\infty} \frac{P(\mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)} \in uA)}{P(\mathbf{X} \in uA)} \geq \liminf_{u\to\infty} \frac{1 - P(\mathbf{X} \in uA^c)^n}{P(\mathbf{X} \in uA)} = n\,,$$

as required. □

**Remark 2.5.11.** In [45], the authors investigated the difference between product and convolutions of distribution functions, as well as the difference between the tail of the convolution and the sum of tails, for several different classes of distribution functions on $\mathbb{R}^d$. In particular, one of the object of study was the tail behavior and convergence rate of the function $R_n(\mathbf{x}) = \overline{F^{*n}}(\mathbf{x}) - n\overline{F}(\mathbf{x})$. If we add the class of functiosn $\mathscr{S}_A$ for some $A \in \mathcal{R}$, we can slightly modify the definition of this function to $R'_n(uA) = F^{*n}(uA) - nF(uA)$, and by Corollary 2.5.10, conclude that $R'_n(uA) = o(1)F(uA)$.

**Remark 2.5.12.** We note at this point that the assumption $F \in \mathscr{S}_A$ is NOT equivalent to the assumption that (2.20) holds for all $n$. In fact, the latter assumption is weaker. To see that, consider the following example. Let $X$ and $Y$ be two independent nonnegative one-dimensional random variables with subexponential distributions, such that $X + Y$ is not subexponential; recall that such random variables exist, see [36]. We construct a bivariate random vector $\mathbf{Z}$ by taking a Bernoulli $(1/2)$ random variable $B$ independent of $X$ and $Y$ and setting $\mathbf{Z} = (X, 0)$ if $B = 0$ and $\mathbf{Z} = (0, Y)$ if $B = 1$. Let $A = \{(x, y) : \max(x, y) > 1\}$. Since the marginal distributions of the bivariate distribution of $\mathbf{Z}$ are, obviously, subexponential, we see by (2.13) that (2.20) holds for all $n \geq 1$. However, for $u > 0$,

$$\overline{F_A}(u) = \frac{1}{2}P(X > u) + \frac{1}{2}P(Y > u)\,,$$

so the distribution $F_A$ is a mixture of the distributions of $X$ and $Y$. By Theorem 2 of [19], any non-trivial mixture of the distributions of $X$ and $Y$ is subexponential if and only if their convolution is. Since, by construction, that convolution is not subexponential, we conclude that $F_A \notin \mathscr{S}$ and $F \notin \mathscr{S}_A$.

In the next proposition we check that the basic properties of one-dimensional subexponential distributions extend to the multivariate case.

**Proposition 2.5.13.** *Let $A \in \mathcal{R}$ and $F \in \mathscr{S}_A$.*

*(a) If $G$ is a distribution on $\mathbb{R}^d$ supported by $[0,\infty)^d$, such that*

$$\lim_{u\to\infty} \frac{F(uA)}{G(uA)} = c > 0,$$

*then $G \in \mathscr{S}_A$.*

*(b) For any $\mathbf{a} \in \mathbb{R}^d$,*

$$\lim_{u\to\infty} \frac{F(uA+\mathbf{a})}{F(uA)} = 1. \tag{2.21}$$

*(c) Let $\mathbf{X}, \mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ be independent random vectors with distribution $F$. For any $\epsilon > 0$, there exists $K > 0$ such that for all $u > 0$ and $n \geq 1$,*

$$\frac{P(\mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)} \in uA)}{P(\mathbf{X} \in uA)} < K(1+\epsilon)^n\,. \tag{2.22}$$

*Proof.* (a) This is an immediate consequence of the univariate subexponentiality of $F_A$ and the corresponding property of one-dimensional subexponential distributions; see e.g. Lemma 4 in [20].

(b) Write $A = \mathbf{b} + G$ as in Remark 2.5.2. As in part (b) of Lemma 2.5.3, we have $u_1 G \subset u_2 G$ if $u_1 > u_2 > 0$. Since $G$ is an increasing set, it follows that there exists $u_1 > 0$ such that for all $u > u_1$ we have $(u+u_1)A \subset uA + \mathbf{a} \subset (u-u_1)A$. Therefore,

$$\begin{aligned}
\overline{F_A}(u+u_1) &= F((u+u_1)A) \\
&\leq F(uA+\mathbf{a}) \\
&\leq F((u-u_1)A) = \overline{F_A}(u-u_1)\,,
\end{aligned}$$

and the claim follows from the one-dimensional long tail property of $F_A$.

(c) The claim follows from Lemma 2.5.9 and the corresponding one-dimensional bound; see e.g. Lemma 3 in [20]. □

**Remark 2.5.14.** In our Definition 2.5.5 of multivariate subexponentiality one can drop the assumption that a distribution is supported by $[0,\infty)^d$. We can check that both Corollary 2.5.10 and Proposition 2.5.13 remain true in this extended case.

Our next step is to show that multivariate regular varying distributions fall within the class $\mathscr{S}_{\mathcal{R}}$ of multivariate subexponential distributions. The definition of non-standard multivariate regular variation for distributions supported by $[0,\infty)^d$ was given in (2.10). Presently we would only consider the standard multivariate regular variation, but allow distributions not necessarily restricted to the first quadrant. In this case one assumes that there is a non-zero Radon measure $\mu$ on $[-\infty,\infty]^d \setminus \{\mathbf{0}\}$, charging only finite points, and a function $b$ on $(0,\infty)$ increasing to infinity, such that

$$tF\big(b(t)\cdot\big) \xrightarrow{v} \mu \tag{2.23}$$

vaguely on $[-\infty,\infty]^d \setminus \{\mathbf{0}\}$. The measure $\mu$ is called the tail measure of $\mathbf{X}$; it has automatically a scaling property: for some $\alpha > 0$, $\mu(uA) = u^{-\alpha}\mu(A)$ for every $u > 0$ and every Borel set $A \in \mathbb{R}^d$, and the function $b$ in (2.23) is regularly varying with exponent $1/\alpha$; see [47]. We say that $F$ (and $\mathbf{X}$) are regularly varying with exponent $\alpha$ and use the notation $F \in MRV(\alpha,\mu)$.

**Proposition 2.5.15.** $MRV(\alpha,\mu) \subset \mathscr{S}_{\mathcal{R}}$.

*Proof.* We start by showing that for any $A \in \mathcal{R}$, $\mu(\partial A) = 0$. Since for any $u > 0$,

$$\mu(\partial(uA)) = \mu(u\partial A) = u^{-\alpha}\mu(\partial A)\,,$$

it is enough to show that for any $u > 1$, $\partial(uA) \cap \partial A = \emptyset$ (indeed, $\mu(\partial A) > 0$ would then imply existence of uncountably many disjoint sets of positive measure).

Suppose, to the contrary, that $\partial(uA) \cap \partial A \neq \emptyset$, and let $\mathbf{x} \in \partial(uA) \cap \partial A$. The set $I_A$ in part (c) of Lemma 2.5.3, has, by construction, the property that $u^{-1}\mathbf{x}$, as an element of $u^{-1}\partial(uA) = \partial A$, satisfies $\mathbf{p}^T u^{-1}\mathbf{x} = 1$ for some $\mathbf{p} \in I_A$. But then $\mathbf{p}^T\mathbf{x} = u > 1$, which says that $\mathbf{x}$ is in $A$, rather than in $\partial A$, which is a subset of $A^c$.

It follows from (2.23) that for any set $A \in \mathcal{R}$,

$$tP\big(\mathbf{X} \in b(t)A\big) \to \mu(A) \in (0, \infty)$$

as $t \to \infty$. Since the function $b$ is regularly varying with exponent $1/\alpha$, we immediately conclude that the distribution function $F_A$ has a regularly varying tail, hence $F_A$ is subexponential. Because $A \in \mathcal{R}$ is arbitrary, it follows that $F \in \cap_{A\in\mathcal{R}}\mathscr{S}_A = \mathscr{S}_{\mathcal{R}}$. □

We proceed with clarifying the relation between the class $\mathscr{S}_{\mathcal{R}}$ we have introduced in this section and the classes $\mathscr{S}(\nu; \mathbf{b})$ and $S(\mathbb{R}^d)$ of Section 2.4. We will also provide several examples of distributions that belong to $\mathscr{S}_{\mathcal{R}}$, as well as sufficient conditions for a distribution to be a member of $\mathscr{S}_{\mathcal{R}}$.

Example 2.4.1, combined with Proposition 2.5.6, show that neither $\mathscr{S}(\nu; \mathbf{b})$ nor $S(\mathbb{R}^d)$ are subsets of $\mathscr{S}_{\mathcal{R}}$. We will present an example to show that $\mathscr{S}_{\mathcal{R}} \not\subset \mathscr{S}(\nu; \mathbf{b})$.

We start with presenting a sufficient condition for a distribution $F$ to be a member of $\mathscr{S}_{\mathcal{R}}$. We assume for the moment that $F$ is supported by $[0, \infty)^d$.

Let $\mathbf{X} \sim F$ be a nonnegative random vector on $\mathbb{R}^d$ such that $P(\mathbf{X} = \mathbf{0}) = 0$. Denote the $L_1$ norm (the choice of norm is does not matter for the subsequent

arguments) of $\mathbf{X}$ by

$$W = ||\mathbf{X}||_1 = \sum_{i=1}^{d} X_i \,, \tag{2.24}$$

and the projection of $\mathbf{X}$ onto the $d$-dimensional unit simplex $\Delta_d$ by

$$I = \frac{\mathbf{X}}{||\mathbf{X}||_1} = \frac{\mathbf{X}}{W} \in \Delta_d \,. \tag{2.25}$$

Let $\nu$ be the distribution of $I$ over $\Delta_d$, and let $(F_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Delta_d}$ be a set of regular conditional distributions of $W$ given $I$. Notice that, if the law $F$ of $\mathbf{X}$ in $\mathbb{R}^d$ has a density $f$ with respect to the $d$-dimensional Lebesgue measure, then a version of $(F_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Delta_d}$ has densities with respect to the one-dimensional Lebesgue measure, given by

$$f_{\boldsymbol{\theta}}(w) = \frac{w^{d-1} f(w\boldsymbol{\theta})}{\int_0^\infty u^{d-1} f(u\boldsymbol{\theta})\, du}, \; w > 0. \tag{2.26}$$

**Proposition 2.5.16.** *Suppose $\mathbf{X}$ is a random vector on $\mathbb{R}^d$ with distribution $F$, supported by $[0,\infty)^d$, such that $P(\mathbf{X} = \mathbf{0}) = 0$. Suppose the marginal distributions $F_i$, $i = 1, \ldots, d$ have dominated varying tails as in* (2.5)*. Further, assume that there is a set of regular conditional distributions $(F_{\boldsymbol{\theta}})_{\boldsymbol{\theta}\in\Delta_d}$ of $W$ given $I$ such that $F_{\boldsymbol{\theta}} \in \mathscr{L}$ for each $\boldsymbol{\theta} \in \Delta_d$ and for some $C, t_0 > 0$,*

$$\frac{\overline{F_{\boldsymbol{\theta}_1}}(2t)}{\overline{F_{\boldsymbol{\theta}_2}}(t)} \leq C \tag{2.27}$$

*for all $t > t_0$ and for all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Delta_d$. Then $F \in \mathscr{S}_{\mathcal{R}}$.*

*Proof.* Let $A \in \mathcal{R}$ be fixed. We first check that if each of the marginal distributions have dominated varying tails, then $F_A$ also has a dominated varying tail. For simplicity we assume that $d = 2$, the proof for higher dimensions is similar. We first take $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ with $a_1, a_2, b_1, b_2 > 0$, such that

$[0, a_1) \times [0, a_2) \subset A^c \cap [0, \infty)^2 \subset [0, b_1) \times [0, b_2)$. Since $F$ is supported by $[0, \infty)^2$, it follows that for $t > 0$, $\overline{F}(t\mathbf{a}) \geq F(tA) = \overline{F_A}(t) \geq \overline{F}(t\mathbf{b})$. Therefore

$$\liminf_{t\to\infty} \frac{\overline{F_A}(2t)}{\overline{F_A}(t)} \geq \liminf_{t\to\infty} \frac{\overline{F}(2t\mathbf{b})}{\overline{F}(t\mathbf{a})} \geq \liminf_{t\to\infty} \frac{\frac{1}{2}\overline{F_1}(2tb_1) + \frac{1}{2}\overline{F_2}(2tb_2)}{\overline{F_1}(ta_1) + \overline{F_2}(ta_2)} > 0,$$

since $F_1$ and $F_2$ have dominated varying tails. By definition $F_A$ also has a dominated varying tail.

Since $\mathscr{L} \cap \mathscr{D} \subset \mathscr{S}$, it now suffices to show that $F_A \in \mathscr{L}$.

For $\boldsymbol{\theta} \in \Delta_d$, let

$$h_{\boldsymbol{\theta}} = \inf\{w > 0 : \, w\boldsymbol{\theta} \in A\} > 0\,, \tag{2.28}$$

Note that $h_{\boldsymbol{\theta}}$ is bounded away from 0. Further, by convexity of $A^c$, $h(\mathbf{e}^{(i)}) < \infty$ for at least one coordinate vector $\mathbf{e}^{(i)}$, $i = 1, \ldots, d$. Since the dominated variation of the marginal tails implies, in particular, that each coordinate of the vector $\mathbf{X}$ is positive with positive probability, we conclude that

$$\nu\big\{\boldsymbol{\theta} \in \Delta_d : \, h_{\boldsymbol{\theta}} < \infty\big\} > 0\,.$$

We conclude that there is $M > 0$ and a measurable set $B \subset \Delta_d$ with $\delta := \nu(B) > 0$, such that

$$1/M \leq h_{\boldsymbol{\theta}} \leq M \;\text{ for all } \boldsymbol{\theta} \in B.$$

Note that for $t > 0$,

$$\overline{F}_A(t) = \int_{\Delta_d} \overline{F_{\boldsymbol{\theta}}}(th_{\boldsymbol{\theta}})\, \nu(d\boldsymbol{\theta}). \tag{2.29}$$

Therefore,

$$\frac{\overline{F_A(t)} - \overline{F_A(t+1)}}{\overline{F_A}(t)} = \frac{\int_{\Delta_d} (\overline{F_{\boldsymbol{\theta}}}(th_{\boldsymbol{\theta}}) - \overline{F_{\boldsymbol{\theta}}}((t+1)h_{\boldsymbol{\theta}}))\, \nu(d\boldsymbol{\theta})}{\int_{\Delta_d} \overline{F_{\boldsymbol{\theta}}}(th_{\boldsymbol{\theta}})\, \nu(d\boldsymbol{\theta})}\,, \tag{2.30}$$

and we wish to show that this quantity goes to $0$ as $t \to \infty$.

By the assumptions, for any fixed $\boldsymbol{\theta}$, $F_{\boldsymbol{\theta}} \in \mathscr{L}$, hence for any fixed $\boldsymbol{\theta}$ such that $h_{\boldsymbol{\theta}} < \infty$,

$$\lim_{t\to\infty} \frac{\overline{F_{\boldsymbol{\theta}}}(th_{\boldsymbol{\theta}}) - \overline{F_{\boldsymbol{\theta}}}((t+1)h_{\boldsymbol{\theta}})}{\overline{F_{\boldsymbol{\theta}}}(th_{\boldsymbol{\theta}})} = 0.$$

Therefore, for a given $\epsilon > 0$, there exists $t_\epsilon > 0$ such that, for all $t > t_\epsilon$, $\nu(S_{t,\epsilon}) < \epsilon$, where

$$S_{t,\epsilon} = \left\{ \boldsymbol{\theta} \in \Delta_d : \, h_{\boldsymbol{\theta}} < \infty \text{ and } \frac{\overline{F_{\boldsymbol{\theta}}}(th_{\boldsymbol{\theta}}) - \overline{F_{\boldsymbol{\theta}}}((t+1)h_{\boldsymbol{\theta}})}{\overline{F_{\boldsymbol{\theta}}}(th_{\boldsymbol{\theta}})} > \epsilon \right\}.$$

Let $\epsilon < (\delta/2)^2$. Then $\nu\big(B \cap S_{t,\epsilon}^c\big) > \big(\nu(S_{t,\epsilon})\big)^{1/2}$. By the definition of the set $B$ and by (2.27), and recalling that $h_{\boldsymbol{\theta}} > 0$ for any $\boldsymbol{\theta} \in \Delta_d$, for some $C_1,\, \tilde{t}_0 > 0$

$$\overline{F_{\boldsymbol{\theta}_1}}(th_{\boldsymbol{\theta}_1}) \le C_1 \overline{F_{\boldsymbol{\theta}_2}}(th_{\boldsymbol{\theta}_2})$$

for any $\boldsymbol{\theta}_1 \in S_{t,\epsilon}$ and any $\boldsymbol{\theta}_2 \in B \cap S_{t,\epsilon}^c$, for all $t > \tilde{t}_0$. Therefore, for $t > t_\epsilon + \tilde{t}_0$,

$$\begin{aligned}
\int_{S_{t,\epsilon}} \overline{F_{\boldsymbol{\theta}}}(th_{\boldsymbol{\theta}}) - \overline{F_{\boldsymbol{\theta}}}((t+1)h_{\boldsymbol{\theta}})\, \nu(d\boldsymbol{\theta}) &\le \int_{S_{t,\epsilon}} \overline{F_{\boldsymbol{\theta}}}(th_{\boldsymbol{\theta}})\, \nu(d\boldsymbol{\theta}) \\
&< \frac{\nu(S_{t,\epsilon})}{\nu(B \cap S_{t,\epsilon}^c)} C_1 \int_{B\cap S_{t,\epsilon}^c} \overline{F_{\boldsymbol{\theta}}}(th_{\boldsymbol{\theta}})\, \nu(d\boldsymbol{\theta}) \\
&< \epsilon^{1/2} C_1 \int_{\Delta_d} \overline{F_{\boldsymbol{\theta}}}(th_{\boldsymbol{\theta}})\, \nu(d\boldsymbol{\theta}).
\end{aligned}$$

Hence, for $t > t_\epsilon + \tilde{t}_0$, the quantity in (2.30) is bounded above by $\epsilon + \epsilon^{1/2} C_1$. Letting $\epsilon \searrow 0$ gives us the desired result. □

We are now ready to give an example showing that $\mathscr{S}_{\mathcal{R}} \not\subset \mathscr{S}(\nu; \mathbf{b})$.

**Example 2.5.17.** Let $0 < |\gamma| \le 1/12$. It is shown in [9] that a legitimate probability distribution $F$, supported by the positive quadrant of $\mathbb{R}^2$, satisfies

$$P(X > x, Y > y) = \frac{1 + \gamma \sin(\log(1+x+y)) \cos(\frac{1}{2}\pi \frac{x-y}{1+x+y})}{1+x+y}, \; x, y \ge 0\,. \tag{2.31}$$

Then

$$P(X > x) = P(Y > x) \sim x^{-1} \text{ as } x \to \infty\,,$$

but $F \notin \mathscr{S}(\nu; \mathbf{b})$; see [9]. Straightforward differentiation gives us the density $f$ of $F$, and one can check that it satisfies

$$\frac{2 - 4\gamma - 3\gamma\pi - \pi^2/4}{(1+x+y)^3} \leq f(x,y) \leq \frac{2 + 4\gamma + 3\gamma\pi + \pi^2/4}{(1+x+y)^3}\,,$$

so by (2.26), we have

$$a\frac{w}{(1+w)^3} \leq f_{\boldsymbol{\theta}}(w) \leq b\frac{w}{(1+w)^3},\ w > 0,$$

for some $0 < a < b < \infty$, independent of $\boldsymbol{\theta}$. It is clear that the conditions of Proposition 2.5.16 are satisfied and, hence, $F \in \mathscr{S}_{\mathcal{R}}$.

Proposition 2.5.16 gives us a way to check that a multivariate distribution belongs to the class $\mathscr{S}_{\mathcal{R}}$, but it only applies to distributions that have, marginally, dominated varying tails. In the remainder of this section we provide sufficient conditions for membership in $\mathscr{S}_{\mathcal{R}}$ that do not require marginals with dominated varying tails. We start with a motivating example.

**Example 2.5.18.** [Rotationally invariant case] Assume that there is a one-dimensional distribution $G$ such that $\overline{F_{\boldsymbol{\theta}}} = \overline{G}$ for all $\boldsymbol{\theta} \in \Delta_d$. Let $A \in \mathcal{R}$, and notice that, in the rotationally invariant case, a random variable $Y_A$ with distribution $F_A$ can be written, in law, as

$$Y_A \stackrel{d}{=} ZH^{-1}, \tag{2.32}$$

with $Z$ and $H$ being independent, $Z$ with the distribution $G$, and $H = h_\Theta$. Here $h$ is defined by (2.28), and $\Theta$ has the law $\nu$ over the simplex $\Delta_d$. Recall that the function $h$ is bounded away from zero, so that the random variable $H^{-1}$ is bounded. If $G \in \mathscr{S}$, then the product in the right hand side is subexponential by Corollary 2.5 in [10]. Hence $F_A \in \mathscr{S}$ for all $A \in \mathcal{R}$, and so $F \in \mathscr{S}_{\mathcal{R}}$.

The rotationally invariant case of Example 2.5.18 can be slightly extended, without much effort, to the case where there is a bounded positive function $\left(a_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Delta_d\right)$ such that $F_{\boldsymbol{\theta}}(\cdot) = G(\cdot/a_{\boldsymbol{\theta}})$ for some $G \in \mathscr{S}$. An argument similar to the one in the example shows that we can still conclude that $F \in \mathscr{S}_{\mathcal{R}}$. In order to achieve more than that, we note that the distribution $F_A$ can be represented, by (2.29), as a mixture of scaled regular conditional distributions. Note also that the product of independent random variables in (2.32) is just a special case of that mixture, to which we have been able to apply Corollary 2.5 in [10]. It is likely to be possible to extend that result to certain mixtures that are more general than products of independent random variables, and thus to obtain additional criteria for membership in the class $\mathscr{S}_{\mathcal{R}}$. We leave serious extensions of this type to future work. A small extension that still steps away from exact products is below, and it takes a result in [10] as an ingredient. We formulate the statement in terms of the distribution of a random variable that only in a certain asymptotic sense looks like a product of independent random variables.

**Theorem 2.5.19.** *Let $(\Omega_i, \mathcal{F}_i, P_i)$, $i = 1, 2$ be probability spaces. Let $Q$ be a random variable defined on the product probability space. Assume that there are nonnegative random variables $X_i$, $i = 1, 2$, defined on $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ correspondingly, such that $X_1$ has a subexponential distribution $F$, and for some $t_0 > 0$ and $C > 0$,*

$$X_1(\omega_1)X_2(\omega_2) - CX_2(\omega_2) \le Q(\omega_1, \omega_2) \le X_1(\omega_1)X_2(\omega_2) + CX_2(\omega_2) \tag{2.33}$$

*a.s. on the set $\{Q(\omega_1, \omega_2) > t_0\}$. Suppose $P(X_2 > 0) > 0$, and let $G$ be the distribution of $X_2$. Suppose that there is a function $a : (0, \infty) \to (0, \infty)$, such that*

1. *$a(t) \nearrow \infty$ as $t \to \infty$;*
2. *$\frac{t}{a(t)} \nearrow \infty$ as $t \to \infty$;*

3. $\lim_{t\to\infty} \frac{\overline{F}(t-a(t))}{\overline{F}(t)} = 1$;

4. $\lim_{t\to\infty} \frac{\overline{G}(a(t))}{P(X_1X_2>t)} = 0$.

*Then $Q$ has a subexponential distribution.*

*Proof.* Let $H$ denote the distribution of $X_1X_2$. It follows by Theorem 2.1 in [10] that $H$ is subexponential. We show that $P(Q > t) \sim \overline{H}(t)$ as $t \to \infty$. This will imply that $Q$ has a subexponential distribution.

We start by checking that

$$\lim_{t\to\infty} \frac{\overline{H}(t-a(t))}{\overline{H}(t)} = 1, \text{ which will imply that } \lim_{t\to\infty} \frac{\overline{H}(t+a(t))}{\overline{H}(t)} = 1 \qquad (2.34)$$

when we substitute $t + a(t)$ for $t$ in the limit, since $a$ is increasing. To verify the limit, suppose first that $X_2 \geq 1$ a.s., and write

$$P\big(t - a(t) < X_1X_2 \leq t\big) \leq P_2(X_2 > a(t))$$
$$+ \int_{\Omega_2} P_1(t/X_2(\omega_2) - a(t)/X_2(\omega_2) < X_1 \leq t/X_2(\omega_2))\, \mathbb{1}\big(X_2(\omega_2) \leq a(t)\big)\, P_2(d\omega_2)$$

The first term in the right hand side is $o(\overline{H}(t))$ by the assumption (4), while the same is true for the second term by the assumption (3), since by the assumption (2), $a(t)/y \leq a(t/y)$ if $y \geq 1$. This proves (2.34) if $X_2 \geq 1$ a.s. and hence, by scaling, if $X_2 \geq \epsilon$ a.s. for some $\epsilon > 0$. An elementary truncation argument then shows that (2.34) holds if $P(X_2 > 0) > 0$.

Note that for $t > t_0$,

$$\begin{aligned} P(Q > t) &\leq P(X_1X_2 + CX_2 > t) \\ &\leq \overline{G}(a(t)) + \overline{H}(t - Ca(t))\,. \end{aligned}$$

This implies that $\limsup_{t\to\infty} P(Q > t)/\overline{H}(t) \leq 1$. The statement $\liminf_{t\to\infty} P(Q > t)/\overline{H}(t) \geq 1$ can be shown in a similar way. □

Despite a limited scope of the extension given in Theorem 2.5.19, it allows one to construct a number of examples of multivariate distributions in $\mathscr{S}_{\mathcal{R}}$ by choosing, for example, $\Omega_2 = \Delta_d$ and $X_2(\boldsymbol{\theta}) = 1/h(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Delta_d$, and selecting a function $Q$ to model additional randomness in the radial direction.

## 2.6 Ruin Probabilities

We circle back to the ruin problem. As mentioned in the introduction, the notion of subexponentiality we introduced in Section 2.5 was designed with insurance applications in mind. In this section we describe such an application more explicitly.

Consider the following extension to the ruin problem in Section 2.2, a renewal model for the reserves of an insurance company with $d$ lines of business. Suppose that claims arrive according to a renewal process $(N_t)_{t\geq 0}$ given by $N_t = \sup\{n \geq 1 : T_n \leq t\}$. The arrival times $(T_n)$ form a renewal sequence

$$T_0 = 0, \quad T_n = Y_1 + \cdots + Y_n \text{ for } n \geq 1, \tag{2.35}$$

where the interarrival times $(Y_i)_{i\geq 1}$ form a sequence of independent and identically distributed positive random variables. We will call a generic interarrival time $Y$. At the arrival time $T_i$ a random vector-valued claim size $\mathbf{X}^{(i)} = \left(X_1^{(i)}, \ldots, X_d^{(i)}\right)$ is incurred, so that the part of the claim going to the $j$th line of business is $X_j^{(i)}$. We assume that the claim sizes $(\mathbf{X}^{(i)})$ are i.i.d. random vectors with a finite mean, and we denote their common law by $F$. We assume further that the claim size process is independent of the renewal process of the claim arrivals. The $j$th line of business collects premium at the rate of $p^j$ per unit of

time. Let $\mathbf{p}$ be the vector of the premium rates, and $\mathbf{X}$ a generic random vector of claim sizes.

Suppose that the company has an initial buffer capital of $u$, out of which the amount of $ub_j$ is allocated to the $j$th line of business, $j = 1, 2, \ldots, d$. Here $b_1, \ldots, b_d$ are positive numbers, $b_1 + \cdots + b_d = 1$. Then $u\mathbf{b}$ denotes the vector for the initial capital buffer allocation. With the above notation, the claim surplus process $(\mathbf{S}_t)_{t\geq 0}$ and the risk reserve process $(\mathbf{R}_t)_{t\geq 0}$ are given by

$$\mathbf{S}_t = \sum_{i=1}^{N_t} \mathbf{X}^{(i)} - t\mathbf{p}, \quad \mathbf{R}_t = u\mathbf{b} - \mathbf{S}_t = u\mathbf{b} + t\mathbf{p} - \sum_{i=1}^{N_t} \mathbf{X}^{(i)},\ t \geq 0\,.$$

The company becomes insolvent (ruined) when the risk reserve process hits a certain ruin set $L \subset \mathbb{R}^d$. Equivalently, ruin occurs when the claim surplus process enters the set $u\mathbf{b} - L$. We will assume that the ruin set satisfies the following condition.

**Assumption 2.6.1.** The ruin set $L$ is an open decreasing ($-L$ is increasing) set such that $\mathbf{0} \in \partial L$, satisfying $L = uL$ for $u > 0$, and such that $L^c$ is convex.

Note that this assumption means that the ruin occurs when the claim surplus process enters the set $uA$, with $A = \mathbf{b} - L \in \mathcal{R}$, as defined in Section 2.5. In fact, the ruin set $L$ can be viewed as being of the form $-G$, as defined in Remark 2.5.2. Examples of such ruin sets are, of course, the sets

$$L = \left\{\mathbf{x} :\ x_j < 0 \text{ for some } j = 1, \ldots, d\right\} \text{ and } L = \left\{\mathbf{x} :\ x_1 + \ldots + x_d < 0\right\},$$

discussed in Section 2.5. A general framework was proposed in [34]. In this framework capital can be transferred between different business lines, but the transfers incur costs, and the solvency set has the form

$$L^c = \left\{\mathbf{x} :\ \mathbf{x} = \sum_{i\neq j} v_{ij}(\pi_{ij}\mathbf{e}^i - \mathbf{e}^j) + \sum_{i=1}^{d} w_i\mathbf{e}^i,\ v_{ij} \geq 0,\ w_i \geq 0\right\}, \tag{2.36}$$

where $\mathbf{e}^1, \ldots, \mathbf{e}^d$ are the standard basis vectors, and $\Pi = (\pi_{ij})_{i,j=1}^d$ is a matrix satisfying

(i) $\pi_{ij} \geq 1$ for $i, j \in \{1, \ldots, d\}$,

(ii) $\pi_{ii} = 1$ for $i \in \{1, \ldots, d\}$,

(iii) $\pi_{ij} \leq \pi_{ik}\pi_{kj}$ for $i, j, k \in \{1, \ldots, d\}$.

In the financial literature, a matrix satisfying the above constraints is called a bid-ask matrix. In our context, the entry $\pi_{ij}$ can be interpreted as the amount of capital that needs to be taken from business line $i$ in order to transfer $1$ unit of capital to business line $j$.

We note that each of the above ruin sets is a cone, i.e. it satisfies $L = uL$ for $u > 0$, as assumed in Assumption 2.6.1.

We maintain the notation $A = \mathbf{b} - L \in \mathcal{R}$. Note that we can write the ruin probability as

$$\begin{aligned} \psi_{\mathbf{b},L}(u) &= P(\mathbf{R}_t \in L \text{ for some } t \geq 0) \\ &= P\left(\sum_{i=1}^{n} \mathbf{X}^{(i)} - Y_i\mathbf{p} \in uA \text{ for some } n \geq 1\right) \\ &= P\left(\sum_{i=1}^{n} \mathbf{Z}^{(i)} \in uA \text{ for some } n \geq 1\right), \end{aligned} \tag{2.37}$$

where $\mathbf{Z}^{(i)} = \mathbf{X}^{(i)} - Y_i\mathbf{p}$, $i = 1, 2, \ldots$. We let $\mathbf{Z}$ denote a generic element of the sequence $(\mathbf{Z}^{(i)})_{i\geq 1}$. We will assume a positive safety loading, an assumption that takes now the form

$$\mathbf{c} = -\mathbb{E}[\mathbf{Z}] > \mathbf{0}\,,$$

see e.g. [1]. The assumption of the finite mean for the claim sizes implies that

$$\theta := \int_0^\infty F\Big([0,\infty)^d + v\mathbf{c}\Big)\, dv < \infty\,,$$

and we can define a probability measure on $\mathbb{R}^d$, supported by $[0,\infty)^d$, by

$$F^I(\cdot) = \frac{1}{\theta}\int_0^\infty F(\cdot + v\mathbf{c})\,dv\,. \tag{2.38}$$

Denote

$$H(u) = \int_0^\infty F(uA + v\mathbf{c})\,dv\,,\ u > 0\,. \tag{2.39}$$

The following is the main result of this section.

**Theorem 2.6.2.** *Suppose that the law $F^I$ is in $\mathscr{S}_A$. Then the ruin probability $\psi_{\mathbf{b},L}$ satisfies*

$$\lim_{u\to\infty}\frac{\psi_{\mathbf{b},L}(u)}{H(u)} = 1. \tag{2.40}$$

**Remark 2.6.3.** Notice, for comparison, that in the univariate case, with the ruin set $L = (-\infty, 0)$ (and $b = 1$) we have $A = (1,\infty)$, and

$$H(u) = \int_0^\infty \overline{F}(u + vc)\,dv = \frac{1}{c}\int_u^\infty \overline{F}(v)\,dv\,.$$

In this case the statement (2.40) agrees with the standard univariate result on subexponential claims; see e.g. Theorem 1.3.8 in [21]. If the claim arrival process is Poisson, then this is (2.2) of Section 2.3.

*Proof of Theorem 2.6.2.* We start by observing that the function $H$ is proportional to the tail of a subexponential distribution $F_A^I$ and, hence, can itself be viewed as the tail of a subexponential distribution. We can and will, for example, simply refer to the "long tail property" of $H$.

We use the "one big jump" approach to heavy tailed large deviations; see e.g. [61], and the first step is to show that

$$\lim_{u\to\infty}\frac{\int_0^\infty P(\mathbf{Z}\in uA + v\mathbf{c})\,dv}{H(u)} = 1\,. \tag{2.41}$$

Indeed, the upper bound in (2.41) follows from the fact that $A$ is increasing. For the lower bound, notice that, by Fatou's lemma, it is enough to prove that that for each fixed $\mathbf{y}$,

$$\lim_{u\to\infty} \frac{\int_0^\infty F(uA + v\mathbf{c} + y\mathbf{p})\, dv}{H(u)} = 1\,.$$

This, however, follows from the fact for sufficiently large $u > 0$, there exists some $u_1 > 0$ such that $(u + u_1)A + v\mathbf{c} \subset uA + v\mathbf{c} + y\mathbf{p}$, and the long tail property of $H$.

We proceed to prove the lower bound in (2.40). Let $\mathbf{S}_n := \sum_{i=1}^n \mathbf{Z}^{(i)}$, $n = 1, 2, \ldots$. Let $\epsilon, \delta$ be small positive numbers, by the Weak Law of Large Numbers, we can choose $K = K_{\epsilon,\delta}$ so large that

$$P\big(\mathbf{S}_n > -(K + n(1+\epsilon))\mathbf{c}\big) > 1 - \delta,\ n = 1, 2, \ldots\,.$$

Define $M_n = \sup\{u > 0 :\ \mathbf{S}_i \in uA \text{ for some } 1 \le i \le n\}$ and $M = \sup\{u > 0 :\ \mathbf{S}_n \in uA \text{ for some } n\}$. For $u > 0$,

$$\begin{aligned}
\psi_{\mathbf{b},L}(u) = P(M > u) &= \sum_{n\ge 0} P(M_n \le u,\, \mathbf{S}_{n+1} \in uA)\\
&\ge \sum_{n\ge 0} P\big(M_n \le u,\, \mathbf{S}_n > -(K + n(1+\epsilon))\mathbf{c},\, \mathbf{Z}^{(n+1)} \in uA + (K + n(1+\epsilon))\mathbf{c}\big)\\
&\ge \sum_{n\ge 0} (1 - \delta - P(M_n > u)) P\big(\mathbf{Z}^{(n+1)} \in uA + K\mathbf{c} + n(1+\epsilon)\mathbf{c}\big)\\
&\ge (1 - \delta - P(M > u)) \sum_{n\ge 0} P\big(\mathbf{Z} \in uA + K\mathbf{c} + n(1+\epsilon)\mathbf{c}\big)\,.
\end{aligned}$$

Rearranging, and using the monotonicity of $A$, we see that

$$\begin{aligned}
\psi_{\mathbf{b},L}(u) &\ge \frac{(1-\delta)\sum_{n\ge 0} P(\mathbf{Z} \in uA + K\mathbf{c} + n(1+\epsilon)\mathbf{c})}{1 + \sum_{n\ge 0} P(\mathbf{Z} \in uA + K\mathbf{c} + n(1+\epsilon)\mathbf{c})}\\
&\sim \frac{1-\delta}{1+\epsilon} \int_0^\infty P\big(\mathbf{Z} \in uA + K\mathbf{c} + v\mathbf{c}\big)\, dv\\
&\sim \frac{1-\delta}{1+\epsilon} \int_0^\infty P\big(\mathbf{Z} \in uA + v\mathbf{c}\big)\, dv,\ u \to \infty\,,
\end{aligned}$$

where the last step is by the long tail property of $F_A^I$. Letting $\delta, \epsilon$ to 0, we have, thus, obtained the lower bound in (2.40). We proceed to prove a matching upper bound.

Fix $0 < \epsilon < 1$. For $r > 0$, we define a sequence $(\tau_n)$ as follows: we set $\tau_0 = 0$, and

$$\tau_1 = \inf\{n \geq 1 : \mathbf{S}_n \in rA - n(1-\epsilon)\mathbf{c}\}.$$

For $m \geq 2$, we set $\tau_m = \infty$ if $\tau_{m-1} = \infty$. Otherwise, let

$$\tau_m = \tau_{m-1} + \inf\{n \geq 1 : \mathbf{S}_{n+\tau_{m-1}} - \mathbf{S}_{\tau_{m-1}} \in rA - n(1-\epsilon)\mathbf{c}\}.$$

If we let $\gamma = P(\tau_1 < \infty)$, then for any $m \geq 1$, $P(\tau_m < \infty) = \gamma^m$. By the positive safety loading assumption, $\gamma \to 0$ as $r \to \infty$. Note that for $u > 0$,

$$\begin{aligned} P\big(\tau_1 < \infty,\, \mathbf{S}_{\tau_1} \in uA\big) &= \sum_{n\geq 1} P(\tau_1 = n,\, \mathbf{S}_n \in uA) \\ &\leq \sum_{n\geq 1} P\big(\mathbf{S}_{n-1} \in rA^c - (n-1)(1-\epsilon)\mathbf{c},\, \mathbf{S}_n \in uA\big)\,. \end{aligned}$$

By part (c) of Lemma 2.5.3, $\mathbf{S}_n \in uA$ if and only if $\sup_{\mathbf{p}\in I_A} \mathbf{p}^T\mathbf{S}_n > u$. Further,

$$\sup_{\mathbf{p}\in I_A} \mathbf{p}^T\mathbf{S}_n \leq \sup_{\mathbf{p}\in I_A} \mathbf{p}^T\big(\mathbf{S}_{n-1} + (n-1)(1-\epsilon)\mathbf{c}\big) + \sup_{\mathbf{p}\in I_A} \mathbf{p}^T\big(\mathbf{Z}^{(n)} - (n-1)(1-\epsilon)\mathbf{c}\big).$$

Let $u > r$. Recalling Lemma 2.5.8, if $\mathbf{S}_{n-1} \in rA^c - (n-1)(1-\epsilon)\mathbf{c}$, then $\sup_{\mathbf{p}\in I_A} \mathbf{p}^T\big(\mathbf{S}_{n-1} + (n-1)(1-\epsilon)\mathbf{c}\big) \leq r$, so for $\sup_{\mathbf{p}\in I_A} \mathbf{p}^T\mathbf{S}_n > u$ to hold, it must be the case that $\sup_{\mathbf{p}\in I_A} \mathbf{p}^T\big(\mathbf{Z}^{(n)} - (n-1)(1-\epsilon)\mathbf{c}\big) > u - r$, implying that $\mathbf{Z}^{(n)} \in (u-r)A + (n-1)(1-\epsilon)\mathbf{c}$.

Summing up, we see that, as $u \to \infty$,

$$\begin{aligned} P\big(\tau_1 < \infty,\, \mathbf{S}_{\tau_1} \in uA\big) &\leq \sum_{n\geq 1} P\big(\mathbf{Z}^{(n)} \in (u-r)A + (n-1)(1-\epsilon)\mathbf{c}\big) \\ &\sim \int_0^\infty P(\mathbf{Z} \in (u-r)A + v(1-\epsilon)\mathbf{c})\, dv \\ &\sim \frac{1}{1-\epsilon} H(u-r)\,. \end{aligned}$$

Letting $\epsilon \to 0$ and using the long tail property of $H$, we obtain

$$\limsup_{u\to\infty} \frac{P\big(\tau_1 < \infty,\, \mathbf{S}_{\tau_1} \in uA\big)}{H(u)} \le 1\,. \tag{2.42}$$

Let $(\mathbf{V}^{(i)})$ be a sequence of independent identically distributed random vectors whose law is the conditional law of $\mathbf{S}_{\tau_1}$ given that $\tau_1 < \infty$. By (2.42), there is a distribution $B$ on $[0,\infty)$ such that $\overline{B}(u) \sim \gamma^{-1}H(u)$ as $u \to \infty$ and

$$P\big(\mathbf{V}^{(1)} \in uA\big) \le \overline{B}(u) \ \text{ for all } u \ge 0.$$

Note, further, that by the definition of the sequence $(\tau_m)$, for every $m \ge 0$, on the event $\{\tau_m < \infty\}$, we have, for $1 \le i < \tau_{m+1}$, $\mathbf{S}_{\tau_m+i} - \mathbf{S}_{\tau_m} \in rA^c - i(1-\epsilon)\mathbf{c} \subset rA^c - (1-\epsilon)\mathbf{c}$. If $\mathbf{S}_{\tau_m} \in (u-r)A^c + (1-\epsilon)\mathbf{c}$, then we have $\mathbf{S}_{\tau_m+i} \in uA^c$. Hence, for the event $\{\mathbf{S}_n \in uA \text{ for some } n\}$ to occur, we must have

$$\mathbf{S}_{\tau_m} \in \big((u-r)A + (1-\epsilon)\mathbf{c}\big) \cup uA \ \text{ for some } m.$$

Therefore, noting that $\big((u-r)A + (1-\epsilon)\mathbf{c}\big) \cup uA \subset (u-r)A$, we can use Lemma 2.5.9 to obtain

$$\begin{aligned}
\psi_{\mathbf{b},L}(u) = P(M > u) &\le \sum_{m\ge1} P\big(\mathbf{S}_{\tau_m} \in \big((u-r)A + (1-\epsilon)\mathbf{c}\big) \cup uA\big) \\
&\le \sum_{m\ge1} \gamma^m P\big(\mathbf{V}^{(1)} + \cdots + \mathbf{V}^{(m)} \in (u-r)A\big) \\
&\le \sum_{m\ge1} \gamma^m \overline{B^{(m)}}(u-r).
\end{aligned}$$

By the assumption, the $H$ is the tail of a subexponential distribution, and, hence, $B$ is subexponential as well. This implies that

$$\lim_{u\to\infty} \frac{\overline{B^{(m)}}(u)}{\overline{B}(u)} = m\,,$$

and that for any $\epsilon > 0$, there exists $K > 0$ such that for all $u > 0$ and $m \ge 1$,

$$\frac{\overline{B^{(m)}}(u)}{\overline{B}(u)} \le K(1+\epsilon)^m.$$

Since we can make $\gamma > 0$ as small as we wish by choosing $u$ and $r$ large, we can use the dominated convergence theorem to obtain

$$\limsup_{u\to\infty} \frac{\psi_{\mathbf{b},L}(u)}{\gamma \overline{B}(u-r)} = \sum_{m\geq 1} \gamma^{m-1} m = \frac{1}{(1-\gamma)^2}.$$

Letting $r \to \infty$, which makes $\gamma \to 0$, we have that

$$\limsup_{u\to\infty} \frac{\psi_{\mathbf{b},L}(u)}{H(u)} \leq 1\,,$$

which is the required upper bound in (2.40). □

We finish this section by returning to the special case of multivariate regularly varying claims. Recall that, by Proposition 2.5.15, the distributions in $MRV(\alpha, \mu)$ are in $\mathscr{S}_{\mathcal{R}}$. The asymptotic behaviour of the ruin probability with the solvency set $L^c$ given by (2.36), and multivariate regularly varying claims with $\alpha > 1$, was determined by [34]. To state their result, notice that the tail measure of a random vector $\mathbf{X}$ (recall (2.23)) is determined up to a scaling by a positive constant, and a different scaling in the tail measure can be achieved by scaling appropriately the function $b$ in (2.23). Let us scale the tail measure $\mu$ in such a way that it assigns unit mass to the complement of the unit ball in $\mathbb{R}^d$. The norm we choose is unimportant, but for consistency with the notation used elsewhere in the paper, let us use the $L_1$ norm. With this convention, we can restate (2.23) as

$$\frac{P(\mathbf{X} \in u\cdot)}{P(\|\mathbf{X}\| > u)} \xrightarrow{v} \mu \tag{2.43}$$

vaguely on $[-\infty, \infty]^d \setminus \{\mathbf{0}\}$. It was shown by [34] that under the assumption (2.43) (and with the solvency set $L^c$ given by (2.36)), the ruin probability satisfies

$$\lim_{u\to\infty} \frac{\psi_{\mathbf{b},L}(u)}{uP(\|\mathbf{X}\| > u)} = \int_0^\infty \mu(\mathbf{b} - L + v\mathbf{c})\, dv. \tag{2.44}$$

We extend the above result to all ruin sets satisfying Assumption 2.6.1. To avoid a degenerate situation (and the resulting complications in the notation) we will assume that $\mu\{\mathbf{x} : x_i > 0\} > 0$ for each $i = 1, \dots, d$.

**Proposition 2.6.4.** *Assume that the ruin set $L$ satisfies Assumption 2.6.1. If the claim sizes satisfy* (2.43) *with $\alpha > 1$, then* (2.44) *holds.*

*Proof.* By Theorem 2.6.2, it suffices to show that

$$\lim_{u\to\infty} \frac{\int_0^\infty P(\mathbf{X} \in uA + v\mathbf{c})\, dv}{uP(\|\mathbf{X}\| > u)} = \int_0^\infty \mu(A + v\mathbf{c})\, dv\,,$$

which we proceed to do. By a change of variables,

$$\frac{\int_0^\infty P(\mathbf{X} \in uA + v\mathbf{c})\, dv}{uP(\|\mathbf{X}\| > u)} = \frac{\int_0^\infty P(\mathbf{X} \in u(A + v\mathbf{c}))\, dv}{P(\|\mathbf{X}\| > u)}\,, \tag{2.45}$$

and for every $v > 0$,

$$\frac{P(\mathbf{X} \in u(A + v\mathbf{c}))}{P(\|\mathbf{X}\| > u)} \to \mu(A + v\mathbf{c})$$

as $u \to \infty$. In the last step we use (2.43), and the fact that the tail measure does not charge the boundary of sets in $\mathcal{R}$, shown in the proof of Proposition 2.5.15. Therefore, we only need to justify taking the limit inside the integral in (2.45). However, by the definition of the set $A$,

$$\frac{P(\mathbf{X} \in u(A + v\mathbf{c}))}{P(\|\mathbf{X}\| > u)} \le \sum_{i=1}^{d} \frac{P(X^{(i)} > ub_i + uvc_i)}{P(X^{(i)} > u)}\,.$$

The non-degeneracy assumption on the measure $\mu$ implies that each $X^{(i)}$ is itself regularly varying with exponent $\alpha$. Therefore, by the Potter bounds, there are finite positive constants $C_i$, $i = 1, \dots, d$, and a number $\varepsilon \in (0, \alpha - 1)$ such that for all $u \ge 1$,

$$\frac{P(X^{(i)} > ub_i + uvc_i)}{P(X^{(i)} > u)} \le C_i(b_i + vc_i)^{-(\alpha-\varepsilon)},\ i = 1, \dots, d\,.$$

Since the functions in the right hand side are integrable, the dominated convergence theorem applies. □

CHAPTER 3

# ESTIMATING THE EXTREMAL INDEX AND THRESHOLD SELECTION

## 3.1 Introduction

We now move our attention to the matter of extremal inference. In this thesis we are mainly concerned with improving and devising techniques involved in extremal inference. A good place to begin is with the matter of threshold selection. Many statistical procedures in extreme value theory depend on a choice of a threshold such that only the observations above that threshold are used for the inference. In the classical Hill estimator of the exponent of regular variation, this corresponds to choosing the number of the upper statistics used to construct the estimator, and in the standard "peaks over threshold" procedures, the term "threshold" even appears in the name; see e.g. [15] and [47]. The inference results often depend on the threshold in a significant way, so a major effort has been invested in choosing the threshold "in the right way"; see e.g. [49], [17], [18], [41]. Many threshold-based extremal inference procedures discard the observations below the threshold, which in most cases amounts to discarding a larger part (indeed, often an overwhelmingly larger part) of the sample. This counterintuitive step reflects the underlying belief that the observations above the threshold carry information about the "tail" of the distribution, while those below the threshold carry information about the "center" of the distribution.

It is reasonable to assume that such a binary rule by necessity neglects a part of the information stored in the original sample that is relevant for extremal inference. An alternative to using a binary rule would be acknowledging that larger observations carry more information about the "extremes" than smaller

observations do, but instead of discarding the latter completely, using them in the extremal inference, with a smaller weight. This idea can be implemented in a number of ways, the most natural of which is to use multiple "thresholds" instead of trying to select the "right" threshold. In this case it is more appropriate to talk about "levels" of observations that are weighted differently, rather than "thresholds".

In this chapter we apply this idea to estimating the extremal index (defined below), but the approach is more general than its application to estimation of the extremal index. It can, in principle, be used in any extremal estimation problem, though the actual implementation may depend significantly on the problem. The chapter is organized as follows. In Section 3.2 we introduce the blocks estimator for the extremal index and observe its shortfalls, followed by motivations for possible improvement. We introduce our improved version of the blocks estimator - the multilevel estimator - using multiple thresholds (levels) and list the assumptions used in the chapter in Section 3.3. Section 3.4 considers the asymptotic behaviour of the various ingredients in our estimator. In Section 3.5 we prove a central limit theorem for the estimator. In Section 3.6 we both propose a procedure to reduce the bias of the estimator as well as present a simulation study and a case study.

## 3.2 The Blocks Estimator

In the proceeding sections we construct a procedure for estimating the *extremal index*, a quantity designed to measure the amount of clustering of the extremes in a stationary sequence. Suppose that $X_1, X_2, \ldots$ is a stationary se-

quence of random variables with a marginal distribution function $F$, and let $M_n = \max(X_1, \ldots, X_n)$, $n = 1, 2, \ldots$. Suppose there exists $\theta \geq 0$ with the following property: for every $\tau > 0$, there is a sequence $(v_n)$ such that $n\overline{F}(v_n) \to \tau$ and $P(M_n \leq v_n) \to e^{-\theta\tau}$ as $n \to \infty$. Then $\theta$ is called the extremal index of the sequence $X_1, X_2, \ldots$; it is automatically in the range $0 \leq \theta \leq 1$; see [21]. The relation of the extremal index to extremal clustering is best observed by considering the exceedances of the stationary sequence over high thresholds. Let $(v_n)$ be a sequence such that $n\overline{F}(v_n) \to \tau$ as $n \to \infty$ for some $\tau > 0$. Then under certain mixing conditions (conditions regulating the dependence relations between elements in the stationary sequence), the point processes of exceedances converge weakly in the space of finite point processes on $[0, 1]$ to a compound Poisson process:

$$N_n = \sum_{i=1}^{n} \delta_{i/n} \mathbb{1}(X_i > v_n) \xrightarrow{d} N = \sum_{i=1}^{\infty} \xi_i \delta_{\Gamma_i}, \tag{3.1}$$

where $\delta_x$ is a point mass at $x$, the points $0 < \Gamma_1 < \Gamma_2 < \ldots$ constitute a homogeneous Poisson process with intensity $\tau\theta$ on $[0, 1]$ which is independent of an i.i.d. positive integer-valued sequence $\{\xi_i\}$; see e.g. [32]. The latter sequence is interpreted as the sequence of the extremal cluster sizes, and the extremal index $\theta$ is, under mild conditions, equal to the reciprocal of the expected cluster size $E\xi$. We will assume that the latter expectation is finite, and the extremal index is positive.

The problem of estimating the extremal index parameter is well-known in literature; references include [31], [56], [22], [42], and [5]. The most common methods of estimation include the blocks method, the runs method, and the inter-exceedance method. We focus on the blocks method in order to demonstrate an application of our idea of using multiple levels of observations.

The blocks method is based on the interpretation of the extremal index as the reciprocal of the expected cluster size of extremes. It is based on choosing a block size $r_n$ much smaller than $n$ and a level (or threshold) $u_n$. Split the $n$ observations $X_1, X_2, \ldots, X_n$ into $k_n = \lfloor n/r_n \rfloor$ contiguous blocks of equal length $r_n$. The blocks estimator is then defined as the reciprocal of average number of exceedances of the level $u_n$ per block among blocks with at least one exceedance. If $M_{i,j}$ denotes $\max\{X_{i+1}, \ldots, X_j\}$ for $i < j$ and $M_j = M_{0,j}$, then the blocks estimator has the form

$$\widehat{\theta}_n = \frac{\sum_{i=1}^{k_n} \mathbb{1}\big(M_{(i-1)r_n, ir_n} > u_n\big)}{\sum_{i=1}^{k_n r_n} \mathbb{1}\big(X_i > u_n\big)} \,. \tag{3.2}$$

Assuming that $r_n \overline{F}(u_n) \to 0$ but $n\overline{F}(u_n) \to \infty$ as $n \to \infty$, and certain mixing conditions, this estimator has been shown to be consistent and asymptotically normal; see [29] and [58].

We observe Monte Carlo approximations to the properties of the blocks estimator to better understand its behavior. We use the ARMAX process for a simulation study. The ARMAX process $(X_i)$ is defined as follows. Let $Z_1, Z_2, \ldots$ be a sequence of i.i.d. unit Fréchet random variables. For $0 < \theta \leq 1$, let $X_1 = Z_1/\theta$, and

$$X_i = \max((1-\theta)X_{i-1}, Z_i), \; i \geq 2 \,. \tag{3.3}$$

It can be shown that the extremal index of such a sequence is $\theta$; see e.g. Chapter 10 of [4].

The blocks estimator (3.2) is applied to $5000$ simulated ARMAX sequences of size $n = 10000$ with $\theta = 0.5$, using a block size of $r_n = 200$. Varying levels of the threshold $u_n$ are used, to include from the largest $100$ order statistics, up to the largest $150$ order statistics of each sequence. The estimates using each level of $u_n$, as well as the accompanying standard errors, are plotted in Figure 3.1.

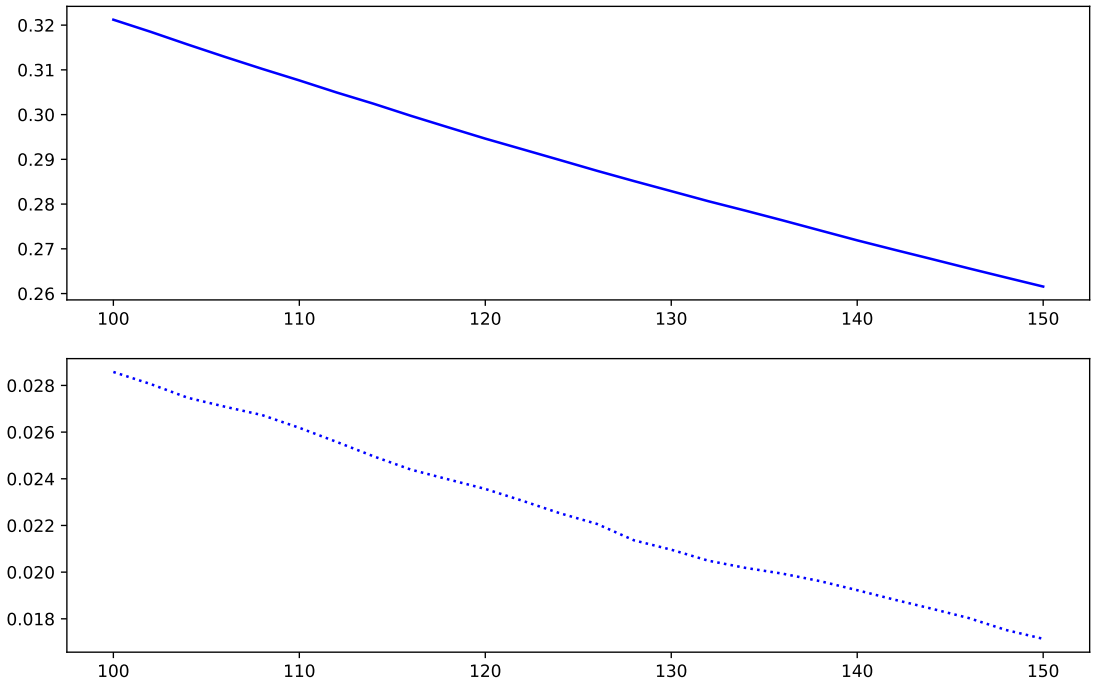


Figure 3.1: Monte Carlo simulation the estimates using the blocks estimator (3.2) (top, solid line) and their associated standard errors (bottom, dotted line) plotted against the number of largest order statistics used in the estimate. Data are simulated from an ARMAX model with $\theta = 0.5$.

From the figure, several observations can be made. One such observation is that there seems to be a large bias associated with the blocks estimator. From the plot of the estimates, even in the best case, there is a bias of around $0.18$ when the true value of $\theta$ is $0.5$. A further observation is that the estimate (as well as the associated standard error) seems to decrease linearly with the number of observations used in the estimator.

We can devise a couple of ideas to improve the blocks estimator using the above observations. Noting that the estimates using fewer observations tend to be closer to the true value (namely having a smaller bias), and yet at the same time, estimates using more observations achieve a smaller variance, we are motivated to therefore somehow combine the estimates using multiple levels of

observations, which can lower the overall variance while keeping still keeping the bias small. Further, we would like to exploit the linearity structure displayed in Figure 3.1 to correct for the bias in the estimator. We will explore both ideas in the remainder of the chapter.

Multiple thresholds in extremal inference - especially in the context of the extremal index - have been used before, such as in [16], [35], and [57]. In [16], for example, estimates of the extremal index based on multiple thresholds were combined together in order to minimize the bias of the estimator. Our idea is different, as we are viewing using multiple levels as a natural means to reduce the variance of an estimator.

## 3.3 The Multilevel Estimator

Let $X_1, \ldots, X_n$ be a stationary sequence of random variables with marginal distribution $F$, and an extremal index $\theta \in (0, 1]$. We now present a version of the blocks estimator (3.2) based on multiple levels. With a block size $r_n$ and the number of blocks $k_n = \lfloor n/r_n \rfloor$ as before, we select now $m$ levels $u_n^1 < \cdots < u_n^m := u_n$, and we view the highest level $u_n^m$ as corresponding to the single level $u_n$ in (3.2). The lower levels $u_n^s$, $s = 1, \ldots, m-1$ are used to reduce the variance of the estimator. The levels are chosen in an "asymptotically balanced" way. Specifically, it will be assumed that, as $n \to \infty$,

$$\frac{\overline{F}(u_n^s)}{\overline{F}(u_n^m)} \to \frac{\tau_s}{\tau_m}, \; s = 1, \ldots, m \tag{3.4}$$

for some $\tau_1 > \cdots > \tau_m > 0$.

Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuously differentiable positive decreasing func-

tion. We will use $f$ as a weight function, and we would like to weigh the exceedances over the level $u_n^s$ by $f(\tau_s/\tau_m)$. The fact that $f$ is decreasing reflects our belief that higher exceedances provide more reliable information about the extremes. We will not assume that the numbers $\tau_1, \ldots, \tau_m$ are known ahead of time, so we will use, in practice, an estimator of the ratio $\tau_s/\tau_m$. Specifically, we will use

$$\widehat{\tau_s/\tau_m} = \frac{\sum_{i=1}^{k_n r_n} \mathbb{1}(X_i > u_n^s)}{\sum_{i=1}^{k_n r_n} \mathbb{1}(X_i > u_n^m)},\ s = 1, \ldots, m\,. \tag{3.5}$$

Then our version of the blocks estimator (3.2) based on multiple levels is

$$\widehat{\theta}_n(f) = \frac{\sum_{s=1}^{m} \left[f\left(\widehat{\tau_s/\tau_m}\right) - f\left(\widehat{\tau_{s-1}/\tau_m}\right)\right] \sum_{i=1}^{k_n} \mathbb{1}\left(M_{(i-1)r_n, ir_n} > u_n^s\right)}{\sum_{s=1}^{m} \left[f\left(\widehat{\tau_s/\tau_m}\right) - f\left(\widehat{\tau_{s-1}/\tau_m}\right)\right] \sum_{i=1}^{k_n r_n} \mathbb{1}\left(X_i > u_n^s\right)}\,, \tag{3.6}$$

with the convention that $f\left(\widehat{\tau_0/\tau_m}\right) = 0$. Note that when $m = 1$, (3.6) corresponds to (3.2).

Consistency and asymptotic normality of this estimator depend, as they do for all other related estimators, on certain mixing-type assumptions. Different sets of such conditions are available in literature. We explain next the conditions that we will use in this chapter. These are based on the setup in [32]. For $1 \leq i \leq j \leq n$, and levels $w_n, w_n'$, let $\mathscr{B}_i^j(w_n, w_n')$ denote the $\sigma$-field generated by the events $\{X_d \leq w_n\}$ and $\{X_d \leq w_n'\}$ for $i \leq d \leq j$. For $n \geq 1$ and $1 \leq l \leq n-1$ define

$$\alpha_{n,l}(w_n, w_n') = \max(|P(A \cap B) - P(A)P(B)| : \\ A \in \mathscr{B}_1^k(w_n, w_n'), B \in \mathscr{B}_{k+l}^n(w_n, w_n'), 1 \leq k \leq n-l)$$

and write $\alpha_{n,l}(w_n) = \alpha_{n,l}(w_n, w_n)$. Similarly, one uses the maximal correlation coefficient

$$\rho_{n,l}(w_n, w_n') = \max(\mathrm{corr}(X, Y) : X \in L^2(\mathscr{B}_1^k(w_n, w_n')), \\ Y \in L^2(\mathscr{B}_{k+l}^n(w_n, w_n')), 1 \leq k \leq n-l)\,,$$

where $L^2(\mathscr{F})$ denotes the space of $\mathscr{F}$-measurable square-integrable random variables. Again, we write $\rho_{n,l}(w_n) = \rho_{n,l}(w_n, w_n)$. Trivially,

$$\rho_{n,l}(w_n, w_n') \geq 4\alpha_{n,l}(w_n, w_n').$$

The sequence $\{X_i\}$ is said to satisfy the condition $\Delta(\{w_n\})$ if $\alpha_{n,l_n}(w_n) \to 0$ as $n \to \infty$ for some sequence $\{l_n\}$ with $l_n = o(n)$. If $\{p_n\}$ is a sequence of integers and $\alpha_{p_n,l_n}(w_n) \to 0$ as $n \to \infty$ for some sequence $\{l_n\}$ with $l_n = o(p_n)$, then we will say that $\{X_i\}$ satisfies the condition $\Delta_{\{p_n\}}(\{w_n\})$.

As mentioned earlier, the condition that $r_n\overline{F}(u_n) \to 0$ but $n\overline{F}(u_n) \to \infty$ as $n \to \infty$ is usually required for asymptotic consistency results. This implicitly uses the traditional assumption that $r_n = o(n)$ as $n \to \infty$. It will be convenient to introduce a specific sequence of the integers $\{p_n\}$, which is an intermediate growth sequence between the sequence of the block size $\{r_n\}$ and the sequence of the sample sizes $\{n\}$. Specifically, let

$$p_n\overline{F}(u_n^s) \to \tau_s, \ s = 1, \ldots, m. \tag{3.7}$$

According to (3.4) one such sequence is $p_n = \lceil \tau_m(\overline{F}(u_n))^{-1} \rceil$, $n = 1, 2, \ldots$.

The following assumptions on the stationary sequence $\{X_i\}$ will be used throughout this chapter, not necessarily all in the same place. Some of the assumptions form stronger versions of other assumptions.

**Assumption** $\Delta'$ There is a sequence $l_n = o(r_n)$ such that $p_n r_n^{-1}\alpha_{n,l_n}(u_n^s) \to 0$ as $n \to \infty$ for each $s = 1, \ldots, m$.

**Assumption** $\mathbf{C_1}$ For each $s = 1, \ldots, m$,

$$\sum_{l=1}^{n} \rho_{n,l}(u_n^s) = o(r_n)$$

as $n \to \infty$, and there is a sequence $l_n = o(r_n)$ such that $p_n r_n^{-1} \rho_{n,l_n}(u_n^s) \to 0$ as $n \to \infty$ for each $s = 1, \ldots, m$.

**Assumption $\mathbf{C_1'}$** For each $s = 1, \ldots, m$,

$$\sum_{l=1}^{n} \rho_{n,l}(u_n^s) = o(r_n^{1/2})$$

as $n \to \infty$, and there is a sequence $l_n = o(r_n)$ such that $p_n r_n^{-1} \rho_{n,l_n}(u_n^s) \to 0$ as $n \to \infty$ for each $s = 1, \ldots, m$.

**Assumption $\mathbf{C_2}$** For each $s, t = 1, \ldots, m$,

$$\sum_{l=1}^{n} \rho_{n,l}(u_n^s, u_n^t) = o(r_n)$$

as $n \to \infty$, and there is a sequence $l_n = o(r_n)$ such that $p_n r_n^{-1} \rho_{n,l_n}(u_n^s, u_n^t) \to 0$ as $n \to \infty$ for each $s, t = 1, \ldots, m$.

**Assumption $\mathbf{C_2'}$** For each $s, t = 1, \ldots, m$,

$$\sum_{l=1}^{n} \rho_{n,l}(u_n^s, u_n^t) = o(r_n^{1/2})$$

as $n \to \infty$, and there is a sequence $l_n = o(r_n)$ such that $p_n r_n^{-1} \rho_{n,l_n}(u_n^s, u_n^t) \to 0$ as $n \to \infty$ for each $s, t = 1, \ldots, m$.

The next group of assumptions deals with convergence of certain counting processes. Let $N_{p_n}^{(u)}$ be the point process on $[0, 1]$ with points $(j/p_n : 1 \leq j \leq p_n,\ X_j > u_n)$. Furthermore, for $w > 0$ we write $N_k(w) = \sum_{i=1}^{k} \mathbb{1}(X_i > w)$.

**Assumption P** $N_{p_n}^{(u)}$ converges weakly in the space of finite point processes on $[0, 1]$.

**Assumption $\mathbf{D_1}$** There exists a probability distribution $(\pi_j)_{j \geq 1}$ on the positive

integers such that for all $1 \leq s \leq m$,

$$P(N_{r_n}(u_n^s) = j | M_{r_n} > u_n^s) \to \pi_j,\ j \geq 1,$$
$$E[N_{r_n}^2(u_n^s) | M_{r_n} > u_n^s] \to \sum_{j=1}^{\infty} j^2 \pi_j < \infty.$$

**Assumption $\mathbf{D_2}$** There exist probability distributions $(\varpi_{s,t}(i,j))_{i\geq 1, j\geq 0}$ on $\mathbb{Z}_+ \times \mathbb{Z}_{\geq 0}$ such that for all $1 \leq s < t \leq m$,

$$P(N_{r_n}(u_n^s) = i, N_{r_n}(u_n^t) = j | M_{r_n} > u_n^s) \to \varpi_{s,t}(i,j),\ i \geq j \geq 0,\ i \geq 1,$$
$$E\big[N_{r_n}(u_n^s) N_{r_n}(u_n^t) | M_{r_n} > u_n^s\big] \to \sum_{i=1}^{\infty} \sum_{j=0}^{i} ij \varpi_{s,t}(i,j) < \infty.$$

**Remark 3.3.1.** It is clear that Assumption $\Delta'$ is implied by Assumption $C_1$ which is, in turn, implied both by Assumption $C_1'$ and by Assumption $C_2$. Further, it follows by Theorem 4.1 of [32] that the first part of Assumption $D_1$ is implied by Assumptions $\Delta'$ and $P$. Note that Assumptions $C_1$, $C_2$ and $D_1$ are identical to those posed [52].

**Remark 3.3.2.** Intuitively the mixing conditions $\Delta'$, $C_1$, $C_1'$, $C_2$, $C_2'$ only impose that the size of the blocks $r_n$ must be "large enough". Many models, such as $m$-dependent sequences, or geometrically mixing sequences can easily satisfy those assumptions.

If Assumption $\Delta'$ holds, then it follows from Theorem 5.1 and Lemma 2.3 of [32] that

$$P(M_{r_n} > u_n^s) \sim \tau_s \theta r_n / p_n \tag{3.8}$$

as $n \to \infty$ for $s = 1, \ldots, m$. If we denote

$$\theta_n(f) = \theta_n(\tau_1, \ldots, \tau_m, f) = \frac{p_n}{r_n} \cdot \frac{\sum_{s=1}^{m} (f(\tau_s/\tau_m) - f(\tau_{s-1}/\tau_m)) P(M_{r_n} > u_n^s)}{\sum_{s=1}^{m} (f(\tau_s/\tau_m) - f(\tau_{s-1}/\tau_m))\tau_s}, \tag{3.9}$$

then $\theta_n(f) \to \theta$ as $n \to \infty$.

Another immediate conclusion from (3.8) is that if Assumptions $\Delta'$, $D_1$ and $D_2$ hold, then for $1 \le s < t \le m$,

$$P(N_{r_n}(u_n^s) = i | M_{r_n} > u_n^t) \to \frac{\tau_s}{\tau_t}\big(\pi_i - \varpi_{s,t}(i,0)\big),\ i \ge 1,$$
$$E\big[N_{r_n}(u_n^s) | M_{r_n} > u_n^t\big] \to \psi_{s,t} := \frac{\tau_s}{\tau_t}\sum_{i=1}^{\infty} i\big(\pi_i - \varpi_{s,t}(i,0)\big)\,. \tag{3.10}$$

## 3.4 Preliminary results

The estimator (3.6) is composed of several extremal statistics. In this section we will take a close look at these and related statistics and derive their asymptotic variances and covariances. The derivations are similar to those in [52]. Let $m_n \to \infty$ be any arbitrary sequence of positive integers such that $m_n r_n \le n$ for all $n$. For each level $u_n^s,\ s = 1, \ldots, m$, denote

$$\widehat{M}_{n,m_n}(u_n^s) = \sum_{i=1}^{m_n} \mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^s) \tag{3.11}$$

and

$$\widehat{\tau}_{n,m_n}(u_n^s) = \sum_{i=1}^{m_n r_n} \mathbb{1}(X_i > u_n^s)\,. \tag{3.12}$$

Note that the estimator (3.6) uses these statistics with $m_n = k_n$. For convenience, we denote

$$\widehat{M}_n(u_n^s) = \sum_{i=1}^{k_n} \mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^s) \tag{3.13}$$

and

$$\widehat{\tau}_n(u_n^s) = \sum_{i=1}^{k_n r_n} \mathbb{1}(X_i > u_n^s)\,. \tag{3.14}$$

We first consider the asymptotic variance of $\widehat{M}_{n,m_n}(u_n^s)$.

**Proposition 3.4.1.** *Let $\{X_i\}$ be a stationary sequence with extremal index $\theta$. Let $(p_n)$ be as in* (3.7)*, and suppose that Assumption $C_1$ holds. Then for $1 \leq s \leq m$, as $n \to \infty$,*

$$\frac{p_n}{m_n r_n} \mathrm{var}(\widehat{M}_{n,m_n}(u_n^s)) \to \tau_s \theta. \tag{3.15}$$

*Proof.* Fix $1 \leq s \leq m$ and write out the variance:

$$\begin{aligned}
\mathrm{var}(\widehat{M}_{n,m_n}(u_n^s)) =& \sum_{i=1}^{m_n} \mathrm{var}(\mathbb{1}(M_{(i-1)r_n,ir_n} > u_n^s)) \\
&+ 2 \sum_{1 \leq i < j \leq m_n} \mathrm{cov}(\mathbb{1}(M_{(i-1)r_n,ir_n} > u_n^s), \mathbb{1}(M_{(j-1)r_n,jr_n} > u_n^s)) \\
=& m_n P(M_{r_n} > u_n^s)(1 - P(M_{r_n} > u_n^s)) \\
&+ 2(m_n - 1)\big(P(M_{r_n} > u_n^s, M_{r_n,2r_n} > u_n^s) - (P(M_{r_n} > u_n^s))^2\big) \\
&+ 2 \sum_{v=2}^{m_n - 1} (m_n - v) \mathrm{cov}(\mathbb{1}(M_{r_n} > u_n^s), \mathbb{1}(M_{vr_n,(v+1)r_n} > u_n^s)) \\
:=& I_{1,n} + I_{2,n} + I_{3,n}\,.
\end{aligned}$$

It follows from (3.8) that

$$\frac{p_n}{m_n r_n} I_{1,n} \to \tau_s \theta$$

as $n \to \infty$. Furthermore,

$$\begin{aligned}
\frac{p_n}{m_n r_n} I_{3,n} \leq& 2 \frac{p_n}{r_n} \mathrm{var}(\mathbb{1}(M_{r_n} > u_n^s)) \sum_{v=2}^{m_n - 1} \rho_{n,(v-1)r_n}(u_n^s) \\
\leq& 2 \frac{p_n}{r_n} \mathrm{var}(\mathbb{1}(M_{r_n} > u_n^s)) \frac{1}{r_n} \sum_{l=1}^{n} \rho_{n,l}(u_n^s) \to 0
\end{aligned}$$

by (3.8) and Assumption $C_1$, so it remains to consider $I_{2,n}$. By (3.8) we only need to show that

$$p_n r_n^{-1} P(M_{r_n} > u_n^s, M_{r_n,2r_n} > u_n^s) \to 0.$$

Note that

$$\begin{aligned}
&P(M_{r_n} > u_n^s, M_{r_n,2r_n} > u_n^s) \\
\leq &P(M_{r_n-l_n} > u_n^s, M_{r_n,2r_n} > u_n^s) + P(M_{l_n} > u_n^s) \\
\leq &P(M_{r_n-l_n} > u_n^s)P(M_{r_n} > u_n^s) + \alpha_{n,l_n}(u_n^s) + P(M_{l_n} > u_n^s) \\
\leq &P(M_{r_n} > u_n^s)^2 + \alpha_{n,l_n}(u_n^s) + P(M_{l_n} > u_n^s).
\end{aligned}$$

Since $l_n = o(r_n)$, and $p_n r_n^{-1}\alpha_{n,l_n}(u_n^s) \to 0$, there is an intermediate sequence $l'_n$ with $l_n = o(l'_n)$ and $l'_n = o(r_n)$, such that $p_n(l'_n)^{-1}\alpha_{n,l_n}(u_n^s) \to 0$. Then as in (3.8),

$$p_n(l'_n)^{-1}P(M_{l'_n} > u_n^s) \to \tau_s\theta,$$

so we have both

$$p_n r_n^{-1}P(M_{l_n} > u_n^s) \leq p_n(l'_n)^{-1}(l'_n r_n^{-1})P(M_{l'_n} > u_n^s) \to 0$$

and

$$p_n r_n^{-1}\alpha_{n,l_n}(u_n^s) \to 0\,.$$

Therefore, the result follows. □

The asymptotic covariance of $\widehat{M}_{n,m_n}(u_n^s)$ and $\widehat{M}_{n,m_n}(u_n^t)$ for $s \neq t$ can be obtained in an identical way (with a slightly different assumption). The proof is omitted.

**Proposition 3.4.2.** *Let $\{X_i\}$ be a stationary sequence with extremal index $\theta$. Let $(p_n)$ be as in* (3.7)*, and suppose that Assumption $C_2$ holds. Then for $1 \leq s < t \leq m$, as $n \to \infty$,*

$$\frac{p_n}{m_n r_n}\mathrm{cov}(\widehat{M}_{n,m_n}(u_n^s), \widehat{M}_{n,m_n}(u_n^t)) \to \tau_t\theta. \tag{3.16}$$

Now we find the variance and covariance of $\widehat{\tau}_{n,m_n}(u_n^s)$ and $\widehat{\tau}_{n,m_n}(u_n^t)$ for $1 \leq s < t \leq m$. We start with the variance.

**Proposition 3.4.3.** *Let $\{X_i\}$ be a stationary sequence with extremal index $\theta$. Suppose that Assumptions $C_1$ and $D_1$ hold. Then as $n \to \infty$, for $1 \leq s \leq m$,*

$$\frac{p_n}{m_n r_n} \mathrm{var}(\widehat{\tau}_{n,m_n}(u_n^s)) \to \tau_s \theta \sum_{j=1}^{\infty} j^2 \pi_j \,. \tag{3.17}$$

*Proof.* We proceed as in Proposition 3.4.1. Using the notation $N_{a,b}(w) = \sum_{a<i\leq b} \mathbb{1}(X_i > w)$ for integers $0 \leq a < b$, we obtain for a fixed $1 \leq s \leq m$,

$$\begin{aligned}
\mathrm{var}(\widehat{\tau}_{n,m_n}(u_n^s)) =& \mathrm{var}(N_{m_n r_n}(u_n^s)) \\
=& \sum_{i=1}^{m_n} \mathrm{var}\big(N_{(i-1)r_n, ir_n}(u_n^s)\big) \\
&+ 2 \sum_{1\leq i<j\leq m_n} \mathrm{cov}\big(N_{(i-1)r_n, ir_n}, N_{(j-1)r_n, jr_n}\big) \\
=& m_n \mathrm{var}(N_{r_n}) + 2(m_n - 1)\mathrm{cov}\big(N_{r_n}, N_{r_n, 2r_n}\big) \\
&+ 2 \sum_{v=2}^{m_n - 1} (m_n - v)\mathrm{cov}\big(N_{r_n} N_{vr_n, (v+1)r_n}\big) \\
:=& I_{1,n} + I_{2,n} + I_{3,n} \,.
\end{aligned}$$

It follows from (3.8) and Assumption $D_1$ that

$$\begin{aligned}
\frac{p_n}{m_n r_n} I_{1,n} \sim& \frac{p_n}{r_n} P(M_{r_n} > u_n^s) E[N_{r_n}^2(u_n^s) | M_{r_n} > u_n^s] \\
&- \frac{p_n}{r_n} \big(P(M_{r_n} > u_n^s)\big)^2 \big(E[N_{r_n}(u_n^s) | M_{r_n} > u_n^s]\big)^2 \\
\to& \tau_s \theta \sum_{j=1}^{\infty} j^2 \pi_j
\end{aligned}$$

as $n \to \infty$. Furthermore,

$$\begin{aligned}
\frac{p_n}{m_n r_n} I_{3,n} \leq& 2 \frac{p_n}{r_n} \mathrm{var}(N_{r_n}) \sum_{v=2}^{m_n - 1} \rho_{n,(v-1)r_n}(u_n^s) \\
\leq& 2 \frac{p_n}{r_n} P(M_{r_n} > u_n^s) E[N_{r_n}^2(u_n^s) | M_{r_n} > u_n^s] \frac{1}{r_n} \sum_{l=1}^{n} \rho_{n,l}(u_n^s) \to 0
\end{aligned}$$

by Assumptions $C_1$ and $D_1$. As far as $I_{2,n}$ is concerned, we only need to show that

$$p_n r_n^{-1} E\big(N_{r_n} N_{r_n,2r_n}\big) \to 0.$$

However,

$$\begin{aligned} E\big(N_{r_n} N_{r_n,2r_n}\big) =& E\big(N_{r_n-l_n} N_{r_n,2r_n}\big) + E\big(N_{r_n-l_n,r_n} N_{r_n,2r_n}\big) \\ \leq& (EN_{r_n})^2 + E(N_{r_n})^2 \rho_{n,l_n}(u_n^s) + E\big(N_{r_n-l_n,r_n} N_{r_n,2r_n}\big). \end{aligned}$$

By Assumptions $C_1$ and $D_1$ and the above calculation, both $k_n(EN_{r_n})^2 \to 0$ and $k_n E(N_{r_n})^2 \rho_{n,l_n}(u_n^s) \to 0$ as $n \to \infty$. Furthermore, by stationarity it is clear that

$$\frac{E(N_{r_n})^2}{E(N_{l_n})^2} \geq \lfloor r_n/l_n \rfloor \to \infty$$

as $n \to \infty$. Therefore,

$$\begin{aligned} k_n E\big(N_{r_n-l_n,r_n} N_{r_n,2r_n}\big) \leq& k_n \left(E(N_{l_n})^2\right)^{1/2} \left(E(N_{r_n})^2\right)^{1/2} \\ \leq& k_n E(N_{r_n})^2 \left(\frac{E(N_{l_n})^2}{E(N_{r_n})^2}\right)^{1/2} \to 0 \end{aligned}$$

as $n \to \infty$. This completes the proof. □

The asymptotic covariance between $\widehat{\tau}_{n,m_n}(u_n^s)$ and $\widehat{\tau}_{n,m_n}(u_n^t)$ for $1 \leq s < t \leq m$ can be found in the same way. Once again, we omit the proof.

**Proposition 3.4.4.** *Let $\{X_i\}$ be a stationary sequence with extremal index $\theta$. Suppose that Assumptions $C_2$ and $D_2$ hold. Then as $n \to \infty$, for $1 \leq s < t \leq m$,*

$$\frac{p_n}{m_n r_n} \mathrm{cov}(\widehat{\tau}_{n,m_n}(u_n^s), \widehat{\tau}_{n,m_n}(u_n^t)) \to \tau_s \theta \sum_{i=1}^{\infty} \sum_{j=0}^{i} ij\varpi_{s,t}(i,j). \tag{3.18}$$

We now address the asymptotic covariances between $\widehat{\tau}$ and $\widehat{M}$. We start with the "diagonal" case.

**Proposition 3.4.5.** *Let $\{X_i\}$ be a stationary sequence with extremal index $\theta$. Suppose that Assumption $C_1'$ holds. Then as $n \to \infty$, for $1 \leq s \leq m$,*

$$\frac{p_n}{m_n r_n} \mathrm{cov}(\widehat{M}_{n,m_n}(u_n^s), \widehat{\tau}_{n,m_n}(u_n^s)) \to \tau_s. \tag{3.19}$$

*Proof.* Fix $1 \leq s \leq m$, we have

$$\mathrm{cov}(\widehat{M}_{n,m_n}(u_n^s), \widehat{\tau}_{n,m_n}(u_n^s)) = \sum_{i=1}^{m_n} \sum_{j=1}^{m_n r_n} \mathrm{cov}(\mathbb{1}(M_{(i-1)r_n, ir_n} \leq u_n^s), \mathbb{1}(X_j \leq u_n^s)).$$

We split the sum into two pieces, $I_{1,n} + I_{2,n}$, depending on whether $(i-1)r_n < j \leq ir_n$ or not. By stationarity,

$$\begin{aligned}\frac{p_n}{m_n r_n} I_{1,n} \sim & \frac{p_n}{r_n} \sum_{i=1}^{r_n} \mathrm{cov}(\mathbb{1}(M_{r_n} \leq u_n^s), \mathbb{1}(X_i \leq u_n^s)) \\ \sim & p_n P(X_1 > u_n^s) P(M_{r_n} \leq u_n^s) \to \tau_s\end{aligned}$$

by (3.7) and (3.8).

Furthermore, we can bound $I_{2,n}$ as follows:

$$|I_{2,n}| \leq 2m_n \sqrt{\mathrm{var}(\mathbb{1}(M_{r_n} \leq u_n^s)) \mathrm{var}(\mathbb{1}(X_1 \leq u_n^s))} \sum_{l=1}^{n} \rho_{n,l}(u_n^s),$$

and the fact that $(p_n/(m_n r_n)) I_{2,n} \to 0$ as $n \to \infty$ follows from (3.7), (3.8) and Assumption $C_1'$. □

The asymptotic behaviour of $\mathrm{cov}(\widehat{M}_{n,m_n}(u_n^s), \widehat{\tau}_{n,m_n}(u_n^t))$ with $1 \leq s < t \leq m$ is similar to the "diagonal" case. The proof of the next proposition is similar to the argument in Proposition 3.4.5 (once we use the appropriate assumption), and is omitted.

**Proposition 3.4.6.** *Let $\{X_i\}$ be a stationary sequence with extremal index $\theta$. Suppose that Assumption $C_2'$ holds. Then as $n \to \infty$, for $1 \leq s < t \leq m$,*

$$\frac{p_n}{m_n r_n} \mathrm{cov}(\widehat{M}_{n,m_n}(u_n^s), \widehat{\tau}_{n,m_n}(u_n^t)) \to \tau_t. \tag{3.20}$$

Finally, we consider the asymptotic behaviour of $\operatorname{cov}(\widehat{M}_{n,m_n}(u_n^t), \widehat{\tau}_{n,m_n}(u_n^s))$ with $1 \leq s < t \leq m$.

**Proposition 3.4.7.** *Let $\{X_i\}$ be a stationary sequence with extremal index $\theta$. Suppose that Assumptions $\Delta'$, $D_1$ and $D_2$ hold. Then as $n \to \infty$, for $1 \leq s < t \leq m$,*

$$\frac{p_n}{m_n r_n} \operatorname{cov}(\widehat{M}_{n,m_n}(u_n^t), \widehat{\tau}_{n,m_n}(u_n^s)) \to \tau_t \theta \psi_{s,t}, \tag{3.21}$$

*where $\psi_{s,t}$ is defined in* (3.10).

*Proof.* As before,

$$\operatorname{cov}(\widehat{M}_{n,m_n}(u_n^t), \widehat{\tau}_{n,m_n}(u_n^s)) = \sum_{i=1}^{m_n} \sum_{j=1}^{m_n r_n} \operatorname{cov}\big(\mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^t), \mathbb{1}(X_j > u_n^s)\big).$$

Once again we split the sum into two pieces, $I_{1,n} + I_{2,n}$, depending on whether $(i-1)r_n < j \leq ir_n$ or not. By stationarity,

$$\begin{aligned}
\frac{p_n}{m_n r_n} I_{1,n} \sim & \frac{p_n}{r_n} \sum_{i=1}^{r_n} \operatorname{cov}(\mathbb{1}(M_{r_n} > u_n^t), \mathbb{1}(X_i > u_n^s)) \\
= & \frac{p_n}{r_n} \sum_{i=1}^{r_n} P(M_{r_n} > u_n^t, X_i > u_n^s) - p_n P(M_{r_n} > u_n^t) P(X_1 > u_n^s) \\
= & \frac{p_n}{r_n} E[N_{r_n}(u_n^s) | M_{r_n} > u_n^t] P(M_{r_n} > u_n^t) - p_n P(M_{r_n} > u_n^t) P(X_1 > u_n^s) \\
\to & \tau_t \theta \psi_{s,t}
\end{aligned}$$

as $n \to \infty$ by (3.7), (3.8) and (3.10). Since $I_{2,n} \to 0$ as before, the proof of the proposition is complete. □

## 3.5 A Central Limit Theorem for the Multilevel Estimator

In this section we establish the asymptotic normality of our multilevel estimator (3.6). We start by checking the consistency of the estimator. For notational

convenience, recall the following two definitions.

$$\widehat{M}_n(u_n^s) = \widehat{M}_{n,k_n}(u_n^s) = \sum_{i=1}^{k_n} \mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^s) \tag{3.22}$$

and

$$\widehat{\tau}_n(u_n^s) = \widehat{\tau}_{n,k_n}(u_n^s) = \sum_{i=1}^{k_n r_n} \mathbb{1}(X_i > u_n^s)\,. \tag{3.23}$$

**Proposition 3.5.1.** *Let $\{X_i\}$ be a stationary sequence with extremal index $\theta$. Suppose that Assumptions $C_1$ and $D_1$ hold. Then as $n \to \infty$,*

$$\widehat{\theta}_n(f) \to_P \theta. \tag{3.24}$$

*Proof.* Note that for $1 \le s \le m$, by (3.8),

$$E\left(\frac{p_n}{n}\widehat{M}_n(u_n^s)\right) = \frac{k_n p_n}{n} P(M_{r_n} > u_n^s) \to \tau_s\theta$$

as $n \to \infty$. Since $\operatorname{var}\big((p_n/n)\widehat{M}_n(u_n^s)\big) \to 0$ by Proposition 3.4.1, it follows that $(p_n/n)\widehat{M}_n(u_n^s) \to_P \tau_s\theta$ as $n \to \infty$.

Similarly, by (3.7) and Proposition 3.4.3 we have $(p_n/n)\widehat{\tau}_n(u_n^s) \to_P \tau_s$ as $n \to \infty$ for $1 \le s \le m$. In particular,

$$\widehat{\tau_s/\tau_m} \to_P \tau_s/\tau_m \text{ for } 1 \le s \le m,$$

and the result follows. □

The next theorem is the main result of this section. It establishes asymptotic normality of the estimator (3.6). It requires an assumption on the rate of convergence in (3.8). We assume that, as $n \to \infty$,

$$\sqrt{n/p_n}\big[(p_n/r_n)P(M_{r_n} > u_n^s) - \tau_s\theta\big] \to 0,\ 1 \le s \le m\,. \tag{3.25}$$

Such an assumption is sometimes associated with a sufficiently large block size $r_n$; see e.g. [52].

Under the notation of Assumptions $D_1$ and $D_2$ we denote

$$\mu_2 := \sum_{j=1}^{\infty} j^2 \pi_j,$$
$$\mu_{s,t} := \sum_{i=1}^{\infty} \sum_{j=0}^{i} ij\varpi_{s,t}(i,j),\ 1 \le s < t \le m.$$

**Theorem 3.5.2.** *Let $\{X_i\}$ be a stationary sequence with extremal index $\theta$. Assume that Assumptions $C_1'$, $C_2$, $C_2'$, $D_1$ and $D_2$ hold. Assume further* (3.25)*. Then as $n \to \infty$,*

$$\sqrt{n/p_n}\big(\widehat{\theta}_n(f) - \theta\big) \to_d \mathcal{N}(0, \sigma^2), \tag{3.26}$$

*where $\sigma^2 = \mathbf{h}^T \mathbf{\Sigma} \mathbf{h}$, with a $(2m) \times (2m)$ covariance matrix $\mathbf{\Sigma}$ and a $2m$-dimensional vector $\mathbf{h}$ defined as follows: for $1 \le s \le t \le m$,*

$$\sigma_{s,t} = \tau_t \theta,$$
$$\sigma_{m+s,m+t} = \tau_s \theta \mu_{s,t},$$
$$\sigma_{s,m+t} = \tau_t,$$
$$\sigma_{t,m+s} = \tau_t \theta \psi_{s,t}\,,$$

*where $\mu_{s,s}$ is taken to be $\mu_2$ for each $s$, while $\psi_{s,t}$ is defined by* (3.10) *for $s < t$ and taken to be $1/\theta$ if $s = t$. Furthermore,*

$$h_s = \frac{f\big(\tau_s/\tau_m\big) - f\big(\tau_{s-1}/\tau_m\big)}{\sum_{t=1}^{m}\big(f\big(\tau_t/\tau_m\big) - f\big(\tau_{t-1}/\tau_m\big)\big)\tau_t}, \quad 1 \le s \le m,$$
$$h_{m+s} = -\frac{\big(f\big(\tau_s/\tau_m\big) - f\big(\tau_{s-1}/\tau_m\big)\big)\theta}{\sum_{t=1}^{m}\big(f\big(\tau_t/\tau_m\big) - f\big(\tau_{t-1}/\tau_m\big)\big)\tau_t}, \quad 1 \le s \le m,$$

*where we set $\tau_0 = \infty$ and $f(\infty) = 0$.*

*Proof.* The argument is similar to that used in Theorem 4.2 of [52]. Notice that

$$\widehat{\theta}_n(f) = h\big((p_n/n)\widehat{M}_n(u_n^1), \ldots, (p_n/n)\widehat{M}_n(u_n^m), (p_n/n)\widehat{\tau}_n(u_n^1), \ldots, (p_n/n)\widehat{\tau}_n(u_n^m)\big),$$
$$\theta = h\big(\tau_1\theta, \ldots, \tau_m\theta, \tau_1, \ldots, \tau_m\big)\,,$$

where $h:\ [0,\infty)^m \times (0,\infty)^m \to [0,\infty)$ is defined by

$$h(x_1,\ldots,x_m,y_1,\ldots,y_m) = \frac{\sum_{s=1}^m \big(f(y_s/y_m) - f(y_{s-1}/y_m)\big)x_s}{\sum_{s=1}^m \big(f(y_s/y_m) - f(y_{s-1}/y_m)\big)y_s}\,.$$

Here and for the remainder of the proof we use the convention $y_0 = \infty$ and $f(\infty) = 0$. Since

$$\nabla h\big(\tau_1\theta,\ldots,\tau_m\theta,\tau_1,\ldots,\tau_m\big) = \mathbf{h}\,,$$

by the delta method we only need to prove that

$$\sqrt{n/p_n}\begin{pmatrix} (p_n/n)\widehat{M}_n(u_n^1) - \tau_1\theta \\ \vdots \\ (p_n/n)\widehat{M}_n(u_n^m) - \tau_m\theta \\ (p_n/n)\widehat{\tau}_n(u_n^1) - \tau_1 \\ \vdots \\ (p_n/n)\widehat{\tau}_n(u_n^m) - \tau_m \end{pmatrix} \to_d \mathcal{N}(\mathbf{0},\mathbf{\Sigma})\,. \tag{3.27}$$

We will, actually, prove the statement

$$\sqrt{n/p_n}\begin{pmatrix} (p_n/n)\big[\widehat{M}_n(u_n^1) - k_nP(M_{r_n} > u_n^1)\big] \\ \vdots \\ (p_n/n)\big[\widehat{M}_n(u_n^m) - k_nP(M_{r_n} > u_n^m)\big] \\ (p_n/n)\widehat{\tau}_n(u_n^1) - \tau_1 \\ \vdots \\ (p_n/n)\widehat{\tau}_n(u_n^m) - \tau_m \end{pmatrix} \to_d \mathcal{N}(\mathbf{0},\mathbf{\Sigma})\,. \tag{3.28}$$

By (3.25) this will imply (3.27).

We present an argument for the case $m = 2$. The argument for larger values of $m$ is only notationally different. Denote by $Z_{n,i}$, $i = 1,2,3,4$ the 4 entries in the vector in the left hand side of (3.28). By the Cramér-Wold device it suffices to show that for any $\mathbf{a} = (a_1,a_2,a_3,a_4)^T \in \mathbb{R}^4$, as $n \to \infty$,

$$a_1Z_{n,1} + a_2Z_{n,2} + a_3Z_{n,3} + a_4Z_{n,4} \to_d \mathcal{N}(0,\mathbf{a}^T\mathbf{\Sigma}\mathbf{a}). \tag{3.29}$$

Denote $m_n = \lfloor n/p_n \rfloor$ and let $h_n = \lfloor k_n/m_n \rfloor$ and write

$$Z_{n,1} = \sqrt{\frac{p_n}{n}} \sum_{i=1}^{h_n} \overline{I}_i(u_n^1) + \mathrm{o}_p(1),\ Z_{n,2} = \sqrt{\frac{p_n}{n}} \sum_{i=1}^{h_n} \overline{I}_i(u_n^2) + \mathrm{o}_p(1)$$
$$Z_{n,3} = \sqrt{\frac{p_n}{n}} \sum_{i=1}^{h_n} \overline{J}_i(u_n^1) + \mathrm{o}_p(1),\ Z_{n,4} = \sqrt{\frac{p_n}{n}} \sum_{i=1}^{h_n} \overline{J}_i(u_n^2) + \mathrm{o}_p(1),$$

where

$$\overline{I}_i(u_n^1) = \sum_{j=(i-1)m_n}^{im_n-1} \big(\mathbb{1}(M_{(j-1)r_n,jr_n} > u_n^1) - P(M_{r_n} > u_n^1)\big),$$
$$\overline{I}_i(u_n^2) = \sum_{j=(i-1)m_n}^{im_n-1} \big(\mathbb{1}(M_{(j-1)r_n,jr_n} > u_n^2) - P(M_{r_n} > u_n^2)\big),$$
$$\overline{J}_i(u_n^1) = \sum_{j=(i-1)m_nr_n}^{im_nr_n-1} \big(\mathbb{1}(X_j > u_n^1) - \tau_1/p_n\big),$$
$$\overline{J}_i(u_n^2) = \sum_{j=(i-1)m_nr_n}^{im_nr_n-1} \big(\mathbb{1}(X_j > u_n^2) - \tau_2/p_n\big).$$

Let $h_n^* \to \infty$ be a sequence of integers with $(h_n^*)^2 = o(h_n)$, $h_n = o\big((h_n^*)^3\big)$. Partition the set $\{1, \ldots, h_n\}$ into subsets of length $h_n^*$ of consecutive integers, with two adjacent such subsets separated by a singleton. The number of subsets of length $h_n^*$ is then $q_n = \lfloor (h_n+1)/(h_n^*+1) \rfloor$. We have

$$\begin{aligned}\sqrt{\frac{p_n}{n}} \sum_{i=1}^{h_n} \overline{I}_i(u_n^1) =& \sqrt{\frac{p_n}{n}} \sum_{j=1}^{q_n} \sum_{i=(j-1)(h_n^*+1)+1}^{j(h_n^*+1)-1} \overline{I}_i(u_n^1) \\ &+ \sqrt{\frac{p_n}{n}} \sum_{j=1}^{q_n-1} \overline{I}_{j(h_n^*+1)}(u_n^1) + \sqrt{\frac{p_n}{n}} \sum_{i=q_n(h_n^*+1)}^{h_n} \overline{I}_i(u_n^1).\end{aligned} \tag{3.30}$$

The variance of the second term is bounded by

$$\frac{p_nq_n}{n} \mathrm{var}(\overline{I}_1(u_n^1)) + \frac{p_nq_n^2}{n} \rho_{n,h_n^*r_n}(u_n^1) \mathrm{var}(\overline{I}_1(u_n^1)).$$

By Proposition 3.4.1 the first entry above does not exceed a constant multiple of

$$\frac{p_nq_n}{n} \frac{m_nr_n}{p_n} \sim \frac{1}{h_n^*} \to 0$$

since $h_n^* \to \infty$. Since Assumption $C_1$ is in force,

$$\rho_{n,h_n^* r_n}(u_n^1) = \frac{1}{h_n^* r_n} h_n^* r_n \rho_{n,h_n^* r_n}(u_n^1) \leq \frac{1}{h_n^* r_n} \sum_{l=1}^{n} \rho_{n,l}(u_n^1) = \mathrm{o}\left(\frac{1}{h_n^*}\right).$$

Therefore, the second entry above does not exceed a constant multiple of

$$\frac{p_n q_n^2}{n} \frac{1}{h_n^*} \frac{m_n r_n}{p_n} \sim \frac{h_n}{(h_n^*)^3} \to 0$$

by the choice of $h_n^*$. Hence it follows that the variance of the second term in (3.30) converges to zero. Further, the variance of the third term in (3.30) is, apart from a multiplicative constant, bounded by

$$\frac{p_n (h_n^*)^2}{n} \mathrm{var}(\overline{I}_1(u_n^1)) \sim \frac{p_n (h_n^*)^2}{n} \frac{m_n r_n}{p_n} \sim \frac{(h_n^*)^2}{h_n} \to 0,$$

once again by the choice of $h_n^*$. Therefore, we can write

$$Z_{n,1} = \frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n} \left( \sqrt{\frac{p_n q_n}{n}} \sum_{i=(j-1)(h_n^*+1)+1}^{j(h_n^*+1)-1} \overline{I}_i(u_n^1) \right) + o_p(1) =: \frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n} \xi_{n,j,1} + o_p(1).$$

Similarly,

$$Z_{n,2} = \frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n} \left( \sqrt{\frac{p_n q_n}{n}} \sum_{i=(j-1)(h_n^*+1)+1}^{j(h_n^*+1)-1} \overline{I}_i(u_n^2) \right) + o_p(1) =: \frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n} \xi_{n,j,2} + o_p(1),$$

$$Z_{n,3} = \frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n} \left( \sqrt{\frac{q_n}{h_n}} \sum_{i=(j-1)(h_n^*+1)+1}^{j(h_n^*+1)-1} \overline{J}_i(u_n^1) \right) + o_p(1) =: \frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n} \xi_{n,j,3} + o_p(1),$$

$$Z_{n,4} = \frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n} \left( \sqrt{\frac{q_n}{h_n}} \sum_{i=(j-1)(h_n^*+1)+1}^{j(h_n^*+1)-1} \overline{J}_i(u_n^2) \right) + o_p(1) =: \frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n} \xi_{n,j,4} + o_p(1).$$

Writing $\xi_{n,j} = a_1 \xi_{n,j,1} + a_2 \xi_{n,j,2} + a_3 \xi_{n,j,3} + a_4 \xi_{n,j,4}$, we conclude that

$$a_1 Z_{n,1} + a_2 Z_{n,2} + a_3 Z_{n,3} + a_4 Z_{n,4} = \frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n} \xi_{n,j} + o_p(1).$$

Notice that for fixed $n$ the elements of the stationary sequence defining each pair of $\xi_{n,i}$ and $\xi_{n,j}$, $i \neq j$, are separated by at least $h_n^* r_n$ entries. Furthermore, by Assumptions $C_1$ and $C_2$,

$$\rho_{n,h_n^* r_n}(u_n^1, u_n^2) = o(1/h_n) = o(1/q_n)\,.$$

Since for any real $\theta$

$$\begin{aligned}
&\left| E\exp\left\{ i\theta \frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n} \xi_{n,j} \right\} - \prod_{j=1}^{q_n} E\exp\left\{ i\theta \frac{1}{\sqrt{q_n}} \xi_{n,j} \right\} \right| \\
&\leq \sum_{k=1}^{q_n} \left| E\exp\left\{ i\theta \frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n-k+1} \xi_{n,j} \right\} \right. \\
&\qquad \left. - E\exp\left\{ i\theta \frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n-k} \xi_{n,j} \right\} E\exp\left\{ i\theta \frac{1}{\sqrt{q_n}} \xi_{n,q_n-k+1} \right\} \right| \\
&\leq q_n \rho_{n,h_n^* r_n}(u_n^1, u_n^2)
\end{aligned}$$

up to a multiplicative constant, the statement (3.29) will follow once we prove that

$$\frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n} Y_{n,j} \to_d \mathcal{N}(0, \mathbf{a}^T \mathbf{\Sigma} \mathbf{a})\,, \tag{3.31}$$

where for each $n$, $Y_{n,j}$, $j = 1, \ldots, q_n$ are i.i.d. random variables with the same law as $\xi_{n,1}$. Since Propositions 3.4.1 - 3.4.7 tell us that $\mathrm{var}(\xi_{n,1}) \to \mathbf{a}^T \mathbf{\Sigma} \mathbf{a}$ as $n \to \infty$, by the Lindeberg-Feller central limit theorem the convergence in (3.31) will follow once we check that for any $\varepsilon > 0$,

$$E\big(\xi_{n,1}^2 \mathbb{1}(|\xi_{n,1}| > \varepsilon q_n^{1/2})\big) \to 0$$

as $n \to \infty$, which reduces to showing that

$$E\big(\xi_{n,1,i}^2 \mathbb{1}(|\xi_{n,1,j}| > \varepsilon q_n^{1/2})\big) \to 0 \tag{3.32}$$

for each $\varepsilon > 0$ and each pair $i, j = 1, 2, 3, 4$. We will check (3.32) for $i = j = 1$. All other combinations of $i, j$ can be treated in a similar way. If $\widehat{M}_n^*(u_n^1)$ is defined

by (3.11) with $m_n$ replaced by $m_n h_n^*$, then we have to check that

$$\frac{p_n q_n}{n} E\left( (\widehat{M}_n^*(u_n^1))^2 \mathbb{1}\big(|\widehat{M}_n^*(u_n^1)| > \varepsilon\sqrt{n/p_n}\big)\right) \to 0\,.$$

While proving Proposition 3.4.1 we decomposed the variance of $\widehat{M}_n^*(u_n^1)$ into a sum of two terms, the second of which is of a smaller order than the first one. Therefore, we only need to prove that

$$\frac{p_n q_n}{n} \sum_{i=1}^{m_n h_n^*} E\Bigg[ \Big( \mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^1) - P(M_{r_n} > u_n^1)\Big)^2 \\ \mathbb{1}\left( \left| \sum_{j=1}^{m_n h_n^*} \big( \mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^1) - P(M_{r_n} > u_n^1)\big)\right| > \varepsilon\sqrt{n/p_n}\right)\Bigg] \to 0$$

and , since $n/p_n \to \infty$, by changing $\varepsilon > 0$ to a smaller positive number, we only need to show that

$$\frac{p_n q_n}{n} \sum_{i=1}^{m_n h_n^*} P\,\Big( M_{(i-1)r_n, ir_n} > u_n^1, \\ \left| \sum_{|j-i|\geq 2} \big( \mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^1) - P(M_{r_n} > u_n^1)\big)\right| > \varepsilon\sqrt{n/p_n}\Big) \to 0\,.$$

Note that the expression in the left hand side above can be bounded by

$$\frac{p_n q_n}{n} \sum_{i=1}^{m_n h_n^*} P\big(M_{(i-1)r_n, ir_n} > u_n^1\big) \\ \times P\left( \left| \sum_{|j-i|\geq 2} \big( \mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^1) - P(M_{r_n} > u_n^1)\big)\right| > \varepsilon\sqrt{n/p_n}\right) \\ + \frac{p_n q_n}{n} m_n h_n^* \alpha_{n, r_n}(u_n^1)\,.$$

The first term above converges to zero as $n \to \infty$ by Proposition 3.4.1, while the second term converges to zero as $n \to \infty$ by Assumption $C_1$. Therefore, the convergence in (3.31) has been established. □

**Remark 3.5.3.** Note that without the assumption (3.25) what Theorem 3.5.2 proves is that

$$\sqrt{n/p_n}\big(\widehat{\theta}_n(f) - \theta_n(f)\big) \to_d \mathcal{N}(0, \sigma^2).$$

The difference $\theta_n(f) - \theta$ is then responsible for the bias of our estimator.

## 3.6 Testing the estimator

This section is devoted to testing the multilevel estimator (3.6) both on simulated data and real data. As in many cases of extremal inference, we should address the question of the bias of the estimator; see, in particular, Remark 3.5.3. One approach of tackling the bias is to build a simple model for it and then estimate it from the data. Following [16], and further to account for the effect of block size $r_n$, we assume that the main terms in the bias of $\widehat{M}_n(u_n^s)/\widehat{\tau}_n(u_n^s)$ as an estimator of $\theta$ are linear in $\tau_s/k_n$ and $1/r_n$, $s = 1, \ldots, m$. Since we estimate $\tau_s$ by a scaled version of the statistics $\widehat{\tau}_n(u_n^s)$, it is natural to use the following bias-corrected version of the multilevel estimator:

$$\widehat{\theta}_n^b(f) = \frac{\sum_{s=1}^m \big[f\big(\widehat{\tau_s/\tau_m}\big) - f\big(\widehat{\tau_{s-1}/\tau_m}\big)\big](\widehat{M}_n(u_n^s) - \widehat{\beta}_1 \frac{\widehat{\tau}_n(u_n^s)^2}{k_n} - \widehat{\beta}_2 \frac{\widehat{\tau}_n(u_n^s)}{r_n})}{\sum_{s=1}^m \big[f\big(\widehat{\tau_s/\tau_m}\big) - f\big(\widehat{\tau_{s-1}/\tau_m}\big)\big]\widehat{\tau}_n(u_n^s)}\,, \quad (3.33)$$

where $\widehat{\beta}_1$, $\widehat{\beta}_2$ are coefficients estimated from the data. We simply use linear regression as follows.

Use the $m$ levels $u_n^1, \ldots, u_n^m$ and $l$ values of block sizes $r_n^1, \ldots, r_n^l$ to compute the values of $\widehat{M}_n(u_n^s, r_n^i)$, $\widehat{\tau}_n(u_n^s, r_n^i)$ and $\widehat{\theta}_n(u_n^s, r_n^i) = \widehat{M}_n(u_n^s, r_n^i)/\widehat{\tau}_n(u_n^s, r_n^i)$ for $s = 1, \ldots, m$, $i = 1, \ldots, l$, where $\widehat{M}_n(u_n^s, r_n^i)$, $\widehat{\tau}_n(u_n^s, r_n^i)$ respectively denote the quantities $\widehat{M}_n(u_n^s)$ and $\widehat{\tau}_n(u_n^s)$ evaluated using block size $r_n^i$. Now fit are regression plane to the response variables $\widehat{\theta}_n(u_n^s, r_n^i)$ using the predictor variables $\big(\widehat{\tau}_n(u_n^s, r_n^i)/k_n^i\,, 1/r_n^i\big)$, $s = 1, \ldots, m$, $i = 1, \ldots, l$, where $k_n^i = \lfloor \frac{n}{r_n^i} \rfloor$. Specifically, we use the least squares coefficients

$$\big(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2\big)^T = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\widehat{\boldsymbol{\theta}}_n\,, \quad (3.34)$$

where

$$\widehat{\boldsymbol{\theta}}_n = \left(\widehat{\theta}_n(u_n^1, r_n^1) \ \ldots \ \widehat{\theta}_n(u_n^m, r_n^l)\right)^T,$$

and

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \widehat{\tau}_n(u_n^1, r_n^1)/k_n^1 & \widehat{\tau}_n(u_n^2, r_n^1)/k_n^1 & \ldots & \widehat{\tau}_n(u_n^m, r_n^l)/k_n^l \\ 1/r_n^1 & 1/r_n^1 & \ldots & 1/r_n^l \end{pmatrix}^T,$$

where $\widehat{\boldsymbol{\theta}}_n$ is a vector of length $ml$ and $\mathbf{X}$ is a matrix of dimension $ml \times 3$. We use $\widehat{\beta}_1, \widehat{\beta}_2$ in (3.34) as desired coefficients in (3.33). Alternatively, one could estimate that coefficient using different levels from the collection $u_n^1, \ldots, u_n^m$.

**Remark 3.6.1.** Simulation results support the assumption that the main terms in the bias of $\widehat{M}_n(u_n^s)/\widehat{\tau}_n(u_n^s)$ as an estimator of $\theta$ are linear in $\tau_s/k_n$ and $1/r_n$, $s = 1, \ldots, m$.

**Remark 3.6.2.** Note that $\widehat{\beta}_0$ in (3.34) is itself an estimator for $\theta$. We have not studied its statistical properties, but it performs well on simulated data.

In the sequel we test the multilevel estimator (3.6) and its bias-corrected version (3.33) on simulated data and on S&P 500 Daily Log Returns.

### 3.6.1 Simulation Study

We have drawn samples from ARMAX processes and the MA($q$) process. The ARMAX process $(X_i)$ is defined as follows. Let $Z_1, Z_2, \ldots$ be a sequence of i.i.d. unit Fréchet random variables. For $0 < \theta \leq 1$, let $X_1 = Z_1/\theta$, and

$$X_i = \max((1-\theta)X_{i-1}, Z_i), \ i \geq 2\,. \tag{3.35}$$

It can be shown that the extremal index of such a sequence is $\theta$; see e.g. Chapter 10 of [4].

The MA($q$) process $(X_i)$ is defined as

$$X_i = p_q Z_{i-q} + p_{q-1} Z_{i-q+1} + \cdots + p_1 Z_{i-1} + Z_i,\ i \geq 1\,, \tag{3.36}$$

with $0 < p_1, \ldots, p_q < 1$, and the noise sequence consisting of i.i.d. Pareto random variables $Z_{-(q-1)}, \ldots, Z_0, Z_1, \ldots$ with

$$P(Z_0 > x) = \begin{cases} 1, & \text{if } x < 1 \\ x^{-\alpha}, & \text{if } x \geq 1, \end{cases} \tag{3.37}$$

for some $\alpha > 0$. It is elementary that for this sequence the extremal index is

$$\theta = \frac{1}{1 + p_1^\alpha + \cdots + p_q^\alpha}\,,$$

see e.g. [54].

We first test the effect of the estimators (3.6) and (3.33) on the ARMAX model using values of $\theta = 0.25, 0.5, 0.75$, and a finite sample length of $n = 10000$. For the estimator, we have chosen a block size of $r_n = 200$, and a weight function of $f(x) = e^{-x}$. We run the experiments for $m = 1, \ldots, 20$, and for each fixed $m$ we choose $u_n^s$ to be equal to the $(101 + 2(m - s))$-th largest order statistic of the sequence, $1 \leq s \leq m$. That is, each level incorporates 2 more observations above it than the level immediately higher does. When computing coefficients for the bias-reduced estimator (3.33), we use $m' = 12$, with $\bar{u}_n^s$ being the $(91+5(m'-s))$-th largest order statistic of the sequence, $1 \leq s \leq m'$, and $l = 25$, with $r_n^i = 10i$, $1 \leq i \leq l$.

We compare the estimators $\widehat{\theta}(f)$ and $\widehat{\theta}^b(f)$ on the basis of their bias, standard error, and root mean squared error. The results computed from $5000$ simulated

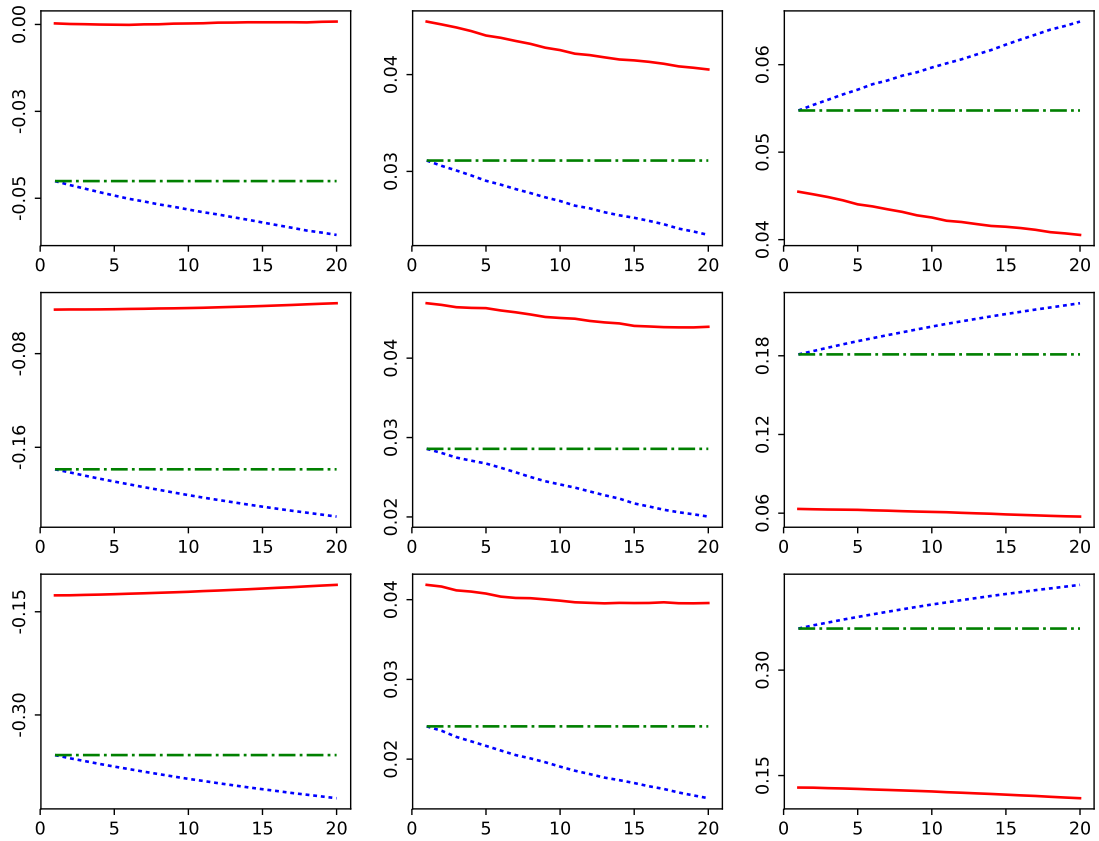


Figure 3.2: Bias (left column), standard error (center column), and root mean squared error (right column) for the naive blocks estimator (3.2) (dot-dash line), the multilevel estimator $\widehat{\theta}(f)$ (dotted line) and the bias-corrected multilevel estimator $\widehat{\theta}^b(f)$ (solid line) plotted against the choice of $m$, number of levels used in the estimators. Data are simulated from ARMAX models with $\theta = 0.25, 0.5, 0.75$ (top to bottom).

sequences are displayed in Figure 3.2. Looking from the top row to the bottom row along the varying values of $\theta = 0.25, 0.5, 0.75$, we see that the plots tell a similar story. As expected for the multilevel estimator $\widehat{\theta}(f)$, the magnitude of the bias increases while the standard error decreases as more levels of observations are incorporated into the estimator. The bias of the bias-corrected version of the estimator, $\widehat{\theta}^b(f)$, seems to be largely insensitive to the choice of $m$, with a decreasing trend both in the standard error and in the root mean squared error. Overall, $\widehat{\theta}^b(f)$ achieves a much better root mean squared error compared to $\widehat{\theta}(f)$,

for all levels $m$ considered. It also outperforms the plain block estimator in the sense of an improved root mean squared error.

We have performed the same analysis for other models, for different values of the highest threshold and for different block sizes. We have also analyzed the non-clustering case $\theta = 1$. Invariably, the qualitative structure seen on Figure 3.2 remained the same. In the remaining experiments in this section we will, therefore, focus on the best performing bias-corrected estimator $\widehat{\theta^b}(f)$ that uses the largest amount of data ($m = 20$ levels).

Our next experiment addresses the effect of the choice of block size $r_n$ on the performance of the estimator $\widehat{\theta^b}(f)$. We again use the ARMAX model with $\theta = 0.25, 0.5, 0.75$ as before. We test the performance of $\widehat{\theta^b}(f)$ using block sizes of $r_n = 40, 50, \ldots, 200$. The root mean squared errors from $5000$ simulated sequences are displayed in Figure 3.3. We see that the choice of the block size does not have a major effect on the root mean squared error.

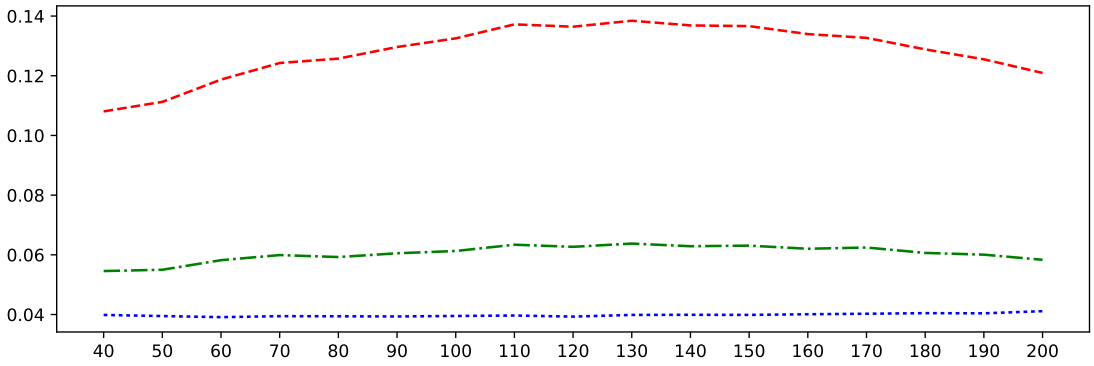


Figure 3.3: Root mean squared error for the estimator $\widehat{\theta^b}(f)$ for true values of $\theta$ being $0.25$ (dotted line), $0.5$ (dot-dash line), and $0.75$ (dash line) plotted against the choice of $r_n$, the size of the blocks used in the estimator. Data are simulated from ARMAX models with $\theta = 0.25, 0.5, 0.75$.

We have also looked at the effect of the block size on the bias and standard

error separately in Figure 3.4. Once again, the standard error is largely insensitive to the choice of the block size. The bias does vary with the block size, but remains invariably small in the absolute value, leading to the root mean squared errors displayed in Figure 3.3.

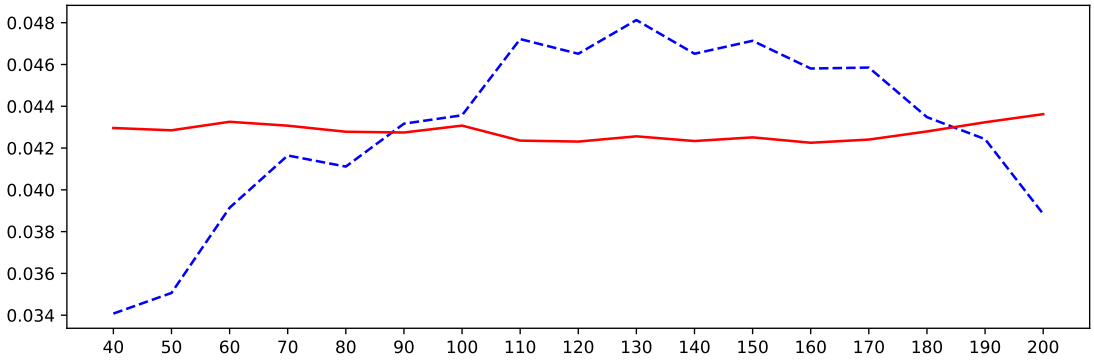


Figure 3.4: Absolute bias (dotted line) and standard error (solid line) for the estimator $\widehat{\theta}^b(f)$ plotted against the choice of $r_n$, the size of the blocks used in the estimator. Data are simulated from an ARMAX model with $\theta = 0.5$.

In the next experiment we fix the the block size to $r_n = 200$ and study the effect of the choice of the weight function. In the setting of the previous experiments we use a second weight function, $f_1(x) = 1/x^{20}$ along with the original weight function $f$. In the relevant range $f_1$ decreases at a much faster rate than $f$. We compare the performance of the estimators $\widehat{\theta}^b(f)$ and $\widehat{\theta}^b(f_1)$. The results are presented in Figure 3.5.

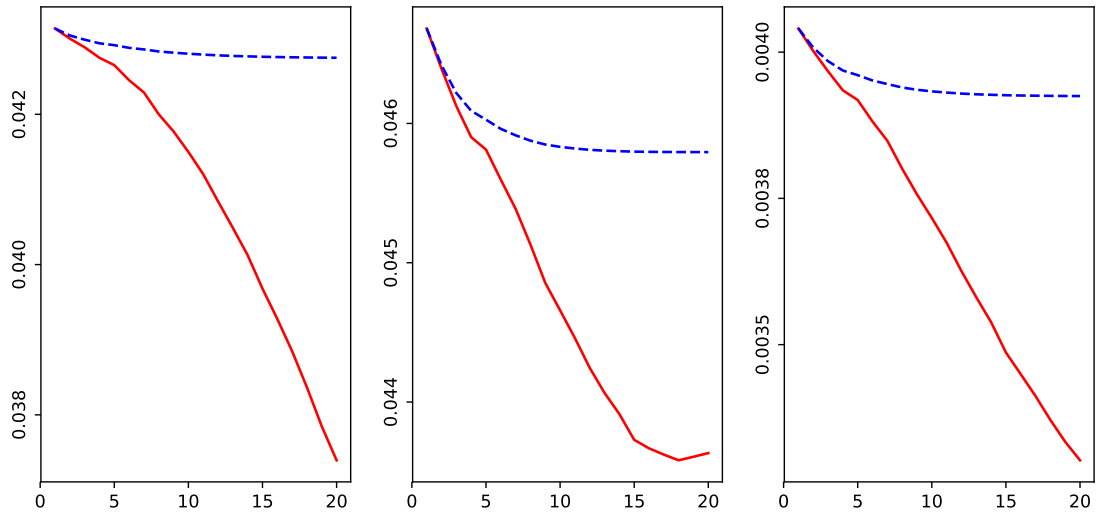


Figure 3.5: Bias (left), standard error (center) and root mean squared error (right) for the bias corrected multilevel estimators $\widehat{\theta}^b(f)$ (solid line) and $\widehat{\theta}^b(f')$ (dotted line) plotted against the choice of $m$, number of levels used in the estimators. Data are simulated from an ARMAX model with $\theta = 0.5$.

As in Figure 3.2 the magnitude of the bias, the standard error, and the root mean squared error of the estimators are all decreasing when $m$, the number of levels used in the estimator, increases. The phenomenon displayed in Figure 3.5 demonstrates that the faster decay of the weight function $f_1$ compared to $f$ leads to smaller contributions from additional levels to the efficiency of the estimator. However, the exact overall effect of the weight function on the estimator is a topic not studied in detail in this paper. It warrants further investigation.

In the previous experiments we have used samples of size $n = 10000$. Sometimes extremal inference has to be performed on data sets of a smaller size, so we have repeated our experiment leading to Figure 3.2 for samples of size $n = 5000$. We only display the results for the ARMAX model with $\theta = 0.5$. We use $r_n = 100$, and $f(x) = e^{-x}$. Once again, we experiment with $m = 1, \ldots, 20$ levels, and for each fixed $m$ we choose $u_n^s$ to be equal to the $(51+m-s)$-th largest

order statistic of the sequence, $1 \leq s \leq m$. When computing coefficients for the bias-reduced estimator, we use $m' = 12$, with $\bar{u}_n^s$ being the $(41 + 3(m' - s))$-th largest order statistic of the sequence, $1 \leq s \leq m'$, and $l = 15$, with $r_n^i = 10i$, $1 \leq i \leq l$. The results from $5000$ simulated sequences are displayed, in Figure 3.6. As expected, the smaller sample size leads to some deterioration in the quality of the estimation in comparison with the large sample size used in Figure 3.2, but the comparison of the estimators and the lessons derived from both figures remain the same.

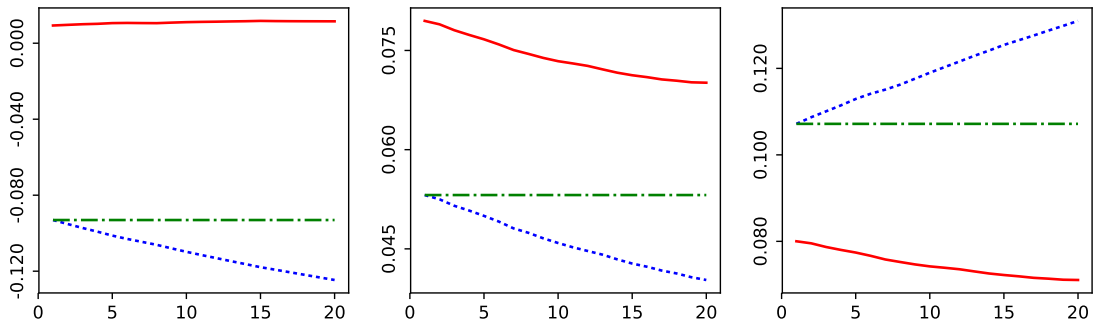


Figure 3.6: Bias (left), standard error (center), and root mean squared error (right) for the naive blocks estimator (3.2) (dot-dash line), the multilevel estimator $\widehat{\theta}(f)$ (dotted line) and the bias-corrected multilevel estimator $\widehat{\theta}^b(f)$ (solid line) plotted against the choice of $m$, number of levels used in the estimators. Data are simulated from an ARMAX model with $\theta = 0.5$.

Finally, we experiment with constructing a multiple threshold version of an estimator different from the plain block estimator. We have chosen the sliding blocks estimator of [52]. We use the ARMAX models with $\theta = 0.25, 0.5, 0.75$. For each simulated sequence, we first compute the optimal threshold as described in [52], then choose $m = 1, \ldots, 20$, where $u_n^m$ corresponds to the level of the optimal threshold, and for each $1 \leq s \leq m$, the level corresponding to $u_n^s$ incorporates $10$ more observations than the level immediately above it. The results from $5000$ simulated sequences are displayed in Figure 3.7. Once again we see that the root

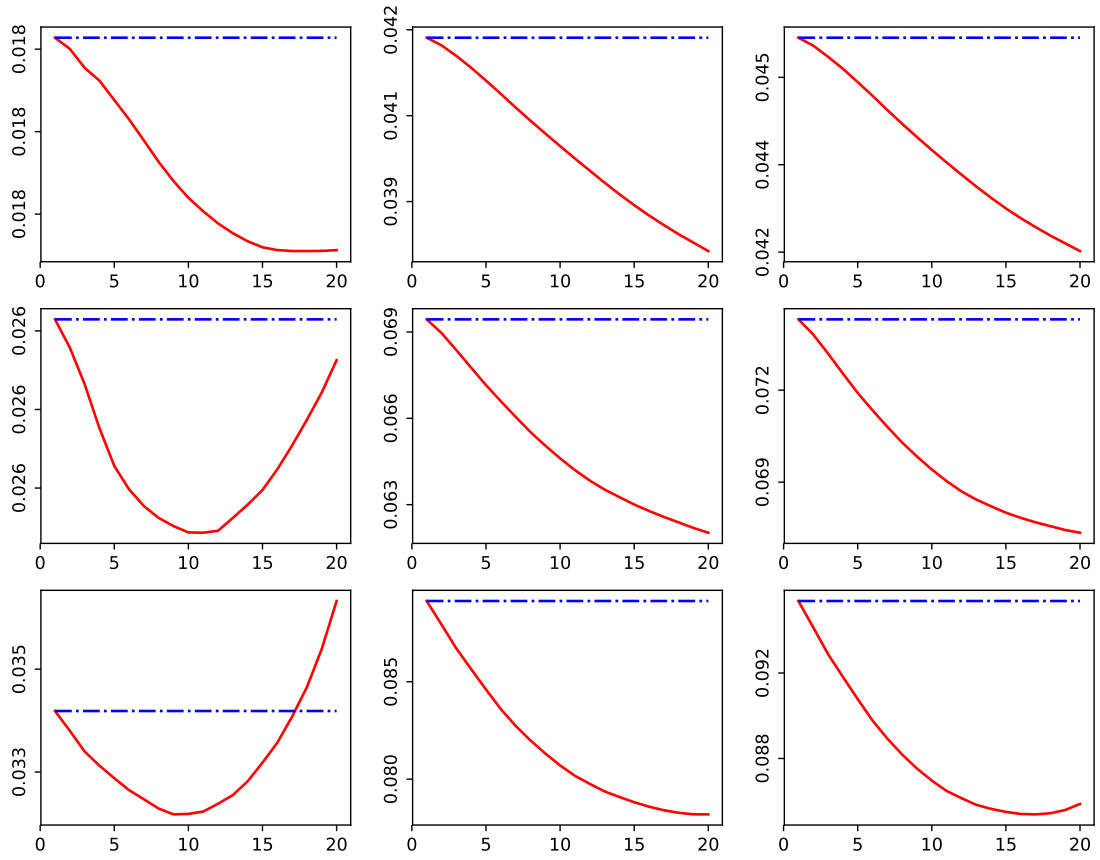


Figure 3.7: Bias (left column), standard error (center column), and root mean squared error (right column) for the sliding blocks estimator (dot-dash line), the multilevel sliding blocks estimator (solid line) plotted against the choice of $m$, number of levels used in the estimators. Data are simulated from ARMAX models with $\theta = 0.25, 0.5, 0.75$ (top to bottom).

mean squared error is almost invariably decreasing with increasing number of levels $m$.

### 3.6.2 S&P 500 Daily Log Returns

We now use the estimators developed in this paper to estimate the extremal index of the losses among the daily log returns for S&P 500 during the ten-year period between 1 January 1990 and 31 December 1999. The log returns

themselves are plotted in Figure 3.8.

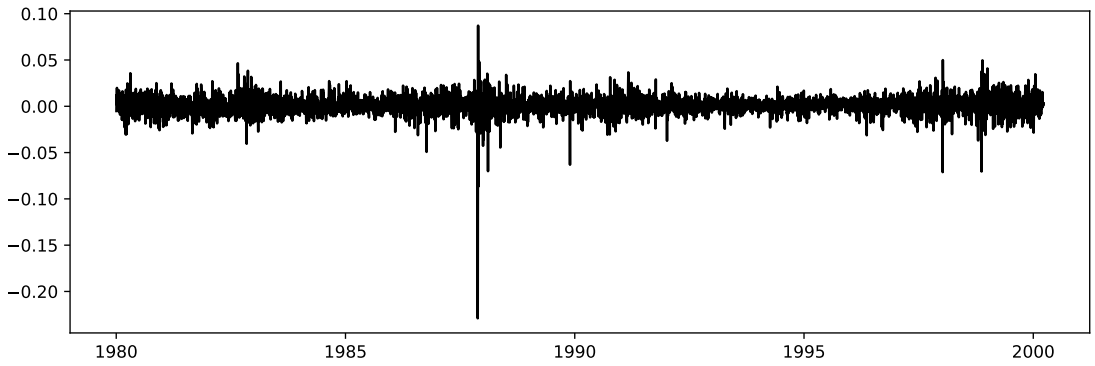


Figure 3.8: Daily Log Returns for S&P 500 from 1980 - 1999

There are $n = 5055$ returns in this data set, and the negative of their values form our sample. We choose $m = 1, \ldots, 20$ and $u_n^s$ to be the $51 + (m - s)$-th largest order statistic, $1 \leq s \leq m$. We choose the block size $r_n = 40$, resulting in $k_n = 126$ blocks. For the weight function we use $f(x) = e^{-x}$. When computing the bias-corrected estimator we use (3.34) with $m' = 12$ levels, $\bar{u}_n^s$ being the $41 + 3(m' - s)$-th largest order statistic in the sample, $1 \leq s \leq m'$, and set $l = 15$, with $r_n^i = 10i$, $1 \leq i \leq l$.

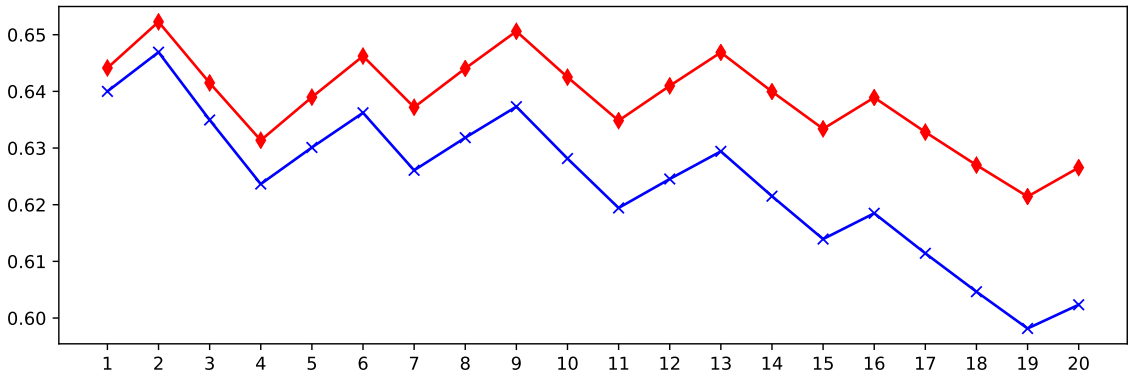


Figure 3.9: The values of the multilevel estimator $\widehat{\theta}(f)$ ('x' marker) and the bias-corrected multilevel estimator $\widehat{\theta}^b(f)$ (diamond marker) plotted against the choice of $m$, number of levels used in the estimators, for the negative daily log returns of S&P 500.

The plots of the two estimators are shown above as a function of the number of levels $m$. We have also evaluated the variability of the estimators by performing a block-level bootstrap. We have not presented the resulting pointwise 1-standard error confidence intervals on Figure 3.9 since this makes the structure of the pointwise estimators harder to see, but the order of magnitude of these intervals is $[0.5, 0.8]$.

CHAPTER 4

# A DISTANCE BASED DECLUSTERING METHOD FOR EXTREMAL INFERENCE

## 4.1 Introduction

In this chapter we approach the problem of extremal inference under dependence. Many extremal estimators implicitly or explicitly assume the independence of input variables. When estimating the parameters of the generalized extreme value distribution, for example, one often uses the maximum likelihood estimator, which requires the inputs to be independent and identically distributed to obtain a tractable likelihood function. In the classical Hill estimator for the exponent of regular variation, one often selects a number of upper statistics for inference, whereby the method of selecting such an "optimal" number of observations often assumes that the input observations are independent; see e.g. [41] and [40]. As a result, when dependence exists among the observations, many such estimators become inapplicable or intractable, such as ones using the maximum likelihood method. Alternatively, for example, in the case of the Hill estimator, one can still guarantee asymptotic consistency under certain conditions; see e.g. [30]. However, in this case, methods of selecting the number of upper statistics in the i.i.d. case can no longer guarantee "optimality" when dependence is present.

We focus on cases when data exhibits serial dependence. Specifically, we will assume that our data $X_1, X_2, \ldots$ is a stationary sequence. Some methods have been proposed to overcome the issue of serial dependence. However, many such methods often make specific model assumptions and are therefore diffi-

cult to generalize; see e.g. [48] and [28]. Such methods usually use all the data above a certain threshold. We believe that there is room for improvement when applying such methods: as seen in Chapter 3, large observations in a stationary sequence tend to cluster together. By taking all values above a certain threshold for an estimator, one may end up using observations within the same cluster, which are highly dependent with each other, and thereby "polluting" the estimator. Our approach, therefore, is to attempt to select observations from different clusters so as to minimize the clustering among the inputs used for inference.

We will apply our approach to the Hill estimator for the index of regular variation. Recall that a distribution function $F$ is regularly varying with index or exponent $\alpha$ if for any $t > 0$,

$$\lim_{x\to\infty} \frac{F(tx)}{F(x)} = t^{-\alpha}\,. \tag{4.1}$$

Data displaying heavy-tailed or regular variation behavior are encountered in many fields such as hydrology, finance, and network traffic. A number of estimators exist for estimating the regular variation index. The best known estimator among those is the Hill estimator, introduced by [27]. It is defined as follows. For $1 \le i \le n$, let $X_{i,n}$ denote the $i$-th smallest order statistic from the sequence $\{X_i\}$. Then the Hill estimator based on the $k$ largest order statistics is defined as

$$H_{k,n} = \frac{1}{k}\sum_{i=0}^{k-1} \log \frac{X_{n-i,n}}{X_{n-k,n}}. \tag{4.2}$$

When $X_1, \ldots, X_n$ are i.i.d., and $k = o(n)$ as $n \to \infty$, then it is known that $H_{k,n}$ is an asymptotically consistent estimator of $\gamma = 1/\alpha$; see e.g. [37]. It is also shown that under suitable mixing conditions, the Hill estimator is also asymptotically consistent for dependent sequences; see e.g. [25], [12], [11].

In practice, the problem of choosing an appropriate number $k$ of upper order statistics is quite difficult. The Hill estimator is usually plotted against a range of $k$, resulting in a "Hill plot", and the value around which the plot stabilizes is chosen to be the Hill estimate. The procedure is difficult use due to high sensitivity of the estimator to the choice of $k$, and often even the Hill plot is visually difficult to interpret. Many efforts have been put into finding a systematic way of selecting an "optimal" number of upper order statistics $k$ for the Hill estimator; see e.g. [26], [17], [41]. However, those methods all work under the assumption that the input variables are i.i.d., and the "optimality" conditions fail when serial dependence is introduced. In [41], the authors devise a procedure to sequentially test the samples $\{\log \frac{X_{n-i,n}}{X_{n-k,n}} : i = 0, 1, \ldots, k-1\}$ for the null hypothesis of exponential distribution, in order to determine the number $k$ of upper order statistics. We will apply our approach to select observations from different clusters, and use the method in [41] on those observations.

This chapter is organized as follows. We build intuition and propose an algorithm for selecting observations, and list the assumptions used in the chapter in Section 4.2. In Section 4.3 we apply this algorithm in conjunction with the method proposed in [41], and show that asymptotic consistency results still hold. In Section 4.4 we present a simulation study.

## 4.2 The Selection Algorithm

In this section we propose our algorithm for data selection. First let us let us take a look at a couple of point processes. Suppose $X_1, X_2, \ldots$ is a stationary sequence of random variables with marginal distribution $F$. For two observa-

tions $X_i, X_j$, we will refer to $|i-j|$ as the *distance* between the two elements. For notational convenience, for the rest of this chapter we will use the simplifying assumption that $F$ is supported by $[0,\infty)$. The case when $F$ is supported by $(-\infty,\infty)$ is similar. Let $(v_n)$ be a sequence such that $n\overline{F}(v_n) \to 1$ as $n \to \infty$.

Recall from Chapter 3 that under certain mixing conditions, the point processes of exceedances converge weakly in the space of finite point processes on $[0,1]$ to a compound Poisson process:

$$N_n = \sum_{i=1}^{n} \delta_{i/n} \mathbb{1}(X_i > v_n) \xrightarrow{d} N = \sum_{i=1}^{\infty} \xi_i \delta_{\Gamma_i}, \tag{4.3}$$

where $\delta_x$ is a point mass at $x$, the points $0 < \Gamma_1 < \Gamma_2 < \dots$ constitute a homogeneous Poisson process with intensity $\theta$ on $[0,1]$ which is independent of an i.i.d. positive integer-valued sequence $\{\xi_i\}$.

Now let us look at a slightly different point process. Let $\tilde{N}_n$ be the point process on $\mathbb{R} - \{0\}$ define as

$$\tilde{N}_n = \sum_{i=1}^{n} \delta_{X_i/v_n}. \tag{4.4}$$

Note that the state space is $\mathbb{R}-\{0\}$ in order to avoid a buildup of mass at $0$ as $n \to \infty$. Then, assuming that the marginal distribution $F$ of $X_1$ is regularly varying with some index $\alpha$, and again under some general mixing conditions, as $n \to \infty$, the point process $\tilde{N}_n$ converges weakly to a point process with representation

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \delta_{P_i Q_{ij}}, \tag{4.5}$$

where $\sum_{i=1}^{\infty} \delta_{P_i}$ is a Poisson process on $(0,\infty)$ with intensity measure $\nu$, where for $x > 0$, $\nu([x,\infty)) = \theta x^{-\alpha}$ with $\theta$ being the extremal index of the sequence $\{X_i\}$; and $\sum_{j=1}^{\infty} \delta_{Q_{ij}}$, $i \geq 1$ are i.i.d. point processes on $(0,1]$ also independent of the Poisson process, with the property that $\max_j Q_{ij} = 1$ for all $i \geq 1$; see e.g. [14] and [39].

The above results, along with results from Chapter 3, give a few important pieces of information that will be useful to us. There are two types of point processes associated with the convergence of $\tilde{N}_n$ in (4.5). The first type the Poisson point process $\sum_{i=1}^{\infty} \delta_{P_i}$ on $(0, \infty)$. Note that this point process has mean measure $\nu([x, \infty)) = \theta x^{-\alpha}$ for $x > 0$. This is a key ingredient in the procedure proposed in [41]: if $X_1, \ldots, X_n$ are i.i.d. and regularly varying, then the largest observations among the sequence resemble the largest points of a Poisson process with power intensity. The other type of point process resulting from the convergence of $\tilde{N}_n$ is of the form $\sum_{j=1}^{\infty} \delta_{Q_{ij}}$ for $i \geq 1$. The properties relevant to us from this set of point processes are that $\sum_{j=1}^{\infty} \delta_{Q_{ij}}$ are i.i.d., and that $\max_j Q_{ij} = 1$ for each $i \geq 1$. In light of this, one can view the convergence of $\tilde{N}_n$ as follows. For each $n$ and each $i \geq 1$, there is a "cluster" of exceedances within the point process $\sum_{i=1}^{n} \delta_{X_i/v_n}$, such that as $n \to \infty$, such a "cluster" converges in distribution to $\sum_{j=1}^{\infty} \delta_{P_i Q_{ij}}$. In turn, since $\max_j Q_{ij} = 1$ for each $i \geq 1$, $P_1, P_2, \ldots$ also correspond to the distributional limits of the largest observations within each "cluster". Therefore by extracting the largest observation within each "cluster", one hopes to produce points that resemble the largest points of a Poisson process with power intensity. This will be the goal of our selection algorithm. Note that in the above discussion, the concept of a "cluster" is not clearly defined. However, from (4.3) we observe that as $n$ tends to infinity, the distances between such "clusters" of exceedances also tend to infinity. Therefore, our algorithm does not need to be able to identify "clusters", but instead, intuitively, it should attempt to select large observations with large distances between each other.

Let $X_1, X_2, \ldots, X_n$ be $n$ contiguous observations from a stationary sequence of random variables with marginal distribution $F$. Let $(u_n)$ be a sequence of thresholds. Recall the blocks estimator (3.2). It is based on choosing some block

size $r_n$, and dividing the $n$ observations into $k_n = \lfloor n/r_n \rfloor$ contiguous blocks of equal length $r_n$. Again denoting $M_{i,j}$ to be $\max\{X_{i+1}, \ldots, X_j\}$ with the convention of $M_j = M_{0,j}$, the blocks estimator is

$$\widehat{\theta}_n = \frac{\sum_{i=1}^{k_n} \mathbb{1}\big(M_{(i-1)r_n, ir_n} > u_n\big)}{\sum_{i=1}^{k_n r_n} \mathbb{1}\big(X_i > u_n\big)}.$$

Let $\overline{\Delta}_n$ denote the set of $X_i$ among the $k_n$ blocks such that $X_i > u_n$. Namely,

$$\overline{\Delta}_n = \{X_i \mid X_i > u_n, 1 \leq i \leq k_n r_n\}. \tag{4.6}$$

Further, let recall from Chapter 3 that for $w > 0$, $N_k(w) = \sum_{i=1}^{k} \mathbb{1}(X_i > w)$. We can denote the size of the set $\overline{\Delta}_n$,

$$N_{k_n r_n}(u_n) = \sum_{i=1}^{k_n r_n} \mathbb{1}\big(X_i > u_n\big) = |\overline{\Delta}_n|. \tag{4.7}$$

We interpret $\overline{\Delta}_n$ as all the observations that form clusters of exceedances. In this case we wish to select from $\overline{\Delta}_n$ the largest observation of each cluster. We proceed as follows.

**Algorithm 1** Distance Data Selection

**for** $l = 1, 2, \ldots$ **do**
  **while** $\exists X_i, X_j \in \overline{\Delta}_n$, such that $X_i < X_j, |i - j| = l$ **do**
    Take such a pair $X_i, X_j$ with the minimal $\min(i, j)$
    **if** $|\overline{\Delta}_n| > \lfloor \widehat{\theta}_n N_{k_n r_n}(u_n) \rfloor$ **then**
      $\overline{\Delta}_n \leftarrow \overline{\Delta}_n \setminus \{X_i\}$
    **else**
      **return** $\overline{\Delta}_n$

To avoid confusion, unless otherwise specified, we will use $\Delta_n$ to refer to the set returned from Algorithm 1, and $\overline{\Delta}_n$ to refer to the original set of exceedances over $u_n$ as in (4.6).

First let us note that $\lfloor \widehat{\theta}_n N_{k_n r_n}(u_n) \rfloor$ is the number of elements remaining in the set $\Delta_n$, the output of the algorithm. Intuitively, if one interprets the extremal index $\theta$ as the reciprocal of the expected extremal cluster size as mentioned in Chapter 3, and $N_{k_n r_n}(u_n)$ as a finite sample approximation to the extremal exceedances, $\lfloor \widehat{\theta}_n N_{k_n r_n}(u_n) \rfloor$ would be an approximation to the number of extremal clusters in the sample. Given the behavior that the distances between extremal clusters tend to infinity as the sample size $n \to \infty$, and the distance-based selection method in Algorithm 1, $\Delta_n$ should be retaining one point from each extremal cluster approximated from $\overline{\Delta}_n$.

The above discussion only offers intuition, and we will need mixing assumptions to guarantee the result. Many of those assumptions will be in a similar vein to those presented in Chapter 3. For completeness' sake we will list them here. In the next section we will apply the above algorithm in the case of the Hill estimator. Certain assumptions will need to apply for asymptotic consistency results. We list those assumptions here as well.

Recall the $\alpha$-coefficient. For a value $u_n > 0$, and $1 \leq i \leq j \leq n$, let $\mathscr{B}_i^j(u_n)$ denote the $\sigma$-field generated by events $\{X_t \leq u_n\}$ for $i \leq t \leq j$. For $n \geq 1$ and $1 \leq l \leq n-1$, the $\alpha$-coefficient is defined as

$$\alpha_{n,l}(u_n) = \max(|P(A \cap B) - P(A)P(B)| : \\ A \in \mathscr{B}_1^k(u_n), B \in \mathscr{B}_{k+l}^n(u_n), 1 \leq k \leq n-l).$$

Again we will say that the sequence $\{X_i\}$ satisfies the condition $\Delta(\{u_n\})$ if $\alpha_{n,l_n}(u_n) \to 0$ as $n \to \infty$ for some sequence $\{l_n\}$ with $l_n = o(n)$. If $p_n$ is a sequence of integers and $\alpha_{p_n,l_n}(u_n) \to 0$ as $n \to \infty$ for some sequence $\{l_n\}$ with $l_n = o(p_n)$, then we will say that $\{X_i\}$ satisfies the condition $\Delta_{\{p_n\}}(\{u_n\})$.

We introduce another mixing coefficient, the $\beta$-coefficient. For the assumptions involved in this chapter, we will define them over slightly different $\sigma$-fields. For a value $u_n > 0$, and $1 \leq i \leq j \leq n$, let $\mathscr{C}_i^j(u_n)$ denote the $\sigma$-field generated by random variables $X_t \mathbb{1}(X_t > u_n)$ for $i \leq t \leq j$. For $n \geq 1$ and $1 \leq l \leq n-1$, the $\beta$-coefficient is defined as

$$\beta_{n,l}(u_n) = \max_k \sup \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} |P(A_i \cap B_j) - P(A_i)P(B_j)|, \tag{4.8}$$

where the supremum is taken over all finite partitions $\{A_1, \ldots, A_I\}$ and $\{B_1, \ldots, B_J\}$ of the probability space, such that $A_i \in \mathscr{C}_1^k(u_n)$ for each $1 \leq i \leq I$ and $B_j \in \mathscr{C}_{k+l}^n(u_n)$ for each $1 \leq j \leq J$, and the maximum is taken over all $1 \leq k \leq n-l$.

One can show that for all $1 \leq l \leq n-1$, $2\alpha_{n,l}(u_n) \leq \beta_{n,l}(u_n)$. For a more comprehensive review and comparison of mixing coefficients, see the survey paper [7].

Similarly to Chapter 3, we would like to ensure the asymptotic consistency of the blocks estimator $\widehat{\theta}_n$. For such consistency results we will require that $r_n \overline{F}(u_n) \to 0$ and $n\overline{F}(u_n) \to \infty$. This implicitly assumes that $r_n = o(n)$ as $n \to \infty$. Again we introduce a sequence $\{p_n\}$ of integers with

$$p_n \overline{F}(u_n) \to \tau \tag{4.9}$$

for some $\tau > 0$. This will imply that $r_n = o(p_n)$ as $n \to \infty$. The following assumptions on the stationary sequence $\{X_i\}$ will be used throughout this chapter.

**Assumption** $\Delta''$ There is a sequence $l_n = o(r_n)$ such that $k_n \alpha_{n,l_n}(u_n) \to 0$ as $n \to \infty$. Further assume that $u_n$ is large enough that $k_n l_n = o(p_n)$ and $r_n^2 k_n / p_n^2 \to 0$ as $n \to \infty$.

**Assumption B** There is a sequence $l_n = o(r_n)$ such that $k_n\beta_{n,l_n}(u_n) \to 0$ as $n \to \infty$.

In Chapter 3 we have already gone over a set of mixing conditions that ensure the asymptotic consistency of the blocks estimator. In particular, we showed that with proper scaling, the numerator and the denominator in the blocks estimator converge "in the right way"; see Proposition 3.5.1. In this chapter, we are not concerned with specific conditions ensuring the asymptotic consistency. Rather, we just require that the ingredients of the blocks estimator are consistent as follows,

**Assumption E** We assume that

$$\frac{p_n}{n} N_{k_n r_n}(u_n) \to_P \tau, \quad \frac{p_n}{n} \sum_{i=1}^{k_n} \mathbb{1}(M_{(i-1)r_n, ir_n} > u_n) \to_P \tau\theta.$$

The next group of assumptions deals with convergence of $N_{p_n}^{(u)}$, the point process on $[0,1]$ with points $(j/p_n : 1 \le j \le p_n,\ X_j > u_n)$.

**Assumption P** $N_{p_n}^{(u)}$ converges weakly in the space of finite point processes on $[0,1]$.

**Assumption D** There exists a probability distribution $(\pi_j)_{j\ge1}$ on the positive integers such that as $n \to \infty$,

$$P(N_{r_n}(u_n) = j | M_{r_n} > u_n) \to \pi_j,\ j \ge 1,$$
$$E[N_{r_n}(u_n) | M_{r_n} > u_n] \to \sum_{j=1}^{\infty} j\pi_j < \infty.$$

Lastly, we will be using the notion of second order regular variation in the next section. We assume that the marginal distribution $F$ of the stationary se-

quance $\{X_i\}$ is regularly varying with exponent $\alpha$. Let $U = \left(\frac{1}{1-F}\right)^{\leftarrow}$ be the generalized inverse function.

**Assumption R** There exists $\rho < 0$ and a function $A$ regularly varying at infinity with exponent $\rho$ such that

$$\lim_{r\to\infty} \frac{\frac{U(rx)}{U(r)} - x^{\gamma}}{A(r)} = x^{\gamma}\frac{x^{\rho}-1}{\rho}, \tag{4.10}$$

where $\gamma = 1/\alpha$. We also assume, without loss of generality, that $A$ is continuous and $|A|$ is eventually decreasing.

**Remark 4.2.1.** Assumption $B$ implies the first part of Assumption $\Delta''$. Assumption $P$ is the same as that in Section 3.3. Assumption $D$ is a weaker version of Assumption $D_1$ in Section 3.3. Assumption $E$ implies that $\widehat{\theta}_n \to_P \theta$. Aside from those listed in Chapter 3, there are several sets of mixing conditions through which this holds; see e.g. [30], [52], [58].

**Remark 4.2.2.** The mixing conditions in $\Delta''$ mainly concern the block size $r_n$, which must be "large enough", and the threshold $u_n$, which must be "low enough". Sequences such that $\beta_{n,l_n}(u_n) \to 0$ are called *absolutely regular* sequences, and the mixing condition in $B$ dictate that the block sizes must be "large enough". Many models that satisfy the assumptions in Chapter 3 - in which the mixing coefficients decay quickly (e.g. $m$-dependent sequences, geometrically mixing sequences) - also satisfy the assumptions posed here.

Based on some of the above assumptions, we can make a claim regarding the minimum distance between the elements of the set $\Delta_n$.

**Proposition 4.2.3.** *If Assumptions $\Delta''$, $P$, and $D$ are satisfied, then as $n \to \infty$,*

$$P(\min_{X_i, X_j \in \Delta_n} |i-j| > r_n) \to 1.$$

*Proof.* We divide $X_1, \ldots, X_n$ into $k_n$ contiguous blocks of size $r_n$. If there exist some $X_i, X_j \in \Delta_n$ with $|i - j| \leq r_n$, then either $X_i, X_j$ are in the same block, or they are in neighboring blocks. First let us look at the possibility that they are in neighboring blocks. If Assumptions $\Delta''$, $P$, and $D$ are satisfied, then by Theorem 5.1 and Lemma 2.3 of [32],

$$P(M_{r_n} > u_n) \sim \tau\theta r_n/p_n. \tag{4.11}$$

So the probability that there are exceedances over $u_n$ in two neighboring blocks is

$$\begin{aligned} P(M_{r_n} > u_n, M_{r_n,2r_n} > u_n) \leq & P(M_{r_n-l_n} > u_n, M_{r_n,2r_n} > u_n) + P(M_{l_n} > u_n) \\ \leq & P(M_{r_n-l_n} > u_n)P(M_{r_n,2r_n} > u_n) \\ & + P(M_{l_n} > u_n) + \alpha_{n,l_n} \\ \leq & P(M_{r_n} > u_n)^2 + P(M_{l_n} > u_n) + \alpha_{n,l_n}. \end{aligned}$$

Hence the probability that there exists two neighboring blocks with exceedances over $u_n$ is bounded above by

$$\begin{aligned} k_n P(M_{r_n} > u_n, M_{r_n,2r_n} > u_n) \leq & k_n(P(M_{r_n} > u_n)^2 + P(M_{l_n} > u_n) + \alpha_{n,l_n}) \\ \sim & k_n(\tau\theta r_n/p_n)^2 + k_n P(M_{l_n} > u_n) + k_n\alpha_{n,l_n}. \end{aligned}$$

By Assumption $\Delta''$, the first and third terms of the above converge to $0$ as $n \to \infty$. And the second term is bounded above by

$$k_n l_n P(X_1 > u_n) \sim k_n l_n \frac{\tau}{p_n},$$

which again converges to $0$ as $n \to \infty$, since $k_n l_n = o(p_n)$ by Assumption $\Delta''$. We conclude that the probability that there are two consecutive blocks with exceedances over $u_n$ converges to $0$.

Now let us consider the probability that there are no two consecutive blocks with exceedances over $u_n$, but $\Delta_n$ contains two elements in the same block. For notational convenience, for $1 \leq i \leq k_n$, let $B_i$ denote the block (set) $\{X_{(i-1)r_n+1}, \ldots, X_{ir_n}\}$. By definition of $\widehat{\theta}_n$ and Algorithm 1, it follows that

$$|\Delta_n| = \lfloor N_{k_n r_n}(u_n) \times \frac{\sum_{i=1}^{k_n} \mathbb{1}(\max B_i > u_n)}{N_{k_n r_n}(u_n)} \rfloor = \sum_{i=1}^{k_n} \mathbb{1}(\max B_i > u_n).$$

(Note here that $\max B_i = M_{(i-1)r_n, ir_n}$.) So if there are two elements $X_i, X_j \in \Delta_n$ in the same block, then by pigeonhole there must be some block $B_k$ such that $\max B_k > u_n$, and $B_k \cap \Delta_n = \emptyset$. Therefore there must be some element $X_m \in B_k$ with $X_m > u_n$ such that $X_m$ was removed from $\overline{\Delta}_n$ during Algorithm 1. Let $X_m$ be the last such element from block $B_k$ that was removed. Then it must be that there was some $X_{m'} \in \Delta_n$ that was "paired" with $X_m$ during the algorithm that resulted in the removal $X_m$. This implies that $|m' - m|$ was the minimum distance between any two elements that were still in the set $\overline{\Delta}_n$. Since there are no consecutive blocks with exceedances over $u_n$, and $X_m$ is the last element to be removed from block $B_k$, it must be that $|m' - m| > r_n$. However, $X_i, X_j$ are never removed and in the same block, so $|i - j| < r_n < |m' - m|$. This is a contradiction. Hence the situation that there are no two consecutive blocks with exceedances, but $\Delta_n$ contains two elements in the same block cannot occur.

Combining the two scenarios above, it follows that $P(\min_{X_i, X_j \in \Delta_n} |i - j| > r_n) \to 1$ as $n \to \infty$. □

In light of Proposition 4.2.3, let us consider the following algorithm.

**Algorithm 2** Block Data Selection

1: $\tilde{\Delta}_n \leftarrow \emptyset$
2: **for** $i = 1, 2, \ldots, k_n$ **do**
3: **if** $M_{(i-1)r_n, ir_n} > u_n$ **then**
4: $\tilde{\Delta}_n \leftarrow \tilde{\Delta}_n \cup \{M_{(i-1)r_n, ir_n}\}$
5: **return** $\tilde{\Delta}_n$

To avoid confusion, we will use $\Delta_n$ to denote the output of Algorithm 1, and $\tilde{\Delta}_n$ to denote the output of Algorithm 2. The following result is immediate from the proof of Proposition 4.2.3.

**Corollary 4.2.4.** *If Assumptions $\Delta''$, $P$, and $D$ are satisfied, then as $n \to \infty$,*

$$P(\Delta_n \neq \tilde{\Delta}_n) \to 0.$$

## 4.3 The Hill Estimator

In this section we will use our selection algorithm in the inference of the regular variation index $\alpha$ of $F$, the marginal distribution function of the sequence $\{X_i\}$. We base our procedure on that proposed in [41]. In the paper, the authors observe that if $V_1 > V_2 > \cdots > V_k$ are the largest points of a Poisson process on $(0, \infty)$ with mean measure $\mu_*(x, \infty) = x^{-\alpha}$, then the set $\{V_i/V_k, i = 1, \ldots, k-1\}$ forms an i.i.d. sample from the Pareto distribution with tail $x^{-\alpha}$, $x > 1$, and therefore taking the logarithms of the elements in the set would result in an i.i.d. sample of exponential random variables with mean $\gamma = 1/\alpha$. Further, if the sequence of random variables $X_1, \ldots, X_n$ are i.i.d. and exhibit regular variation, then the largest order statistics should resemble points of a Poisson process with power intensity. It is therefore logical to test the samples

$\{\log \frac{X_{n-i,n}}{X_{n-k,n}} : i = 0, \ldots, k-1\}$ for the null hypothesis of exponential distribution, and select the largest $k$ at which the null hypothesis is not rejected as the choice of sample fraction $k$ used in the Hill estimator.

When the sequence $\{X_i\}$ exhibits serial dependence, we modify the above procedure. Recall the set $\Delta_n$, the output of our selection algorithm, from Section 4.2. For notational convenience, let $L_n$ denote the number of elements in $\Delta_n$, and let $\bar{X}_1, \ldots, \bar{X}_{L_n}$ denote the random variables in the set $\Delta_n$. Further let $\bar{X}_{1,L_n} \leq \cdots \leq \bar{X}_{L_n,L_n}$ denote the order statistics from $\Delta_n$. Set

$$Q_{k,n} = \frac{\sqrt{k}}{2} \left( \frac{\frac{1}{k} \sum_{i=0}^{k-1} \left( \log \frac{\bar{X}_{L_n-i,L_n}}{\bar{X}_{L_n-k,L_n}} \right)^2}{\left( \frac{1}{k} \sum_{i=1}^{k-1} \log \frac{\bar{X}_{L_n-i,L_n}}{\bar{X}_{L_n-k,L_n}} \right)^2} - 2 \right). \tag{4.12}$$

This is the moment statistic for the test of exponentiality. It can be shown that its large sample distribution under the null hypothesis of exponentiality is the standard normal distribution; see [13]. The authors in [41] then set an increasing series of critical values for the test statistic, and we will do so here similarly. Set $\omega$ to be some critical value, e.g. a $99\%$ quantile with respect to the limiting standard normal distribution. And let $\phi_n \to \infty$ be an increasing sequence as $n \to \infty$. Set

$$U_n = \min(\inf \left\{ k : 1 \leq k \leq L_n, |Q_{k,n}| \geq \omega \sqrt{\frac{\phi_n}{k}} \right\}, L_n). \tag{4.13}$$

We then use the $U_n$ upper order statistics from $\Delta_n$ to estimate $\gamma = 1/\alpha$ using the Hill estimator:

$$H_{U_n,n} = \frac{1}{U_n} \sum_{i=0}^{U_n-1} \log \frac{\bar{X}_{L_n-i,L_n}}{\bar{X}_{L_n-U_n,L_n}}. \tag{4.14}$$

We will show that under mild conditions, $H_{U_n,n} \to_P \gamma$. Many of the arguments will be similar to those in [41]. We proceed as follows. For $n \geq 1$, $j = 1, 2$,

define

$$M^j_{\phi_n,n}(t) = \begin{cases} 0, & \text{if } 0 \le t \le \frac{1}{\phi_n} \\ \frac{1}{\lfloor \phi_n t \rfloor} \sum_{i=0}^{\lfloor \phi_n t \rfloor - 1} \left( \log \frac{\bar{X}_{L_n - i, L_n}}{\bar{X}_{L_n - \lfloor \phi_n t \rfloor, L_n}} \right)^j, & \text{if } \frac{1}{\phi_n} \le t \le \frac{L_n}{\phi_n} \\ \frac{1}{L_n} \sum_{i=0}^{L_n - 1} \left( \log \frac{\bar{X}_{L_n - i, L_n}}{\bar{X}_{1, L_n}} \right)^j, & \text{if } t > \frac{L_n}{\phi_n}. \end{cases}$$

We will be working with spaces of type $D[0,\infty)$, $D^2[0,\infty)$, $D[\delta,\infty)$, and $D^2[\delta,\infty)$ for $\delta > 0$. The $D^2$ spaces will be endowed with the $J_1$ topology. See e.g. [59] for a comprehensive review of the $J_1$ topology.

**Theorem 4.3.1.** *Suppose Assumptions $\Delta''$, $B$, $E$, $P$, $D$, and $R$ are satisfied. Let $\{\phi_n\}$ be an increasing sequence with $\phi_n \to \infty$ and $\phi_n = o(n^{\frac{2|\rho|}{1+2|\rho|}})$, $\phi_n = o(n/p_n)$ as $n \to \infty$, then*

$$\sqrt{\phi_n} t \begin{pmatrix} \frac{M^1_{\phi_n,n}(t)}{\gamma} - 1 \\ \frac{M^2_{\phi_n,n}(t)}{\gamma^2} - 2 \end{pmatrix} \Rightarrow \begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix}$$

*in $D^2[0,\infty)$, where $((W_1(t), W_2(t)), t \ge 0)$ is a two-dimensional zero mean Brownian motion with covariance matrix*

$$\begin{bmatrix} 1 & 4 \\ 4 & 20 \end{bmatrix}.$$

*Proof.* Since we are only dealing with elements $X_i$ of $X_1, \ldots, X_n$ such that $X_i > u_n$, we can without loss of generality replace $X_1, \ldots, X_n$ with $X_1 \mathbb{1}(X_1 > u_n), \ldots, X_n \mathbb{1}(X_n > u_n)$.

We begin by using another sequence $X'_1, \ldots, X'_n$ that is somewhat "coupled" to $X_1, \ldots, X_n$.

By Berbee's Lemma (see e.g. Theorem 5.1 of [51]), for a sequence $Y_1, \ldots, Y_n \ge 0$, and for each $1 \le i \le k_n - 1$, there exists a sequence $\tilde{Y}^i_1, \ldots, \tilde{Y}^i_n$ with the same

joint distribution as $Y_1, \ldots, Y_n$ that is independent of $Y_1, \ldots, Y_{(i-1)r_n}$, and

$$P(Y_k \neq \tilde{Y}_k^i \text{ for some } k > ir_n) \leq \beta_{n,r_n}^Y(0),$$

where in this case $\beta_{n,r_n}^Y(0)$ is the $\beta$-coefficient defined with respect to the sequence $Y_1, \ldots, Y_n$ as in (4.8). We will call such a sequence $\tilde{Y}_1^i, \ldots, \tilde{Y}_n^i$ a coupling sequence of $Y_1, \ldots, Y_n$ independent of $Y_1, \ldots, Y_{(i-1)r_n}$. Note here we only assumed that $Y_1, \ldots, Y_n \geq 0$ for notational convenience.

In light of Berbee's Lemma, we claim that there exists a sequence $X_1', \ldots, X_n'$ such that for $i = 2, \ldots, k_n$, $X_{(i-1)r_n+1}', \ldots, X_n'$ is independent of $X_1', \ldots, X_{(i-1)r_n}'$, and $P(X_k \neq X_k' \text{ for some } k) \leq k_n \beta_{n,r_n}(u_n)$.

In fact, if for any arbitrary sequence $Y_1, \ldots, Y_n$ and for any $1 \leq i \leq k_n - 1$, such a coupling sequence in Berbee's Lemma can be constructed, then one can construct $X_1', \ldots, X_n'$ as follows.

---

1: **for** $i = 1, 2, \ldots, k_n - 1$ **do**
2: Construct a coupling sequence $\tilde{X}_1^i, \ldots, \tilde{X}_n^i$ of $X_1, \ldots, X_n$ independent of $X_1, \ldots, X_{(i-1)r_n}$
3: $(X_1, \ldots, X_n) \leftarrow (X_1, \ldots, X_{ir_n}, \tilde{X}_{ir_n+1}^i, \ldots, \tilde{X}_n^i)$
4: **return** $(X_1, \ldots, X_n)$

---

We denote the output of the above algorithm by $X_1', \ldots, X_n'$. By construction, for $i = 2, \ldots, k_n$, $X_{ir_n+1}', \ldots, X_n'$ is independent of $X_1', \ldots, X_{(i-1)r_n}'$. Further, if $X_k \neq X_k'$ for some $k$, it must be that on some iteration $i$ in the above algorithm, the elements $\tilde{X}_{ir_n+1}^i, \ldots, \tilde{X}_n^i$ used for "replacements" differ from those that they are replacing, and hence

$$P(X_k \neq X_k' \text{ for some } k) \leq \sum_{i=1}^{k_n-1} P(X_k \neq \tilde{X}_k^i \text{ for some } k > ir_n) \leq k_n \beta_{n,r_n}(u_n).$$

Now consider performing Algorithm 1 on the sequence $X_1', \ldots, X_n'$, and denote the output set of the algorithm by $\Delta_n'$. Also perform Algorithm 2 on the

sequence $X'_1, \ldots, X'_n$, and denote the output set of the algorithm by $\tilde{\Delta}'_n$.

Denote $J_n$ the event that $\Delta_n = \tilde{\Delta}'_n$ and that no two elements of $\Delta_n$ come from neighboring blocks of size $r_n$. Note that by construction, conditioned on $J_n$, the elements in $\tilde{\Delta}'_n$ are independently and identically distributed.

Now, with straightforward algebra one sees

$$P(\Delta_n = \tilde{\Delta}'_n) \geq 1 - P(\Delta_n \neq \tilde{\Delta}_n) - P(\tilde{\Delta}_n \neq \tilde{\Delta}'_n).$$

Note that by Corollary 4.2.4, $P(\Delta_n \neq \tilde{\Delta}_n) \to 0$ as $n \to \infty$. Further, the event $\tilde{\Delta}_n \neq \tilde{\Delta}'_n$ is a subset of the event that there are new exceedances over the threshold $u_n$ in the sequence $X'_1, \ldots, X'_n$ that are not in the sequence $X_1, \ldots, X_n$, or vice versa. Hence $P(\tilde{\Delta}_n \neq \tilde{\Delta}'_n) \leq P(X_k \neq X'_k \text{ for some } k) \leq k_n \beta_{n,r_n}(u_n) \to 0$ as $n \to \infty$ by Assumption $B$. It follows that $P(\Delta_n = \tilde{\Delta}'_n) \to 1$ as $n \to \infty$. Combined with Proposition 4.2.3, one sees that $P(J_n) \to 1$ as $n \to \infty$.

Now, define $M'^j_{\phi_n,n}(t)$, $j = 1, 2$, with respect to the set $\tilde{\Delta}'_n$ similarly to $M^j_{\phi_n,n}(t)$, $j = 1, 2$. First note that as $n \to \infty$,

$$P(M'^j_{\phi_n,n}(t) \neq M^j_{\phi_n,n}(t) \text{ for some } t \geq 0) \leq P(\Delta_n \neq \tilde{\Delta}'_n) \to 0.$$

Hence as $n \to \infty$,

$$M'^j_{\phi_n,n}(t) - M^j_{\phi_n,n}(t) \to_P 0 \tag{4.15}$$

uniformly for $t \geq 0$.

Conditioned on the event $J_n$, denote the elements of $\tilde{\Delta}'_n$ as $\tilde{X}'_1, \ldots, \tilde{X}'_{L_n}$.

The rest of this proof and the remaining proofs are very similar to those in [41] with a few nuances. We will present them for the sake of completeness.

Fix $0 < \epsilon < |\rho|$. Conditioned on $J_n$, as in Lemma 3.5.5 in [15], there are i.i.d. Pareto(1)-like random variables $Y_1, \ldots, Y_{L_n}$ (specifically $\tilde{X}'_i = U(Y_i)$ for $1 \le i \le L_n$, hence $Y_i$'s are Pareto(1) conditioned on $X_i > u_n$), a function $A_0 \sim A$, and $r_0 > 0$ with the property that

$$\begin{aligned}
&\log\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-k,L_n}}\right) + A_0(Y_{L_n-k,L_n})\frac{1}{\rho}\left(\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-k,L_n}}\right)^{\rho} - 1\right) \\
&- \epsilon|A_0(Y_{L_n-k,L_n})|\frac{1}{\rho}\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-k,L_n}}\right)^{\rho+\epsilon} \\
\le &\frac{1}{\gamma}\log\left(\frac{\tilde{X}'_{L_n-i,L_n}}{\tilde{X}'_{L_n-k,L_n}}\right) \\
\le &\log\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-k,L_n}}\right) + A_0(Y_{L_n-k,L_n})\frac{1}{\rho}\left(\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-k,L_n}}\right)^{\rho} - 1\right) \\
&+ \epsilon|A_0(Y_{L_n-k,L_n})|\frac{1}{\rho}\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-k,L_n}}\right)^{\rho+\epsilon}
\end{aligned} \tag{4.16}$$

if $Y_{L_n-k,L_n} > r_0$.

Hence for fixed $T > 0$ and again conditioned on $\Delta_n = \tilde{\Delta}'_n$, eventually for large $n$, we have

$$\begin{aligned}
\frac{\lfloor \phi_n t \rfloor}{\sqrt{\phi_n}}\left(\frac{M'^{j}_{\phi_n,n}(t)}{\gamma} - 1\right) \le &\frac{1}{\sqrt{\phi_n}}\sum_{i=0}^{\lfloor \phi_n t \rfloor - 1}\left(\log\frac{Y_{L_n-i,L_n}}{Y_{L_n-\lfloor \phi_n t \rfloor,L_n}} - 1\right) \\
&+ \frac{1}{\rho}\sqrt{\phi_n}A_0(Y_{L_n-\lfloor \phi_n t \rfloor,L_n})\frac{1}{\phi_n}\sum_{i=0}^{\lfloor \phi_n t \rfloor - 1}\left(\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-\lfloor \phi_n t \rfloor,L_n}}\right)^{\rho} - 1\right) \\
&+ \frac{1}{\rho}\sqrt{\phi_n}|A_0(Y_{L_n-\lfloor \phi_n t \rfloor,L_n})|\frac{1}{\phi_n}\sum_{i=0}^{\lfloor \phi_n t \rfloor - 1}\left(\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-\lfloor \phi_n t \rfloor,L_n}}\right)^{\rho+\epsilon}\right)
\end{aligned}$$

for every $t \in [0, T]$. Note that all the terms on the right hand side above (which we interpret as 0 for $0 \le t \le \frac{1}{\phi_n}$) are in $D[0, T]$. Let us denote the second and third terms by $W_n^{(2)}(t)$ and $W_n^{(3)}(t)$, respectively. We start with showing that as $n \to \infty$,

$$\sup_{0 \le t \le T} |W_n^{(2)}(t)| \to_P 0. \tag{4.17}$$

Since $\frac{A_0(Y_{L_n-\lfloor \phi_n t \rfloor, L_n})}{A(Y_{L_n-\lfloor \phi_n t \rfloor, L_n})} \to 1$ almost surely uniformly in $t \in [0,T]$, we will replace $A_0$ by $A$ in this calculation. Further, $|A|$ is eventually decreasing by Assumption $R$, so far large enough $n$,

$$|A(Y_{L_n-\lfloor \phi_n t \rfloor, L_n})| \leq |A(Y_{L_n-\lfloor \phi_n T \rfloor, L_n})|$$

for all relevant $t$. Per Lemma 2.2.3 in [15], if $U_{1,m} \leq \cdots \leq U_{m,m}$ are the order statistics from i.i.d. uniform random variables, then

$$\frac{m}{l_m} U_{l_m+1,m} \to_P 1$$

as $m \to \infty$ for some sequence $l_m \to \infty$, $l_m = o(m)$. Now, considering $\tilde{X}_i' = U(Y_i)$ for $1 \leq i \leq L_n$, we can interpret $Y_i$'s as i.i.d. Pareto(1) random variables conditioned on the corresponding $\tilde{X}_i' > u_n$. Hence writing $F'$ as the distribution of $Y_1, \ldots, Y_{L_n}$, for large $x$, it follows that

$$\overline{F'}(x) = \frac{1/x}{\overline{F}(u_n)}.$$

Further, we have that for $k \geq 0$,

$$Y_{L_n-k,L_n} =_d (F')^{\leftarrow}(1 - U_{k+1,L_n}).$$

Namely, for $k \geq 0$,

$$\frac{1}{\overline{F}(u_n) Y_{L_n-k,L_n}} =_d U_{k+1,L_n}.$$

Since $\phi_n = o(\frac{n}{p_n})$ as $n \to \infty$, which implies that $\phi_n / L_n \to_P 0$ as $n \to \infty$, hence

$$\frac{\lfloor \phi_n T \rfloor \overline{F}(u_n)}{L_n} Y_{L_n-\lfloor \phi_n t \rfloor, L_n} \to_P 1.$$

Since $A$ is regularly varying,

$$\frac{A(Y_{L_n-\lfloor \phi_n t \rfloor, L_n})}{A(\frac{L_n}{\lfloor \phi_n T \rfloor \overline{F}(u_n)})} \to_P 1.$$

Putting everything together, we see that in order to show (4.17), it is enough the prove that for any $\zeta > 0$, as $n \to \infty$,

$$P\left(\sup_{0\le t\le T}\left|\sqrt{\phi_n}A(\frac{L_n}{\lfloor\phi_n T\rfloor\overline{F}(u_n)})\frac{1}{\phi_n}\sum_{i=0}^{\lfloor\phi_n t\rfloor-1}\left(\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-\lfloor\phi_n t\rfloor,L_n}}\right)^{\rho}-1\right)\right|>\zeta\right)\to 0.$$

Combining the fact that $\frac{p_n}{n}N_{k_n r_n}(u_n) \to_P \tau$, $\hat{\theta}_n \to_P \theta$, $p_n\overline{F}(u_n) \to \tau$, and $L_n \le \hat{\theta}_n N_{k_n r_n}(u_n)$, we have that

$$\sqrt{\phi_n}A(\frac{L_n}{\lfloor\phi_n T\rfloor\overline{F}(u_n)}) \to_P 0$$

as $n \to \infty$, by the growth assumption on the sequence $\{\phi_n\}$. Now, let $Z$ represent a generic Pareto(1) random variable. Let $\mu_\rho = 1/(1-\rho) = EZ^\rho$. Then

$$\begin{aligned}
&P\left(\sup_{0\le t\le T}\left|\sqrt{\phi_n}A(\frac{L_n}{\lfloor\phi_n T\rfloor\overline{F}(u_n)})\frac{1}{\phi_n}\sum_{i=0}^{\lfloor\phi_n t\rfloor-1}\left(\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-\lfloor\phi_n t\rfloor,L_n}}\right)^{\rho}-1\right)\right|>\zeta\right)\\
\le& P\left(\sup_{0\le t\le T}\left|\sqrt{\phi_n}A(\frac{L_n}{\lfloor\phi_n T\rfloor\overline{F}(u_n)})\frac{1}{\phi_n}\sum_{i=0}^{\lfloor\phi_n t\rfloor-1}\left(\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-\lfloor\phi_n t\rfloor,L_n}}\right)^{\rho}-\mu_\rho\right)\right|>\zeta/2\right)\\
&+P\left(\sup_{0\le t\le T}\left|\sqrt{\phi_n}A(\frac{L_n}{\lfloor\phi_n T\rfloor\overline{F}(u_n)})\frac{1}{\phi_n}\lfloor\phi_n t\rfloor\frac{\rho}{1-\rho}\right|>\zeta/2\right).
\end{aligned}$$

By above, it follows that for large enough $n$,

$$\begin{aligned}
&P\left(\sup_{0\le t\le T}\left|\sqrt{\phi_n}A(\frac{L_n}{\lfloor\phi_n T\rfloor\overline{F}(u_n)})\frac{1}{\phi_n}\sum_{i=0}^{\lfloor\phi_n t\rfloor-1}\left(\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-\lfloor\phi_n t\rfloor,L_n}}\right)^{\rho}-1\right)\right|>\zeta\right)\\
\le& P\left(\sup_{0\le t\le T}\left|\sqrt{\phi_n}A(\frac{L_n}{\lfloor\phi_n T\rfloor\overline{F}(u_n)})\frac{1}{\phi_n}\sum_{i=0}^{\lfloor\phi_n t\rfloor-1}\left(\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-\lfloor\phi_n t\rfloor,L_n}}\right)^{\rho}-\mu_\rho\right)\right|>\zeta/2\right).
\end{aligned}$$

Now, the process above is a step function with jumps at multiples of $\frac{1}{\phi_n}$, so the largest value of the process occurs at one of those steps. Therefore, the above probability does not exceed

$$\sum_{j=1}^{\lfloor\phi_n T\rfloor}P\left(\sup_{0\le t\le T}\left|\sqrt{\phi_n}A(\frac{L_n}{\lfloor\phi_n T\rfloor\overline{F}(u_n)})\frac{1}{\phi_n}\sum_{i=0}^{j-1}\left(\left(\frac{Y_{L_n-i,L_n}}{Y_{L_n-\lfloor\phi_n t\rfloor,L_n}}\right)^{\rho}-\mu_\rho\right)\right|>\zeta/2\right).$$

Since $\overline{F}(u_n)Y_i =_d Z$ for $1 \le i \le L_n$, by the Renyi representation, for fixed $j$, $\left\{\frac{Y_{L_n-i,L_n}}{Y_{L_n-j,L_n}}\right\}_i =_d \{Z_{j-i,j}\}_i$, where $Z_0, \ldots, Z_{j-1}$ are i.i.d. Pareto(1) random variables. Hence by Chebyshev's inequality,

$$\begin{aligned}
&\sum_{j=1}^{\lfloor \phi_n T \rfloor} P\left( \sup_{0\le t\le T} \left| \sqrt{\phi_n} A(\frac{L_n}{\lfloor \phi_n T \rfloor \overline{F}(u_n)}) \frac{1}{\phi_n} \sum_{i=0}^{j-1} \left( \left( \frac{Y_{L_n-i,L_n}}{Y_{L_n-\lfloor \phi_n t \rfloor, L_n}} \right)^\rho - \mu_\rho \right) \right| > \zeta/2 \right) \\
&\quad \le \sum_{j=1}^{\lfloor \phi_n T \rfloor} \left( \sqrt{\phi_n} A(\frac{L_n}{\lfloor \phi_n T \rfloor \overline{F}(u_n)}) \right)^2 \frac{4}{\zeta^2 \phi_n^2} E\left[ \sum_{i=0}^{j-1} \left( \left( \frac{Y_{L_n-i,L_n}}{Y_{L_n-\lfloor \phi_n t \rfloor, L_n}} \right)^\rho - \mu_\rho \right) \right]^2 \\
&\quad = \sum_{j=1}^{\lfloor \phi_n T \rfloor} \left( \sqrt{\phi_n} A(\frac{L_n}{\lfloor \phi_n T \rfloor \overline{F}(u_n)}) \right)^2 \frac{4}{\zeta^2 \phi_n^2} \left( j \mathrm{Var}(Z^\rho) \right) \\
&\quad = \left( \sqrt{\phi_n} A(\frac{L_n}{\lfloor \phi_n T \rfloor \overline{F}(u_n)}) \right)^2 \frac{4}{\zeta^2} \mathrm{Var}(Z^\rho) \frac{1}{\phi_n^2} \sum_{j=1}^{\lfloor \phi_n T \rfloor} j \to 0
\end{aligned}$$

as $n \to \infty$, by the growth assumption of the sequence $\{\phi_n\}$. This proves (4.17). Similarly one can show that as $n \to \infty$,

$$\sup_{0\le t\le T} |W_n^{(3)}(t)| \to_P 0. \tag{4.18}$$

Similarly applying the corresponding lower bounds, and taking note of (4.15), we have that

$$\frac{\lfloor \phi_n t \rfloor}{\sqrt{\phi_n}} \left( \frac{M^1_{\phi_n,n}(t)}{\gamma} - 1 \right) - \frac{1}{\sqrt{\phi_n}} \sum_{i=0}^{\lfloor \phi_n t \rfloor - 1} \left( \log \frac{Y_{L_n-i,L_n}}{Y_{L_n-\lfloor \phi_n t \rfloor, L_n}} - 1 \right) \to_P 0$$

for $t \in [\delta, \infty)$ for any $\delta > 0$, uniformly on compact intervals. Next, recall that $\overline{F}(u_n)Y_1$ is distributed as a Pareto(1) random variable, and therefore $\log(\overline{F}(u_n)Y_1)$ is a standard exponential random variable, so the differences $\log Y_{L_n-i,L_n} - \log Y_{L_n-i-1,L_n}$, $i = 0, \ldots, L_n$ are independent exponential random variables with means $1/(i+1)$, $i = 0, \ldots, L_n$. Therefore, denoting the $i$th of those exponential random variables by $E_i/(i+1)$, we see that for $k = 1, \ldots, L_n$,

$$\sum_{i=0}^{k-1} \log \frac{Y_{L_n-i,L_n}}{Y_{L_n-k,L_n}} = \sum_{i=0}^{k-1} E_i.$$

Therefore,

$$\frac{\lfloor \phi_n t \rfloor}{\sqrt{\phi_n}} \left( \frac{M^1_{\phi_n,n}(t)}{\gamma} - 1 \right) - \frac{1}{\sqrt{\phi_n}} \sum_{i=0}^{\lfloor \phi_n t \rfloor - 1} (E_i - 1) \to_P 0$$

for $t \in [\delta, \infty)$, uniformly on compact intervals. Squaring (4.16) and repeating the argument also gives us

$$\frac{\lfloor \phi_n t \rfloor}{\sqrt{\phi_n}} \left( \frac{M^2_{\phi_n,n}(t)}{\gamma} - 1 \right) - \frac{1}{\sqrt{\phi_n}} \sum_{i=0}^{\lfloor \phi_n t \rfloor - 1} \left( E_i^2 - 2 \right) \to_P 0$$

for $t \in [\delta, \infty)$, uniformly on compact intervals.

By Theorem 12.6.1 and Remark 12.6.2 in [59], the statement of the proposition will follow once we check that

$$\frac{1}{\sqrt{\phi_n}} \begin{pmatrix} \sum_{i=0}^{\lfloor \phi_n t \rfloor - 1} (E_i - 1) \\ \sum_{i=0}^{\lfloor \phi_n t \rfloor - 1} (E_i^2 - 2) \end{pmatrix} \Rightarrow \begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix}$$

in $D[\delta, \infty)$, where $((W_1(t), W_2(t)), t \geq \delta)$ is a two-dimensional zero mean Brownian motion with covariance matrix

$$\begin{bmatrix} 1 & 4 \\ 4 & 20 \end{bmatrix}.$$

This is, however, an immediate consequence of the multivariate version of Donsker's theorem; see e.g. Theorem 4.3.5 in [59]. □

Now, with the above theorem, and recalling that

$$U_n = \min(\inf \left\{ k : 1 \leq k \leq L_n, |Q_{k,n}| \geq \omega \sqrt{\frac{\phi_n}{k}} \right\}, L_n),$$

we have the following result.

**Theorem 4.3.2.** *Suppose Assumptions $\Delta''$, $B$, $E$, $P$, $D$, and $R$ are satisfied. Let $\omega > 0$ and $\{\phi_n\}$ an increasing sequence with $\phi_n \to \infty$ and $\phi_n = o(n^{\frac{2|\rho|}{1+2|\rho|}})$, $\phi_n = o(n/p_n)$ as $n \to \infty$. Then $\frac{U_n}{\phi_n} \Rightarrow \tau_\omega$, where $\tau_\omega$ is the first time a standard Brownian motion hits $\pm\omega$.*

*Proof.* We begin by showing that for any $\delta > 0$,

$$\sqrt{t}Q_n^*(t) \Rightarrow \frac{1}{2}(W_2(t) - 2W_1(t)) \tag{4.19}$$

in $D[\delta, \infty)$, where $((W_1(t), W_2(t)), t \geq \delta)$ is the two-dimensional Brownian motion from the proposition above, and

$$Q_n^*(t) = \begin{cases} Q_{\lfloor \phi_n t \rfloor, n}, & \text{if } \delta \leq t \leq \frac{L_n}{\phi_n} \\ Q_{L_n, n}, & \text{if } t > \frac{L_n}{\phi_n}. \end{cases}$$

By Theorem 16.7 of [6], we have to prove convergence in $D[\delta, T]$ for each $\delta < T < \infty$. Fix $T > \delta$. Straightforward algebra shows that

$$\begin{aligned} \sqrt{t}Q_n^*(t) =& \frac{\sqrt{\phi_n}t}{2} \frac{(M^2_{\phi_n,n}(t)/\gamma^2 - 2) - 4(M^1_{\phi_n,n}(t)/\gamma - 1)}{(M^1_{\phi_n,n}(t))^2/\gamma^2} \\ & - \sqrt{\phi_n}t \frac{(M^1_{\phi_n,n}(t)/\gamma - 1)^2}{(M^1_{\phi_n,n}(t))^2/\gamma^2} := V_n^{(1)}(t) - V_n^{(2)}(t), \end{aligned}$$

$\delta \leq t \leq T$, while for $0 \leq t \leq 1/\phi_n$, we define both $V_n^{(1)}(t) = V_n^{(2)}(t) = 0$. Call

$$D_n^{(1)}(t) = \frac{\sqrt{\phi_n}t}{2} \left[ (M^2_{\phi_n,n}(t)/\gamma^2 - 2) - 4(M^1_{\phi_n,n}(t)/\gamma - 1) \right], \delta \leq t \leq T.$$

Since the limiting process in Theorem 4.3.1 is continuous, the weak convergence holds also in the uniform topology on $[\delta, T]$, and addition is continuous in the topology. By Proposition 4.3.1 and the continuous mapping theorem,

$$D_n^{(1)}(t) \Rightarrow \frac{1}{2}(W_2(t) - 4W_1(t))$$

in $D[\delta, T]$. Furthermore by Proposition 4.3.1,

$$M^1_{\phi_n,n}(t)/\gamma \to_P 1$$

uniformly on $[\delta, T]$. By Theorem 3.1 in [6] we can conclude that

$$V_n^{(1)}(t) \Rightarrow \frac{1}{2}(W_2(t) - 4W_1(t))$$

in $D[\delta, T]$. Similarly we can show that

$$V_n^{(2)}(t) \to_P 0$$

uniformly on $[\delta, T]$. So we can once again apply Theorem 3.1 in [6] to obtain (4.19).

For fixed $0 < x < \lfloor L_n/\phi_n \rfloor$, and we can write

$$P(U_n \leq \phi_n x) = P\left( |Q_{k,n}| \geq \omega \sqrt{\frac{\phi_n}{k}} \text{ for some } 1 \leq k \leq \phi_n x \right).$$

Therefore for $0 < \delta < x$ we have

$$\begin{aligned} P\left( |Q_{k,n}| \geq \omega \sqrt{\frac{\phi_n}{k}} \text{ for some } \phi_n\delta \leq k \leq \phi_n x \right) &\leq P(U_n \leq \phi_n x) \\ &\leq P\left( |Q_{k,n}| \geq \omega \sqrt{\frac{\phi_n}{k}} \text{ for some } \phi_n\delta \leq k \leq \phi_n x \right) \\ &+ P\left( |Q_{k,n}| \geq \omega \sqrt{\frac{\phi_n}{k}} \text{ for some } 1 \leq k \leq \phi_n\delta \right). \end{aligned}$$

We will show that for any $0 < \delta < x$,

$$\begin{aligned} &\lim_{n\to\infty} P\left( |Q_{k,n}| \geq \omega \sqrt{\frac{\phi_n}{k}} \text{ for some } \phi_n\delta \leq k \leq \phi_n x \right) \\ =&P\left( \sup_{\delta \leq t \leq x} |B(t)| \geq \omega \right), \end{aligned} \tag{4.20}$$

where $(B(t), t \geq 0)$ is a standard Brownian motion, while

$$\lim_{\delta\to 0} \limsup_{n\to\infty} P\left( |Q_{k,n}| \geq \omega \sqrt{\frac{\phi_n}{k}} \text{ for some } 1 \leq k \leq \phi_n\delta \right) = 0. \tag{4.21}$$

It will following from above relations that

$$P(U_n \leq \phi_n x) \to P\left( \sup_{0 \leq t \leq x} |B(t)| \geq \omega \right) = P(\tau_\omega \leq x),$$

with $\phi_n = o(n/p_n)$, and therefore $\phi_n/L_n \to_P 0$, the above holds for all $x > 0$, which is what we need for the statement of the theorem.

Observe that for any $\delta > 0$,

$$\begin{aligned} & P\left(|Q_{k,n}| \geq \omega\sqrt{\frac{\phi_n}{k}} \text{ for some } \phi_n\delta \leq k \leq \phi_n x\right) \\ &= P\left(|\sqrt{t}Q^*_{k,n}| \geq \sqrt{t}\omega\sqrt{\frac{\phi_n}{\lfloor \phi_n t \rfloor}} \text{ for some } \delta \leq t \leq x\right) \\ &= P\left(|\sqrt{t}Q^*_{k,n}| \geq \omega(1+o(1)) \text{ for some } \delta \leq t \leq x\right) \end{aligned}$$

(with the same $o(1)$ for all relevant $t$). Now (4.20) follows from (4.19) and the continuity of the supremum distribution of the Brownian motion.

Now, for a fixed $n$ and $k$, choose some $t$ such that $\lfloor \phi_n t \rfloor = k$, and write

$$M^j_{k,n} := M^j_{\phi_n,n}(t)$$

for $j = 1, 2$. In order to show (4.21), we start with showing that, for any $\delta > 0$, both

$$\begin{aligned} &\inf_{1\leq k\leq \phi_n\delta} M^1_{k,n} \text{ is stochastically bounded away from } 0, \text{ and} \qquad (4.22) \\ &\sup_{1\leq k\leq \phi_n\delta} M^1_{k,n} \text{ is stochastically bounded away from } \infty. \end{aligned}$$

To see this, recall that in the proof of Proposition 4.3.1,

$$(M^1_{k,n}, 1 \leq k \leq \phi_n\delta) = (M^{1,Y}_{k,n}\gamma + W_{k,n}, 1 \leq k \leq \phi_n\delta),$$

where $((M^{1,Y}_{k,n})$ is composed of Pareto-like random variables while

$$\sup_{1\leq k\leq \phi_n\delta} |W_{k,n}| \to_P 0.$$

Further, we've also verified that

$$(M^{1,Y}_{k,n}, 1 \leq k \leq \phi_n\delta) =_d (\frac{1}{k}\sum_{i=1}^{k} E_i, 1 \leq k \leq \phi_n\delta),$$

where $E_1, E_2, \ldots$ are i.i.d. standard exponential random variables. Hence (4.22) follows from the law of large numbers.

We continue by writing

$$P\left(|Q_{k,n}| \geq \omega\sqrt{\frac{\phi_n}{k}} \text{ for some } 1 \leq k \leq \phi_n\delta\right)$$
$$= P\left(k\left|\frac{M^2_{k,n} - 2(M^1_{k,n})^2}{2(M^1_{k,n})^2}\right| \geq \omega\sqrt{\phi_n} \text{ for some } 1 \leq k \leq \phi_n\delta\right).$$

Hence using (4.22), in order to prove (4.21), it is enough to show that

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\left(|R_{k,n}| \geq \omega\sqrt{\phi_n} \text{ for some } 1 \leq k \leq \phi_n\delta\right) = 0, \tag{4.23}$$

where

$$R_{k,n} = k\frac{M^2_{k,n} - 2(M^1_{k,n})^2}{2(M^1_{k,n})^2}, k = 1, \ldots, n.$$

Straightforward algebra shows that the above probability is

$$P\left(\frac{k}{2}\left|\left(\frac{M^2_{k,n}}{\gamma^2} - 2\right) - 4\left(\frac{M^1_{k,n}}{\gamma} - 1\right)\right.\right.$$
$$\left.\left.-2\left(\frac{M^1_{k,n}}{\gamma} - 1\right)^2\right| \cdot \frac{\gamma^2}{(M^1_{k,n})^2} \geq \omega\sqrt{\phi_n} \text{ for some } 1 \leq k \leq \phi_n\delta\right).$$

Now from the proof of Proposition 4.3.1,

$$\frac{\gamma^2}{(M^1_{k,n})^2} \to_P 1$$

uniformly in $k = 1, \ldots, n$. And by (4.22), we can write

$$P\left(|R_{k,n}| \geq \omega\sqrt{\phi_n} \text{ for some } 1 \leq k \leq \phi_n\delta\right)$$
$$\leq P\left(\frac{k}{2}\left(\left|\frac{M^2_{k,n}}{\gamma^2} - 2\right| + 4\left|\frac{M^1_{k,n}}{\gamma} - 1\right|\right)(1 + K_n) \geq \omega\sqrt{\phi_n} \text{ for some } 1 \leq k \leq \phi_n\delta\right),$$

where $K_n \to_P 0$ as $n \to \infty$. So

$$\lim_{n\to\infty} P\left(\frac{k}{2}\left(\left|\frac{M^2_{k,n}}{\gamma^2} - 2\right| + 4\left|\frac{M^1_{k,n}}{\gamma} - 1\right|\right)(1 + K_n) \geq \omega\sqrt{\phi_n} \text{ for some } 1 \leq k \leq \phi_n\delta\right)$$
$$= \lim_{n\to\infty} P\left(\frac{k}{2}\left(\left|\frac{M^2_{k,n}}{\gamma^2} - 2\right| + 4\left|\frac{M^1_{k,n}}{\gamma} - 1\right|\right) \geq \omega\sqrt{\phi_n} \text{ for some } 1 \leq k \leq \phi_n\delta\right).$$

Lastly, by Proposition 4.3.1,

$$\begin{aligned}
&P\left(\frac{k}{2}\left(\left|\frac{M_{k,n}^2}{\gamma^2}-2\right|+4\left|\frac{M_{k,n}^1}{\gamma}-1\right|\right)\geq\omega\sqrt{\phi_n}\text{ for some }1\leq k\leq\phi_n\delta\right)\\
&\quad=P\left(\sup_{0\leq t\leq\delta}\sqrt{\phi_n}t\left(\left|\frac{M_{\lfloor\phi_n t\rfloor,n}^2}{\gamma^2}-2\right|+4\left|\frac{M_{\lfloor\phi_n t\rfloor,n}^1}{\gamma}-1\right|\right)\geq 2\omega\right)\\
&\quad\to P(\sup_{0\leq t\leq\delta}|B(t)|>2\omega),
\end{aligned}$$

where $(B(t), t \geq 0)$ is some Brownian motion, and (4.21) follows. This completes the proof. □

Now we can present the main theorem of this section. Recall that $\gamma = 1/\alpha$, the reciprocal of the regular variation index $\alpha$.

**Theorem 4.3.3.** *Suppose Assumptions $\Delta''$, $B$, $E$, $P$, $D$, and $R$ are satisfied. Let $\omega > 0$ and $\{\phi_n\}$ be an increasing sequence with $\phi_n \to \infty$ and $\phi_n = o(n^{\frac{2|\rho|}{1+2|\rho|}})$, $\phi_n = o(n/p_n)$ as $n \to \infty$. Let $U_n$ be given as in (4.13). Then the Hill estimator based on the $U_n$ upper statistics is consistent, namely,*

$$H_{U_n,n} = \frac{1}{U_n}\sum_{i=0}^{U_n-1}\log\frac{\bar{X}_{L_n-i,L_n}}{\bar{X}_{L_n-U_n,L_n}} \to_P \gamma$$

*as $n \to \infty$. Furthermore,*

$$\sqrt{\phi_n}\left(\frac{H_{U_n,n}}{\gamma}-1\right)\Rightarrow\frac{G}{(\tau_\omega)^{1/2}}, \tag{4.24}$$

*where $G$ is a standard normal random variable indepedent of the first hitting time $\tau_\omega$.*

*Proof.* The idea is to use a random stopping technique in a weak convergence context. The formulation we will use is the one given in Theorem 2.2.1 in [55]. If, for each $n$, $(X_n(t), t \geq 0)$ is a càdlàg process, and $\tau_n$ is a nonnegative random

variable, such that for all $0 \leq a < b < \infty$,

$$\left(\tau_n, \sup_{t\in[a,b)} X_n(t)\right) \Rightarrow \left(\tau_0, \sup_{t\in[a,b)} X_0(t)\right), \tag{4.25}$$
$$\left(\tau_n, \inf_{t\in[a,b)} X_n(t)\right) \Rightarrow \left(\tau_0, \inf_{t\in[a,b)} X_0(t)\right)$$

for some continuous process $(X_0(t), t \geq 0)$ and a nonnegative random variable $\tau_0$, then $X_n(\tau_n) \Rightarrow X_0(\tau_0)$.

Note that it is enough to prove the weak convergence (4.24) as the consistency result follows automatically. Note that (4.24) is equivalent to

$$\sqrt{\phi_n}\frac{1}{U_n}\sum_{i=0}^{U_n-1}\left(\frac{1}{\gamma}\log\frac{\bar{X}_{L_n-i,L_n}}{\bar{X}_{L_n-U_n,L_n}} - 1\right) \Rightarrow \frac{G}{(\tau_\omega)^{1/2}}. \tag{4.26}$$

We will show that for any $\delta > 0$,

$$\sqrt{\phi_n}\frac{1}{U_n \vee \phi_n\delta}\sum_{i=0}^{\lfloor U_n\vee\phi_n\delta\rfloor-1}\left(\frac{1}{\gamma}\log\frac{\bar{X}_{L_n-i,L_n}}{\bar{X}_{L_n-\lfloor U_n\vee\phi_n\delta\rfloor,L_n}} - 1\right) \Rightarrow \frac{G}{(\tau_\omega \vee \delta)^{1/2}}. \tag{4.27}$$

Then (4.26) will follow from (4.21).

Note that the expression in the left hand side of (4.27) results from a substitution of the random time

$$\tau_n = \max\left(\delta, \frac{U_n}{\phi_n}\right)$$

into the càdlàg process

$$V_n(t) = \sqrt{\phi_n}\frac{1}{\phi_n t}\sum_{i=0}^{\lfloor\phi_n t\rfloor-1}\left(\frac{1}{\gamma}\log\frac{\bar{X}_{L_n-i,L_n}}{\bar{X}_{L_n-\lfloor\phi_n t\rfloor,L_n}} - 1\right), t \geq \delta.$$

According to (4.25) and to self-similarity of the Brownian motion, it is enough to check that for all $\delta \leq a < b < \infty$,

$$\left(\max\left(\delta, \frac{U_n}{\phi_n}\right), \sup_{t\in[a,b)} V_n(t)\right) \Rightarrow \left(\tau_\omega \vee \delta, \sup_{t\in[a,b)} \frac{B(t)}{t}\right), \tag{4.28}$$

where $\tau_\omega$ is as before, the first hitting time of a standard Brownian motion, independent of another standard Brownian motion $B$. The proof for the analogue statement where the suprema is replaced by the infima is similar.

In order to show (4.28), it is enough to show that for any $x \geq \delta$ and $y \geq 0$,

$$P\left(U_n \leq \phi_n x,\ \sup_{t\in[a,b)} V_n(t) \leq y\right) \to P\left(\tau_\omega \leq x,\ \sup_{t\in[a,b)} \frac{B(t)}{t} \leq y\right).$$

This statement will follow once we check that for every $0 < \delta' < x$,

$$P\left(\phi_n\delta' \leq U_n \leq \phi_n x,\ \sup_{t\in[a,b)} V_n(t) \leq y\right) \to P\left(\sup_{\delta'\leq t\leq x} |B_1(t)| \geq \omega,\ \sup_{t\in[a,b)} \frac{B(t)}{t} \leq y\right),$$

where $B$ and $B_1$ are independent standard Brownian motions, which, in turn, will be implied by the statment

$$\begin{aligned} &P\left(\sup_{\delta'\leq t\leq x} |\sqrt{t}Q_n^*(t)| \geq \omega,\ \sup_{t\in[a,b)} V_n(t) \leq y\right) \\ \to &P\left(\sup_{\delta'\leq t\leq x} |B_1(t)| \geq \omega,\ \sup_{t\in[a,b)} \frac{B(t)}{t} \leq y\right). \end{aligned} \tag{4.29}$$

To this end note that

$$V_n(t) = \frac{1}{t}\left[\sqrt{\phi_n}t\left(\frac{M^1_{\phi_n,n}(t)}{\gamma} - 1\right)\right],$$

and the map $(f(t), t \geq \delta) \to (f(t)/t, t \geq \delta)$ is continuous on $D[\delta, \infty)$. Therefore, the argument leading to (4.19) applies, and it gives joint convergence

$$\begin{bmatrix} \sqrt{t}Q_n^*(t), t \geq \delta' \\ V_n(t), t \geq \delta' \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{1}{2}(W_2(t) - 4W_1(t)), t \geq \delta' \\ \frac{W_1(t)}{t}, t \geq \delta' \end{bmatrix},$$

where $W_1$ and $W_2$ are as in Proposition 4.3.1. Computing correlations shows that $W_1$ and $(W_2 - 4W_1)/2$ are independent standard Brownian motions. Therefore (4.29) follows and the proof of the theorem is complete. □

## 4.4 Testing the Estimator - Simulation Study

In this section we will test the estimator (4.14) on simulated data. For convenience we will refer to (4.14) as the distance based modified Hill estimator. Now, instead of using the distance based selection Algorithm 1, one can use the blocks based Algorithm 2; replacing the elements of $\Delta_n$ in (4.14) by elements of $\tilde{\Delta}_n$, one can produce a different estimator. We will call this the blocks based modified Hill estimator. One can see from the previous chapter that the asymptotic consistency results still hold for the blocks based modified Hill estimator. We will test the properties of both the above estimators, denoted Modified Hill$_D$ and Modified Hill$_B$, respectively, for distance based and blocks based.

It is difficult to find a benchmark against which the performance of (4.14) can be measured. This is due to the fact that, in practice, the choice of the number $k$ of upper statistic is often found through visual inspection of the Hill plot. Further, we are not able to find any systemic approaches in literature to determine the choice of $k$ when data exhibits serial dependence. In light of this, for benchmarking purposes, we will use $k = \sqrt{n}$ upper order statistics for the Hill estimator. Again for convenience we will refer to this benchmark estimator as the regular Hill estimator, or just the Hill estimator.

We have drawn samples from the ARCH(1) process and the MA($q$) process. Recall that the MA($q$) process $(X_i)$ is defined as

$$X_i = p_q Z_{i-q} + p_{q-1} Z_{i-q+1} + \cdots + p_1 Z_{i-1} + Z_i,\ i \geq 1\,, \tag{4.30}$$

with $0 < p_1, \ldots, p_q < 1$, and the noise sequence consisting of i.i.d. Pareto ran-

dom variables $Z_{-(q-1)}, \ldots, Z_0, Z_1, \ldots$ with

$$P(Z_0 > x) = \begin{cases} 1, & \text{if } x < 1 \\ x^{-\alpha}, & \text{if } x \geq 1, \end{cases} \tag{4.31}$$

for some $\alpha > 0$. In this case the regular variation index of the sequence is $\alpha$, and the extremal index is

$$\theta = \frac{1}{1 + p_1^\alpha + \cdots + p_q^\alpha}\,.$$

The ARCH($p$) process is defined by the equations

$$X_i = \sigma_i Z_i \quad i \geq 1\,, \tag{4.32}$$

where $\{Z_i\}$ is a sequence of i.i.d. standard normal random variables, and $\sigma_i$ obeys the relation

$$\sigma_i^2 = \beta + \sum_{j=1}^{p} \lambda_i X_{i-j}^2. \tag{4.33}$$

We will focus on the ARCH(1) process, whose index of regular variation can be numerically determined from the parameters. Some of those numerical solutions can be found in [21].

The extremal index of an ARCH(1) process can be computed as follows. First we look at the absolute value process $(|X_i|)$. Its extremal index is

$$\theta_{|X|} = \lim_{k\to\infty} E\left( |Z_1|^\alpha - \max_{j=2,\ldots,k+1} \left| Z_j^2 \prod_{i=2}^{j} A_i \right|^{\alpha/2} \right)_+ \Big/ E|Z_1|^\alpha\,,$$

where $\alpha$ is the index of regular variation, $A_i = \lambda_1 Z_{i-1}^2$ for $i \geq 1$, and $x_+$ denotes $\max(x, 0)$ for $x \in \mathbb{R}$. The extremal index for the sequence $(X_i)$ is

$$\theta = 2\theta_{|X|}(1 - \tilde{\Pi}(0.5))\,,$$

where $\tilde{\Pi}$ is the probability generating function of the compounding probabilities of a compound Poisson process; see e.g. [38].

We test the effect of the estimator (4.14) against different values of $\alpha$ using different types of processes. We used a finite sample length of $n = 10000$. For the estimator (4.14), we choose $u_n$ to correspond to the $(\lfloor\sqrt{n}\rfloor + 1)$-th largest observation in the sample. Namely, our selection algorithm chooses observations from the benchmark set of upper order statistics. Note that in this case $\sqrt{n} = 100$, meaning we are using the top $1\%$ of the data. We choose a block size of $r_n = 50$ when computing the blocks estimator for the extremal index. We choose $\omega$ to correspond to the $95\%$ quantile of a standard normal distribution, and finally, set $\phi_n = (\log n)^2$. The Monte Carlo mean and root mean squared error of $10000$ simulated instances of the estimators are presented below.

| Method | | Hill | | Modified Hill$_B$ | | Modified Hill$_D$ | |
|---|---|---|---|---|---|---|---|
| Process | $\alpha$ | Mean | RMSE | Mean | RMSE | Mean | RMSE |
| ARCH | 1.08 | 1.18 | 0.264 | 1.15 | 0.228 | 1.14 | 0.200 |
| ARCH | 2.0 | 2.07 | 0.324 | 2.04 | 0.304 | 2.02 | 0.268 |
| ARCH | 2.3 | 2.35 | 0.341 | 2.31 | 0.326 | 2.29 | 0.295 |
| MA | 1.0 | 1.04 | 0.148 | 1.02 | 0.153 | 1.01 | 0.132 |
| MA | 2.0 | 2.22 | 0.362 | 2.13 | 0.325 | 2.12 | 0.288 |
| MA | 2.5 | 2.87 | 0.523 | 2.74 | 0.450 | 2.71 | 0.404 |

Table 4.1: Mean and root mean squared error of the regular Hill estimator and the modified Hill estimators. Data are generated using ARCH and Moving Average models with different values of $\alpha$ and a sequence length of $n = 10000$.

We see that in each instance the distance based modified Hill estimator outperforms both the blocks based modified Hill estimator as well as the regular Hill estimator, in terms of having both a smaller bias and a smaller root mean squared error. Compared to the regular Hill estimator, the blocks based modified Hill estimator produces a smallr bias in every case. However for smaller

values of $\alpha$ for the moving average process, the blocks based modified Hill estimator produces too large of a standard error, resulting in a higher root mean squared error than the regular Hill estimator. Simulated sequences with sizes $n = 5000$ and $n = 20000$ are also used to test the estimators, with results having the same qualitative structure. Those results are presented in Appendix A. This suggests that our distance based selection algorithm that systemically removes clustered observations does indeed help improve the Hill estimator, and in some cases significantly so. Further, even though asymptotically, the distance based and the blocks based selection algorithms produce the same set with probability 1, the distance based modified Hill estimator still outperforms the blocks based Hill estimator in finite samples.

Next let us test the sensitivity of our estimators with respect to the threshold $u_n$. For this we will focus on an ARCH(1) process with $\alpha = 2$. We keep the same setup as the previous exercise, but vary the threshold $u_n$ to correspond to the $(\lfloor i\% \cdot n \rfloor + 1)$-th largest observation in the sample, for $i = 0.5, 0.6, \ldots, 1.5$. Note that the previous exercise corresponds to the case when $i = 1$.

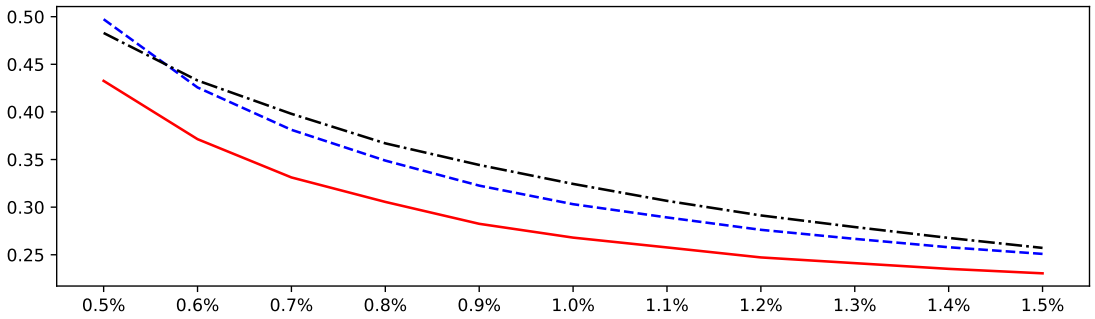


Figure 4.1: Root mean squared error for the distance based modified Hill estimator (solid line), the blocks based modified Hill estimator (dotted line), and the regular Hill estimator (dot-dash line) plotted against percentage of largest order statistics used for estimation. Data are simulated from an ARCH(1) model with $\alpha = 2$.

The Monte Carlo root mean squared error of $10000$ simulated instances of the estimators are presented above. One sees that again, the distance based modified Hill estimator outperforms the other two estimators at all levels of thresholds; whereas the blocks based modified Hill estimator outperforms the regular Hill estimator when more data is used. For all three estimators, the general trend of the root mean squared errors is decreasing as the threshold $u_n$ decreases. However, (not shown) the trend reverses and the root mean squared errors start to increase as the threshold $u_n$ continues to decrease, due to the increasing magnitude of the bias. Similar graphs for the bias and the standard errors of the estimators are included in Appendix A.

Lastly we will test the sensitivity of the estimators with respect to the block size $r_n$. We again use the ARCH(1) process with $\alpha = 2$. Keeping the same setup as our first exercise, we vary the block size $r_n = 20, 30, \ldots, 80$. The Monte Carlo root mean squared error of $10000$ simulated instances of the estimators are presented below.

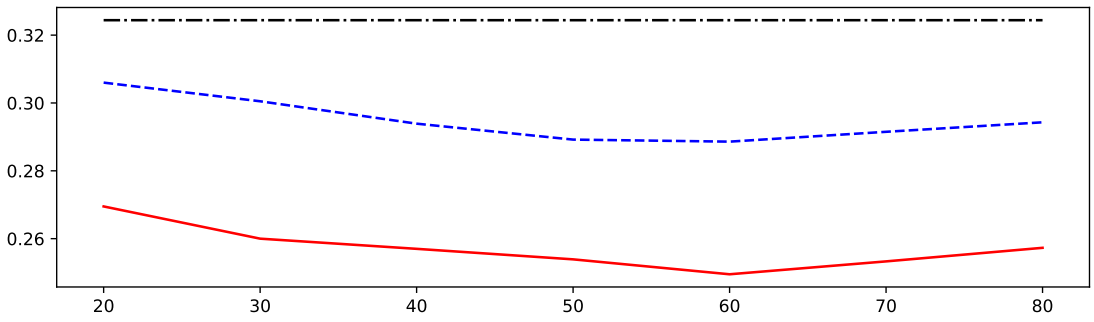


Figure 4.2: Root mean squared error for the distance based modified Hill estimator (solid line), the blocks based modified Hill estimator (dotted line), and the regular Hill estimator (dot-dash line) plotted against $r_n$, the block size used for extremal index estimation. Data are simulated from an ARCH(1) model with $\alpha = 2$.

We see that there is no discernable trend in the relationship between the

modified Hill estimators and the changing block sizes. In fact, the estimators seem significantly less sensitive to the varying block sizes $r_n$ than they are to the varying choice of threshold $u_n$. Similar graphs for the bias and the standard errors of the estimators are included in Appendix A.

APPENDIX A

# ADDITIONAL SIMULATION STUDY FOR THE MODIFIED HILL ESTIMATORS

| Method | | Hill | | Modified Hill$_B$ | | Modified Hill$_D$ | |
|---|---|---|---|---|---|---|---|
| Process | $\alpha$ | Mean | RMSE | Mean | RMSE | Mean | RMSE |
| ARCH | 1.08 | 1.21 | 0.320 | 1.18 | 0.274 | 1.16 | 0.233 |
| ARCH | 2.0 | 2.10 | 0.392 | 2.05 | 0.337 | 2.01 | 0.289 |
| ARCH | 2.3 | 2.37 | 0.410 | 2.31 | 0.371 | 2.27 | 0.321 |
| MA | 1.0 | 1.06 | 0.183 | 1.04 | 0.176 | 1.03 | 0.150 |
| MA | 2.0 | 2.27 | 0.446 | 2.17 | 0.381 | 2.16 | 0.334 |
| MA | 2.5 | 2.94 | 0.633 | 2.79 | 0.527 | 2.77 | 0.468 |

Table A.1: Mean and root mean squared error of the regular Hill estimator and the modified Hill estimator. Data are generated using ARCH and Moving Average models with different values of $\alpha$ and a sequence length of $n = 5000$.

| Method | | Hill | | Modified Hill$_B$ | | Modified Hill$_D$ | |
|---|---|---|---|---|---|---|---|
| Process | $\alpha$ | Mean | RMSE | Mean | RMSE | Mean | RMSE |
| ARCH | 1.08 | 1.15 | 0.223 | 1.13 | 0.191 | 1.12 | 0.168 |
| ARCH | 2.0 | 2.05 | 0.275 | 2.02 | 0.259 | 2.02 | 0.235 |
| ARCH | 2.3 | 2.34 | 0.292 | 2.30 | 0.280 | 2.29 | 0.251 |
| MA | 1.0 | 1.03 | 0.122 | 1.01 | 0.128 | 1.01 | 0.115 |
| MA | 2.0 | 2.17 | 0.296 | 2.11 | 0.274 | 2.10 | 0.250 |
| MA | 2.5 | 2.82 | 0.436 | 2.70 | 0.380 | 2.70 | 0.355 |

Table A.2: Mean and root mean squared error of the regular Hill estimator and the modified Hill estimator. Data are generated using ARCH and Moving Average models with different values of $\alpha$ and a sequence length of $n = 20000$.

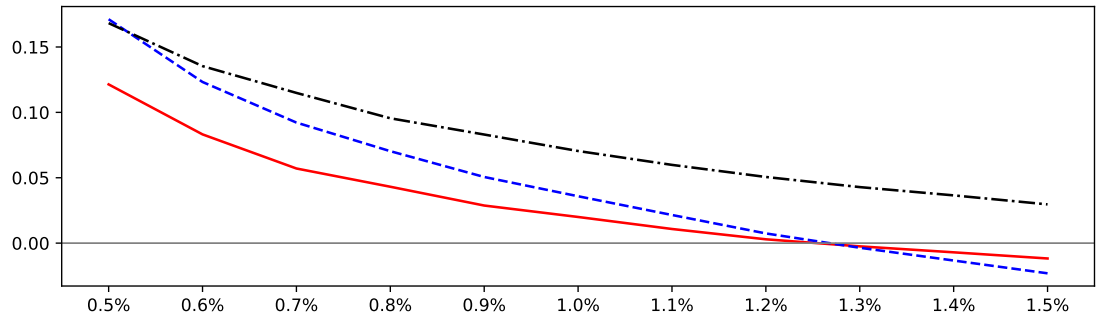


Figure A.1: Bias for the distance based modified Hill estimator (solid line), the blocks based modified Hill estimator (dotted line), and the regular Hill estimator (dot-dash line) plotted against percentage of largest order statistics used for estimation. Data are simulated from an ARCH(1) model with $\alpha = 2$.

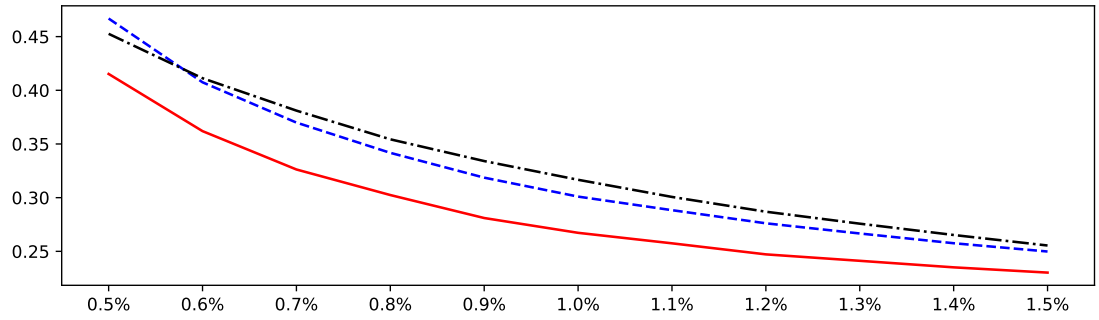


Figure A.2: Standard error for the distance based modified Hill estimator (solid line), the blocks based modified Hill estimator (dotted line), and the regular Hill estimator (dot-dash line) plotted against percentage of largest order statistics used for estimation. Data are simulated from an ARCH(1) model with $\alpha = 2$.

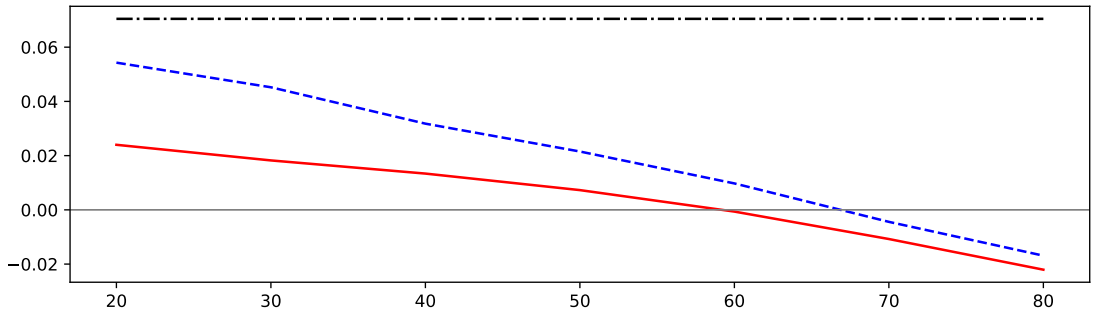


Figure A.3: Root mean squared error for the distance based modified Hill estimator (solid line), the blocks based modified Hill estimator (dotted line), and the regular Hill estimator (dot-dash line) plotted against $r_n$, the block size used for extremal index estimation. Data are simulated from an ARCH(1) model with $\alpha = 2$.

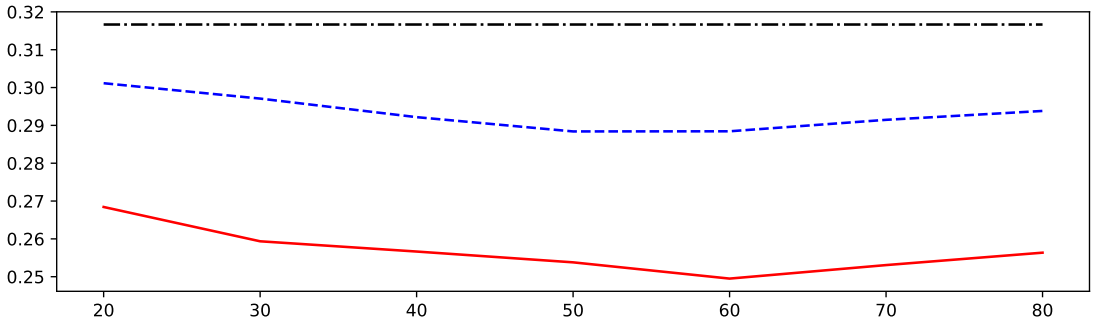


Figure A.4: Root mean squared error for the distance based modified Hill estimator (solid line), the blocks based modified Hill estimator (dotted line), and the regular Hill estimator (dot-dash line) plotted against $r_n$, the block size used for extremal index estimation. Data are simulated from an ARCH(1) model with $\alpha = 2$.

## BIBLIOGRAPHY


[1] S. Asmussen. *Ruin Probabilities*. World Scientific Publishing Company, 2000.

[2] A. Baltrunas, E.A.M. Omey, and S. Van Gulck. Hazard rates and subexponential distributions. *Publications de l'Institut Mathématique*, 80 (94):29–46, 2006.

[3] B. Basrak, R. Davis, and T. Mikosch. A characterization of multivariate regular variation. *Annals of Applied Probability*, 12:908–920, 2002.

[4] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels. *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.

[5] B. Berghaus and A. Bücher. Weak convergence of a pseudo maximum likelihood estimator for the extremal index. Preprint. Available at: `https://arxiv.org/pdf/1608.01903.pdf`, 2017.

[6] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, second edition, 1999.

[7] R.C. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2:107–144, 2005.

[8] V.P. Chistyakov. A theorem on sums of independent random variables and its applications to branching random processes. *Theory of Probability and its Applications*, 9:640–648, 1964.

[9] D.B.H. Cline and S. Resnick. Multivariate subexponential distributions. *Stochastic Processes and Their Applications*, 42:49–72, 1992.

[10] D.B.H. Cline and G. Samorodnitsky. Subexponentiality of the product of independent random variables. *Stochastic Processes and Their Applications*, 49:75–98, 1994.

[11] S. Csörgo, P. Deheuvels, and D. M. Mason. Kernel estimates of the tail index of a distribution. *The Annals of Statistics*, pages 1050–1077, 1985.

[12] S. Csörgo and D. M. Mason. Central limit theorems for sums of extreme values. *Mathematical Proceedings of the Cambridge Philosophical Society*, 98:547–558, 1985.

[13] R.C. Dahiya and J. Gurland. Goodness of fit tests for the gamma and exponential distributions. *Technometrics*, 14:791–801, 1972.

[14] R.A. Davis and T. Hsing. Point processes for partial sum convergence for weakly dependent random variables with infinite variance. *The Annals of Probability*, 23:879–917, 1995.

[15] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, New York, 2006.

[16] H. Drees. Bias correction for estimators of the extremal index. 2011. Preprint. Available at: `https://arxiv.org/pdf/1107.0935.pdf`.

[17] H. Drees and E. Kaufmann. Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, 75:149–172, 1998.

[18] D. Dupuis. Exceedances over high thresholds: a guide to threshold selection. *Extremes*, 1:251–261, 1998.

[19] P. Embrechts and C.M. Goldie. On closure and factorization properties of subexponential distributions. *Journal of Australian Mathematical Society, Series A*, 29:243–256, 1980.

[20] P. Embrechts, C.M. Goldie, and N. Veraverbeke. Subexponentiality and infinite divisibility. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 49:335–347, 1979.

[21] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin, 1997.

[22] C. Ferro and J. Segers. Inference for clusters of extreme values. *Journal of the Royal Statistical Society. Series B (Methodological)*, 65:545–556, 2003.

[23] S. Foss, T. Konstantopoulos, and S. Zachary. The principle of a single big jump: discrete and continuous time modulated random walks with heavy-tailed increments. *Journal of Theoretical Probability*, 20:581–612, 2007.

[24] C. Goldie. Subexponential distributions and dominated-variation tails. *Journal of Applied Probability*, 15:440–442, 1978.

[25] P. Hall. On some simple estimates of an exponent of regular variation. *J. Roy. Stat. Assoc.*, 44:37–42, 1982. Series B.

[26] P. Hall. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of multivariate analysis*, 32(2):177–203, 1990.

[27] B.M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3:1163–1174, 1975.

[28] J.B. Hill. On tail index estimation for dependent, heterogeneous data. *Econometric Theory*, 26(5):1398–1436, 2010.

[29] T. Hsing. Estimating the parameters of rare events. *Stochastic Processes and their Applications*, 37:117–139, 1991.

[30] T. Hsing. On tail index estimation using dependent data. *The Annals of Statistics*, 19:1547–1569, 1991.

[31] T. Hsing. Extremal index estimation for a weakly dependent stationary sequence. *The Annals of Statistics*, 21:2043–2071, 1993.

[32] T. Hsing, J. Hüsler, and M.R. Leadbetter. On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, 78:97–112, 1988.

[33] H. Hult and F. Lindskog. Extremal behavior for regularly varying stochastic processes. *Stochastic Processes and Their Applications*, 115:249–274, 2005.

[34] H. Hult and F. Lindskog. Heavy-tailed insurance portfolios: buffer capital and ruin probabilities. Technical Report, 2006.

[35] F. Laurini and J. A. Tawn. New estimators for the extremal index and other cluster characteristics. *Extremes*, 6(3):189–211, 2003.

[36] J.R. Leslie. On the non-closure under convolution of the subexponential family. *Journal of Applied Probability*, 26:58–66, 1989.

[37] D. Mason. Laws of large numbers for sums of extreme values. *Annals of Probability*, 10:756–764, 1982.

[38] T. Mikosch and C. Stărică. Limit thery for the sample autocorrelations and extremes of a garch(1,1) process. *The Annals of Statistics*, 28:1427–1451, 2000.

[39] T. Mori. Limit distributions of two-dimensional point processes generated by strong mixing sequences. *Yokohama Mathematical Journal*, 25:155–168, 1977.

[40] L. Németh and A. Zempléni. Regression estimator for the tail index. 2017. Preprint. Available at: `http://arxiv.org/pdf/1708.04815.pdf`.

[41] T. Nguyen and G. Samorodnitsky. Tail inference: where does the tail begin? *Extremes*, 15:437–461, 2012.

[42] P. J. Northrop. An efficient semiparametric maxima estimator of the extremal index. *Extremes*, 18(4):585–603, 2015.

[43] E.A.M. Omey. Subexponential distribution functions in $\mathbb{R}^d$. *Journal of Mathematical Sciences*, 138:5434–5449, 2006.

[44] E.A.M. Omey, F. Mallor, and J. Santos. Multivariate subexponential distributions and random sums of random vectors. *Advances in Applied Probability*, 38:1028–1046, 2006.

[45] E.A.M. Omey and R. Vesilo. The difference between the product and the convolution product of distribution functions in $\mathbb{R}^d$. *Publications de l'Institut Mathematique*, 89 (103):19–36, 2011.

[46] E.J.G. Pitman. Subexponential distribution functions. *Journal of Australian Mathematical Society, Series A*, 29:337–347, 1980.

[47] S. Resnick. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York, 2007.

[48] S. Resnick and C. Stărică. Tail index estimation for dependent data. *Annals of Applied Probability*, 8:1156–1183, 1998.

[49] S. Resnick and C. Stărică. Smoothing the Hill estimator. *Advances in Applied Probability*, 20:271–293, 1997.

[50] S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York, 1987.

[51] E. Rio. Inequalities and limit theorems for weakly dependent sequences. 2013.

[52] C. Robert, J. Segers, and C. Ferro. A sliding block estimator for the extremal index. *Electronic Journal of Statistics*, 3:993–1020, 2009.

[53] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 2015.

[54] M. Scotto, K. Turkman, and C. Anderson. Extremes of some sub-sampled time series. *Journal of Time Series Analysis*, 24, 2003.

[55] D.S. Silvestrov. *Limit theorems for randomly stopped stochastic processes*. Springer, London, 2004.

[56] R.L. Smith and I. Weissman. Estimating the extremal index. *ournal of the Royal Statistical Society. Series B (Methodological)*, 56:515–528, 1994.

[57] M Süveges. Likelihood estimation of the extremal index. *Extremes*, 10(1-2):41–55, 2007.

[58] I. Weissman and S. Novak. On blocks and runs estimators of the extremal index. *Journal of Statistical Planning and Inference*, 66:281–288, 1998.

[59] W. Whitt. *Stochastic-Process Limits. An Introduction to Stochastic-Process Limits and Their Applications to Queues*. Springer, New York, 2002.

[60] E. Willekens. *Higher order theory for subexponential distributions*. PhD thesis, K.U. Leuven, 1986. In Dutch.

[61] S. Zachary. A note on Veraverbeke's theorem. *Queueing Systems*, 46:9–14, 2004.