Doctoral Thesis

University of Trento

School of Social Sciences

PhD Program in Economics and Management

Heavy-tailed Phenomena and Tail Index Inference

a dissertation submitted to the doctoral school of economics and management in partial fulfillment of the requirements for the Doctoral degree (Ph.D.) in Economics and Management

Mofei Jia

December 2014

Supervisor:

Prof. Emanuele Taufer
Università degli Studi di Trento

Internal Evaluation Commission:

Prof. Simos Meintanis
National and Kapodistrian University of Athens

Prof. Pier Luigi Novi Inverardi
Università degli Studi di Trento

Final Committee:

Presidente:
Prof. Roberto Benedetti
Università degli Studi "G. D'Annunzio" di
Chieti-Pescara

Componente:
Prof. Alessandro Innocenti
Università degli Studi di Siena

Segretario:
Dott. Marco Faillo
Università degli Studi di Trento

Acknowledgements

This thesis would never have been done without the assistance of many people. I am really grateful to all the people who have helped me during this period. First of all, I would like to express my sincere gratitude to my supervisor Professor Emanuele Taufer, who introduced me to the field of my PhD research and provided me with many opportunities. From him, I have acquired not only academic knowledge, but also the ability and attitude necessary to carry out my research.

I am grateful to Professor Nikolai N. Leonenko, who was of great support during my visit at Cardiff University. I would also like to thank Professor Stefano Schiavo, who kindly provided the data I used for my PhD research. Special thanks to Professor Christopher L. Gilbert, who informed me of the call for PhD positions in the University of Trento. Thanks to Professor Svetlozar (Zari) T. Rachev for suggestions and communications via email. I am also grateful to my colleague Simone Pfuderer, who kindly provided me with a lot of help in my studies and life in Trento.

Finally, all thanks to my mother and father for all their love, understanding, and unconditional support.

Abstract

This thesis focuses on the analysis of heavy-tailed distributions, which are widely applied to model phenomena in many disciplines. The definition of heavy tails based on the theory of regular variation highlights the importance of the tail index, which indicates the existence of moments and characterises the rate at which the tail decays. Two new approaches to make inference for the tail index are proposed.

The first approach employs a regression technique and constructs an estimator of the tail index. It exploits the fact that the behaviour of the characteristic function near the origin reflects the behaviour of the distribution function at infinity. The main advantage of this approach is that it utilises all observations to constitute each point in the regression, not just extreme values. Moreover, the approach does not rely on prior information on the starting point of the tail behaviour of the underlying distribution and shows excellent performance in a wide range of cases: Pareto distributions, heavy-tailed distributions with a non-constant slowly varying factor, and composite distributions with heavy tails.

The second approach is motivated by the asymptotic properties of a special moment statistic, the so-called partition function. This statistic considers blocks of data and is generally used in the context of multifractality. Due to the interplay between the weak law of large numbers and the generalised central limit theorem, the asymptotic behaviour of the partition function is strongly affected by the existence of moments even for weakly dependent samples. Via a quantity, the scaling function, a graphical method to identify the existence of heavy tails is proposed. Moreover, the plot of the scaling function allows one to make inference for the underlying distribution: with infinite variance, finite variance with tail index larger than two, or all moments finite. Furthermore, since the tail index is reflected at the breakpoint of the plot of the scaling function, this gives the possibility to estimate the tail index.

Both these two approaches use the entire distribution, not just the tail, to analyse the tail behaviour. This sheds a new light on the analysis of heavy-tailed distributions. At the end of this thesis, these two approaches are used to detect power laws in empirical data sets from a variety of fields and contribute to the debate on whether city sizes are better approximated by a power law or a log-normal distribution.

Keywords: Heavy tails, Tail index, Regular variation, Partition function, Scaling function, Power law, Distribution of city sizes

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Normal distributions and the central limit theorem

Normal distributions are widely applied when probabilistically modelling phenomena in the real world, ranging from the natural sciences to the social sciences. One reason for assuming a normal distribution is that its bell shape fits empirical data fairly well, especially for data clustering around an average value. For instance, the average height of an adult female in the US is around 165 cm and the height of each individual rarely deviates substantially far from this value. The mean and variance characterise the distribution of the height well because the probability of meeting an adult female with twice or half as tall as the mean height is infinitesimal.

The central limit theorem (CLT) is another reason for using a normal distribution. Roughly speaking, the CLT states that the average of a sufficiently large number of independent and identically distributed (i.i.d.) random variables, with a finite mean and a finite variance, approximately has a normal distribution. Furthermore, normal distributions have some nice mathematical properties. For instance, the sum of two independent normally distributed variables is still normally distributed.

1.2 Heavy-tailed phenomena

However, there are many phenomena whose distributions deviate from the assumption of normal distributions. For instance, in the US, the average population of a place (city, town, or village) in 2010 was 7877[1]. However, quite a few places, such as New York, Los Angeles, Chicago and etc, have inhabitants more than 100,000, i.e., more than 100 times the average size. Another example one might encounter is damage from hurricanes. The estimated total property damage of Hurricane Katrina (2005) is $108 billion (2005 USD), nearly twice the damage of Hurricane Sandy (2012), the second costliest US hurricane. The cost of Hurricane Katrina is nearly triple the damage brought by Hurricane Andrew (1992), the costliest hurricane in the US prior to 2005. Similar examples can be found in daily returns of financial assets, the intensities of earthquakes, transmission rates of files and file sizes stored on a server, and fire insurance losses.

All the above phenomena have some common properties, i.e., the behaviour of the data is dominated by large values and the probability of exhibiting a huge value is relatively big. This kind of empirical data is described as heavy-tailed distributed, or following a power-law distribution. The description "heavy tails" or "power law", used frequently in the literature, refers to the fact that the tail probability decays to 0 at a constant power rate of the value of the observation x in contrast to an exponential rate of x (e.g., a normal distribution). This is opposite to many phenomena which can be easily characterised by their average values. Therefore, sometimes the commonly adopted assumption that the random variables under investigation follow a normal distribution is highly questionable. With regard to modern finance, Mina & Xiao (2001) (p.25) note:

"However, it has often been argued that the true distribution of returns (even after standardizing by the volatility) implies a larger probability of extreme returns than that implied from the normal distribution."

Heavy-tailed distributions discussed in this thesis are defined according

[1]This figure is derived from the tabulates in the US Census 2010.

to the theory of regularly varying functions, which behave asymptotically like power functions. A function G is regularly varying at infinity, if for every $x > 0$,

$$\lim_{t\to\infty} \frac{G(tx)}{G(t)} = x^{-\alpha}.$$

If $\alpha = 0$, G is the so-called slowly varying function, which is generically denoted by $L(x)$. Thus, the heavy-tailed distribution, or the power-law distribution, is defined as

$$P(X > x) = x^{-\alpha}L(x), \quad as \quad x \to \infty, \tag{1.1}$$

where $\alpha > 0$ (Feller, 1967). Examples of such distributions are Pareto, Student's t, Cauchy, F, and stable distributions. Some standard heavy-tailed distributions with specific forms of $L(x)$ (Wang & Tsai, 2009) are listed in Table 1.1. The tail index α is always positive in the analysis of heavy-tailed phenomena. Moreover, it is a commonly used parameter to describe the behaviour of the tail of a distribution: the smaller α the slower the decay of $P(X > x)$ to 0 as $x \to \infty$, and thus the more likely to generate extreme values.

To illustrate the marked difference between a power-law distribution and a normal distribution, the tail probability of each distribution is analysed here. Suppose N and H are random variables from the standard normal distribution with density function $\psi(x)$, and a power-law distribution, the standard Cauchy distribution with $\alpha = 1$, respectively. As $x \to \infty$, the right-tail probabilities are

$$P(N > x) \sim \frac{\psi(x)}{x} \sim \frac{C_1}{x}e^{-x^2/2}, \tag{1.2}$$

by Mill's ratio and

$$P(H > x) \sim C_2 x^{-\alpha}, \tag{1.3}$$

where $\alpha = 1$, respectively.

Figure 1.1 plots the two tail probabilities against the values of x. It is obvious that the tail of the standard normal distribution goes to 0 much

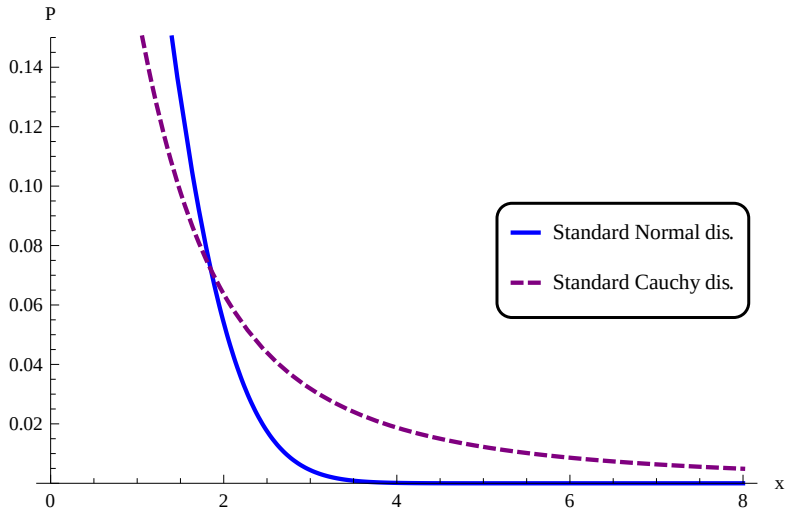
Table 1.1: A list of some special heavy-tailed distributions

Distribution	Probability density function	$\alpha > 0$	The function $L(x)$
Pareto	$\propto x^{-k-1}$	k	$\propto 1$
Generalized Pareto	$\propto (1 - kx\sigma^{-1})^{1/k-1}$	$-1/k$	$\propto (x^{-1} - k\sigma^{-1})^{1/k-1}$
Burr	$\propto x^{c-1}/(1 + x^c)^{k+1}$	ck	$\propto (x^{-c} + 1)^{-(k+1)}$
Student's t	$(1 + x^2 k^{-1})^{-(k+1)/2}$	k	$(x^{-2} + k^{-1})^{-(k+1)/2}$
F	$\propto x^{k_1/2-1}(k_1 + k_2 x)^{-(k_1+k_2)/2}$	$k_2/2$	$(k_1 x^{-1} + k_2)^{-(k_1+k_2)/2}$

The notation $\propto$ refers to "proportional to a constant".

faster than does the tail of the Cauchy distribution. Mathematically, for any positive α, C_1, and C_2, the ratio of the tail probability in (1.2) to that in (1.3) goes to 0 as $x \to \infty$. It is straightforward that modelling using power-law distributions instead of normal distributions generates a much higher probability of larger values and thus captures the important feature of heavy-tailed phenomena.

Figure 1.1: Tail probabilities of the standard normal and Cauchy distributions



However, modelling using power-law distributions is very different from modelling using classical normal distributions. A number of classical statistics and their inferences are established based on averages and moments of samples. If the tail probability follows (1.3) with $\alpha > 0$, moments with order higher than or equal to α do not exist. This follows since

$$E(X^\rho) = \int_0^\infty x^{\rho-1} P(X > x)\,\mathrm{d}x \approx \int_1^\infty x^{\rho-1} x^{-\alpha}\,\mathrm{d}x \begin{cases} < \infty, & \text{if } \rho < \alpha, \\ = \infty, & \text{if } \rho \geq \alpha, \end{cases}$$

where $\int f(x) \approx \int g(x)$ means that the limiting behaviours of both integrals are the same, i.e., either convergent or divergent (Resnick, 2006).

What will happen to a statistical world relying heavily on moments if

the moments do not exist? As a minimum, the classical CLT does not apply to the cases with tail index $\alpha \leq 2$, since the finite variance assumption is violated. Instead, one appeals to the generalised central limit theorem (GCLT), which says that the only possible non-trivial resulting limits are stable distributions for infinite variance models. Another statement of the GCLT is that the normalised sum of i.i.d. random variables belongs to the domain of attraction of a stable distribution. A random variable X is in the domain of attraction of a stable distribution Z if constants $a_n > 0$, b_n exist such that

$$a_n(X_1 + X_2 + \cdots + X_n) - b_n \xrightarrow{d} Z$$

holds when $X_1, X_2, X_3, \ldots$ are i.i.d. copies of X (Rachev, 2003). Here let $\xrightarrow{d}$ denote convergence in distribution.

1.3 Statistical inference for heavy-tailed phenomena: the tail index α

The research question in this thesis is, thus, how to analyse heavy tails by statistical method. Generally, the following two questions should attract attention when heavy tails are suspected:

1. identify the existence of heavy tails, and then
2. estimate the tail index α of the underlying distribution.

To investigate these two questions, numerous graphical and estimation methods have been proposed in the literature: log-log plot, quantile-quantile (QQ) plot, Pickands estimation, Hill estimation, etc. In the rest of this section, a number of commonly used techniques to detect heavy tails and methods of estimating the tail index are presented.

Perhaps the simplest tool for detecting heavy tails is the tail-probability plot in a log-log scale, which dates back to Pareto (1896) who studied the distribution of income. Taking the logarithm of both sides of (1.1), with $L(x)$ replaced with a real non-zero constant C, it can be seen that a power-law

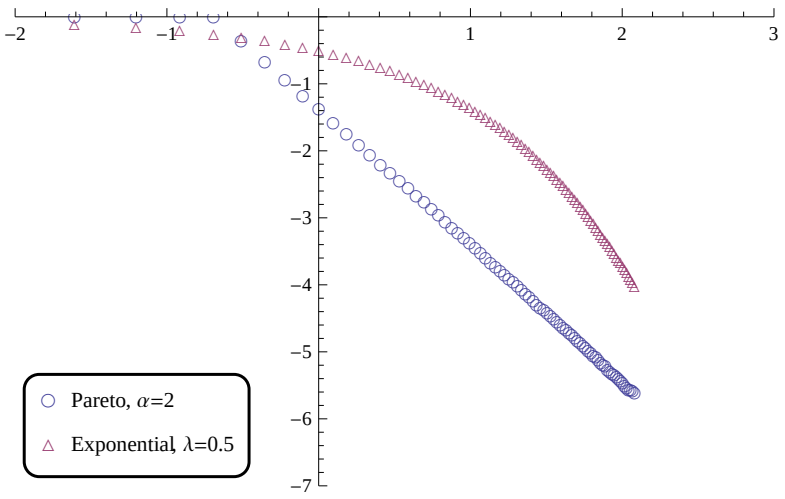
distribution satisfies the relation

$$\ln P(X > x) = \ln C - \alpha \ln x.$$

This suggests that the log-log plot of the tail probability for a power-law distribution is a straight line, whose absolute slope is the tail index α.

However, the approximately straight line may also be caused by some non-power-law distributions. Figure 1.2 shows the log-log plots for two data sets drawn from Pareto and exponential distributions, respectively. Although over the entire distribution these two plots are very different, both plots approximately follow a straight line for $\ln x > 1.5$. Hence, it may be intrinsically difficult to discern a power-law distribution by the log-log plot of the tail probability, if the threshold x from which onwards a power law holds is unknown.

Figure 1.2: Log of tail probabilities versus log of inputs for Pareto and exponential distributions, $n = 10^5$



A graphical technique called the QQ plot is another exploratory method often applied to analyse heavy tails. For a sample $X_1, X_2, \ldots, X_n$, set

$$X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$$

for the order statistics indexed smallest to largest. Plot the quantiles

$$\left\{\left(-\ln\left(1-\frac{i}{n+1}\right), \ln X_{i,n}\right),\ 1 \leq i \leq n\right\}.$$

If the distribution has a heavy tail, or at least has approximately a heavy tail, the plot should be roughly linear with slope $1/\alpha$. The essence of this method is the standard QQ plot of log-transformed data on exponential quantiles. This graphical method is exploited to define a QQ estimator, which is based on the k upper-order statistics chosen by visual observation of the linear portion of the QQ plot (Kratz & Resnick, 1996). This QQ estimator is usually highly sensitive to the choice of k.

The Pickands estimator which uses three kinds of upper order statistics (Pickands III, 1975) and the moment estimator based on the "moments" of the Hill estimator (Dekkers et al., 1989) are other methods for deciding whether a distribution is heavy-tailed or not. Both these two methods estimate γ, a parameter used to characterise the family of extreme-value distributions, which is equal to $1/\alpha$ if $\gamma > 0$. If $\gamma \leq 0$, it indicates that the assumption of a heavy-tailed distribution is inappropriate.

An extensive literature deals with the estimation of the tail index α using the Hill estimator (Hill, 1975). The Hill estimator is defined as follows: let $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$ denote the order statistics of the nonnegative observations $X_1, X_2, \cdots X_n$. The Hill estimator based on k upper order statistics is

$$\hat{\alpha}_k^{Hill} = \left(\frac{1}{k}\sum_{i=1}^{k} \ln \frac{X_{(i)}}{X_{(k+1)}}\right)^{-1}.$$

The consistency of the Hill estimator for i.i.d., weakly dependent, or linear data were subsequently shown by Mason (1982), Hsing (1991), and Resnick & Stărică (1995), respectively. Its asymptotic normality was discussed by Hall (1982); Haeusler & Teugels (1985); de Haan & Resnick (1998), etc. However, the Hill estimator is subject to the same difficulty as the QQ estimator, namely the choice of k. Several adaptive selection methods have

been proposed (see Hall & Welsh (1985); Beirlant et al. (1996) and references therein). Furthermore, unlike the Pickands and moment estimators, the Hill estimator is uninformative on whether the assumption of heavy tails is appropriate or not.

Most of the aforementioned estimators are constructed from the extreme order statistics based on the tail probability (1.1). This tail probability function is only specified in the neighbourhood of infinity. It is reasonable to expect that only a relatively small proportion of the upper order statistics is used in the estimation. The estimation is therefore surprisingly difficult even when large samples are available.

As an alternative, some other statistics, rather than upper order statistics, are used to build up estimators of the tail index. Several examples are listed here: the estimators proposed by Politis (2002) using diverging statistics, by Meerschaert & Scheffler (1998) using the sample variance and sample size, and by McElroy & Politis (2007) using over subsets of the whole data set.

1.4 Motivation and structure

This thesis introduces two new techniques which utilise the entire data set to detect heavy tails and estimate the tail index. Both techniques start with the assumption about the tail probability of the distribution. However, in order to avoid the use of extreme order statistics in the estimation, they consider different aspects rather than the cumulative distribution function (CDF). The first approach builds on the relation between the characteristic function (CF) and the CDF. The second approach is constructed from the partition function (PF), a special kind of moment statistic. In the asymptotic mean squared error sense, no estimator can dominate others as studied in de Haan & Peng (1998). Hence, the estimators proposed in this thesis are not designed to outperform others in all contexts. Instead, the purpose of this thesis is to shed a new light on considering the entire sample, not just the tail section, in the analysis of heavy-tailed phenomena.

This thesis consists of three main chapters. Chapter 2 suggests an estimator based on the empirical characteristic function (ECF) using a

regression technique. This regression estimator is motivated by the two-point-ratio estimator introduced by Welsh (1986), which exploits the relationship between the distribution function at infinity and the characteristic function near the origin (Pitman, 1968). The ECF is also utilised by Koutrouvelis (1980) to estimate the four parameters for stable distributions, whose characteristic function is specified. His estimation is based on a log-log regression on a variant of the logarithm of the ECF $\phi_n(t)$ and variable t. However, he does not discuss the theoretical properties of the proposed estimators. The properties are investigated only through a simulation study. Furthermore, the relation between the CF and its empirical counterpart with the tail behaviour of the underlying distribution has been exploited by Meintanis & Koutrouvelis (1990) and Donatos & Meintanis (1996). For the regression estimator proposed in this thesis, theoretical properties such as consistency and asymptotic normality are obtained. Simulation studies and an empirical example are subsequently presented to illustrate the theoretical findings.

Chapter 3 uses the PF to analyse heavy tails. The PF is a sample moment statistic often used in the context of multifractality. The asymptotic behaviour of the PF is strongly influenced by the tail of the underlying distribution. The scaling function links the PF and the tail behaviour. A graphical method to detect heavy tails and estimation methods of the tail index based on the scaling function and the PF are proposed. To some extent, the underlying idea of these methods is based on the asymptotic properties of sums, which is analogous to the estimation method proposed by Meerschaert & Scheffler (1998). However, the methods introduced in this thesis are more general and work not only for the i.i.d. variables but also for weakly dependent samples.

Chapter 4 focuses on applications of the two methods presented in Chapters 2 and 3. In order to evaluate the performance, several empirical data sets in various fields, especially data sets on the distribution of city sizes, are analysed. There is a heated ongoing debate on whether the distribution of city sizes is fitted better by a power-law tail or a log-normal tail. Besides the curiosity of the nature of the underlying distribution, there

are theoretical reasons for investigating this issue: distinct implications are derived by different laws. For instance, Zipf's law, a power law with $\alpha = 1$, is suggested by Gabaix (1999), while the log-normal distribution is consistent with the proportionate growth process proposed by Gibrat (1931). Hence, two potential approaches are proposed to distinguish power-law tails from log-normal tails. Chapter 5 summarises and concludes.

Chapter 2

Regression estimation of the tail index

2.1 Introduction

Heavy-tailed distributions occur in a wide range of situations where extreme events are more likely to happen than they would under a Gaussian distribution. Insurance losses, financial log returns, hyperlinks on the World Wide Web, intensities of earthquakes, the population of cities are all examples that follow heavy-tailed distributions.

One definition of distributions with heavy tails is based on the theory of regularly varying functions. Roughly speaking, regularly varying functions resemble those functions which behave like power functions at infinity.

Definition 2.1. *(Feller, 1967) A measurable function $G : (0, \infty)$ is regularly varying at infinity with exponent $-\alpha$ $(0 < \alpha < \infty)$ if for every $x > 0$,*

$$\lim_{t \to \infty} \frac{G(tx)}{G(t)} = x^{-\alpha},$$

where α is called the index of regular variation, or the tail index. We refer to α-varying functions as RV_α. If $\alpha = 0$, G is said to vary slowly, which is generically denoted by $L(x)$. In other words, G varies regularly iff it is of the form $x^{-\alpha}L(x)$.

The problem of estimating the tail index of a distribution has a long history. Pareto (1896) uses the absolute slope of the log-log plot of the tail probability to determine the tail index of the distribution of income. However, the accuracy of this estimator entirely depends on the upper percentage of the data used to make the log-log plot and the starting point of the tail convergence behaviour for the underlying distribution (Fofack & Nolan, 1999).

At present, a range of estimators for the unknown tail index exist. A simple and old estimator is the Pickands estimator, a linear combination of log-spacing order statistics (Pickands III, 1975). Improvements of this estimator have been sought and discovered by scholars (e.g., Falk, 1994; Segers, 2005). The Hill (1975) estimator based on the upper order statistics and their asymptotic properties can be considered as the classical tail index estimator. Extensions of the Hill estimator have been proposed: for instance, the family of j-moment ratio estimators (Danielsson et al., 1996) and the smoothing Hill estimator (Resnick & Stărică, 1997). Maximum likelihood estimation, which relies on numerical optimisation, has been considered by Smith (1987). The moment estimator introduced by Dekkers et al. (1989) utilises empirical moments of the limiting distribution.

Estimators of the tail index are usually based on extreme order statistics and their asymptotic properties. As an alternative, estimators proposed by Politis (2002) and McElroy & Politis (2007) are based on the growth of appropriately chosen diverging statistics. The inspiration for this approach is the introduction of an estimator based on the ratio of the logarithm of sample variance to the logarithm of sample size (Meerschaert & Scheffler, 1998).

In general, for a clear distinction between tail weights, very large samples are required. The reason is that the distinction rests on a very small percentage of the empirical data, possibly less than 0.01 per cent (Heyde & Kou, 2004). For instance, if the 99.9 per cent or higher quantile is used in the distinction, one would require nearly 50,000 observations even at a 95 per cent confidence level. Therefore, an estimator based on extreme order statistics is not reliable when large samples are not available. For

instance, the "Hill horror plot" (Embrechts et al., 1997), displaying the Hill estimate against the number of order statistics used for the construction of the Hill estimate, shows that the Hill estimator may have poor behaviour, even under conditions where the estimation method applies. It is not surprising that the literature on the estimation of the tail index is subject to considerable uncertainty and controversy.

To avoid the use of extreme order statistics, the characteristic function (CF) can be applied in the estimation. The idea is based on the relation between the behaviour of the distribution function at infinity and the behaviour of the CF near the origin (Pitman, 1968). In this approach, the whole sample is utilised. For instance, Welsh (1986) investigates simple estimators by using two-point ratio estimation around the origin. However, in the absence of prior information, the practical way of choosing these two points near the origin is not clear. A variant of the empirical characteristic function (ECF) is used by Koutrouvelis (1980) to estimate the parameters for stable distributions by regression techniques. His results rely on simulations but the theoretical properties of the estimators have not been investigated yet. Further research to exploit the relation between the CF and its empirical counterpart with the tail behaviour of the underlying distribution has been carried out by Meintanis & Koutrouvelis (1990) and Donatos & Meintanis (1996).

In this chapter, we present a regression method to estimate the tail index α utilising the ECF. This method is based on a linear regression on the logarithm of the real part of the ECF near the origin. The theoretical properties of the estimator, including its bias, variance and asymptotic distribution, are derived. Moreover, this method is applied under a semi-parametric assumption about the tails of the distribution rather than a fully parametric assumption as that in Koutrouvelis (1980).

This chapter is organised as follows. In Section 2.2, we outline the framework of the methodology and describe the proposed estimator. The theoretical properties of this estimator and potential implementation problems are discussed in Section 2.3. Section 2.4 summarises the performance of the proposed estimator compared with that of the Hill

estimator in a simulation study. An empirical example is presented in Section 2.5. A short summary is provided in Section 2.6. Proofs of theoretical details are deferred to Section 2.7.

Throughout this thesis, we let $\xrightarrow{d}$ denote convergence in distribution. Let $a_n = O(b_n)$ if the ratio $|a_n/b_n|$ is bounded for large n. Let $a_n = o(b_n)$ if the ratio $|a_n/b_n|$ converges to zero. Also, $X_n = o_p(1)$ means that X_n converges in probability to zero. Finally, Let $X_n = O_p(1)$ mean, for every $\eta > 0$, a constant $K(\eta)$ and an integer $n(\eta) \leq n$ exist to make $P\{|X_n| \leq K(\eta)\} \geq 1-\eta$ hold.

2.2 Methodology

We provide some formal definitions first. Suppose we observe n independent random variables $X_1, X_2, \ldots, X_n$ with distribution function $F(x)$. The tail sum, $H(x)$, is defined for $x \geq 0$ by

$$H(x) = 1 - F(x) + F(-x),$$

where $1 - F(x)$ and $F(-x)$ represent the upper and lower tails, respectively. We assume that the tail sum of F is regularly varying at infinity, but otherwise arbitrary, i.e., the tail sum satisfies the assumption that

$$H(x) = x^{-\alpha}L(x), \quad as \quad x \to \infty, \tag{2.1}$$

where $\alpha > 0$ and $L(x)$ is slowly varying at infinity.

Remark 2.1. *Here, we suppose that both the upper and lower tails are regularly varying at infinity with index α_1 and α_2, respectively. If the two tails have the same index, i.e., $\alpha_1 = \alpha_2 = \alpha$, then the tail sum $H(x)$ will regularly vary with index α. If the two tails have different indices, the one with the smaller index (the tail probability decays to zero more slowly) dominates in the tail sum, i.e., $\alpha = \min(\alpha_1, \alpha_2)$. Moreover, if X is a non-negative random variable, i.e., $F(x) = 0$ for $x < 0$, the tail sum $H(x)$ is equal to $1 - F(x)$. In this case, the tail sum just reflects the upper tail.*

Let ϕ denote the CF of X, i.e., for all real t,

$$\phi(t) = E[e^{itx}] = \int_{-\infty}^{\infty} e^{itx}\,\mathrm{d}F(x),$$

where i is the imaginary unit. Furthermore, let

$$\phi(t) = U(t) + iV(t), \tag{2.2}$$

where $U(t) = \int_{-\infty}^{\infty} \cos(tx)\,\mathrm{d}F(x)$ is the real part of the CF, and $V(t) = \int_{-\infty}^{\infty} \sin(tx)\,\mathrm{d}F(x)$ is the imaginary part. Integrating (2.2) by parts (more details in the proof of Theorem 2.1), we obtain

$$1 - U(t) = t\int_{0}^{\infty} \sin(tx)H(x)\,\mathrm{d}x. \tag{2.3}$$

This means that the behaviour of the tail sum $H(x)$ depends only on $U(t)$, the real part of the CF.

By Pitman (1968), for an infinite-variance distribution whose $H(x)$ is of index $0 < \alpha < 2$ as $x \to \infty$,

$$1 - U(t) \sim s(\alpha)H(1/t) = s(\alpha)L(1/t)t^{\alpha}, \quad as \quad t \downarrow 0,$$

where the function $s(q)$ is defined as follows.

$$s(q) = \begin{cases} \frac{\pi/2}{\Gamma(q)\sin(q\pi/2)}, & \text{if } q > 0, \\ 1, & \text{if } q = 0. \end{cases}$$

$s(q)$ is finite for any q which is not an even positive integer.

For the special case that $H(x)$ is of index $\alpha = 2$ as $x \to \infty$, we obtain

$$1 - U(t) \sim t^2 \int_{0}^{1/t} xH(x)\,\mathrm{d}x, \quad as \quad t \downarrow 0. \tag{2.4}$$

The term $\int_0^{1/t} xH(x)\,\mathrm{d}x$ will be analysed in detail later.

In the case $\alpha > 2$, the distribution has a finite second moment μ_2, i.e.,

$$\mu_2 = \int_{-\infty}^{\infty} x^2 \mathrm{d}F(x) = -\int_0^{\infty} x^2 \mathrm{d}H(x) = \int_0^{\infty} 2xH(x)\mathrm{d}x < \infty. \tag{2.5}$$

By a Taylor expansion,

$$\begin{aligned} t\int_0^{\infty} \sin(tx)H(x)\mathrm{d}x &= t^2 \int_0^{\infty} xH(x)\mathrm{d}x + O(t^4) \\ &= \frac{1}{2}\mu_2 t^2 + o(t^2), \quad as \quad t \downarrow 0. \end{aligned}$$

Therefore,

$$1 - U(t) = \frac{1}{2}\mu_2 t^2 + o(t^2), \quad as \quad t \downarrow 0.$$

In summary, as $t \downarrow 0$, we have a relation of the form

$$1 - U(t) \sim \begin{cases} s(\alpha)L(1/t)t^{\alpha}, & \text{if } 0 < \alpha < 2, \\ t^2 \int_0^{1/t} xH(x)\,\mathrm{d}x, & \text{if } \alpha = 2, \\ \frac{1}{2}\mu_2 t^2, & \text{if } \alpha > 2. \end{cases}$$

Taking the logarithm of both sides of the above equation, as $t \downarrow 0$, we obtain

$$\ln(1 - U(t)) \sim \begin{cases} \ln\left(s(\alpha)L(1/t)\right) + \alpha \ln t, & \text{if } 0 < \alpha < 2, \\ \ln \int_0^{1/t} xH(x)\,\mathrm{d}x + 2\ln t, & \text{if } \alpha = 2, \\ \ln \frac{\mu_2}{2} + 2\ln t, & \text{if } \alpha > 2. \end{cases} \tag{2.6}$$

In order to define an estimator of α in this chapter, relationships in (2.6) are exploited.

Since the real part of the CF, $U(t)$, is generally unknown, we simply replace it with the real part of the ECF, i.e., $U_n(t) = \frac{1}{n}\sum_{i=1}^{n} \cos(X_i t)$. We evaluate (2.6) at points $t_1, t_2, \ldots, t_m$ around the origin and rearrange it to obtain

$$\ln\left(1 - U_n(t_j)\right) \sim \ln C' + \alpha' \ln t_j + \ln \frac{1 - U_n(t_j)}{1 - U(t_j)}, \quad j = 1, 2, \ldots, m, \tag{2.7}$$

where the values of C' and α' depend on which case of (2.6) is considered.

Set

$$y_i = \ln(1 - U_n(t_j)),\ Z_j = \ln t_j \text{ and } \epsilon_j = \ln \frac{1 - U_n(t_j)}{1 - U(t_j)}$$

and note the formal similarity of (2.7) to a simple linear regression. The proposed estimator of α is the least squares estimator of the slope coefficient in the least squares regression data $(y_j, Z_j),\ j = 1, 2, \ldots, m$. To obtain the ECF and then each point in the regression, all the observations in the sample, rather than only a few extreme order statistics, are used. In the next section we will carefully evaluate the properties of this estimation strategy by discussing the choice of points $t_1, t_2, \ldots, t_m$ and analysing the impact on the errors ϵ_j and the "constant" C'.

Remark 2.2. *This estimation method has some analogy to the method used by Geweke & Porter-Hudak (1983) in the estimation of long memory time series models. In that paper, the relation between the spectral density function at the origin and the covariance function at infinity is exploited.*

Since the size of the bias of the proposed estimator depends critically on the behaviour of "$L(x)$" in (2.1), some expansions must be elaborated upon if we are to describe the bias. In this chapter, we follow the same assumption as Hall (1982), i.e., let

$$L(x) = C[1 + Dx^{-\beta} + o(x^{-\beta})].$$

Stated formally:

Assumption 2.1.

$$H(x) = Cx^{-\alpha}[1 + Dx^{-\beta} + o(x^{-\beta})], \quad as \quad x \to \infty, \tag{A1}$$

where $C > 0, \alpha > 0, \beta > 0$ and D is a non-zero real number.

Remark 2.3. *Some classes of distributions which satisfy (A1) are displayed below (Hall & Welsh, 1985):*

1. *Stable distributions with index $1 < \alpha < 2$. In this case, we have $\beta/2 < \alpha \leq \beta$, a relationship which also holds for $\alpha = 1$.*

2. *Extreme value distributions with* $F(x) = e^{-x^{-\alpha}}$, $x > 0$ *or stable distributions with index* $0 < \alpha < 1$. *In this case, we have* $\alpha = \beta$.

3. *Powers of "smooth" distributions* X. *It means that if* $X = Y^{-1/\alpha}$, *then* Y *admits a Taylor series expansion of at least three terms about the origin. In this case, we also have* $\alpha = \beta$.

Analogous to Pitman (1968) and Welsh (1986), we formally connect the CF near the origin and the distribution function at infinity in the following theorem:

Theorem 2.1. *Suppose* $(A1)$ *holds with* $0 < \alpha < 2$ *and there exists a non-negative integer* p *such that* $2p < \alpha + \beta < 2p + 2$. *Then as* $t \downarrow 0$,

$$1 - U(t) = Cs(\alpha)t^{\alpha} + D_1 t^{\gamma} + o(t^{\gamma}), \tag{2.8}$$

where $\gamma = \min\{\alpha + \beta, 2\}$ *and the constant* $D_1 = CDs(\alpha + \beta)$ *if* $\alpha + \beta < 2$, *otherwise* D_1 *is not specified unless the form of the remainder term in* $(A1)$ *is known.*

Remark 2.4. *Although relation* (2.8) *only applies to the case* $0 < \alpha < 2$, *it can be easily extended to the general case* $\alpha > 0$. *In more detail, if the distribution of* X *satisfies* $(A1)$ *with tail index* $\alpha > 0$, *for any* $w > \alpha/2$, X^w *has a distribution function satisfying* (2.8) *with tail index* $0 < \alpha/w < 2$. *Moreover, if* $w > (\alpha + \beta)/2$, *then* $(\alpha + \beta)/w < 2$ *and thus* D_1 *is specified. Therefore, it is convenient to focus on the case* $0 < \alpha < 2$ *first. The most important class of distributions satisfying* $(A1)$ *is the class of stable distributions.*

Under $(A1)$, as $t \downarrow 0$, for $0 < \alpha < 2$, relation (2.6) is rewritten as

$$\begin{aligned} \ln(1 - U(t)) &= \alpha \ln t + \ln[Cs(\alpha)] + \ln\left[1 + \frac{D_1}{Cs(\alpha)} t^{\gamma-\alpha} + o(t^{\gamma-\alpha})\right] \\ &= \alpha \ln t + \ln[Cs(\alpha)] + \frac{D_1}{Cs(\alpha)} t^{\gamma-\alpha} + o(t^{\gamma-\alpha}), \end{aligned}$$

where the second step is obtained by a Taylor expansion on t close to zero.

In summary, the final form of the regression is

$$y = C_\alpha + \alpha \ln t + \ln \frac{1 - U_n(t)}{1 - U(t)}, \quad as \quad t \downarrow 0. \tag{2.9}$$

Here, the "constant" of this regression is

$$C_\alpha = \ln [Cs(\alpha)] + \frac{D_1}{Cs(\alpha)} t^{\gamma - \alpha} + o(t^{\gamma - \alpha}).$$

The regression estimator of α is given by

$$\hat{\alpha} = \frac{\sum_{j=1}^{m} a_j y_j}{S_{zz}}, \tag{2.10}$$

where $a_j = Z_j - \bar{Z} = \ln t_j - \frac{1}{m} \sum_{k=1}^{m} \ln t_k$, and $S_{zz} = \sum_{i=1}^{m} (Z_i - \bar{Z})^2 = \sum_{i=1}^{m} a_i^2$. The intercept is calculated by $\hat{C_\alpha} = \bar{y} - \hat{\alpha} \bar{Z}$.

2.3 Theoretical properties and extensions

This section starts by analysing the theoretical properties of the proposed estimator. Some extensions of this regression method are given afterwards.

2.3.1 Theoretical properties

The error term can be rewritten as

$$\varepsilon_j = \ln [1 + \frac{U(t_j) - U_n(t_j)}{1 - U(t_j)}].$$

The definition and convergence properties of the ECF show that

$$E\{U_n(t_i) - U(t_i)\}\{U_n(t_j) - U(t_j)\} = (2n)^{-1}\{U(t_i + t_j) + U(t_i - t_j) - 2U(t_i)U(t_j)\}.$$

As $t_j \downarrow 0$ and with relation (2.8), we obtain

$$\begin{aligned} Var\{[1-U(t_j)]^{-1}[U(t_j)-U_n(t_j)]\} &= \frac{[U(2t_j)+U(0)-2U(t_j)^2]}{2n[1-U(t_j)]^2} \\ &= \frac{(2-2^{\alpha-1})}{Cs(\alpha)nt_j{}^{\alpha}}\{1+o(1)\}. \end{aligned} \tag{2.11}$$

To make sure this variance is finite, nt_j^{α} needs to diverge as $n \to \infty$ for all t_j's. Heuristically, one needs to evaluate the CF at the origin. Let $t_j = j/n^{\eta}$ with $j = 1, 2, ..., m = n^{\delta}$ for some $0 < \delta < \eta < 1$ be the general form to select all points in the regression. In order to make $nt_j^{\alpha} \to \infty$ hold as $n \to \infty$ for all $0 < \alpha < 2$, the condition $0 < \delta < \eta \leq 1/2$ is necessary. Moreover, to have as many points as possible in the regression, $t_j = j/\sqrt{n}$ with $j = 1, 2, ..., m = n^{\delta}$ and $0 < \delta < 1/2$ seems to be a sensible choice (this issue will be further discussed in Section 2.3.2). The closer t_j to zero, the higher the variance of the corresponding term is and the higher variability presents in the regression.

Furthermore, the term $\frac{U(t_j)-U_n(t_j)}{1-U(t_j)}$ can be treated as a random variable with mean zero and variance (2.11). A Taylor expansion of $\ln(1+x)$ equal to $x + o(x)$ as $x \to 0$ allows us to get rid of the logarithm in the error term (e.g., Theorem 14.4-1 in Bishop et al. (2007)), i.e.,

$$\varepsilon_j = \frac{U(t_j)-U_n(t_j)}{1-U(t_j)} + o_p((nt_j^{\alpha})^{-\frac{1}{2}}). \tag{2.12}$$

Based on the discussion above, the following conclusion is derived:

Lemma 2.1. *Suppose* $(A1)$ *holds with* $0 < \alpha < 2$*, as* $n \to \infty$*, for* $t_i > t_j > 0$ *and* $\lambda = t_i/t_j$*:*

$$\begin{aligned} E(\varepsilon_j) &= 0. \\ Var(\varepsilon_j) &= (nt_j^{\alpha}Cs(\alpha))^{-1}(2-2^{\alpha-1})\{1+o(1)\}. \\ Cov(\varepsilon_i,\varepsilon_j) &= \begin{cases} \frac{\{2\lambda^{\alpha}+2-(\lambda+1)^{\alpha}-(\lambda-1)^{\alpha}\}}{(2nt_i^{\alpha}Cs(\alpha))}\{1+o(1)\}, & \text{if } \lambda < \infty, \\ (nt_i^{\alpha}Cs(\alpha))^{-1}\{1+o(1)\}, & \text{if } \lambda \to \infty. \end{cases} \end{aligned}$$

Note that the dependence between two errors decays as the distance between them increases.

With Lemma 2.1, asymptotic expressions of the bias and variance of the estimator are obtained:

Theorem 2.2. *Suppose* $(A1)$ *holds with* $0 < \alpha < 2$*, and* $\hat{\alpha}$ *is defined in* (2.10) *with* $t_j = j/\sqrt{n}$ *for* $j = 1, 2, ..., m = n^\delta$ *with* $0 < \delta < 1/2$*. Then as* $n \to \infty$*,*

$$E(\hat{\alpha}) - \alpha = \frac{D_1(\gamma - \alpha)}{Cs(\alpha)(\gamma - \alpha + 1)^2} n^{(\delta - 1/2)(\gamma - \alpha)} \{1 + o(1)\}.$$

Since $\gamma > \alpha$, the bias goes to zero as $n \to \infty$. For the same α and γ, the bias goes to zero faster as $n \to \infty$ for a smaller δ. The size of the bias is jointly determined by the δ we choose and C_α, the constant in (2.9).

The variance of the estimator is given in the following theorem:

Theorem 2.3. *Suppose* $(A1)$ *holds with* $0 < \alpha < 2$*, and* $\hat{\alpha}$ *is defined in* (2.10) *with* $t_j = j/\sqrt{n}$ *for* $j = 1, 2, ..., m = n^\delta$ *with* $0 < \delta < 1/2$*. Then as* $n \to \infty$*,*

$$Var(\hat{\alpha}) = O\{n^{\frac{\alpha}{2} - 1 - \delta\alpha}\}.$$

Here, the exponent of n, $\frac{\alpha}{2} - 1 - \delta\alpha$, is always less than 0 in the regression. Therefore, the variance of $\hat{\alpha}$ converges regardless of the value of δ.

Combining Theorems 2.2 and 2.3, the mean squared error (MSE) is derived as

$$MSE(\hat{\alpha}) = O\{n^{2(\delta - 1/2)(\gamma - \alpha)}\} + O\{n^{\frac{\alpha}{2} - 1 - \delta\alpha}\}, \tag{2.13}$$

as $n \to \infty$, under $(A1)$ with $0 < \alpha < 2$. Similarly, the MSE goes to zero as $n \to \infty$ for all $0 < \delta < 1/2$. Therefore, $\hat{\alpha}$ is a consistent estimator of α for all $0 < \alpha < 2$.

With regard to the explicit size of the MSE, it is determined by the term which prevails on the right-side of (2.13). The difference between these two orders is $2\delta\gamma - \gamma - \delta\alpha + \frac{\alpha}{2} + 1$. If this value is less than zero, i.e., $\delta < \frac{1}{2} - \frac{1}{2\gamma - \alpha}$, the second term dominates in the MSE. More precisely, if $\gamma = 2$, $2\gamma - \alpha$ is always between 2 and 4 for all $0 < \alpha < 2$ and the solution of $\delta < \frac{1}{2} - \frac{1}{2\gamma - \alpha}$

exists under the condition $0 < \delta < 1/2$. If $\gamma = \alpha + \beta$, a combination of small α and β can make $\frac{1}{2} - \frac{1}{2\gamma - \alpha}$ negative, which conflicts with the condition $\delta > 0$. In this case, the first term determines the size of the MSE.

The asymptotic normality of the estimator is stated in the following theorem:

Theorem 2.4. *Suppose* $(A1)$ *holds with* $0 < \alpha < 2$, *and* $\hat{\alpha}$ *is defined in* (2.10) *with* $t_j = j/\sqrt{n}$ *for* $j = 1, 2, ..., m = n^{\delta}$ *with* $0 < \delta < 1/2$. *Then as* $n \to \infty$,

$$n^{\frac{1}{2} - \frac{\alpha}{4} + \frac{\delta\alpha}{2}} (\hat{\alpha} - E(\hat{\alpha})) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \tag{2.14}$$

where $\sigma^2 < \infty$ *is a real constant (given in the proof).*

From this theorem and Theorem 2.2 we deduce that, as $n \to \infty$,

$$n^{\frac{1}{2} - \frac{\alpha}{4} + \frac{\delta\alpha}{2}} (\hat{\alpha} - \alpha) = \mathcal{N}(0, \sigma^2) + \frac{D_1(\gamma - \alpha)}{Cs(\alpha)(\gamma - \alpha + 1)^2} n^{\delta\gamma - \frac{\gamma}{2} - \frac{\delta\alpha}{2} + \frac{\alpha}{4} + \frac{1}{2}} \{1 + o(1)\}.$$

If the second term on the right-hand side of the above equation is finite, i.e., $\delta\gamma - \frac{\gamma}{2} - \frac{\delta\alpha}{2} + \frac{\alpha}{4} + \frac{1}{2} \leq 0$ (as analysed in the MSE part), the distributional information about $\hat{\alpha} - \alpha$ is derived:

At the point $\delta = \frac{1}{2} - \frac{1}{2\gamma - \alpha}$ if it exists, as $n \to \infty$,

$$n^{\frac{1}{2} - \frac{\alpha}{4} + \frac{\delta\alpha}{2}} (\hat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, \sigma^2) + \frac{D_1(\gamma - \alpha)}{Cs(\alpha)(\gamma - \alpha + 1)^2}.$$

Otherwise, in the event that the condition $0 < \delta < \frac{1}{2} - \frac{1}{2\gamma - \alpha}$ holds, as $n \to \infty$,

$$n^{\frac{1}{2} - \frac{\alpha}{4} + \frac{\delta\alpha}{2}} (\hat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Remark 2.5. *As discussed in the MSE part, the condition of the asymptotic normality of* $n^{\frac{1}{2} - \frac{\alpha}{4} + \frac{\delta\alpha}{2}} (\hat{\alpha} - \alpha)$ *always holds if* $\gamma = 2$. *However, if* $\gamma = \alpha + \beta$, *whether the condition holds depends on both the values of* α *and* β. *Therefore, the general conclusion about the asymptotic normality of* $n^{\frac{1}{2} - \frac{\alpha}{4} + \frac{\delta\alpha}{2}} (\hat{\alpha} - E(\hat{\alpha}))$ *is provided in Theorem 2.4.*

2.3.2 Implementation of the regression estimation method

During the application of our estimation strategy, we encounter five main problems: the choice of δ; the selection of t_j, the estimation of the standard error of $\hat{\alpha}$, the case $\alpha = 2$, and the case $\alpha > 2$. These five problems are analysed in turn.

The choice of δ

In the sense of the MSE, the "optimal" rate of convergence equates the order of the squared bias with the order of asymptotic variance. Suppose $(A1)$ holds with $0 < \alpha < 2$ and as $n \to \infty$, $bias^2\{\hat{\alpha}\} \approx n^{2(\delta-1/2)(\gamma-\alpha)}$ and $Var\{\hat{\alpha}\} \approx n^{\frac{\alpha}{2}-1-\delta\alpha}$, which implies $\delta = \frac{1}{2} - \frac{1}{2\gamma-\alpha}$. Hence, the "optimal" rate of convergence can be achieved with t_j of the order of magnitude of $n^{-1/(2\gamma-\alpha)}$.

As mentioned in the analysis of the MSE and of asymptotic normality, the existence of $\delta = \frac{1}{2} - \frac{1}{2\gamma-\alpha}$ needs to be carefully considered. Again, for $\gamma \neq 2$ with a combination of small α and β, the "optimal" rate of convergence does not exist and the condition of asymptotic normality for $\hat{\alpha}$ does not hold either. For the three groups of distributions in Remark 2.3, a rough idea about the value of $\frac{1}{2} - \frac{1}{2\gamma-\alpha}$ can be obtained. For $0 < \alpha < 1$, $\alpha = \beta$ and then $0 < 2\gamma - \alpha = 3\alpha < 3$, finally this value should be less than $1/6$. Hence, if $\alpha < 2/3$, the "optimal" rate does not exist. For $1 \leq \alpha < 2$, $\alpha \leq \beta$ then $\gamma = 2$ and $2 < 2\gamma - \alpha = 4 - \alpha \leq 3$. In this case, $0 < \frac{1}{2} - \frac{1}{2\gamma-\alpha} \leq \frac{1}{6}$ and the "optimal" rate exists. Generally, $\frac{1}{2} - \frac{1}{2\gamma-\alpha}$ is not larger than $1/6$ for these three groups of distributions.

The selection of t_j

From a practical point of view, if we always set $t_j = j/\sqrt{n}$, too few points are included in the regression for small sample sizes. Therefore, in order to have more points in the regression, we introduce the following rule to select

point t_j:

$$t_j = \begin{cases} \frac{j}{n},\ j = 1, 2, ..., m = n^{\delta},\ 0 < \delta < 1, & \text{if } n < N, \\ \frac{j}{\sqrt{n}},\ j = 1, 2, ..., m = n^{\delta},\ 0 < \delta < 1/2, & \text{otherwise.} \end{cases} \tag{2.15}$$

The threshold level N is an integer chosen by the researcher. This practical rule is applied and compared with the theoretical rule in the simulation study. From the simulation results, it seems reasonable and sensible to have this specific small sample rule in practice.

The estimation of the standard error of $\hat{\alpha}$

Theorem 2.3 gives the variance of $\hat{\alpha}$ for large samples. A natural, simple procedure is to use the heteroscedasticity-and-autocorrelation-consistent (HAC) standard error, or Newey-West standard error, denoted by s_{HAC}, to replace the standard error of $\hat{\alpha}$ (for further detailed computations, see formula (5) and theorem 2 in Newey & West (1987)). The reason is that if the least squares intercept $\hat{C_\alpha}$ converges in probability to the population intercept, the residuals e_i from the usual ordinary least squares (OLS) arithmetic become asymptotically equivalent to the error terms ϵ_i and s^2_{HAC} consistently estimates the variance of $\hat{\alpha}$.

Therefore, the next step is to check whether the least squares intercept $\hat{C_\alpha}$ converges in probability to the population intercept. Since $\hat{C_\alpha} = \bar{y} - \hat{\alpha}\bar{Z}$, this will occur if $(\hat{\alpha} - \alpha)\bar{Z} = o_p(1)$. As argued above, as $n \to \infty$,

$$\bar{Z} = \ln m - 1 + (\ln m)/(2m) + O(1/m) - \ln \sqrt{n},$$

so $\lim_{n\to\infty} \bar{Z}/\ln(m/\sqrt{n}) = 1$. The asymptotic standard error of $\hat{\alpha}$ is proportional to $n^{\frac{\alpha}{4}-\frac{1}{2}-\frac{\delta\alpha}{2}}$ and

$$\lim_{n\to\infty} n^{\frac{\alpha}{4}-\frac{1}{2}-\frac{\delta\alpha}{2}}(\ln \frac{m}{\sqrt{n}}) = 0.$$

Hence $\hat{C_\alpha}$ is consistent for the population intercept for all $0 < \alpha < 2$ and $\delta > 0$. Therefore, the condition of using s_{HAC} as a consistent estimator for

the standard error of $\hat{\alpha}$ is guaranteed.

The case $\alpha = 2$

As stated in (2.4), $1 - U(t) \sim t^2 \int_0^{1/t} xH(x)\,\mathrm{d}x$ for $\alpha = 2$, as $t \downarrow 0$. Since $H(x) \in RV_2$, it is obvious that $xH(x) \in RV_1$. Then either the integral is $\int_0^{1/t} xH(x)\,\mathrm{d}x < \infty$ or $\int_0^{1/t} xH(x)\,\mathrm{d}x = \infty$. Moreover, by Karamata's theorem (Resnick, 2006), if $G \in RV_1$, then $\int_0^{\lambda} G(x)\,\mathrm{d}x \in RV_0$. In the case of $\alpha = 2$,

$$\int_0^{\lambda} G(x)\,\mathrm{d}x = \int_0^{1/t} xH(x)\,\mathrm{d}x \in RV_0.$$

This integral is denoted as $L_1(1/t)$. Thus, one can estimate α if

$$\lim_{t \downarrow 0} L_1(1/t) < \infty. \tag{2.16}$$

For instance, condition (2.16) is satisfied for any function with a finite non-zero limit.

By taking the logarithm and plugging $U_n(t)$ into (2.4), we obtain

$$\ln\left(1 - U_n(t)\right) \sim \ln L_1(1/t) + 2\ln t + \ln \frac{1 - U_n(t)}{1 - U(t)}, \quad as \quad t \downarrow 0. \tag{2.17}$$

As $t \downarrow 0$, the expected value of $\frac{U(t_j) - U_n(t_j)}{1 - U(t_j)}$ is still zero and the variance of it is equal to $O\{(nt_j^2)^{-1}\}$. The statistical properties of the error term $\epsilon' = \ln \frac{1 - U_n(t)}{1 - U(t)}$ are obtained as well: as $t \downarrow 0$,

$$E(\epsilon'_j) = 0,\ Var(\epsilon'_j) = O\{(nt_j^2)^{-1}\},\ \text{and}\ Cov(\epsilon'_i, \epsilon'_j) = O\{(nt_i^2)^{-1}\}$$

with $t_i > t_j$, using the same criterion to choose t_j, i.e.,

$$\frac{j}{\sqrt{n}},\ j = 1, 2, ..., m = n^{\delta},\ 0 < \delta < 1/2.$$

The expected value of the estimated slope in (2.17) depends on $L_1(1/t)$. However, the estimated slope approaches the true value as $n \to \infty$ because $L_1(1/t)$ converges to a finite constant as $n \to \infty$. The variance of the

estimator is equal to $O(n^{-2\delta})$. Consistency and asymptotic normality still hold. The Newey-West standard error s_{HAC} can be used as a consistent estimator of the standard error of the estimated α.

The case $\alpha > 2$

All the results for $\alpha = 2$ can be extended to the case $\alpha > 2$, i.e., a distribution with finite variance. The only change needed is to replace $L_1(1/t)$ with μ_2, the second moment of the underlying distribution. No further assumption is required because μ_2 is finite.

It is worth noting that, even for the case $0 < \alpha < 2$, assumption $(A1)$ could be relaxed. The estimator $\hat{\alpha}$ is consistent as long as the corresponding slowly varying function $L(x)$ has a finite non-zero limit as $x \to \infty$. Accordingly the term C in Section 2.3.1 has to be replaced with the finite limit.

2.4 Testing the estimator on simulated data

In this section, we evaluate the performance of the proposed estimator on simulated data. Four groups of data are considered. The first group consists of symmetric α-stable distributions which satisfy $(A1)$. The second group considers several distributions which do not follow $(A1)$ as extensions: Student's t-distributions, Pareto distributions, normal distributions, and exponential distributions. The third group is about a distribution with a non-constant slowly varying tail. The last group considers a composite distribution whose central part fits a normal distribution but tails are of Pareto form (DuMouchel, 1983).

Although many tail index estimators exist in the literature, the proposed estimator is compared with the Hill estimator in this chapter. The Hill estimator is chosen because it is usually regarded as a benchmark due to its small asymptotic variance. Moreover, the Hill estimator is a representative of estimators based on the specified form of the distribution function at infinity. Hence, it can be used as a reasonable comparison with the estimator

proposed here based on the CF near the origin. Finally, a complete and general comparison of estimators of the tail index is hard to make because of different second-order conditions (de Haan & Ferreira, 2006).

The Hill estimator is defined as follows. Let $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$ denote the order statistics of the sample $X_1, X_2, \cdots X_n$, and let k_n be a sequence of positive integers satisfying $1 \leq k_n < n$, $\lim_{n\to\infty} k_n = \infty$, and $\lim_{n\to\infty} (k_n/n) = 0$. The Hill estimator based on k_n upper order statistics is

$$\hat{\alpha}_{k_n} = \left(\frac{1}{k_n} \sum_{i=1}^{k_n} \ln \frac{X_{(i)}}{X_{(k_n+1)}} \right)^{-1}.$$

In order to improve the performance of the Hill estimator, the absolute values of samples are used in the estimation.

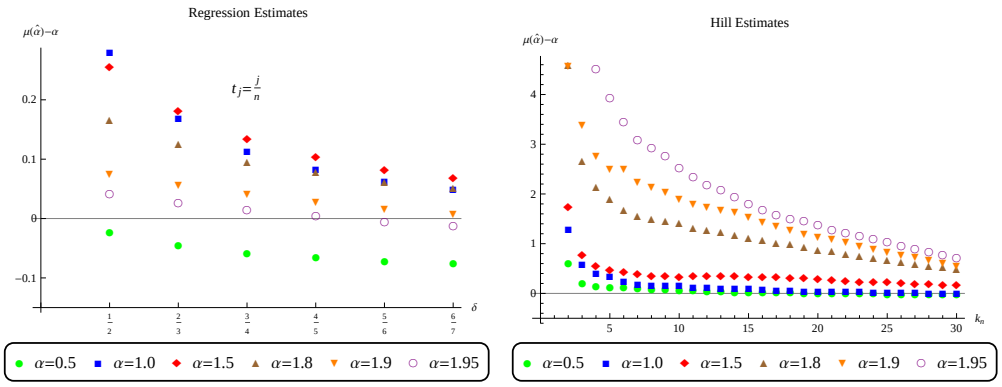
Simulation results focus on data generated from symmetric α-stable distributions. We ran simulations with sample size $n = 100$, $n = 1000$, and $n = 5000$ for symmetric α-stable distributions. For the other three groups only sample size $n = 1000$ is analysed. The results reported here are obtained from 250 iterations.

Let $N = 1000$ in (2.15) be the threshold for the selection of t_j. For $n = 100$, the small sample version, the rule $t_j = j/n$ is applied. For $n = 5000$, the large sample version, the rule $t_j = j/\sqrt{n}$ is used. For $n = 1000$, in order to make a comparison, we show the results of both $t_j = j/n$ and $t_j = j/\sqrt{n}$. Let δ be equal to $1/2, 2/3, 3/4, 4/5, 5/6, 6/7$ for all simulations where $t_j = j/n$. Let δ be equal to $1/6, 1/4, 1/3$ for all simulations where $t_j = j/\sqrt{n}$.

2.4.1 Symmetric α-stable distributions

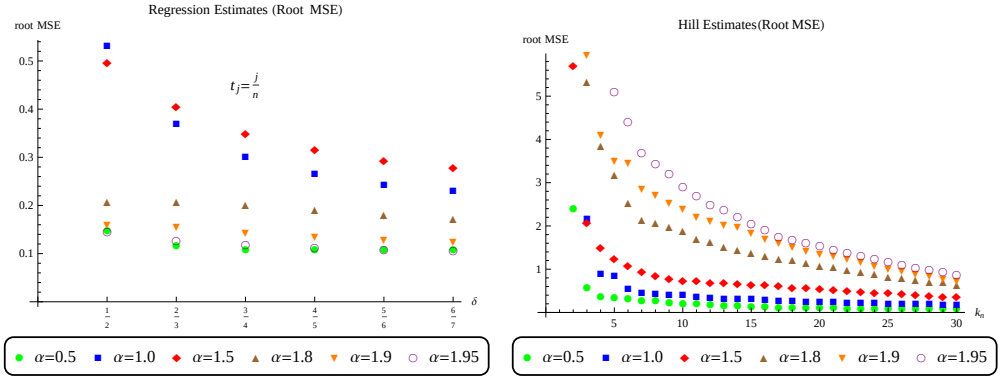
For symmetric α-stable distributions, we set $\alpha = 0.5, 1.0, 1.5, 1.8, 1.9, 1.95$. The simulation results are presented in Figures 2.1 to 2.6. For each sample size, the difference between the mean of regression estimates and α and the root MSE are plotted against δ on the left-hand side. Meanwhile, the difference between the mean of Hill estimates and α and the root MSE are plotted against the number of upper order statistics k_n for the construction of the Hill estimate on the right-hand side.

Figure 2.1: Estimates for independent symmetric α-stable distributions, $n = 100$



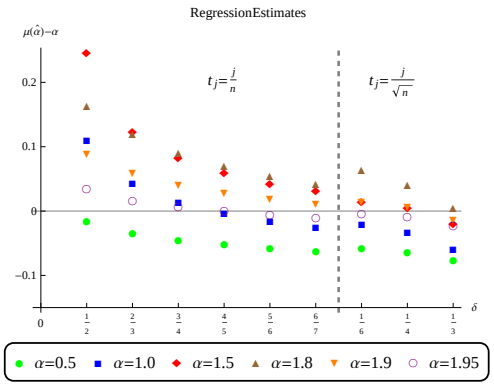
(a) RE: $\mu(\hat{\alpha}) - \alpha$ (b) Hill: $\mu(\hat{\alpha}) - \alpha$

Figure 2.2: Root MSE for independent symmetric α-stable distributions, $n = 100$

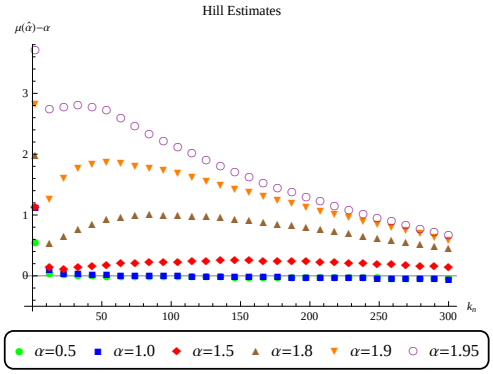


(a) RE: root MSE (b) Hill: root MSE

Figure 2.3: Estimates for independent symmetric α-stable distributions, $n = 1000$

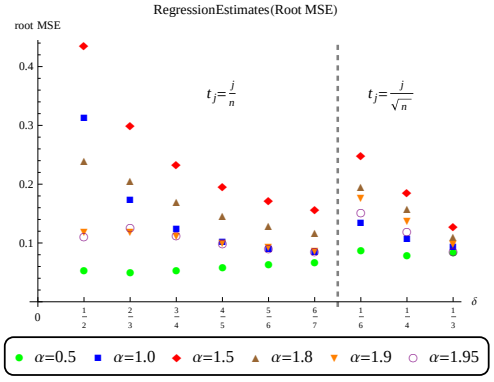


(a) RE: $\mu(\hat{\alpha}) - \alpha$

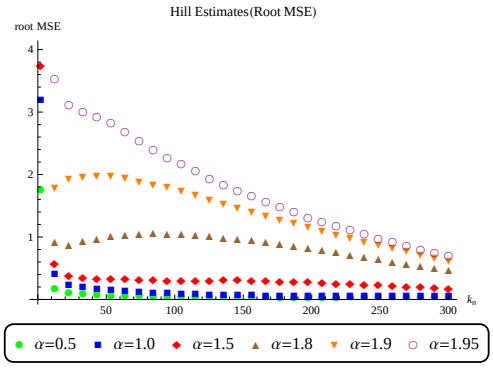


(b) Hill: $\mu(\hat{\alpha}) - \alpha$

Figure 2.4: Root MSE for independent symmetric α-stable distributions, $n = 1000$

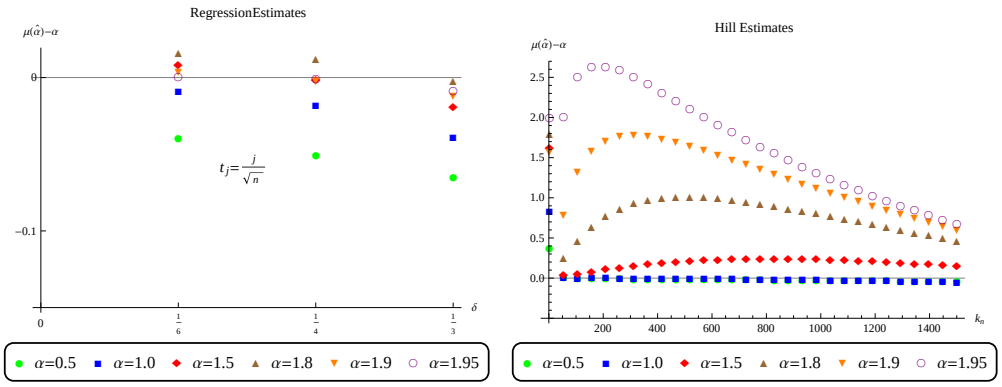


(a) RE: root MSE



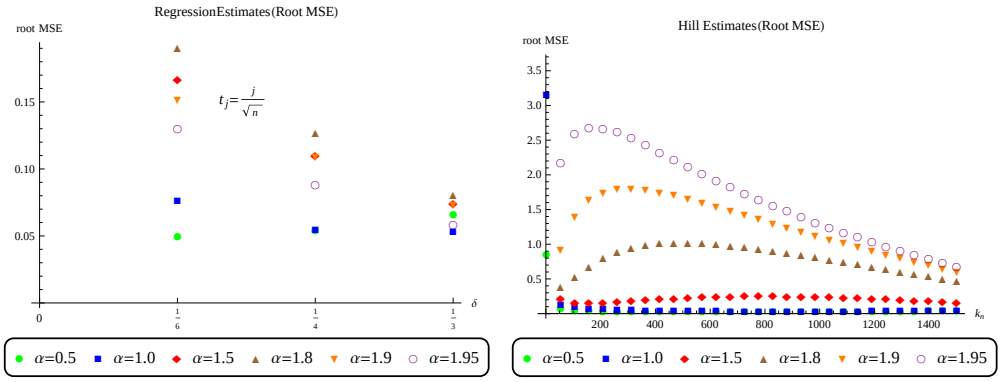
(b) Hill: root MSE

Figure 2.5: Estimates for independent symmetric α-stable distributions, $n = 5000$



(a) RE: $\mu(\hat{\alpha}) - \alpha$ (b) Hill: $\mu(\hat{\alpha}) - \alpha$

Figure 2.6: Root MSE for independent symmetric α-stable distributions, $n = 5000$



(a) RE: root MSE (b) Hill: root MSE

For $\alpha \leq 1$, the performance of the proposed estimator $\hat{\alpha}$ is generally similar to that of the Hill estimator. However, the accuracy of the Hill estimator relies heavily on the number of high order statistics included in the estimation. In order to obtain the optimal number of high order statistics, some adaptive procedures which require more computing effort are needed (Beirlant et al., 2004). While as shown by the small value of the root MSE, the performance of the regression estimator is quite reliable if relatively large δ is chosen.

For $1 < \alpha < 2$, the regression estimator performs much better than the Hill estimator does, especially when α is very close to 2. The Hill estimator is highly inaccurate no matter how many order statistics are included in the estimation when α approaches 2. By contrast, the root MSE of the regression estimator is always quite small regardless of the value of δ.

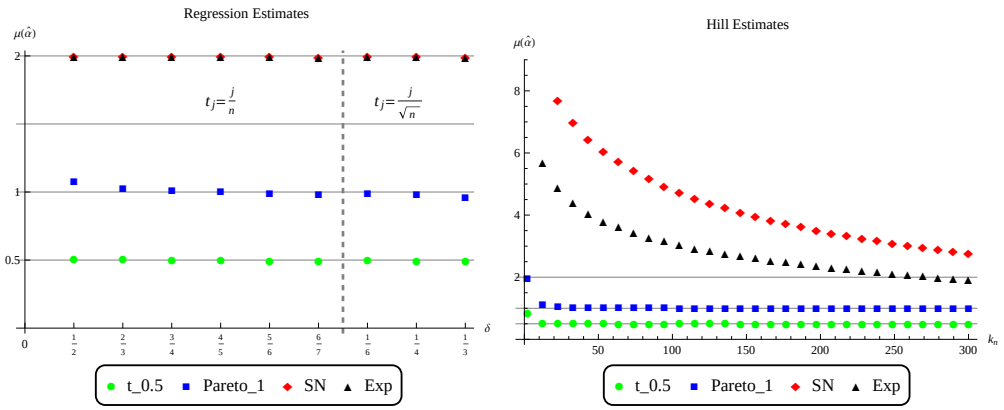
We also note that the selection methods $t_j = j/n$ and $t_j = j/\sqrt{n}$ work well for $n = 100$ and $n = 5000$, respectively. For $n = 1000$, the accuracy of the regression estimator changes slightly as the selection method changes. Both selection methods show the same trend, namely that the regression estimator approaches the true value as δ increases. Therefore, it is reasonable to use $t_j = j/n$ instead of $t_j = j/\sqrt{n}$ when sample size is small. Moreover, the convergence of both the regression and Hill estimators is investigated. For both methods, estimators converge to the true value and root MSEs reduce as sample size increases.

2.4.2 Several alternative distributions

In the second group, the following distributions are analysed: Student's t-distributions ($\alpha = 0.5$), Pareto distributions ($\alpha = 1$ and $x_{min} = 1$), standard normal distributions, and exponential distributions ($\lambda = 2$). With respect to the last two distributions, we expect our regression estimator to be around 2 and the Hill estimator should be quite large due to the absence of heavy tails. Therefore, we only report the mean values of regression estimates (Figure 2.7a) and of Hill estimates (Figure 2.7b) here.

From Figure 2.7 we see that, for Student's t-distributions and Pareto distributions, the regression estimates and Hill estimates are very close to the true value of α. For distributions with finite variance, the regression method always returns a value almost equal to 2 as expected. By contrast, the Hill estimator decreases as the sample fraction increases for finite variance distributions. Indeed, the Hill estimator changes substantially from more than 8 to slightly above 2. Therefore it is quite difficult to find the appropriate percentage included in the estimation, let alone to obtain inference on whether heavy tails exist or not.

Figure 2.7: Estimates for several alternative distributions, $n = 1000$



(a) RE: $\mu(\hat{\alpha})$ (b) Hill: $\mu(\hat{\alpha})$

2.4.3 A distribution with a non-constant slowly varying tail

We now turn to the third comparison, where $L(x)$ is not a constant. We consider a distribution whose survival function is defined as

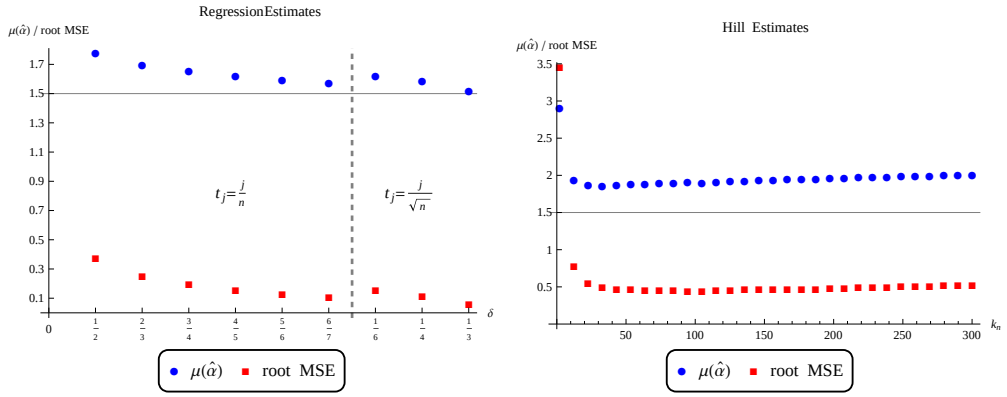
$$\overline{F}(x) = 1 - F(x) = \frac{e^{\frac{3}{2}}}{x^{\frac{3}{2}} \ln x}, \quad x \geq e. \tag{2.18}$$

Since the Hill estimator is designed mainly for Pareto distributions, we introduce the above example in which the Hill estimator may not know how to make correct inference apart from the slowly varying factor $\ln x$ (Resnick, 2006). Due to the non-constant slowly varying function in (2.18), the Hill estimator behaves poorly compared with the proposed estimator in this case. The mean and root MSE of the regression estimates against δ are presented in Figure 2.8a, while the mean and root MSE of the Hill estimates against k_n are presented in Figure 2.8b.

Figure 2.8b shows that the Hill estimator is highly sensitive to the non-constant slowly varying function. The mean of Hill estimates is far away from the true value. The root MSE of Hill estimates is around 0.5. In Figure 2.8a, the effect of the non-constant slowly varying function on the regression

estimator is not as dramatic as that on the Hill estimator. The mean of the regression estimates gets closer to the true value as δ increases and the root MSE is generally less than 0.2. It seems that the regression estimator retains its accuracy regardless of the presence of some kind of non-constant slowly varying function.

Figure 2.8: Estimates for distributions with non-constant slowly varying tails, $n = 1000$



(a) RE (b) Hill

2.4.4 Distributions with composite density functions

Finally, the proposed estimation method is extended to distributions with composite density functions. In general, this kind of distribution has a distribution function with finite variance in the center and follows a regularly varying function in the tails. In this example, we use a density function which has Pareto tails but matches a normal density function in the center (DuMouchel, 1983). With $0 < 1/\gamma < 2$, the corresponding density function denoted by $f_\gamma(x)$ is defined as

$$f_\gamma(x) = \begin{cases} c\frac{1}{\sqrt{2\pi}}e^{-x^2/2}, & \text{if } |x| < 1, \\ c\frac{\Phi(-1)}{\sigma^{-\gamma}}\left[\frac{1+\gamma(|x|-1)}{\sigma^{-\gamma}}\right]^{-\gamma^{-1}-1}, & \text{if } |x| \geq 1, \end{cases} \tag{2.19}$$

where $\sigma = \Phi(-1)\sqrt{2\pi e}$ to make the density function continuous at points $|x| = 1$, and $c = 1/(\Phi(1) - \Phi(-1) + 2\int_1^\infty \frac{\Phi(-1)}{\sigma^{-\gamma}}[\frac{1+\gamma(|x|-1)}{\sigma^{-\gamma}}]^{-\gamma^{-1}-1}\,\mathrm{d}x)$. The tail index of this distribution is equal to $1/\gamma$.

Using (2.3) and due to the symmetry of this density function, the following relation is derived

$$I - U(t) \sim t\int_0^1 \sin(tx)H_1(x)\mathrm{d}x + t\int_1^\infty \sin(tx)H_2(x)\mathrm{d}x, \quad as \quad t \downarrow 0,$$

where $H_1(x)$ is the tail sum with respect to the normal distribution part and $H_2(x) \in RV_{1/\gamma}$. By a Taylor expansion, the first term on the right-hand side can be rewritten as

$$t\int_0^1 \sin(tx)H_1(x)\mathrm{d}x = t^2\int_0^1 xH_1(x)\mathrm{d}x + O(t^4), \quad as \quad t \downarrow 0,$$

where the coefficient of t^2 is bounded in the range from 0 to 1. Moreover, as $t \downarrow 0$,

$$\begin{aligned} t\int_1^\infty \sin(tx)H_2(x)\mathrm{d}x &= \int_t^\infty \sin(x)H_2(x/t)\mathrm{d}x \\ &\sim \int_t^\infty (x/t)^{-1/\gamma}\sin(x)\mathrm{d}x \\ &= s(1/\gamma)t^{1/\gamma}. \end{aligned}$$

Hence, the final result is

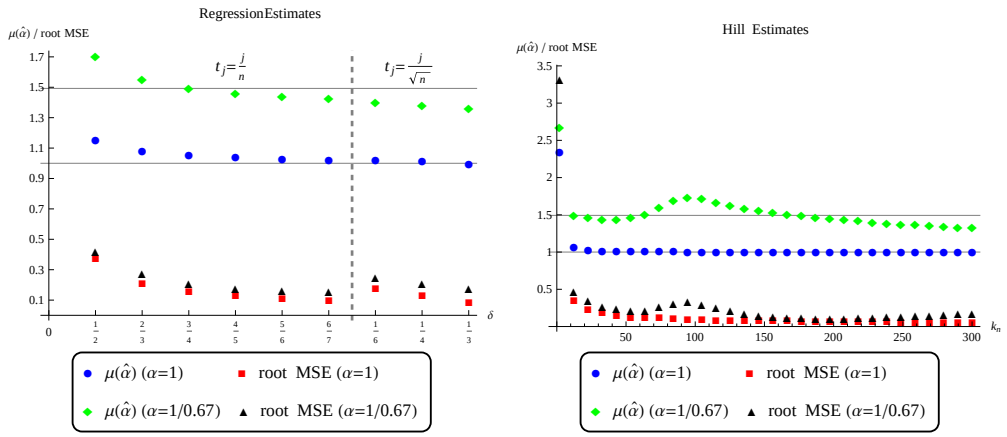
$$1 - U(t) \sim s(1/\gamma)t^{1/\gamma} + t^2\int_0^1 xH_1(x)\mathrm{d}x + O(t^4), \quad as \quad t \downarrow 0.$$

The term with $t^{1/\gamma}$ dominates the limiting behaviour of $1 - U(t)$ for $1/\gamma < 2$ as $t \downarrow 0$. Hence, the regression estimation method can be applied to this kind of composite distribution.

To evaluate the performance of the regression method, we generated data with the density function defined in (2.19) with $\gamma = 0.67$ and 1. In Figure 2.9a, the mean and root MSE of regression estimates against δ are presented. The mean and root MSE of Hill estimates against k_n are shown in Figure

2.9b.

Figure 2.9: Estimates for distributions with composite density functions, $n = 1000$



(a) RE (b) Hill

In Figure 2.9, for $\alpha = 1$, the Hill estimator generally performs better. However, the accuracy of the regression estimator is comparable to that of the Hill estimator, since the regression estimates are not far away from the true value and root MSEs are less than 0.3 for large δ by $t_j = j/n$ and all δ's by $t_j = j/\sqrt{n}$. For $\alpha = 1/0.67 \approx 1.49$, the regression estimator works slightly better from the point of view of the mean and root MSE.

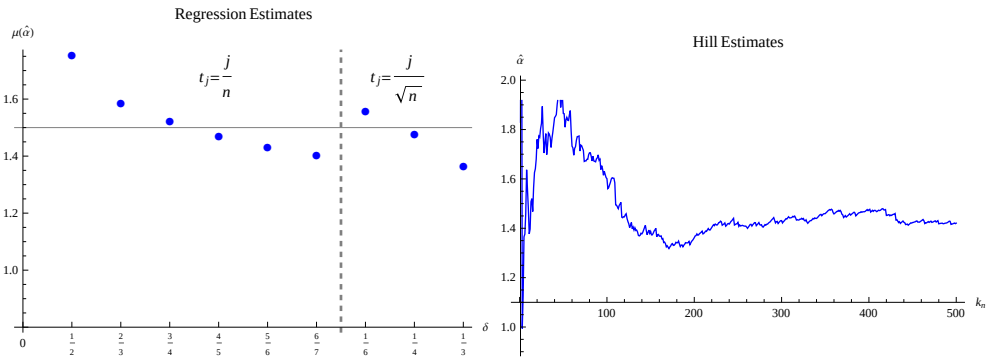
2.5 An empirical application

In this section, the proposed estimation method is applied to an empirical data set: Danish fair insurance claims. The data set consists of 2167 claims in millions of Danish Krone (1985 prices) from the years 1980 to 1990 inclusive. The claims comprise damage to furniture and personal property, damage to buildings, and loss of profits. The data are analysed in McNeil (1996)[1]. The Hill plot in Figure 2.10b indicates that the Hill estimator is relatively stable around 1.4 for k_n larger than 200. For the regression method, we present the

[1]Data source: http://www.macs.hw.ac.uk/ mcneil/data.html

results in Figure 2.10a. The regression estimator suggests that the tail index is around 1.5, which is nearly consistent with the result obtained from the Hill method.

Figure 2.10: Estimates for Danish fair insurance claims



(a) RE (b) Hill

2.6 Conclusion

In this chapter, we present an approach to estimate the tail index α for independent random variables. This approach exploits the reflection of the distribution function at infinity on the CF near the origin. The method performs an OLS regression near the origin which is based on the cosine part of the ECF. Therefore, the entire sample is used to calculate every point in the regression. Consistency and asymptotic normality are obtained under some conditions. Simulations and comparisons illustrate the great potential of the proposed method. In order to apply this method in practice, different selection methods of regression points with respect to t_j are proposed for small and large samples, respectively. Our simulation results show that the practical methods work quite well. One main advantage of this regression method is that the accuracy of the estimator is not very sensitive to the choice of δ, and consistency holds regardless of the value of δ. Another advantage is that it works well even when α is close to 2.

2.7 Proofs

Some preliminary lemmas are needed for the proofs.

Lemma 2.2. *Assume that* $0 < \beta < 1$*, then as* $m \to \infty$ *(Chan et al., 1995),*

$$\frac{1}{m}\sum_{j=1}^{m}(\ln j)^2 - \Big(\frac{1}{m}\sum_{j=1}^{m}\ln j\Big)^2 = 1 + O\{m^{-1}(\ln m)^2\}.$$

$$\frac{1}{m}\sum_{j=1}^{m} j^{-\beta}\ln j - \Big(\frac{1}{m}\sum_{j=1}^{m} j^{-\beta}\Big)\Big(\frac{1}{m}\sum_{j=1}^{m}\ln j\Big) = \frac{-\beta}{(1-\beta)^2}m^{-\beta} + O(m^{-1}\ln m).$$

Remark 2.6. *For the first equation, if* $\beta = 1$*, the result is* $-(\ln m)^2/2m + O(\ln m/m)$*; if* $1 < \beta < 2$*, the result is* $O(\ln m/m)$*.*

Lemma 2.3. *For* $0 < \beta < 2$*, then as* $m \to \infty$*,*

$$\frac{1}{m}\sum_{j=1}^{m}\ln j = \ln m - 1 + \frac{\ln m}{2m} + O(1/m). \tag{2.20}$$

$$\frac{1}{m}\sum_{j=1}^{m} j^{\beta}\ln j - \Big(\frac{1}{m}\sum_{j=1}^{m}\ln j\Big)\Big(\frac{1}{m}\sum_{j=1}^{m} j^{\beta}\Big) = \frac{\beta}{(\beta+1)^2}m^{\beta} + O(m^{\beta-1}\ln m). \tag{2.21}$$

For $0 < \beta < 1$*, as* $m \to \infty$*,*

$$\sum_{j=1}^{m} j^{-\beta} = \frac{m^{-\beta+1}}{-\beta+1} + O(1). \tag{2.22}$$

Proof. Lemma 2.3 follows from the Euler-Maclaurin formula for asymptotic expansions of infinite series (Abramowitz & Stegun, 1965). The formula is

displayed as follows.

$$\sum_{n=a}^{b} f(n) \sim \int_a^b f(x)\,\mathrm{d}x + \frac{f(a)+f(b)}{2} + \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k)!}\left(f^{(2k-1)}(b) - f^{(2k-1)}(a)\right), \tag{2.23}$$

where $B_1 = -1/2$, $B_2 = 1/6$, $B_3 = 0$, $B_4 = -1/50$, $B_6 = 1/42$, $B_7 = 0$, $B_8 = -1/30$, ..., are the Bernoulli numbers.

For (2.20), as $m \to \infty$, according to (2.23),

$$\frac{1}{m}\sum_{j=1}^{m} \ln j \sim \frac{1}{m}\Big\{ \int_1^m \ln x\,\mathrm{d}x + \frac{\ln m}{2} + \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k)!}\left(f_1^{(2k-1)}(m) - f_1^{(2k-1)}(1)\right)\Big\}. \tag{2.24}$$

The first term on the right-hand side of (2.24) is

$$\begin{aligned} \int_1^m \ln x\,\mathrm{d}x &= x\ln x|_1^m - \int_1^m x\,\mathrm{d}\ln x \\ &= m\ln m - \int_1^m 1\,\mathrm{d}x \\ &= m\ln m - m + 1. \end{aligned} \tag{2.25}$$

To get the last term on the right-hand side (2.24), first we calculate that

$$\begin{aligned} f_1(x) &= \ln x \\ f_1^{(1)}(x) &= x^{-1} \\ f_1^{(2)}(x) &= -x^{-2} \\ f_1^{(3)}(x) &= 2x^{-3} \\ &\vdots \\ f_1^{(2k-1)}(x) &= (2k-2)!x^{-(2k-1)}. \end{aligned}$$

Then calculate the difference between derivatives with the same order at

different points:

$$\begin{aligned} f_1^{(1)}(m) - f_1^{(1)}(1) &= m^{-1} - 1 \\ f_1^{(3)}(m) - f_1^{(3)}(1) &= 2(m^{-3} - 1) \\ &\vdots \\ f_1^{(2k-1)}(m) - f_1^{(2k-1)}(1) &= (2k-2)!(m^{-(2k-1)} - 1). \end{aligned} \tag{2.26}$$

Finally, we plug in (2.25) and (2.26) into (2.24) and obtain

$$\begin{aligned} \frac{1}{m}\sum_{j=1}^{m} \ln j &\sim \frac{1}{m}\Bigg\{ m\ln m - m + 1 + \frac{\ln m}{2} \\ &\qquad + \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k)(2k-1)}(m^{-(2k-1)} - 1)\Bigg\} \\ &= \ln m - 1 + \frac{\ln m}{2m} + O(1/m). \end{aligned} \tag{2.27}$$

With regard to (2.21), as $m \to \infty$, the first term on the left-hand side is

$$\begin{aligned} \frac{1}{m}\sum_{j=1}^{m} j^{\beta}\ln j &\sim \frac{1}{m}\Bigg\{ \int_1^m x^{\beta}\ln x\,\mathrm{d}x + m^{\beta}\ln m/2 \\ &\qquad + \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k)!}\left(f_2^{(2k-1)}(m) - f_2^{(2k-1)}(1)\right)\Bigg\}. \end{aligned} \tag{2.28}$$

With similar procedures, we get

$$\begin{aligned} \int_1^m x^{\beta}\ln x\,\mathrm{d}x &= \frac{1}{\beta+1}\int_1^m \ln x\,\mathrm{d}x^{\beta+1} \\ &= \frac{1}{\beta+1}\left\{ x^{\beta+1}\ln x\Big|_1^m - \int_1^m x^{\beta+1}\,\mathrm{d}\ln x\right\} \\ &= \frac{1}{\beta+1}\left\{ m^{\beta+1}\ln m - \frac{x^{\beta+1}}{\beta+1}\Big|_1^m\right\} \\ &= \frac{1}{\beta+1}\left\{ m^{\beta+1}\ln m - \frac{m^{\beta+1}}{\beta+1} + \frac{1}{\beta+1}\right\}. \end{aligned} \tag{2.29}$$

$$
\begin{aligned}
f_2(x) &= x^\beta \ln x \\
f_2^{(1)}(x) &= \beta x^{\beta-1} \ln x + x^{\beta-1} \\
f_2^{(2)}(x) &= \beta(\beta-1)x^{\beta-2} \ln x + \beta x^{\beta-2} + (\beta-1)x^{\beta-2} \\
&= \beta(\beta-1)x^{\beta-2} \ln x + (2\beta-1)x^{\beta-2} \\
f_2^{(3)}(x) &= \beta(\beta-1)(\beta-2)x^{\beta-3} \ln x + \beta(\beta-1)x^{\beta-3} + (2\beta-1)(\beta-2)x^{\beta-3} \\
&= \beta(\beta-1)(\beta-2)x^{\beta-3} \ln x + (3\beta^2-6\beta+2)x^{\beta-3} \\
&\vdots
\end{aligned}
$$

$$
\begin{aligned}
f_2^{(1)}(m) - f_2^{(1)}(1) &= \beta m^{\beta-1} \ln m + m^{\beta-1} - 1 \\
f_2^{(3)}(m) - f_2^{(3)}(1) &= \beta(\beta-1)(\beta-2)m^{\beta-3} \ln m + (3\beta^2-6\beta+2)(m^{\beta-3}-1) \\
&\vdots
\end{aligned} \tag{2.30}
$$

With (2.29) and (2.30), the result of (2.28) is

$$
\begin{aligned}
\frac{1}{m}\sum_{j=1}^{m} j^\beta \ln j \;\sim\; & \frac{1}{m}\Bigg\{ \frac{1}{\beta+1}\left(m^{\beta+1}\ln m - \frac{m^{\beta+1}}{\beta+1} + \frac{1}{\beta+1}\right) + \frac{m^\beta \ln m}{2} \\
& + \frac{1/6}{2}\left(\beta m^{\beta-1}\ln m + m^{\beta-1} - 1\right) \\
& + \frac{-1/50}{4!}\Big[\beta(\beta-1)(\beta-2)m^{\beta-3}\ln m \\
& + (3\beta^2-6\beta+2)(m^{\beta-3}-1)\Big] + O(1)\Bigg\} \\
\sim\; & \frac{1}{m}\Bigg\{ \frac{1}{\beta+1}\left(m^{\beta+1}\ln m - \frac{m^{\beta+1}}{\beta+1}\right) \\
& + \frac{m^\beta \ln m}{2} + o(m^\beta \ln m)\Bigg\}.
\end{aligned} \tag{2.31}
$$

Similarly, the last term on the left-hand side of (2.21) is

$$\begin{aligned}\frac{1}{m}\sum_{j=1}^{m} j^{\beta} &\sim \frac{1}{m}\Bigg\{\int_{1}^{m} x^{\beta}\,\mathrm{d}x + \frac{m^{\beta}+1}{2} \\ &\quad + \sum_{k=1}^{\infty}\frac{B_{2k}}{(2k)!}\left(f_3^{(2k-1)}(m) - f_3^{(2k-1)}(1)\right)\Bigg\} \\ &= \frac{1}{m}\Bigg\{\frac{m^{\beta+1}-1}{(\beta+1)} + \frac{m^{\beta}+1}{2} \\ &\quad + \sum_{k=1}^{\infty}\frac{B_{2k}}{(2k)!}\left(f_3^{(2k-1)}(m) - f_3^{(2k-1)}(1)\right)\Bigg\}. \end{aligned} \tag{2.32}$$

$$\begin{aligned} f_3(x) &= x^{\beta} \\ f_3^{(1)}(x) &= \beta x^{\beta-1} \\ f_3^{(2)}(x) &= \beta(\beta-1)x^{\beta-2} \\ f_3^{(3)}(x) &= \beta(\beta-1)(\beta-2)x^{\beta-3} \\ &\vdots \end{aligned}$$

$$\begin{aligned} f_3^{(1)}(m) - f_3^{(1)}(1) &= \beta\left(m^{\beta-1}-1\right) \\ f_3^{(3)}(m) - f_3^{(3)}(1) &= \beta(\beta-1)(\beta-2)\left(m^{\beta-3}-1\right) \\ &\vdots \end{aligned} \tag{2.33}$$

Plugging in (2.33) to (2.32),

$$\begin{aligned}\frac{1}{m}\sum_{j=1}^{m} j^{\beta} &\sim \frac{1}{m}\Bigg\{\frac{m^{\beta+1}-1}{(\beta+1)} + \frac{m^{\beta}+1}{2} + \frac{1/6}{2}\beta\left(m^{\beta-1}-1\right) \\ &\quad + \frac{-1/50}{4!}\beta(\beta-1)(\beta-2)\left(m^{\beta-3}-1\right) + O(1)\Bigg\} \\ &= \frac{m^{\beta}}{\beta+1} + \frac{m^{\beta-1}}{2} + O(m^{\beta-2}). \end{aligned} \tag{2.34}$$

Therefore, combining the results in (2.27), (2.31), and (2.34), the following result is obtained by simplification

$$\begin{aligned}
&\left(\frac{1}{m}\sum_{j=1}^{m} j^{\beta}\ln j\right) - \left(\frac{1}{m}\sum_{j=1}^{m}\ln j\right)\left(\frac{1}{m}\sum_{j=1}^{m} j^{\beta}\right) \\
&\quad\sim \left[\frac{1}{\beta+1}\left(m^{\beta}\ln m - \frac{m^{\beta}}{\beta+1}\right) + \frac{m^{\beta-1}\ln m}{2} + o(m^{\beta-1}\ln m)\right] \\
&\qquad - \left[\ln m - 1 + \frac{\ln m}{2m} + O(1/m)\right]\left[\frac{m^{\beta}}{\beta+1} + \frac{m^{\beta-1}}{2} + O(m^{\beta-2})\right] \\
&\quad= \beta(\beta+1)^{-2}m^{\beta} + O(m^{\beta-1}\ln m). \qquad (2.35)
\end{aligned}$$

For (2.22), as $m \to \infty$, according to (2.23),

$$\begin{aligned}
\sum_{j=1}^{m} j^{-\beta} &\sim \int_{1}^{m} x^{-\beta}\,\mathrm{d}x + \frac{m^{-\beta}+1}{2} + \sum_{k=1}^{\infty}\frac{B_{2k}}{(2k)!}\left(f_4^{(2k-1)}(m) - f_4^{(2k-1)}(1)\right) \\
&= \frac{m^{-\beta+1}-1}{-\beta+1} + \frac{m^{-\beta}+1}{2} + \sum_{k=1}^{\infty}\frac{B_{2k}}{(2k)!}\left(f_4^{(2k-1)}(m) - f_4^{(2k-1)}(1)\right).
\end{aligned}$$

$$\begin{aligned}
f_4(x) &= x^{-\beta} \\
f_4^{(1)}(x) &= -\beta x^{-\beta-1} \\
f_4^{(2)}(x) &= -\beta(-\beta-1)x^{-\beta-2} \\
f_4^{(3)}(x) &= -\beta(-\beta-1)(-\beta-2)x^{-\beta-3} \\
&\vdots
\end{aligned}$$

$$\begin{aligned}
f_4^{(1)}(m) - f_4^{(1)}(1) &= -\beta\left(m^{-\beta-1}-1\right) \\
f_4^{(3)}(m) - f_4^{(3)}(1) &= -\beta(-\beta-1)(-\beta-2)\left(m^{-\beta-3}-1\right) \\
&\vdots
\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{j=1}^{m} j^{-\beta} &\sim \frac{m^{-\beta+1}-1}{-\beta+1}+\frac{m^{-\beta}+1}{2}+o(1)\\ &= \frac{m^{-\beta+1}}{-\beta+1}+O(1). \qquad (2.36)\end{aligned}$$

(2.27), (2.35), and (2.36) complete the proof of Lemma 2.3. □

Lemma 2.4. *As $m \to \infty$,*

$$\begin{aligned}&\sum_{j=1}^{m}(\ln j-\frac{1}{m}\sum_{k=1}^{m}\ln k)^2 j^{-\beta}\\ &= \begin{cases}(\frac{2}{(1-\beta)^3}+\frac{1}{1-\beta}-\frac{2}{(1-\beta)^2})m^{1-\beta}\{1+o(1)\}, & \text{if } 0<\beta<1,\\ \frac{1}{3}(\ln m)^3\{1+o(1)\}, & \text{if } \beta=1,\\ O\{(\ln m)^2\}, & \text{if } 1<\beta<2.\end{cases} \qquad (2.37)\\ &\sum_{j=1}^{m}(\ln j-\frac{1}{m}\sum_{k=1}^{m}\ln k)^2 j^{\beta}\\ &= \begin{cases}m+O((\ln m)^2), & \text{if } \beta=0,\\ (\frac{2}{(1+\beta)^3}+\frac{1}{1+\beta}-\frac{2}{(1+\beta)^2})m^{1+\beta}\{1+o(1)\}, & \text{if } 0<\beta<1.\end{cases} \qquad (2.38)\end{aligned}$$

Proof. Lemma 2.4 is proved according to (2.23) as well. The term on the left-hand side of (2.37) can be expanded as

$$\begin{aligned}\sum_{j=1}^{m}(\ln j-\frac{1}{m}\sum_{k=1}^{m}\ln k)^2 j^{-\beta} &= \sum_{j=1}^{m}(\ln j)^2 j^{-\beta}+\frac{1}{m^2}(\sum_{k=1}^{m}\ln k)^2\sum_{j=1}^{m}j^{-\beta}\\ &\quad -\frac{2}{m}\sum_{k=1}^{m}\ln k\sum_{j=1}^{m}j^{-\beta}\ln j\\ &\sim \sum_{j=1}^{m}(\ln j)^2 j^{-\beta}+(\ln m-1+o(1))^2\sum_{j=1}^{m}j^{-\beta}\\ &\quad -2(\ln m-1+o(1))\sum_{j=1}^{m}j^{-\beta}\ln j,\end{aligned}$$

where the second step is derived by plugging in (2.20).[2]

Splitting the cases:

•$0 < \beta < 1$ By (2.23),

$$\begin{aligned}\sum_{j=1}^{m}(\ln j)^2 j^{-\beta} &\sim \int_1^m (\ln x)^2 x^{-\beta}\,\mathrm{d}x + \frac{(\ln m)^2 m^{-\beta}}{2} + o((\ln m)^2 m^{-\beta}) \\ &\sim \frac{1}{1-\beta}\int_1^m (\ln x)^2\,\mathrm{d}x^{1-\beta} + O((\ln m)^2 m^{-\beta}) \\ &\sim \frac{1}{1-\beta}(\ln m)^2 m^{1-\beta} - \frac{2}{(1-\beta)^2}m^{1-\beta}\ln m \\ &\quad + \frac{2}{(1-\beta)^3}m^{1-\beta} + O(m^{1-\beta}\ln m). \end{aligned} \tag{2.39}$$

By (2.22),

$$\begin{aligned}&(\ln m - 1 + o(1))^2 \sum_{j=1}^{m} j^{-\beta} \\ &\quad\sim (\ln m - 1 + o(1))^2 \frac{m^{1-\beta}}{1-\beta} \\ &\quad\sim \frac{1}{1-\beta}m^{1-\beta}(\ln m)^2 - \frac{2}{1-\beta}m^{1-\beta}\ln m + \frac{m^{1-\beta}}{1-\beta} + o(m^{1-\beta}).\end{aligned} \tag{2.40}$$

According to (2.23),

$$\begin{aligned}&-2(\ln m - 1 + o(1))\sum_{j=1}^{m} j^{-\beta}\ln j \\ &\quad\sim -2(\ln m - 1 + o(1))\left(\int_1^m x^{-\beta}\ln x\,\mathrm{d}x + O(m^{-\beta}\ln m)\right) \\ &\quad\sim -2(\ln m - 1 + o(1))\Big(\frac{m^{1-\beta}\ln m}{1-\beta} - \frac{m^{1-\beta}}{(1-\beta)^2} \\ &\qquad + \frac{1}{(1-\beta)^2} + O(m^{-\beta}\ln m)\Big).\end{aligned} \tag{2.41}$$

[2]To simply the proof, only the first two orders of each component in all the equations about Lemma 2.4 are considered.

Combining (2.39), (2.40), and (2.41), we derive

$$\sum_{j=1}^{m}(\ln j - \frac{1}{m}\sum_{k=1}^{m}\ln k)^2 j^{-\beta} \sim (\frac{2}{(1-\beta)^3} + \frac{1}{1-\beta} - \frac{2}{(1-\beta)^2})m^{1-\beta}\{1+o(1)\}. \quad (2.42)$$

•$\beta = 1$ With similar procedures,

$$\begin{aligned}\sum_{j=1}^{m}(\ln j)^2 j^{-1} &\sim \int_1^m (\ln x)^2 x^{-1}\,\mathrm{d}x + O((\ln m)^2 m^{-1}) \\ &\sim \frac{(\ln m)^3}{3} + O((\ln m)^2 m^{-1}). \quad (2.43)\end{aligned}$$

$$\begin{aligned}(\ln m - 1 + o(1))^2 \sum_{j=1}^{m} j^{-1} &\sim (\ln m - 1 + o(1))^2(\ln m + \frac{1}{2} + o(1)) \\ &\sim (\ln m)^3 + O((\ln m)^2). \quad (2.44)\end{aligned}$$

$$\begin{aligned}&-2(\ln m - 1 + o(1))\sum_{j=1}^{m} j^{-1}\ln j \\ &\sim -2(\ln m - 1 + o(1))\left(\int_1^m x^{-1}\ln x\,\mathrm{d}x + O(m^{-1}\ln m)\right) \\ &\sim -(\ln m)^2(\ln m - 1 + o(1))^2. \quad (2.45)\end{aligned}$$

With (2.43), (2.44), and (2.45),

$$\sum_{j=1}^{m}(\ln j - \frac{1}{m}\sum_{k=1}^{m}\ln k)^2 j^{-1} \sim \frac{1}{3}(\ln m)^3\{1+o(1)\}. \quad (2.46)$$

•$1 < \beta < 2$

Similar to (2.39),

$$\begin{aligned}\sum_{j=1}^{m}(\ln j)^2 j^{-\beta} &\sim \int_1^m (\ln x)^2 x^{-\beta}\,\mathrm{d}x + O((\ln m)^2 m^{-\beta}) \\ &\sim \frac{2}{(1-\beta)^3} + O(1). \end{aligned} \tag{2.47}$$

$$\begin{aligned}&(\ln m - 1 + o(1))^2 \sum_{j=1}^{m} j^{-\beta} \\ &\quad\sim (\ln m - 1 + o(1))^2 (\int_1^m m^{-\beta}\,\mathrm{d}x + \frac{1}{2} + O(1)) \\ &\quad\sim (\ln m - 1 + o(1))^2 (\frac{-1}{(1-\beta)} + \frac{1}{2} + O(1)). \end{aligned} \tag{2.48}$$

Analogous to (2.41),

$$\begin{aligned}&-2(\ln m - 1 + o(1)) \sum_{j=1}^{m} j^{-\beta} \ln j \\ &\quad\sim -2(\ln m - 1 + o(1)) \left(\int_1^m x^{-\beta} \ln x\,\mathrm{d}x + O(m^{-1}\ln m) \right) \\ &\quad\sim (\ln m - 1 + o(1))^2 (\frac{1}{(1-\beta)^2} + O(1)). \end{aligned} \tag{2.49}$$

Therefore, by (2.47), (2.48), and (2.49),

$$\sum_{j=1}^{m} (\ln j - \frac{1}{m}\sum_{k=1}^{m} \ln k)^2 j^{-\beta} \sim O((\ln m)^2). \tag{2.50}$$

Thus (2.42), (2.46), and (2.50) complete the proof of (2.37).

With regard to (2.38), splitting the cases:

•$\beta = 0$ By Lemma 2.2,

$$\begin{aligned}\sum_{j=1}^{m} (\ln j - \frac{1}{m}\sum_{k=1}^{m} \ln k)^2 &= \sum_{j=1}^{m} (\ln j)^2 - \frac{1}{m}(\sum_{k=1}^{m} \ln k)^2 \\ &= m + O((\ln m)^2). \end{aligned}$$

•$0 < \beta < 1$ Substituting β for $-\beta$ in the exponent in (2.39), (2.40), (2.41), and (2.42), all the results still hold.

This completes the proof of Lemma 2.4. □

Before proving the theorems, another form of $s(q)$ is provided here. Suppose for some non-negative integer p,

$$s(q) = \begin{cases} \int_0^\infty y^{-q} \sin y \, \mathrm{d}y, & \text{if } 0 < q < 2, \\ \int_0^\infty y^{-q} \{\sin y - \sum_{i=1}^{p} (-1)^{i-1} \frac{y^{2i-1}}{(2i-1)!}\} \, \mathrm{d}y, & \text{if } 2p < q < 2p+2, \ p > 0. \end{cases} \tag{2.51}$$

Proof of Theorem 2.1. The characteristic function $\phi(t)$ can be rewritten as

$$\begin{aligned} \phi(t) &= \int_{-\infty}^{0} e^{itx} \, \mathrm{d}F(x) + \int_{0}^{\infty} e^{itx} \, \mathrm{d}[F(x) - 1] \\ &= \int_{-\infty}^{0} \cos(tx) \, \mathrm{d}F(x) + \int_{0}^{\infty} \cos(tx) \, \mathrm{d}[F(x) - 1] \\ &\quad + i\Big\{ \int_{-\infty}^{0} \sin(tx) \, \mathrm{d}F(x) + \int_{0}^{\infty} \sin(tx) \, \mathrm{d}[F(x) - 1] \Big\}. \end{aligned}$$

Integrating by parts,

$$\begin{aligned} \phi(t) &= F(0) + t \int_{-\infty}^{0} \sin(tx) F(x) \, \mathrm{d}x \\ &\quad -(F(0) - 1) + t \int_{0}^{\infty} [F(x) - 1] \sin(tx) \, \mathrm{d}x \\ &\quad -it\Big\{ \int_{-\infty}^{0} \cos(tx) F(x) \, \mathrm{d}x + \int_{0}^{\infty} \cos(tx) [F(x) - 1] \, \mathrm{d}x \Big\}. \end{aligned}$$

Therefore, we derive

$$\begin{aligned} U(t) - 1 &= t \int_{-\infty}^{0} \sin(tx) F(x) \, \mathrm{d}x + t \int_{0}^{\infty} [F(x) - 1] \sin(tx) \, \mathrm{d}x \\ &= -t \int_{0}^{\infty} \sin(tx) F(-x) \, \mathrm{d}x + t \int_{0}^{\infty} [F(x) - 1] \sin(tx) \, \mathrm{d}x \\ &= -t \int_{0}^{\infty} \sin(tx) H(x) \, \mathrm{d}x. \end{aligned} \tag{2.52}$$

If $q = 0$, i.e., $0 < \alpha + \beta < 2$, combined with $(A1)$ and (2.51), it is straightforward to show that

$$1 - U(t) = Cs(\alpha)t^{\alpha} + CDs(\alpha+\beta)t^{\alpha+\beta} + o(t^{\alpha+\beta}).$$

For $q > 0$, i.e., $2q < \alpha + \beta < 2q + 2$, we split (2.52) into two parts:

$$1 - U(t) = t\int_0^1 \sin(tx)H(x)\,\mathrm{d}x + \int_t^{\infty} \sin xH(x/t)\,\mathrm{d}x.$$

As $t \downarrow 0$, the Maclaurin series for $\sin(tx)$ shows

$$\left| t\int_0^1 \sin(tx)H(x)\,\mathrm{d}x \right| = t^2\int_0^1 xH(x)\,\mathrm{d}x + O(t^4).$$

With $(A1)$ and $s(\alpha+\beta)$ defined in (2.51), for $q > 0$,

$$\begin{aligned}
&\int_t^{\infty} \sin xH(x/t)\,\mathrm{d}x \\
&\quad = Ct^{\alpha}\int_t^{\infty} x^{-\alpha}\sin x\,\mathrm{d}x \\
&\qquad + CDt^{\alpha+\beta}\int_t^{\infty} x^{-\alpha-\beta}\Big\{\sin x - \sum_{i=1}^{q}(-1)^{i-1}\frac{x^{2i-1}}{(2i-1)!}\Big\}\,\mathrm{d}x \\
&\qquad + CD\sum_{i=1}^{q}\frac{(-1)^{i-1}}{(2i-1)!}t^{2i}\int_1^{\infty} x^{2i-1-\alpha-\beta}\,\mathrm{d}x + o(t^2).
\end{aligned}$$

Then, for $q > 0$,

$$1 - U(t) = Cs(\alpha)t^{\alpha} + D_1t^2 + o(t^2),$$

where D_1 is not specified. This completes the proof of Theorem 2.1. □

Proof of Theorem 2.2. The difference between $\hat{\alpha}$ and α is

$$\hat{\alpha} - \alpha = \frac{D_1}{Cs(\alpha)}\frac{\sum_{j=1}^m a_j t_j^{\gamma-\alpha}}{S_{zz}} + \frac{\sum_{j=1}^m a_j o(t_j^{\gamma-\alpha})}{S_{zz}} + \frac{\sum_{j=1}^m a_j\epsilon_j}{S_{zz}}. \qquad (2.53)$$

As $m \to \infty$ and $n \to \infty$, by Lemma 2.2,

$$\begin{aligned} S_{zz} &= m\Big\{\frac{1}{m}\sum_{j=1}^{m}(\ln j)^2 - \frac{1}{m^2}(\sum_{j=1}^{m}\ln j)^2\Big\} \\ &= m\{1 + O\{m^{-1}(\ln m)^2\}\}. \end{aligned} \tag{2.54}$$

As $m \to \infty$ and $n \to \infty$, by Lemma 2.3,

$$\begin{aligned} \sum_{j=1}^{m} a_j {t_j}^{\gamma-\alpha} &= mn^{-\frac{\gamma-\alpha}{2}}\Big\{\frac{1}{m}\sum_{j=1}^{m} j^{\gamma-\alpha}\ln j - \Big(\frac{1}{m}\sum_{j=1}^{m}\ln j\Big)\Big(\frac{1}{m}\sum_{j=1}^{m} j^{\gamma-\alpha}\Big)\Big\} \\ &= mn^{-\frac{\gamma-\alpha}{2}}\Big[(\gamma-\alpha)(\gamma-\alpha+1)^{-2}m^{\gamma-\alpha} + O(m^{\gamma-\alpha-1}\ln m)\Big]. \end{aligned}$$

The result follows from $m = n^{\delta}$. □

Proof of Theorem 2.3. The variance of $\hat{\alpha}$ only comes from the error terms, and it is given by

$$Var(\hat{\alpha}) = \frac{1}{S_{zz}^2}\sum_{j=1}^{m} a_j^2 Var(\varepsilon_j) + \frac{1}{S_{zz}^2}\sum_{j=1}^{m}\sum_{i=j+1}^{m} a_i a_j Cov(\varepsilon_i, \varepsilon_j). \tag{2.55}$$

As $m \to \infty$ and $n \to \infty$, according to Lemma 2.1 and Lemma 2.2, the first term and the second term of the right-hand side of (2.55) can be written as

$$\frac{1}{S_{zz}^2}\sum_{j=1}^{m} a_j^2 Var(\varepsilon_j) \sim \frac{n^{\frac{\alpha}{2}-1}}{S_{zz}^2}\sum_{i=1}^{m} a_j^2 j^{-\alpha} \sim \frac{n^{\frac{\alpha}{2}-1}}{m^2}\sum_{j=1}^{m}(\ln j - \frac{1}{m}\sum_{k=1}^{m}\ln k)^2 j^{-\alpha}$$

and

$$\begin{aligned}
&\frac{2}{S_{zz}^2}\sum_{i=2}^{m}\sum_{j=1}^{i-1}a_i a_j Cov(\varepsilon_i,\varepsilon_j)\\
&\quad\sim \frac{n^{\frac{\alpha}{2}-1}}{S_{zz}^2}\sum_{i=2}^{m}\sum_{j=1}^{i-1}a_i a_j i^{-\alpha}\\
&\quad\sim \frac{n^{\frac{\alpha}{2}-1}}{m^2}\sum_{i=2}^{m}(\ln i-\frac{1}{m}\sum_{k=1}^{m}\ln k)i^{-\alpha}\sum_{j=1}^{i-1}(\ln j-\frac{1}{m}\sum_{k=1}^{m}\ln k)\\
&\quad\sim \frac{n^{\frac{\alpha}{2}-1}}{m^2}\sum_{i=2}^{m}(\ln i-\frac{1}{m}\sum_{k=1}^{m}\ln k)(i-1)i^{-\alpha}(\ln (i-1)-\frac{1}{m}\sum_{k=1}^{m}\ln k)\\
&\quad\sim \frac{n^{\frac{\alpha}{2}-1}}{m^2}\sum_{i=1}^{m}(\ln i-\frac{1}{m}\sum_{k=1}^{m}\ln k)^2 i^{-\alpha+1},
\end{aligned}$$

respectively. Combining these two parts and by Lemma 2.4, the variance of $\hat{\alpha}$ is

$$Var(\hat{\alpha})=\begin{cases}\frac{n^{\frac{\alpha}{2}-1}}{m^2}(O(m^{1-\alpha})+O(m^{2-\alpha})), & \text{if } 0<\alpha<1,\\ \frac{n^{\frac{\alpha}{2}-1}}{m^2}(O((\ln m)^3)+O(m)), & \text{if } \alpha=1,\\ \frac{n^{\frac{\alpha}{2}-1}}{m^2}(O((\ln m)^2)+O(m^{2-\alpha})), & \text{if } 1<\alpha<2.\end{cases}\tag{2.56}$$

It is straightforward that the second term in the parentheses for each case on the right-hand side of (2.56) is the leading term of the order of the variance. Therefore, the order of the variance of $\hat{\alpha}$ is $n^{\frac{\alpha}{2}-1}m^{-\alpha}$ for all $0<\alpha<2$. This completes the proof of Theorem 2.3. □

Proof of Theorem 2.4. From (2.53), $\hat{\alpha}-E(\hat{\alpha})$ is equal to $\frac{\sum_{j=1}^{m}a_j\epsilon_j}{S_{zz}}$. Combined with (2.12), it can be rewritten as

$$\begin{aligned}
\hat{\alpha}-E(\hat{\alpha}) &= \sum_{j=1}^{m}\frac{a_j\left(\frac{U_n(t_j)-U(t_j)}{1-U(t_j)}+o_p((nt_j^{\alpha})^{-\frac{1}{2}})\right)}{S_{zz}}\\
&= \frac{1}{n}\frac{\sum_{j=1}^{m}a_j\sum_{i=1}^{n}\frac{[\cos(X_i t_j)-U(t_j)]}{[1-U(t_j)]}}{S_{zz}}+\frac{\sum_{j=1}^{m}a_j o_p((nt_j^{\alpha})^{-\frac{1}{2}})}{S_{zz}}.
\end{aligned}\tag{2.57}$$

For the second term on the right-hand side of (2.57), by Lemma 2.2, as $m \to \infty$ and $n \to \infty$,

$$\begin{aligned}
&\sum_{j=1}^{m} a_j o_p((nt_j^{\alpha})^{-\frac{1}{2}}) \\
&\qquad = mn^{\frac{\alpha-2}{4}} \Big\{ \Big(\frac{1}{m} \sum_{j=1}^{m} j^{-\alpha/2} \ln j \Big) - \Big(\frac{1}{m} \sum_{j=1}^{m} \ln j \Big) \Big(\frac{1}{m} \sum_{j=1}^{m} j^{-\alpha/2} \Big) \Big\} \\
&\qquad = mn^{\frac{\alpha-2}{4}} \Big[-\frac{\alpha}{2} (1 - \frac{\alpha}{2})^{-2} o_p(m^{-\frac{\alpha}{2}}) + o_p\{O(m^{-1} \ln m)\} \Big].
\end{aligned}$$

With (2.54), we obtain that

$$\frac{\sum_{j=1}^{m} a_j o_p((nt_j^{\alpha})^{-\frac{1}{2}})}{S_{zz}} = o_p\{n^{\frac{\alpha-2}{4}} m^{-\frac{\alpha}{2}}\} = o_p\{n^{\frac{\alpha-2-2\alpha\delta}{4}}\}.$$

For the first term on the right-hand side of (2.57), we change the order of the summations. Then it is rewritten as

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{a_j(\cos(X_i t_j) - U(t_j))}{S_{zz}[1 - U(t_j)]}.$$

Define the random variable

$$W_{i,m} = \sum_{j=1}^{m} \frac{a_j(\cos(X_i t_j) - U(t_j))}{S_{zz}[1 - U(t_j)]}.$$

Note that $W_{i,m}$'s with $i = 1, 2, \ldots, n$ for a given m are i.i.d. random variables with null mean and variance

$$\lim_{n \to \infty} Var(W_{i,m}) = O\{n^{\alpha/2 - \delta\alpha}\}.$$

The variance is derived by noting that, for $t_k > t_j$ and $m = n^\delta$, as $n \to \infty$,

$$\begin{aligned} &Cov\Big(\frac{\cos(X_i t_k) - U(t_k)}{1 - U(t_k)}, \frac{\cos(X_i t_j) - U(t_j)}{1 - U(t_j)}\Big) \\ &= E\Big(\frac{[\cos(X_i t_k) - U(t_k)][\cos(X_i t_j) - U(t_j)]}{[1 - U(t_k)][1 - U(t_j)]}\Big) \\ &= \frac{U(t_k + t_j) + U(t_k - t_j) - 2U(t_k)U(t_j)}{2[1 - U(t_k)][1 - U(t_k)]}, \end{aligned}$$

which is equal to $nCov(\varepsilon_k, \varepsilon_j)$. Similarly, as $n \to \infty$,

$$Var(\frac{\cos(X_i t_j) - U(t_j)}{1 - U(t_j)}) = nVar(\epsilon_j).$$

Next, for $m = n^\delta$, define the random variable

$$W^*_{i,n^\delta} = \frac{1}{n^{\frac{\alpha}{4} - \frac{\delta\alpha}{2}}} W_{i,n^\delta}$$

and let

$$\sigma^2 = \lim_{n\to\infty} Var(W^*_{i,n^\delta}).$$

Then denote

$$S_n = \sum_{i=1}^{n} W^*_{i,n^\delta}$$

and

$$\sigma_n^2 = Var(\sum_{i=1}^{n} W^*_{i,n^\delta}) = \sum_{i=1}^{n} Var(W^*_{i,n^\delta}) = nVar(W^*_{1,n^\delta}) = n\sigma^2_{W^*}.$$

The next step is to show that the Lindeberg condition, a sufficient condition, for the central limit theorem holds. The sequence of W^*_{i,n^δ} satisfies the Lindeberg condition (see details in Karr (1993)) if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{i=1}^{n} E[W^{*2}_{i,n^\delta}; \{|W^*_{i,n^\delta}| > \epsilon\sigma_n\}] = 0. \tag{2.58}$$

Since

$$E[W^{*2}_{i,n^\delta}; \{|W^*_{i,n^\delta}| > \epsilon\sigma_n\}] = E[W^{*2}_{1,n^\delta}; \{|W^*_{1,n^\delta}| > \epsilon\sigma_n\}],$$

the term inside the limit on the left-hand side of (2.58) can be written as

$$\begin{aligned}\frac{1}{\sigma_n^2}\sum_{i=1}^{n} E[W_{i,n^\delta}^{*2};\{|W_{i,n^\delta}^{*}|>\epsilon\sigma_n\}] &= \frac{1}{n\sigma_{W^*}^2} nE[W_{1,n^\delta}^{*2};\{|W_{1,n^\delta}^{*}|>\epsilon\sigma_{W^*}\sqrt{n}\}] \\ &= \frac{1}{\sigma_{W^*}^2}E[W_{1,n^\delta}^{*2};\{|W_{1,n^\delta}^{*}|>\epsilon\sigma_{W^*}\sqrt{n}\}].\end{aligned}$$

Note that, by Chebyshev's inequality, as $n \to 0$, for every $\epsilon > 0$,

$$P(|W_{1,n^\delta}^{*}| \geq \epsilon\sigma_{W^*}\sqrt{n}) \leq \frac{Var(W_{1,n^\delta}^{*})}{\epsilon^2\sigma_{W^*}^2 n} = \frac{\sigma_{W^*}^2}{\epsilon^2\sigma_{W^*}^2 n} \to 0.$$

Since

$$W_{i,n^\delta}^{*2}\mathbb{1}\{|W_{i,n^\delta}^{*}|>\epsilon\sigma_n\} \leq W_{i,n^\delta}^{*2}$$

and $E(W_{i,n^\delta}^{*2}) < \infty$, by the dominated convergence theorem, the Lindeberg condition (2.58) holds. Therefore, the sequence of W_{i,n^δ}^{*} satisfies the central limit theorem, i.e., as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i^* \xrightarrow{d} \mathcal{N}(0,\sigma^2).$$

By Slutsky's theorem, as $n \to \infty$,

$$\frac{\sqrt{n}}{n^{\frac{\alpha}{4}-\frac{\delta\alpha}{2}}}(\hat{\alpha} - E(\hat{\alpha})) \xrightarrow{d} \mathcal{N}(0,\sigma^2).$$

This completes the proof of Theorem 2.4. □

Chapter 3

Tail index inference via the empirical scaling function

—Based on the asymptotic properties of the partition function

3.1 Introduction

Distributions with heavy tails are of considerable importance in financial modelling. The history of heavy tails in finance dates back to Benoit Mandelbrot's fundamental work in the 1960s. He conjectured that the variation of speculative prices follows the so-called "stable Paretian" law, for instance cotton prices investigated in Mandelbrot (1963). Student's t-distributions have been considered notably in the literature on modelling the return distributions of financial assets (see Heyde & Leonenko (2005) and references therein).

Since there are different types and definitions of heavy tails, in this chapter, the heavy-tailed distribution is clarified as follows. The distribution of a random variable X is defined as heavy tailed with index $\alpha > 0$ if it has a regularly varying tail, i.e.,

$$P(|X| > x) = \frac{L(x)}{x^{\alpha}}, \quad |x| \to \infty,$$

where $L(x)$ is the so-called slowly varying function, satisfying $L(tx)/L(x) \to 1$ as $|x| \to \infty$, for any $t > 0$ (Embrechts et al., 1997). The existence of the moments of the random variable is determined fully by the tail index α. More precisely, $E(|X|^q)$ is finite only when $q < \alpha$, otherwise this expected value goes to infinity. Due to the importance of α, a new technique to make inference for α is investigated here.

In the literature, estimators of the tail index are usually based on extreme order statistics and their asymptotic properties: for instance, the Pickands estimator (Pickands III, 1975), the Hill estimator (Hill, 1975) and the moment estimator (Dekkers et al., 1989). Various extensions and refinements have been proposed and a nice survey on the comparison of these estimators has been done in de Haan & Peng (1998) (see references therein).

Alternatively, estimators based on the diverging properties of some other statistics are presented. An estimator based on the asymptotics of the sum for $0 < \alpha < 2$ is introduced in the pioneering work of Meerschaert & Scheffler (1998). The asymptotic behaviour of the sum for heavy-tailed data depends on the tail index α, which is related to the extreme values of the underlying distribution somehow. More precisely, this method is based on the fact that sample variance diverges at a rate depending on α. For $\alpha > 2$, a power-transformation of data is needed. Estimators presented in Politis (2002) and McElroy & Politis (2007) exploit this idea.

The underlying assumption of the estimators mentioned above is that random variables are independent distributed. Consequently, most of these estimation methods are not applicable to dependent cases and the performance of these estimators are questionable. While, estimators under dependent structures are also proposed. For instance, Hill (2010) introduces an estimator considering stationary and strong-mixing data. In this chapter, the important question of relaxing the independent condition for estimating α is considered and a similar assumption is introduced.

A novel graphical method applied as an exploratory tool to conjecture heavy tails and the range of the tail index of the underlying distribution is proposed here. Based on the graphical method, estimation methods are also established for the tail index. This approach does not rely on upper

order statistics or their asymptotic properties. Indeed, it is motivated by the scaling property of risky asset returns. The scaling function, denoted by $\tau(q)$, is always applied to the turbulence and multifractal theory to check multifractality in data (Frisch, 1980; Mandelbrot et al., 1997). It is defined by the relation

$$E(|X(t)|^q) = c(q)t^{\tau(q)}, \tag{3.1}$$

where $\tau(q)$ takes into account the influence of time t on moment q. In practice, the scaling function is typically estimated by partitioning the data into blocks and calculating the partition function (defined later).

The limiting behaviour of $\tau(q)$ is influenced by heavy tails, namely the tail index α. For instance, Heyde (2009) has shown that the plots of empirical scaling functions for independently and heavy-tailed distributed random variables with index $\alpha > 2$ are initially linear and ultimately concave. For $\alpha < 2$, the bilinear property of the plot of the empirical scaling function arises naturally from the infinite moments of fractional Lévy motion (Heyde & Sly, 2008).

In this chapter, the relation between the tail index and the scaling function is exploited for a more general range of α under a weakly dependent condition. This relation builds on the asymptotic properties of the partition function, whose blocking structure enables us to extract more information on the tail property. To some extent, this idea goes on the same line as that in Meerschaert & Scheffler (1998).

We present the asymptotic properties of the partition function in Section 3.2. In Section 3.3 we apply these results to make inference for the tail index. Various simulations are conducted in Section 3.4 to evaluate the performance of the proposed methods. An application on exchange rates is presented as well. A summary and further discussions are provided in Section 3.5.

3.2 Asymptotic properties of the partition function

The instrument to estimate the scaling function is the partition function, also called the empirical structure function, which is a special kind of moment statistic. The relationship between the tail index and the scaling function is based on the partition function. Suppose a sample $X_1, \ldots, X_n$ comes from a strictly stationary stochastic process $\{X_t, t \in \mathbb{Z}_+\}$ (discrete time) or $\{X_t, t \in \mathbb{R}_+\}$ (continuous time) which has a heavy-tailed marginal distribution with tail index α. The partition function is defined as the following quantity:

$$S_q(n,t) = \frac{1}{\lfloor n/t \rfloor} \sum_{i=1}^{\lfloor n/t \rfloor} \left| \sum_{j=1}^{\lfloor t \rfloor} X_{t(i-1)+j} \right|^q , \tag{3.2}$$

where $q > 0$ and $1 \leq t \leq n$. Roughly speaking, the partition function considers the average of different orders of moment statistic on consecutive blocked data. The special case, i.e., the empirical q-th absolute moment, is obtained by setting $t = 1$. To allow the size of blocks to grow as sample size increases, let t be equal to n^s. By doing this, we can consider the limiting behaviour of $\ln S_q(n, n^s)/\ln n$, which indicates the rate of divergence of $S_q(n, n^s)$.

Before presenting the theorem about the asymptotic properties of $S_q(n, n^s)$, some definitions are introduced first.

1. The strong mixing, or α-mixing coefficient: for two sub-σ-algebras, $\mathcal{A} \subset \mathcal{F}$ and $\mathcal{B} \subset \mathcal{F}$ on the same complete probability space $(\Omega, \mathcal{F}, P)$, the strong mixing coefficient is defined as

$$a(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(A \cap B) - P(A)P(B)|.$$

2. Strong mixing, or α-mixing, property with an exponentially decaying rate: a process $\{X_t,\ t \geq 0\}$ has a strong mixing property with an

exponentially decaying rate if

$$a(\tau) = \sup_{t \geq 0} a(\mathcal{F}_t, \mathcal{F}^{t+\tau}) \to 0, \quad as \quad \tau \to \infty,$$

and meanwhile $a(\tau) = O(e^{-b\tau})$ holds for some $b > 0$, where $\mathcal{F}_t = \sigma\{X_s, s \leq t\}$ and $\mathcal{F}^{t+\tau} = \sigma\{X_s, s \geq t + \tau\}$.

The next theorem establishes the rate of growth. Basically, this theorem summarises the result of the interplay between the generalised central limit theorem and the weak law of large numbers. The explanation and proof of this theorem are presented in Grahovac et al. (2014).

Theorem 3.1. *Suppose $\{X_t, t \in \mathbb{Z}_+\}$ is a strictly stationary sequence that has a strong mixing property with an exponentially decaying rate and has a heavy-tailed marginal distribution with tail index $\alpha > 0$. Assume $E(X_i) = 0$ for $\alpha > 1$. Then for $q > 0$ and every $s \in (0, 1)$, as $n \to \infty$,*

$$\frac{\ln S_q(n, n^s)}{\ln n} \xrightarrow{p} R_\alpha(q, s) := \begin{cases} \frac{sq}{\alpha}, & \text{if } q \leq \alpha \text{ and } \alpha \leq 2, \\ s + \frac{q}{\alpha} - 1, & \text{if } q > \alpha \text{ and } \alpha \leq 2, \\ \frac{sq}{2}, & \text{if } q \leq \alpha \text{ and } \alpha > 2, \\ \max\left\{s + \frac{q}{\alpha} - 1, \frac{sq}{2}\right\}, & \text{if } q > \alpha \text{ and } \alpha > 2, \end{cases} \tag{3.3}$$

where $\xrightarrow{p}$ refers to convergence in probability.

Remark 3.1. *The zero-expectation assumption in this theorem is not a restriction indeed since one can always demean a sequence to satisfy it. Special cases of this theorem have been proved in Sly (2005) and Chechkina & Gonchara (2000).*

3.3 Applications in heavy-tailed phenomena

In this section the results in Theorem 3.1 are extended to the scaling function. Based on the scaling function, we propose a graphical method for detecting heavy tails and identifying whether the tail index α is larger than

2 or not. More precisely, the idea is to make inference for α from the asymptotic behaviour of $\ln S_q(n, n^s)/\ln n$ for a range of values of s and q. Two regression methods are also presented as extensions of the graphical method.

3.3.1 Scaling function

According to a variant of (3.1), for fixed $q > 0$, the scaling function $\tau(q)$ at a single point q is estimated by regression $y_s = \ln S_q(n, n^s)/\ln n$ on $x_s = s$, for a range of values of $s \in (0, 1)$. From the well known formula for the slope of the regression line, we get an estimate $\hat{\tau}_N(q)$ based on $N-1$ equidistant points in $(0, 1)$, i.e., $s = \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}$,

$$\hat{\tau}_N(q) = \frac{\sum_{i=1}^{N-1} \frac{i}{N} y_{\frac{i}{N}} - \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{i}{N} \sum_{j=1}^{N-1} y_{\frac{i}{N}}}{\sum_{i=1}^{N-1} \left(\frac{i}{N}\right)^2 - \frac{1}{N-1} \left(\sum_{i=1}^{N-1} \frac{i}{N}\right)^2}.$$

Notice that, by choosing enough regression points, $\tau(q)$ can be estimated arbitrarily precisely. Therefore, the continuous form of the previous equation is obtained by letting $N \to \infty$,

$$\begin{aligned} \lim_{N\to\infty} \hat{\tau}_N(q) &= \frac{\int_0^1 s \frac{\ln S_q(n,n^s)}{\ln n} ds - \int_0^1 s ds \int_0^1 \frac{\ln S_q(n,n^s)}{\ln n} ds}{\int_0^1 s^2 ds - \left(\int_0^1 s ds\right)^2}, \\ &= 12 \int_0^1 s \frac{\ln S_q(n, n^s)}{\ln n} ds - 6 \int_0^1 \frac{\ln S_q(n, n^s)}{\ln n} ds. \end{aligned}$$

As $n \to \infty$ and with Theorem 3.1, we obtain the asymptotic plot of $\tau(q)$: for $\alpha \leq 2$,

$$\tau(q) = \begin{cases} \frac{q}{\alpha}, & \text{if } 0 < q \leq \alpha, \\ 1, & \text{if } q > \alpha. \end{cases} \tag{3.4}$$

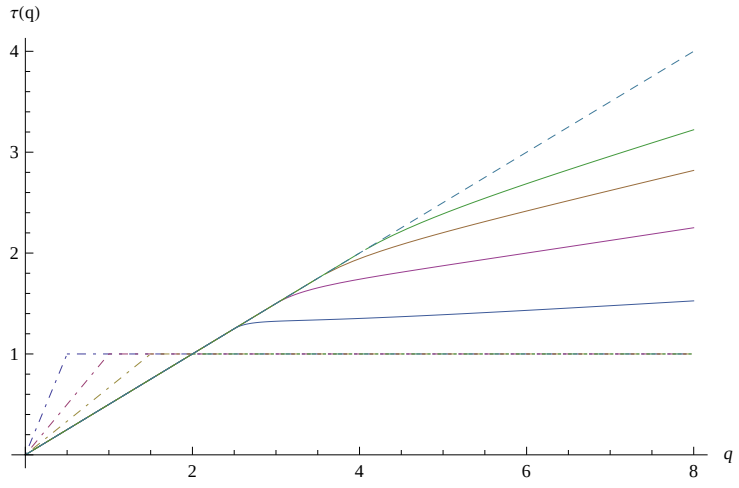
The asymptotic plot of $\tau(q)$ in (3.4) is bilinear with slope $1/\alpha > 1/2$ for $q \leq \alpha$ and then a horizontal line for $q > \alpha$. Therefore, when $\alpha \leq 2$, there will be a sharp slope change at the point where q is equal to the tail index in the graph.

For $\alpha > 2$, $R_\alpha(q, s)$ is not linear in s due to the maximum term for $q > \alpha$ in (3.3). The reason is that for large q, the estimated $\tau(q)$ is particularly sensitive to extreme values as $\max |X_i|^q \approx \sum |X_i|^q$. Moveover, a small s makes the rate of convergence to normal much slower. However, it is approximately linear. If we assume $R_\alpha(q, s) \approx \tau(q)s + c(q)$, the previous approach gives

$$\tau(q) = \begin{cases} \frac{q}{2}, & \text{if } 0 < q \leq \alpha, \\ \frac{q}{2} + \frac{2(\alpha-q)^2(2\alpha+4q-3\alpha q)}{\alpha^3(2-q)^2}, & \text{if } q > \alpha. \end{cases} \tag{3.5}$$

Then, the asymptotic plot of $\tau(q)$ is concave and appears approximately bilinear with slope $1/2$ for $q \leq \alpha$ and slowly decreasing for $q > \alpha$.

Figure 3.1: Plots of scaling function $\tau(q)$ against moment q



The baseline is shown by a dashed line. The case $\alpha \leq 2$ ($\alpha = 0.5, 1.0, 1.5$) and $\alpha > 2$ ($\alpha = 2.5, 3.0, 3.5, 4.0$) are shown by dot-dashed and solid lines, respectively.

When α is large, i.e., $\alpha \to \infty$, it follows from (3.5) that $\tau(q) = q/2$. This case corresponds to data from a distribution with all moments finite, e.g., an independent normally distributed sample. This line will be referred to as the baseline. It is worth noting that the asymptotical shape of the scaling

function calculated for heavy-tailed data significantly depends on the value of the tail index α. Plots of the asymptotic scaling functions for a range of values of α are shown in Figure 3.1. The baseline is shown by a dashed line. The case $\alpha \leq 2$ ($\alpha = 0.5, 1.0, 1.5$) and $\alpha > 2$ ($\alpha = 2.5, 3.0, 3.5, 4.0$) are shown by dot-dashed and solid lines, respectively.

3.3.2 A graphical method

Since the tail index strongly affects the shape of the scaling function, it motivates to use a graphical method based on the scaling function to conjecture the tail index of the underlying distribution. In particular, the asymptotic results indicate that a sharp difference exists between the plots of distributions with infinite variance ($\alpha \leq 2$) and the others ($\alpha > 2$).

In practice, for a finite sample and chosen N, the empirical scaling function $\hat{\tau(q)}$ for fixed $q > 0$ is estimated from

$$\ln S_q(n, n^s)/\ln n = \ln c(q) + \tau(q)s \tag{3.6}$$

by ordinary least squares for a range of s. Repeating this for a range of values of q gives the plot of the empirical scaling function $\hat{\tau(q)}$ against q.

By examining the plot and comparing it with the baseline, one can make inference about the tails of the underlying distribution. If $\hat{\tau(q)}$ is above the baseline for $q < 2$ and nearly horizontal afterwards then true α is probably less than 2. By examining the point where the graph breaks, one can roughly estimate the interval containing α. If $\hat{\tau(q)}$ coincides with the baseline for $q < 2$ and diverges from it for $q > 2$, then true α is probably greater than 2. The point at which deviation starts can be treated as an estimator of α. This establishes a graphical method for distinguishing two cases, i.e., $\alpha \leq 2$ and $\alpha > 2$.

If the graph coincides with the baseline, then we can suspect that the data do not exhibit heavy tails and the moments are finite for the considered range of q. This establishes a useful method to distinguish between heavy tails and non-heavy tails. We illustrate how the method works on simulated and real world examples in the next section.

3.3.3 Regression estimation methods

In this subsection, we define two regression methods to estimate the tail index. Following the assumptions of Theorem 3.1, estimators defined here work for stationary strong mixing samples.

The first method (Method 1) is related to the graphical method and relies on the estimated scaling function. Since both (3.4) and (3.5) can be treated approximately as a piecewise linear function and α is the breakpoint of the piecewise function, we estimate α by regressing $\hat{\tau}(q)$ on a range of values of q. More precisely, the estimator of α is derived from the function

$$\hat{\tau(q)} = \begin{cases} \beta_{00} + \beta_{01} q, & \text{if } 0 < q \leq \alpha, \\ \beta_{00} + \beta_{01}\alpha + \beta'_{01}(q - \alpha), & \text{if } q > \alpha, \end{cases}$$

by minimising the sum of squared residuals. Therefore, nonlinear least squares regression techniques, for instance Levenberg-Marquardt algorithm (Smyth, 2002), can be used to fit the model to given data.

Method 1 may have two main potential problems. The first one is that the estimated scaling function $\hat{\tau(q)}$ instead of the asymptotic scaling function $\tau(q)$ is used in the regression, which may cause bias in the estimation. The second one is that for $\alpha > 2$, the breakpoint may be not so easy to identify, since the result in (3.5) is just approximately linear. Moreover, from Figure 3.1, we can find that the change of the slope is not so obvious for large α. Therefore, we purpose another method which may improve the accuracy of the estimation.

The second method (Method 2) utilises the partition function at some fixed point s_0 to avoid estimating the empirical scaling function by regression. If we fix $s = s_0 \in (0, 1)$ and reconsider the limit in (3.3), we can find for $\alpha \leq 2$,

$$R_{\alpha \leq 2}(q, s_0) = \begin{cases} \frac{s_0}{\alpha} q, & \text{if } q \leq \alpha, \\ s_0 - 1 + \frac{1}{\alpha} q, & \text{if } q > \alpha. \end{cases} \tag{3.7}$$

The breakpoint of this piecewise linear function is $q_{bp} \leq 2$.

For $\alpha > 2$,

$$R_{\alpha>2}(q, s_0) = \begin{cases} \frac{s_0}{2}q, & \text{if } q \leq \alpha, \\ \max\left\{s_0 - 1 + \frac{1}{\alpha}q, \frac{s_0}{2}q\right\}, & \text{if } q > \alpha. \end{cases} \tag{3.8}$$

With regard to the max term in the second segment, if there is a change of slope, it should occur at the point

$$q_{bp} = \frac{s_0 - 1}{\frac{s_0}{2} - \frac{1}{\alpha}} = \frac{2(1 - s_0)}{\frac{2}{\alpha} - s_0},$$

where $s_0 \in (0, 2/\alpha)$. Therefore, we can rewrite (3.8) as

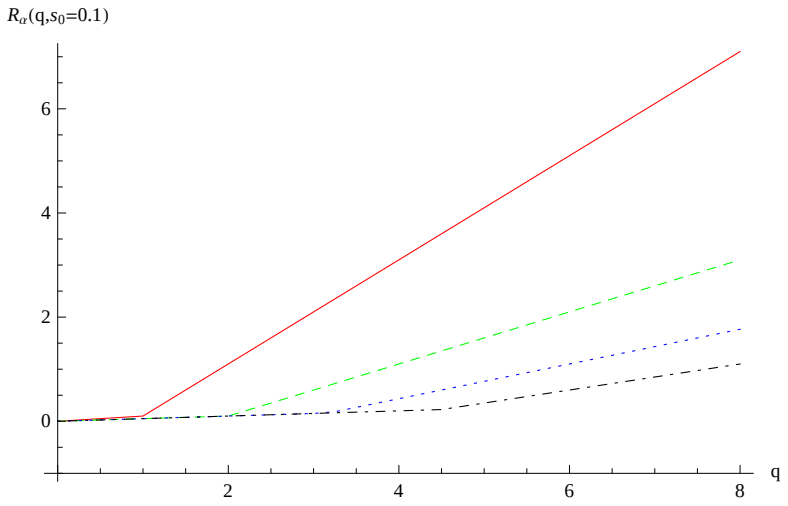
$$R_{\alpha>2}(q, s_0) = \begin{cases} \frac{s_0}{2}q, & \text{if } q \leq q_{bp}, \\ s_0 - 1 + \frac{1}{\alpha}q, & \text{if } q > q_{bp}. \end{cases} \tag{3.9}$$

Comparing (3.7) with (3.9), we find the slope of the second part of the broken line is always equal to $1/\alpha$ no matter whether α is larger than 2 or not. Therefore, we obtain our estimator of α by regressing $\ln S_q(n, n^{s_0})/\ln n$ on a range of values of q. Due to continuity at the breakpoint, the function is given by

$$\frac{\ln S_q(n, n^{s_0})}{\ln n} = \begin{cases} \beta_{10} + \beta_{11}q, & \text{if } 0 < q \leq q_{bp}, \\ \beta_{10} + \beta_{11}q_{bp} + \beta'_{11}(q - q_{bp}), & \text{if } q > q_{bp}. \end{cases} \tag{3.10}$$

Then the estimator is the reciprocal of the slope for the second part of the piecewise function in (3.10), i.e., $\hat{\alpha} = 1/\hat{\beta'_{11}}$. By Theorem 3.1, this estimator is consistent. It is worth mentioning that the change of the slope is sharper for smaller s_0 as shown in Figure 3.2. As mentioned above, if we consider α less than 6, we should choose $s_0 \in (0, 1/3)$. Hence, we set $s_0 = 0.1$ in Figure 3.2. However, this method may have a similar drawback that the slope change is not substantial for large α.

Figure 3.2: Plots of $R_\alpha(q, s_0)$ against q with $s_0 = 0.1$



The case $\alpha = 1, 2, 3, 4$ are shown by solid, dashed, dotted, and dot-dashed lines from left to right.

3.4 Simulations

We provide a simulation study to illustrate how the graphical and estimation procedures defined in the last section work in practice. The data used in the Monte Carlo study are either i.i.d. or dependently distributed. We describe the cases considered here in the next subsection. It is worth mentioning that the conclusion from the simulation is quite general and not generated by an unusual sample selection.

3.4.1 Data

Independent and identically distributed random variables

We first consider i.i.d. random variables. We choose three types of heavy-tailed distributions in our study: stable distributions, Student's t-distributions, and Pareto distributions. The corresponding characteristic function or probability density function for each type of distributions is introduced below.

First, recall that a random variable X is said to follow a stable distribution with index of stability $\alpha \in (0, 2)$ (except for the normal distribution $\alpha = 2$) if it has a characteristic function of the following form:

$$\begin{aligned} \phi_X(\zeta) &= E(\exp\{i\zeta X\}) \\ &= \begin{cases} \exp\left\{-\sigma^\alpha|\zeta|^\alpha\left[1 - i\beta\text{sgn}(\zeta)\tan\frac{\alpha\pi}{2} + i\zeta\mu\right]\right\}, & \text{if } \alpha \neq 1, \\ \exp\left\{-\sigma|\zeta|\left[1 - i\beta\frac{2}{\pi}\text{sgn}(\zeta)\ln|\zeta| + i\zeta\mu\right]\right\}, & \text{if } \alpha = 1, \end{cases} \end{aligned}$$

where $\sigma \in (0, \infty)$ is the scale parameter, $\beta \in [-1, 1]$ the skewness parameter, and $\mu \in (-\infty, \infty)$ the location parameter. Let $S_\alpha(\beta, \sigma, \mu)$ denote the stable distribution. Additionally, when $\mu = 0$, X is strictly stable. If $\beta = 0$, X is symmetric. When $1 < \alpha < 2$, $E(X) = \mu$. It can be proved that with $-1 < \beta \leq 1$ and $0 < \alpha < 2$, as $x \to \infty$,

$$P(X > x) \sim c_\alpha(1 + \beta)x^{-\alpha},$$

where $c_\alpha = \Gamma(\alpha)(\sin\frac{\pi\alpha}{2})/\pi$. It is obvious that stable distributions follow a power law under the conditions mentioned above.

Second, we consider Student's t-distribution $T(\nu, \delta, \mu)$ with the probability density function

$$\text{student}[\nu, \delta, \mu](x) = \frac{\Gamma(\frac{\nu+1}{2})}{\delta\sqrt{\pi}\Gamma(\frac{\nu}{2})}\left(1 + \left(\frac{x-\mu}{\delta}\right)^2\right)^{-\frac{\nu+1}{2}}, \qquad x \in \mathbb{R}, \quad (3.11)$$

(the so-called symmetric scaled Student's t-distribution), where $\delta > 0$ is the scaling parameter, ν the tail parameter (usually called degrees of freedom) and $\mu \in \mathbb{R}$ the location parameter. It follows that both the left-hand and the right-hand tails of Student's t-distribution density (3.11) decrease as $|x|^{-\nu-1}$, i.e., this distribution has heavy tails. Moreover, the expectation $E(X) = \mu$ for $\nu > 1$ and the variance $Var(X) = \frac{\delta^2}{\nu-2}$ is finite for $\nu > 2$.

Finally, if X is a random variable from a Pareto (Type I) distribution,

then the survival function of X is given by

$$P(X > x) = \begin{cases} \left(\frac{x_{min}}{x}\right)^{\alpha}, & \text{if } x \geq x_{min}, \\ 1, & \text{if } x < x_{min}, \end{cases}$$

where $x_{min} > 0$ is the minimum possible value of X, and $\alpha > 0$ is the tail index. We denote the Pareto distribution as $Pa(\alpha, x_{min})$. For $\alpha > 1$, $E(X) = \frac{\alpha x_{min}}{\alpha - 1}$.

Remark 3.2. *The increments of homogenous Lévy processes can be considered here as well: for instance, Lévy-Student processes, Lévy-stable processes ($0 < \alpha < 2$), and Wiener processes (or standard Brownian motions). Indeed, suppose a discretely observed sample $Y_1, \ldots, Y_n$ are drawn from a homogenous Lévy process $\{Y_t\}$, $t > 0$. Denote one step increments as $X_i = Y(i) - Y(i-1)$, then the quantity*

$$\frac{1}{\lfloor n/t \rfloor} \sum_{i=1}^{\lfloor n/t \rfloor} \left| Y_{i\lfloor t \rfloor} - Y_{(i-1)\lfloor t \rfloor} \right|^q,$$

is equivalent to that in (3.2) *and $X_1, \ldots, X_n$ are i.i.d.. Hence, the stationary independent increments of homogenous Lévy processes satisfy the assumptions of Theorem 3.1.*

Dependent data

According to the assumptions of Theorem 3.1, our methods can not only be applied to the case of independent random variables, but also to dependent cases. In order to generate appropriate data sets, we introduce Student diffusion processes and Ornstein-Uhlenbeck (OU) type processes with heavy-tailed marginal distributions.

To define the Student diffusion, we introduce the stochastic differential equation (SDE) first:

$$\mathrm{d}X_t = -\theta\,(X_t - \mu)\,\mathrm{d}t + \sqrt{\frac{2\theta\delta^2}{\nu - 1}\left(1 + \left(\frac{X_t - \mu}{\delta}\right)^2\right)}\,\mathrm{d}B_t, \quad t \geq 0, \quad (3.12)$$

(see Bibby et al. (2005) and Heyde & Leonenko (2005)), where $\nu > 1,\ \delta > 0, \mu \in \mathbb{R},\ \theta > 0$, and $B = \{B_t, t \geq 0\}$ is the standard Brownian motion. SDE (3.12) admits a unique ergodic Markovian weak solution $X = \{X_t, t \geq 0\}$ which is a diffusion process with a symmetric scaled Student's t-distribution defined in (3.11). The Student diffusion is the diffusion process which solves the SDE (3.12). This diffusion is strictly stationary if $X_0 \sim T(\nu, \delta, \mu)$.[1]

According to Leonenko & Šuvak (2010), the Student diffusion is a strong mixing process with an exponentially decaying rate, i.e., there exists $\lambda > 0$ such that

$$\alpha(\tau) \leq \frac{1}{4} e^{-\lambda t}.$$

For the simulation of paths of the Student diffusion process $X = \{X_t, t \geq 0\}$ with known values of parameters, we use the Milstein scheme, which has strong and weak orders of convergence both equal to one (Iacus, 2008).

Following Heyde & Leonenko (2005) two OU type processes are considered here: Student OU type processes and α-stable OU type processes. Recall that a stochastic process $X = \{X_t, t \geq 0\}$ is said to be of the OU type if it satisfies a SDE of the form

$$\mathrm{d}X_t = -\lambda X_t \mathrm{d}t + \mathrm{d}L_{\lambda t}, \quad t \geq 0, \tag{3.13}$$

where $L = \{L_t, t \geq 0\}$ is the background driving Lévy process (BDLP) and $\lambda > 0$.

There exists a strictly stationary stochastic process X_t, $t \in \mathbb{R}$, which has a marginal t-distribution $T(\nu, \delta, \mu)$ with density function (3.11) and BDLP, L, such that (3.13) holds for arbitrary $\lambda > 0$. This stationary process X is referred to as the Student OU-type process. Moreover, the cumulant

[1] For $\nu > 1$, the conditional expectation is

$$E[X_{s+t} | X_s = x] = x e^{-\theta t} + \mu(1 - e^{-\theta t}),$$

where μ is the expectation of the invariant distribution. The autocorrelation function of the Student diffusion is explicit for $\nu > 2$, in which the invariant distribution has finite variance.

$$\rho(t) = corr(X_{t+s}, X_s) = e^{-\theta t}, \quad t \geq 0, \quad s \geq 0, \quad \nu > 2.$$

transform of BDLP, L, can be expressed as

$$\kappa_{L(1)}(\zeta) = \log E\{e^{i\zeta L(1)}\} = i\zeta\mu - \delta|\zeta|\frac{K_{\nu/2-1}(\zeta\mu)}{K_{\nu/2}(\zeta\mu)}, \quad \zeta \in \mathbb{R}, \ \ \zeta \neq 0,$$

where K is the modified Bessel function of the third kind and $\kappa_{L(1)}(0) = 0$ (Heyde & Leonenko, 2005).[2] Since for $0 < p < \nu$, the p-moment of $T(\nu, \delta, \mu)$ is finite, according to Masuda (2004), Student OU type processes are strong mixing with an exponential decaying mixing coefficient. For the Student OU process, the exact law of the increments of the BDLP is unknown, therefore, we use the approach introduced by Taufer & Leonenko (2009) to simulate discrete Student OU processes. This approach circumvents the problem of simulating the jumps of the BDLP and is easily applicable when an explicit expression of the cumulant transform is available (for more details see Taufer & Leonenko (2009))[3].

The α-stable OU type process with parameter $\lambda > 0$ and $0 < \alpha < 2$ introduced by Doob (1942) is the solution of the SDE (3.13), with $L = \{L_t, t \geq 0\}$ as the standard α-stable Lévy motion (Janicki & Weron, 1994). An OU process with any operator-stable marginal distribution is strong mixing with an exponentially decaying rate (Masuda, 2004). Since the distribution of increments for the BDLP, L, is known in this case, we

[2]If $\nu > 1$ then $E(X_t) = \mu$, and if $\nu > 2$ then the correlation function is given by

$$\rho(\tau) = corr(X_{t+\tau}, X_t) = e^{-\lambda|\tau|}, \quad \tau \in \Re.$$

[3]In this case, $\kappa(\zeta)$ can be transformed from the characteristic function of $T(\mu, \delta, \nu)$, which is expressed as

$$\phi_{T(\mu,\delta,\nu)}(\zeta) = e^{i\zeta\mu}\frac{K_{\nu/2}(\delta|\zeta|)}{\Gamma(\nu/2)}(\delta|\zeta|)^{\nu/2}2^{1-\nu/2}, \quad \zeta \in \Re\ .$$

The key point of this simulation is to generate ϵ_t, $t = 1, 2, ..., n$ in a variation of (3.13)

$$X_t = e^{-\lambda}X_{t-1} + \epsilon_t, \quad \epsilon_t = \int_{(0,1]} e^{-\lambda(t-s)}\mathrm{d}L(\lambda s),$$

and $X_0 \sim T(\nu, \delta, \mu)$. ϵ_t is a sequence of i.i.d. random variables, whose distribution F is not generally available in an analytical form. With this approach, a numerical version of F is obtained via the inversion of the characteristic function of ϵ_t, which can be directly worked out from $\kappa(\zeta)$.

consider the Euler's scheme of simulation by replacing differentials in (3.13) with differences.

3.4.2 Plots of empirical scaling functions

In the simulation, ten samples, each with $n = 1000$ i.i.d. observations were generated for distributions: $S_{1.5}(0, 1, 0)$, $T(2, \sqrt{2}, 0)$, $T(4, 2, 0)$, $Pa(0.5, 3)$, $Pa(1.5, 3)$ and $N(0, 1)$. Sets of s and q were generated from 0.10 to 0.90 in steps of 0.035, and from 0.25 to 10 in steps of 0.25, respectively. After calculating $\ln S_q(n, n^s)/\ln n$ for each sample, $\hat{\tau}(q)$ was estimated by using (3.6). All scaling functions are estimated by regression on nearly 30 points. There is no strong evidence that a different number of points in the regression improves the estimation substantially.

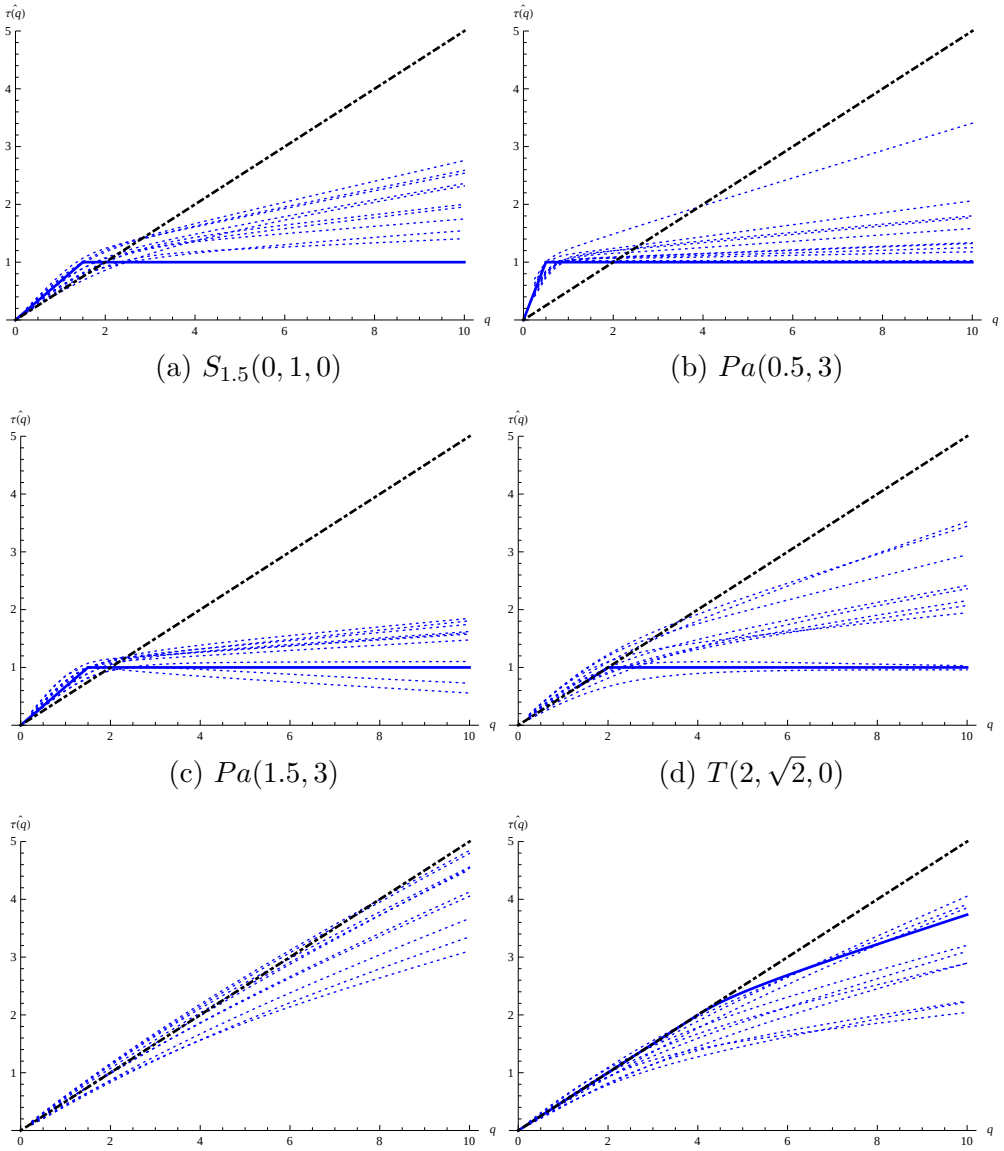
For $Pa(1.5, 3)$, the sample mean is subtracted to adjust in the calculation of $\ln S_q(n, n^s)/\ln n$ due to the zero-expectation assumption. Finally, we plot the estimated scaling functions (dotted) against q together with the corresponding asymptotic plot (solid) and the baseline $q/2$ (dot-dashed) as shown in Figure 3.3.

The plots show a clear-cut behaviour in most cases. It is clear from shapes of the empirical scaling functions that the variance is infinite from Figure 3.3a to 3.3c, since the plots lie above the baseline $\tau(q) = q/2$ then become nearly horizontal at the point $q = \alpha$. The plots for the normal case in Figure 3.3e nearly coincide with the baseline $\tau(q) = q/2$, thus, one can doubt the existence of heavy tails in this case. In Figure 3.3f, the plots for $T(4, 2, 0)$ diverge from $q/2$ gradually after $q = \nu = 4$. For the critical case $\alpha = 2$ as shown in Figure 3.3d, the plots are close to the horizontal line for q larger than some value as the theorem predicts, although the property is not so conspicuous as that for α strictly less than 2.

We next generated 10 sample paths for each of the following processes: (a) a stationary Student diffusion with parameters $\mu = 0$, $\delta = 2$, $\theta = 2$, and $\nu = 3$ at time interval $[0, T]$, $T = 0.35n$; (b) a Student OU process with autoregression parameter $\lambda = 1$ and t-marginal distribution $T(4, 1, 0)$; (c) an α-stable OU process with marginal distribution $S_{1.5}(0, 1, 0)$ and

autoregression parameter $\lambda = 1$.

Figure 3.3: Plots of $\hat{\tau}(q)$ against q for i.i.d. variables



(a) $S_{1.5}(0, 1, 0)$ (b) $Pa(0.5, 3)$

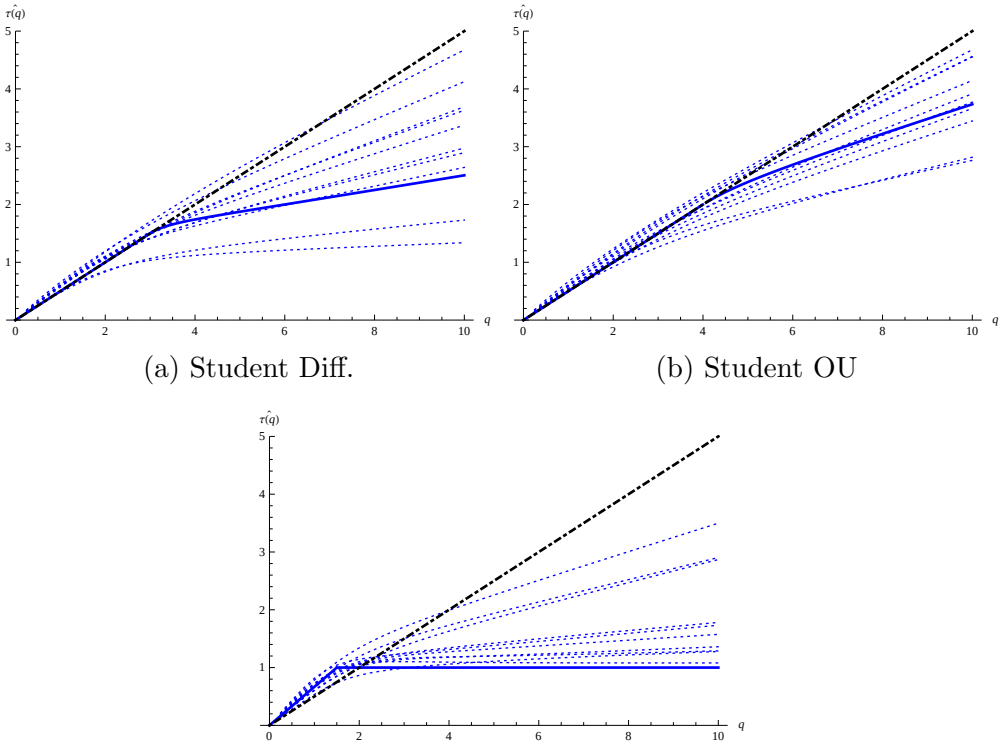
(c) $Pa(1.5, 3)$ (d) $T(2, \sqrt{2}, 0)$

(e) $N(0, 1)$ (f) $T(4, 2, 0)$

Plots of the empirical scaling functions for samples are shown by dotted lines together with the corresponding asymptotic plot (solid line) and the baseline (dot-dashed line).

Figure 3.4 plots the results of the empirical scaling functions for dependent cases. We get results similar to those from the independent cases. It means that the shape of the scaling function is not influenced by the presence of weak form dependence. In short, the plot of the empirical scaling function is quite close to, rather than ideally coincides with, its asymptotic form in most cases.

Figure 3.4: Plots of $\hat{\tau}(q)$ against q for dependent cases



(a) Student Diff.

(b) Student OU

(c) α-stable OU

Plots of the empirical scaling functions for samples are shown by dotted lines together with the corresponding asymptotic plot (solid line) and the baseline (dot-dashed line).

3.4.3 Regression estimation methods

In order to evaluate the performance of the two estimation methods, we compare our estimators with the Hill estimator. Let $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$ denote the order statistics of the sample $X_1, X_2, \cdots X_n$, and k_n be a sequence of positive integers satisfying $1 \leq k_n < n$, $\lim_{n\to\infty} k_n = \infty$, and $\lim_{n\to\infty} (k_n/n) = 0$. The Hill estimator based on k_n upper order statistics is

$$\hat{\alpha}_{k_n} = \left(\frac{1}{k_n} \sum_{i=1}^{k_n} \ln \frac{X_{(i)}}{X_{(k_n+1)}} \right)^{-1}.$$

For each data introduced in Section 3.4.2 we generated 250 samples with $n = 1000$ observations. We set s in the range $(0, 0.5)$ with step size 0.04 in Method 1 and denote the estimator as RE_1. In Method 2, we set the size of the block equal to 3, therefore, $s_0 = \ln 3 \approx 0.16$. The estimator derived from Method 2 is denoted as RE_2. The sample mean is subtracted in order to satisfy the zero-mean assumption in both estimation methods. For the Hill estimator, we choose three groups of extreme order statistics representing low, moderate, and high levels, i.e., using 5%, 10% and 20% of the upper tail. They are denoted by $H05$, $H10$ and $H20$, respectively.

Tables 3.1 and 3.2 report the difference between the mean of estimates $\hat{\alpha}$ and the true value of α and the empirical mean squared error (MSE) for the estimator with different methods for i.i.d. and dependent variables, respectively.

From these two tables, we find that in general the estimator derived from Method 2 performs better than that from Method 1 does, except for the case of non-zero mean (maybe because of the unappropriate use of the sample mean in the demeaning procedure) and the Student diffusion in dependent cases. Therefore, we focus on the comparison between Method 2 and the Hill estimator. For the Hill estimator, it is obvious that the accuracy of the Hill estimator varies considerably according to the chosen percentage of high order statistics.

For i.i.d. variables, the Hill estimator performs better for Pareto distributions, for which the Hill estimator is introduced. For stable

Table 3.1: Comparison of $\mu(\hat{\alpha}) - \alpha$ and MSE for different estimation methods: i.i.d. variables, $n = 1000$

Method	$S_{1.5}(0,1,0)$		$Pa(0.5,3)$		$Pa(1.5,3)$		$T(2,\sqrt{2},0)$		$T(4,2,0)$	
	$\mu(\hat{\alpha}) - \alpha$	MSE	$\mu(\hat{\alpha}) - \alpha$	MSE	$\mu(\hat{\alpha}) - \alpha$	MSE	$\mu(\hat{\alpha}) - \alpha$	MSE	$\mu(\hat{\alpha}) - \alpha$	MSE
RE_1	0.3428	0.2520	0.5434	0.2994	-0.0048	0.0536	0.1392	0.1686	-0.8342	1.1527
RE_2	0.1681	0.1246	-0.0581	0.0075	-0.3459	0.1520	-0.0218	0.1246	-0.6287	0.6737
$H05$	0.2650	0.1272	0.0160	0.0058	-0.0096	0.0457	-0.1395	0.0882	-1.0573	1.2670
$H10$	0.2468	0.0888	0.0053	0.0026	-0.0046	0.0263	-0.2840	0.1092	-1.5691	2.5003
$H20$	0.0054	0.0096	0.0026	0.0012	-0.0071	0.0119	-0.6071	0.3766	-2.2403	5.0321

Table 3.2: Comparison of $\mu(\hat{\alpha}) - \alpha$ and MSE for different estimation methods: dependent cases, $n = 1000$

Method	Sta. OU $S_{1.5}(0,1,0;1)$		Stu. OU $T(4,1,0;1)$		Stu. diff. $T(3,2,0;2)$	
	$\mu(\hat{\alpha}) - \alpha$	MSE	$\mu(\hat{\alpha}) - \alpha$	MSE	$\mu(\hat{\alpha}) - \alpha$	MSE
RE_1	0.4192	0.3193	-0.9343	1.5541	-0.3725	0.4622
RE_2	0.1419	0.1151	0.2769	0.6905	-0.8723	0.8588
$H05$	0.2786	0.2009	-1.1392	1.4216	-0.3964	0.4099
$H10$	0.2605	0.1270	-1.5938	2.5841	-0.7263	0.6037
$H20$	0.0102	0.0173	-2.2644	5.1419	-1.2702	1.6331

distributions, the accuracy of our estimator is comparable to that of the Hill estimator. In this case, it is generally better than $H05$ and slightly worse than $H10$ and $H20$. However, the differences are not conspicuous. For Student's t distributions with $\alpha = 2$, our estimator performs similarly to $H05$ and $H10$. For Student's t-distributions with $\alpha \geq 2$, the Hill estimator is quite inaccurate no matter how many order statistics are used. Moreover, the Hill estimator always shows some downward bias. By contrast, our estimator performs relatively better and is quite close to the true value of α.

For dependent data, the MSE of our estimator is less than those of the Hill estimator for the Student OU processes with $\alpha > 2$. Similar to the results from independent cases, the Hill estimator produces downward bias when $\alpha > 2$. For α-stable OU processes, our estimator performs better than $H05$ and $H10$ and a little worse than $H20$. For Student diffusion processes, our estimator is competitive to $H05$ and $H10$ and much better than $H20$. For $\alpha < 2$, it shows a trend that the more upper order statistics used in the Hill estimation, the more accurate the Hill estimator is. The opposite trend is presented for the Hill estimator for $\alpha \geq 2$. Therefore, it is quite different to find a rule on how to choose an appropriate percentage for the Hill estimation. For our estimator, it retains its accuracy regardless of the presence of dependence.

3.4.4 An application in exchange rates

Here we present an application of our methods on exchange rates. Our illustration is concentrated on the graphical method while the estimates of the tail index are provided as inferences.

As mentioned in the introduction of this chapter, distributions of logarithmic asset returns in financial markets sometimes exhibit heavy tails. Three data sets are considered here. The first data set contains 1538 spots (C.E.T. 2.15 pm) exchange rates of the euro against the US dollar (EUR/USD) from the European Central Bank during the period

$2007 - 2012$[4]. The second data set contains 1516 spot exchange rates of the US dollar against the Chinese yuan (USD/CNY) during the period $2007 - 2012$ observed by the Bank's Foreign Exchange Desk in the London interbank market around 4pm[5]. The third data set contains 187 daily noon buying rates of the US dollar against the Thai baht (USD/THB) from July 01, 1997 to March 31, 1998 in New York for cable transfers payable in foreign currencies[6].

Figure 3.5 presents each exchange rate sequence in the selected period. Plots of the log density against the normalised log return are presented in Figure 3.6. The normalised log return is calculated according to the mean and standard deviation in a finite variance case. The mean and absolute deviation is used for normalisation where variance does not exist (checked by the proposed graphical method). The plot is compared with the log density plot of the standard normal distribution (denoted by a solid line). Moreover, for the cases of the EUR/USD and USD/CNY, the log density plot of Student's t-distribution with $\alpha = 4$ is shown by a dashed line as well. For the USD/THB case, the dashed line is the log density plot of a stable distribution with $\alpha = 1.5$. Figure 3.7 shows the plot of the estimated $\tau(q)$ against the moment q by a solid line compared with the baseline drawn by a dotted line. Data are normalised due to the zero-expectation assumption.

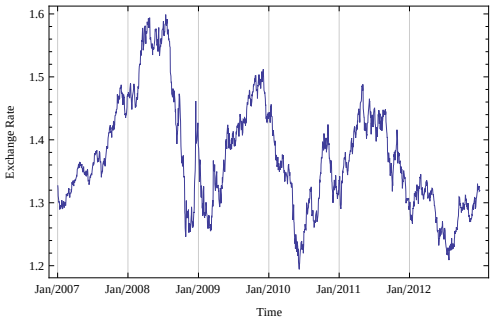
In Figure 3.5a, the EUR/USD presents moderate fluctuations, which is a common property of most floating currencies; previous researches suggest that the tail index is usually between 3 and 5 in this case (Heyde & Liu, 2001; Hurst & Platen, 1997). The comparison of the log density plots is shown in Figure 3.6a. We find that the tails of the log density plot for the data is hyperbolic, which is the property of Student's t-distribution, in contrast to a parabola for a normal distribution. This finding is confirmed by the plot in Fig 3.7a: the slope of $\hat{\tau}(q)$ is really close to $1/2$ for small q and then it decreases afterwards, in the meantime, the plot diverges from the baseline. It is obvious that α is greater than 2 in this case. With our regression

[4]Data source: European Central Bank, http://sdw.ecb.europa.eu

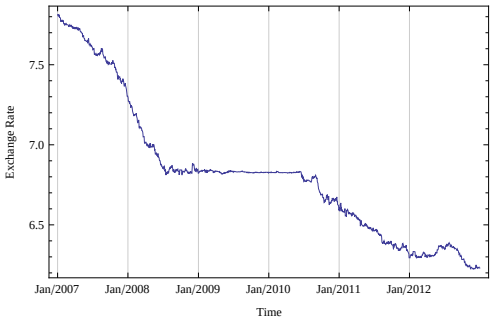
[5]Data source: Bank of England, http://www.bankofengland.co.uk

[6]Data source: Economic Research, Federal Reserve Bank of St. Louis, http://research.stlouisfed.org

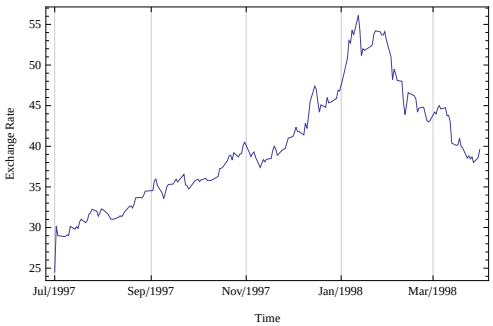
Figure 3.5: Plots of exchange rate sequences



(a) EUR/USD

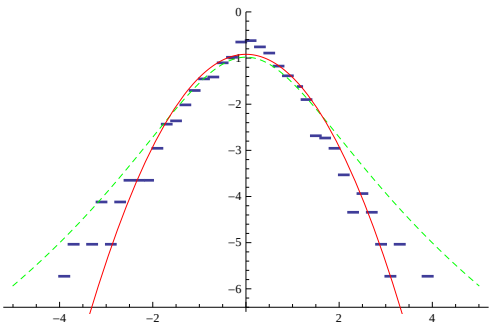


(b) USD/CNY

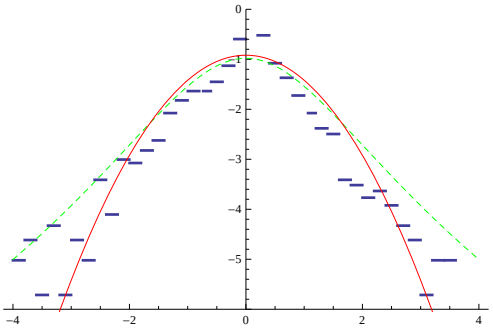


(c) USD/THB

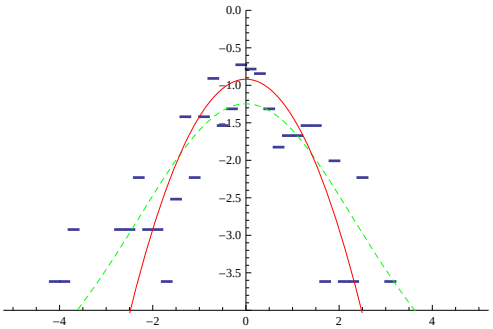
Figure 3.6: Log density plots



(a) EUR/USD



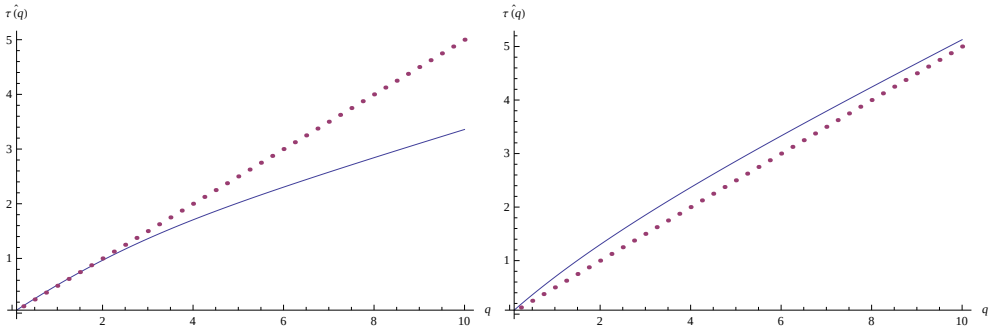
(b) USD/CNY



(c) USD/THB

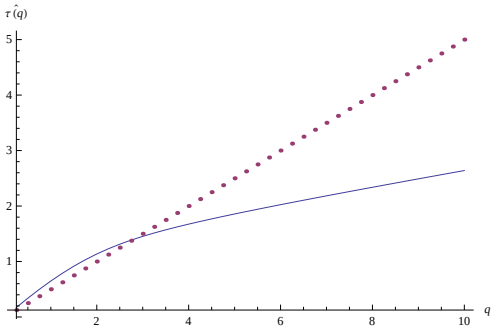
The log density plot of the standard normal distribution is shown by a solid line in each figure. The dashed line is used to show the log density plot for Student's t-distribution with 4 degrees of freedom in the first two graphs and a stable distribution with $\alpha = 1.5$ in the last graph.

Figure 3.7: Plots of $\hat{\tau}(q)$ against q for exchange rate data sets



(a) EUR/USD

(b) USD/CNY



(c) USD/THB

The empirical scaling function is shown by a solid line. The baseline $q/2$ is shown by a dotted line.

methods, we get $RE_1 = 3.4645$ and $RE_2 = 4.7821$, compared with the Hill estimates: $H05 = 3.3432$, $H10 = 2.9225$, and $H20 = 1.9648$. As shown in the simulation study, the Hill estimator severely underestimates the value of the tail index.

In Figure 3.5b, the USD/CNY shows slight fluctuations even during the financial crisis in recent years. In Figure 3.6b, the right tail is more close to a normal distribution. However, there are some deviations from the normal distribution in the left tail. In Figure 3.7b, the plot of $\hat{\tau}(q)$ and the baseline almost coincide, which means the distribution of the USD/CNY is probably not heavy-tailed. The slight departure from the baseline in Figure 3.7b may be caused by the deviations in the left tail as shown in Figure 3.6b. These deviations result in some odd estimates with our methods, i.e., $RE_1 = 3.2898$ and $RE_2 = 4.0623$. Hill estimates are also reported: $H05 = 2.4432$, $H10 = 1,8629$, and $H20 = 1.1722$.

If we consider the exchange rate regime the Chinese government following in this period, we can understand better why this phenomenon happens. Since 2005, the Chinese government began to allow the CNY to float, which means the CNY was no longer pegged to the USD. However, this floating is not totally flexible but restricted around a base rate with narrow fluctuations. Moreover, this base rate is fixed based on a basket of foreign currencies decided by the government. Furthermore, due to the financial crisis in 2008, the CNY is somehow "fixed" unofficially again since it was "repegged" to the USD. Actually, the Chinese government tried to control its exchange rate fluctuations in a reasonable region.

Finally, the USD/THB exhibits high volatility in Figure 3.5c. There is an extreme depreciation at the beginning of July, 1997. From July 1997 to March 1998, Thailand was strongly affected by the Asian financial crisis. Consequently, the Baht was forced to float due to the lack of foreign reserves and it depreciated over 40 per cent during this period. In Figure 3.6c, the plot sharply departs from that of a normal distribution. According to our methods, we find the evidence of heavy tails in Figure 3.7c. The slope of $\hat{\tau}(q)$ is above $1/2$ when q is smaller than 2. Then the curve becomes nearly horizontal, which indicates that the tail index is probably less than 2. The

estimates obtained in this case are: $RE_1 = 2.4732$, $RE_2 = 2.4962$, $H05 = 2.5354$, $H10 = 2.6353$, and $H20 = 1.7716$. Our estimates are not smaller than 2 as expected from the graphical method. A possible explanation is that too few extreme values included in the sample, as presented in Figure 3.5c, make the cut-off behaviour of the plot not sharp enough to obtain an accurate estimate.

3.5 Summary and discussion

In this chapter the partition function and its asymptotic properties are presented first. The result shows that the limit behaviour of the partition function is strongly affected by the existence of moments for the underlying distribution. Moreover, the existence of moments is directly indicated by the tail index of heavy-tailed data. This motivates to apply the partition function to the context of heavy-tailed analysis. Therefore, a graphical method via the scaling function to detect heavy tails is proposed. Indeed, the plot of the scaling function not only informs the existence of heavy tails but also reflects the tail behaviour at a single point. More precisely, there is a reflection between the breakpoint of the scaling function and the tail index α, which allows us to establish estimation methods of α.

Simulation and empirical results indicate the great potential of the proposed methods. The current work could be further extended into two directions. One is to establish a hypothesis test on the existence of heavy tails based on the graphical method. The other is to clarify the nonlinear behaviour of the plot of the empirical scaling function caused by multifractatlity and infinite low order moments.

Chapter 4

Power laws in empirical data

4.1 Overview

Many natural and man-made phenomena approximately follow a remarkable regularity, i.e., a power law, for instance, the population of cities, the number of links to web sites, and the intensity of wars. Mathematically, a random variable X obeys a power law if its distribution satisfies, at least in the upper tail,

$$P(X > x) = kx^{-\alpha},$$

where $k > 0$ and the tail index $\alpha > 0$. This power law is generally only valid for values greater than some threshold x. Thus, in practice, the analysis of power laws focuses on the tail of the distribution of given data.

The application of the power law to empirical data sets was initially proposed by Pareto (1896) for the distribution of personal incomes. He suggested that the number of people with income more than some x is proportional to $1/x^{\alpha}$, where $\alpha \approx 1.7$. A special case of power laws is Zipf's law where the exponent $\alpha \approx 1$. The city-size literature suggests that Zipf's law generally holds for the largest cities in a country (see Gabaix & Ioannides (2004) and references therein). These early studies focus only on the largest cities in a country (e.g., the largest Metropolitan Statistical Areas (MSAs) in the US).

However, a recent development by Eeckhout (2004) extends the analysis

to the entire range of city sizes and demonstrates that the size of all populated places in the US follows a log-normal distribution. There is an ongoing debate about whether the distribution of city sizes is best approximated by a power law or a log-normal distribution. Levy (2009) uses graphical analysis and finds significant deviations from a log-normal distribution in the top range of the largest cities. Several potential problems with Levy's studies are pointed by Eeckhout (2009): e.g., the graphical method is based on visual inspection of the log-log plot to identify the presence of a power law. More recently, Rozenfeld et al. (2011) use an algorithm to construct "cities" from microdata, denoted "clusters". They find that Zipf's law holds, to a good degree of approximation, both in the US and in Great Britain (GB). However, by using the populated places aggregated from 6,127,259 blocks in Census 2010, Bee et al. (2013) conclude that the power-law assumption in the upper tail of the distribution of the US city sizes is questionable. These latest studies imply that the power-law behaviour is less robust than previously claimed based on the uniformly most powerful unbiased test (del Castillo & Puig, 1999) or the switching model (Ioannides & Skouras, 2013).

Understanding the underlying probability law of city sizes helps to shed light on understanding the worldwide urban structure. Various economic models have been introduced based on different thoughts of empirical regularities. One school of thought introduces simple models with economic microfoundations that result in Zipf's law (e.g., Gabaix (1999)). The other school advances models that suggest log-normality. Log-normal distributions are consistent with growth rates of city populations that do depend on the city size[1] (e.g., Eeckhout (2004)).

The main purpose of this chapter is to analyse the behaviour of the upper tail of the distribution of city sizes based on data that cover all city sizes. If a power law is detected, the tail index is estimated. The techniques introduced in the previous two chapters are employed. The first approach is based on a regression technique via the empirical characteristic function

[1]A stochastic growth process which is proportionate gives rise to an asymptotically log-normal distribution (Kapteyn, 1903). This regularity is predicated by Gibrat's law: "The law of proportionate effect will therefore imply that the logarithms of the variable will be distributed following the [normal distribution] (Gibrat, 1931)."

(ECF) introduced in Chapter 2, denoted regression approach (RA). The second approach relies on a graphical method based on the partition function proposed in Chapter 3, denoted graphical approach (GA). Both these two approaches can identify power laws without the difficulty of identifying the start of the tail. In other words, the behaviour in the tail will be identified by these two approaches without the need to determine a starting point of the tail. We consider four data sets derived with different definitions of "city".

In order to assess the performance of the RA, several other empirical data sets from other fields are also analysed. The rest of this chapter is organised as follows. Section 4.2 introduces the RA and GA. The empirical results on the city-size data based on these two approaches are presented in Section 4.3. Examples from other fields are studied in Section 4.4. This chapter concludes with a brief discussion in Section 4.5. All technical details are deferred to Appendix 4.A.

4.2 Two approaches to detect power laws

In the literature, the power law which the distribution of city sizes follows is specified as a Pareto law. Mathematically, the probability density function of a Pareto-law distribution is

$$f(x) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}},\ x > \theta,\ \theta > 0,\ \alpha > 0. \tag{4.1}$$

It is straightforward to show that a Pareto tail obeys the relation

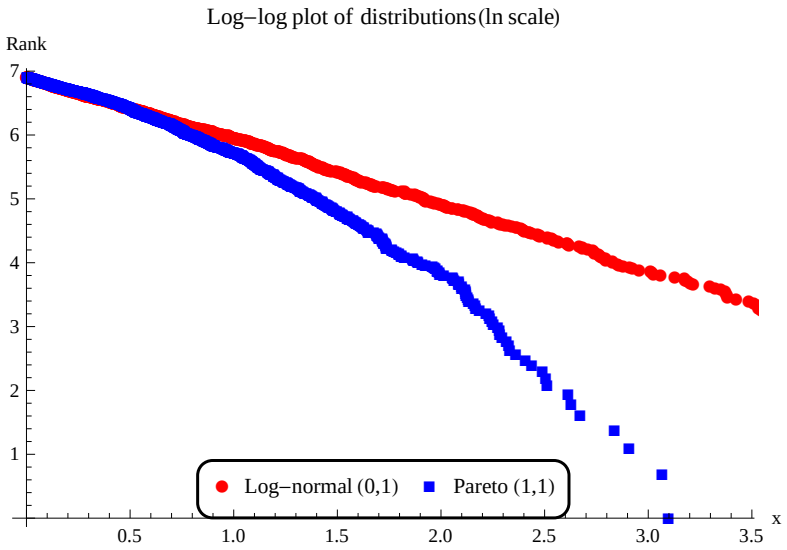
$$\ln F(x) = constant - \alpha \ln x,$$

by taking the logarithm of both sides of the cumulative density function (CDF) derived from (4.1). It implies that the log-log plot of a Pareto tail is a straight line above some threshold x.

Unfortunately, this log-log plot is unreliable in distinguishing a Pareto tail from that of other distributions. The straight line of the log-log plot may

be caused by a log-normal distribution as shown in Figure 4.1. Therefore, it is not sufficient to distinguish a Pareto tail from that of other distributions only from a straight line observed in the plot, in spite of its popularity and simplicity. It is necessary to find a more appropriate and robust method to identify the Pareto tail. Two possible approaches based on previous chapters are presented.

Figure 4.1: Log-log plot of log-normal and Pareto distributions: an example, $n = 1000$



4.2.1 The regression approach

The regression approach (RA) is based on the ECF set out in Chapter 2. As proved in Section 2.4.4, a distribution with a Pareto tail satisfies the assumption that the tail is regularly varying at infinity, otherwise arbitrary, i.e.,

$$P(X > x) = Cx^{-\alpha}[1 + Dx^{-\beta} + o(x^{-\beta})], \quad as \quad x \to \infty, \tag{4.2}$$

where $C > 0, \alpha > 0, \beta > 0$ and D is a real number. Based on the relation between the characteristic function (CF) and the distribution function, for a

Pareto tail with index $0 < \alpha < 2$,

$$1 - U(t) \sim s(\alpha)t^{\alpha}, \quad as \quad t \downarrow 0,$$

where $U(t) = \int_{-\infty}^{\infty} \cos(tx)\,\mathrm{d}F(x)$ is the real part of the CF.

If the Pareto law holds for the distribution of city sizes, the estimator of the tail index derived from the regression below should be around one:

$$\ln\left(1 - U_n(t)\right) = C_{\alpha} + \alpha \ln t + \ln \frac{1 - U_n(t)}{1 - U(t)}, \quad as \quad t \downarrow 0, \tag{4.3}$$

where C_{α} is a "constant" depending on α and $U_n(t) = \frac{1}{n}\sum_{i=1}^{n} \cos(X_i t)$ is the real part of the ECF. If a log-normal distribution holds, the estimator of α should be quite close to two due to its finite variance. Each point in the regression (4.3) is calculated using all observations, not just the "extremes", in the given sample.

To show how the RA works, a further example, a composite distribution is used. This is motivated by that the body of the distribution of city sizes (the part except the tail) is well approximated by a log-normal distribution (Levy, 2009). The composite log-normal-Pareto distribution of which the "body" is log-normal and the "tail" is Pareto captures the property of the distribution of city sizes. This distribution was initially applied to insurance for modelling actuarial data to handle large loss payments (Cooray & Ananda, 2005).

The composite density is derived from

$$f_C(x) = \begin{cases} cf_1(x), & \text{if } 0 < x < \theta, \\ cf_2(x), & \text{if } \theta \leq x < \infty, \end{cases}$$

where c is a normalising constant, $f_1(x)$ is a log-normal density and $f_2(x)$ is a Pareto density, i.e.,

$$f_1(x) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right), \quad x > 0,\ \sigma > 0,\ \mu \in \mathbb{R}, \tag{4.4}$$

where (μ, σ) are the mean and standard deviation of $\ln x$ and $f_2(x) = f(x)$.

Imposing the necessary conditions of continuity and differentiability at θ, namely, $\int_0^\infty f(x)\,\mathrm{d}x = 1$, $f_1(\theta) = f_2(\theta)$, and $f_1'(\theta) = f_2'(\theta)$, Cooray & Ananda (2005) derive the smooth density function of the so-called composite log-normal Pareto distribution

$$f_{LP}(x) = \begin{cases} \frac{\alpha\theta^\alpha}{(1+\Phi(k))x^{\alpha+1}} \exp\left(-\frac{\alpha^2}{2k^2}\ln^2\left(\frac{x}{\theta}\right)\right), & \text{if } 0 < x < \theta, \\ \frac{\alpha\theta^\alpha}{(1+\Phi(k))x^{\alpha+1}}, & \text{if } \theta \le x < \infty, \end{cases} \tag{4.5}$$

where $\Phi(.)$ is the CDF of the standard normal distribution and $k \approx 0.372238898$. Indeed, this value of k is the positive solution of the equation $\exp(-k^2) = 2\pi k^2$.[2] Therefore, the tail of this composite distribution is of the form

$$P(X > x) \sim x^{-\alpha},$$

which is a special case of (4.2) with $D = 0$.

Table 4.1: Mean and root MSE for composite log-normal-Pareto distributions

α	n	$\delta = 1/4$		$\delta = 1/3$		$\delta = 1/2$	
		Mean	root MSE	Mean	root MSE	Mean	root MSE
$\alpha = 0.5$	1000	0.4979	0.0481	0.4937	0.0349	0.4726	0.0351
	5000	0.5014	0.0252	0.4997	0.0173	0.4932	0.0124
$\alpha = 0.8$	1000	0.7874	0.0644	0.7786	0.0492	0.7277	0.0775
	5000	0.6862	0.1206	0.7108	0.0939	0.7506	0.0523
$\alpha = 1.0$	1000	0.9734	0.0839	0.9592	0.0703	0.8860	0.1191
	5000	0.9929	0.0625	0.9860	0.0436	0.9547	0.0504
$\alpha = 1.2$	1000	1.1591	0.1123	1.1303	0.1003	1.0311	0.1735
	5000	1.3395	0.1642	1.2467	0.0773	1.1481	0.0605
$\alpha = 1.5$	1000	1.4124	0.1455	1.3739	0.1499	1.2338	0.2702
	5000	1.4747	0.1252	1.4456	0.0980	1.3622	0.1427

Simulations of the composite distributions with parameters $\alpha = 0.5, 0.8, 1.0, 1.2, 1.5$ and $\theta = 50$ were run for i.i.d. variables with

[2]The number of the parameters of the composite distribution has been reduced to two: $\theta > 0$ and $\alpha > 0$, while the other two parameters can be calculated: $\sigma = k/\alpha$, $\mu = \ln\theta - \alpha\sigma^2$ and the constant is $c = 1/(1+\Phi(k))$.

$n = 1000$ and 5000. Point t_j in the regression is chosen by the rule

$$t_j = j/\sqrt{n},\ j = 1, 2, \ldots, m = n^{\delta},\ where\ 0 < \delta < 1/2.$$

The mean of the estimates and the root MSE based on 250 iterations are displayed in Table 4.1. The results indicate that the estimator obtained by the RA is consistent. In general, the estimate approaches the true value when t is close to zero, because the relation works best when the CF is near the origin.

4.2.2 The graphical approach

The graphical approach (GA) applied in this chapter is based on the technique introduced in Chapter 3. It relies on the plot of the scaling function based on the asymptotic properties of the partition function. More precisely, the partition function is defined as

$$S_q(n, n^s) = \frac{1}{\lfloor n^{1-s} \rfloor} \sum_{i=1}^{\lfloor n^{1-s} \rfloor} \left| \sum_{j=1}^{\lfloor n^s \rfloor} X_{\lfloor n^s \rfloor (i-1)+j} \right|^q, \tag{4.6}$$

where $q > 0$ and $s \in (0, 1)$.

For a sequence of i.i.d. random variables X drawn from a log-normal distribution, for each $s \in (0, 1)$,

$$\frac{\ln S_q(n, n^s)}{\ln n} \xrightarrow{p} \frac{sq}{2}, \tag{4.7}$$

as $n \to \infty$. Thus, the asymptotic plot of the scaling function $\tau(q)$ is a straight line with the slope equal to $1/2$ with respect to the moment q (proofs of (4.7) and (4.8) are given in Appendix 4.A), i.e.,

$$\tau(q) = \frac{q}{2}. \tag{4.8}$$

For a sequence of i.i.d. random variables X drawn from a distribution

with a Pareto tail with $0 < \alpha \leq 2$, for each $s \in (0, 1)$,

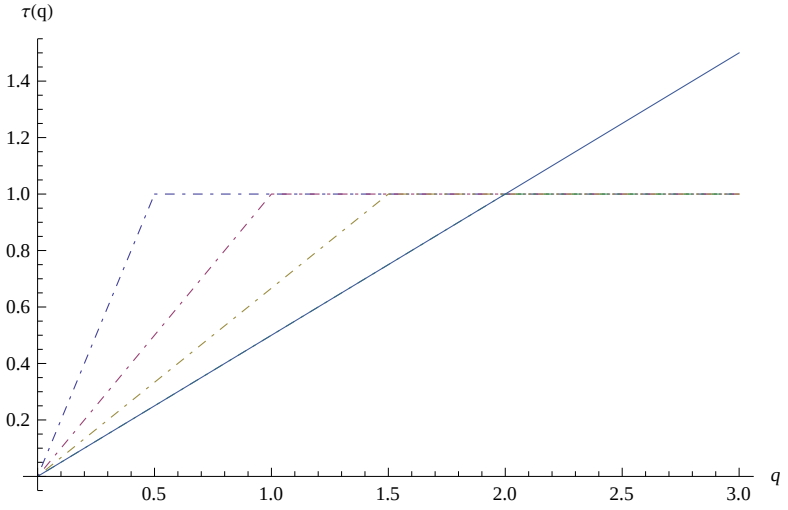
$$\frac{\ln S_q(n, n^s)}{\ln n} \xrightarrow{p} \begin{cases} \frac{sq}{\alpha}, & \text{if } 0 < q \leq \alpha, \\ s + \frac{q}{\alpha} - 1, & \text{if } q > \alpha, \end{cases}$$

as $n \to \infty$. Hence, the asymptotic plot of the scaling function is bilinear and depends on the value of the tail index α (more details in Sections 3.2 and 3.3), i.e., for all $0 < \alpha \leq 2$, as $n \to \infty$,

$$\tau(q) = \begin{cases} \frac{q}{\alpha}, & \text{if } 0 < q \leq \alpha, \\ 1, & \text{if } q > \alpha. \end{cases}$$

The asymptotic plots of the scaling functions for log-normal and Pareto distributions are presented in Figure 4.2. Plots of Pareto tails with $\alpha = 0.5, 1.0, 1.5$ and the log-normal tail are denoted by dot-dashed and solid lines, respectively. These two families of distributions are easily distinguished in this figure, especially for small α.

Figure 4.2: Asymptotic plots of $\tau(q)$



Plots of Pareto tails with $\alpha = 0.5, 1.0, 1.5$ and the log-normal tail are denoted by dot-dashed and solid lines, respectively.

Based on simulation results, plots of the empirical scaling functions for log-normal distributions, denoted $LN(\mu, \sigma^2)$, with different parameters together with the asymptotic plot, $\tau(q) = q/2$, are shown in Figure 4.3. The plots are based on i.i.d. variables with $n = 1000$. The scaling parameter σ of the log-normal distribution, rather than the location parameter μ, influences the accuracy of the plot of the empirical scaling function. This can be explained by the fact that for large σ the log-normal distribution behaves like a power tail with a slowly increasing "tail index". As an illustration, the probability density function in (4.4) is rewritten as (Malevergne et al., 2011):

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \frac{1}{x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\mu^2}{2\sigma^2}} \cdot x^{-1+\frac{u}{\sigma^2}-\frac{\ln x}{2\sigma^2}},$$

where the "tail index"

$$\alpha(x) = \frac{1}{2\sigma^2} \ln(\frac{x}{e^{2\mu}}).$$

For σ large enough, $\alpha(x)$ increases so slowly that it seems constant even with x shifting several decades.

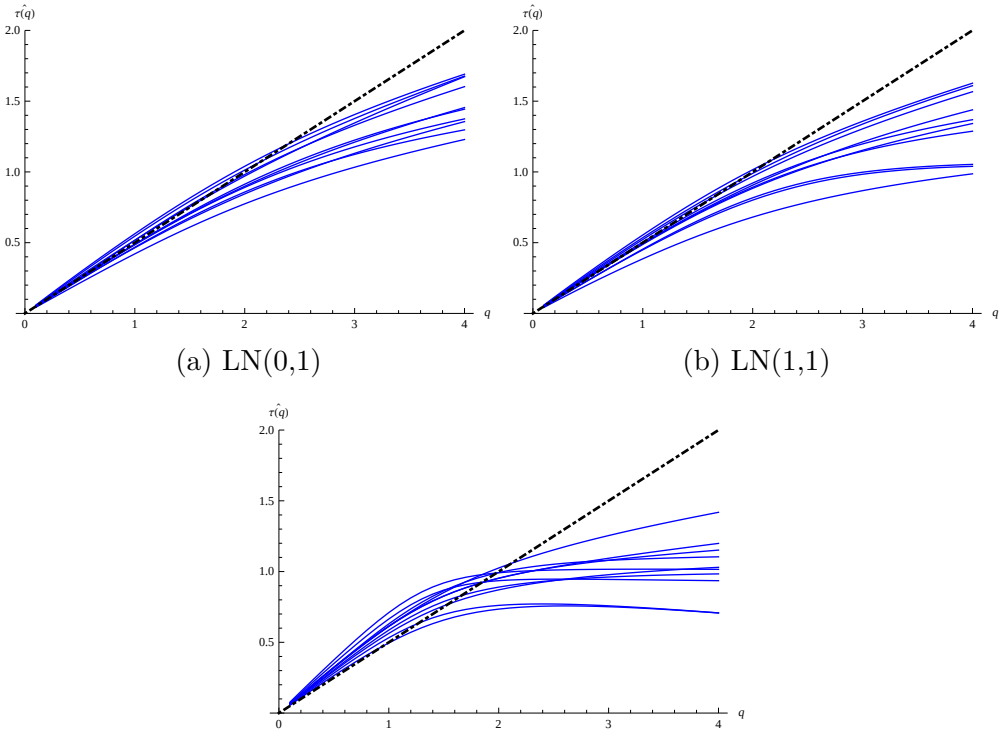
To attenuate this effect, it is suggested to "standardise" the log-normal distribution. The standardisation is obtained by a power transformation of empirical data. It is easy to show that a γ-power transformation of a Pareto tail with index α is still a Pareto tail with a transformed index α/γ. A γ-power transformation of a log-normal distribution $LN(\mu, \sigma^2)$ is still a log-normal distribution $LN(\gamma\mu, \gamma^2\sigma^2)$. Indeed, this power transformation does not change the class of distributions which original data belong to.

The proposed GA is based on the measurement of the distance between the plots of the empirical and asymptotic scaling functions under the null hypothesis via a Monte Carlo procedure[3]. The "distance" here is defined as the maximum distance D between the empirical and asymptotic plots of the scaling functions, which is similar to the Kolmogorov-Smirnov (KS) statistic (Press et al., 1992):

$$D = \max |\tau(q) - \hat{\tau}(q)|. \tag{4.9}$$

[3]A Monte Carlo procedure is used since the distribution of the statistic of the test is unknown.

Figure 4.3: Plots of log-normal distributions, $n = 10000$



(a) LN(0,1)

(b) LN(1,1)

(c) LN(0,4)

Plots of the empirical scaling functions for samples are shown by solid lines together with the asymptotic plot (dot-dashed line).

The p-value of the test is defined as the fraction of the distances of comparable synthetic data sets that are larger than the empirical distance. The larger p-value, the more plausibly the distribution fits to the data.

In detail, the procedure is as follows.

1. Fit the empirical data to a Pareto tail using the Hill estimator involving sequential statistical testing. The estimates of the lower bound $\hat{x}_{min}$ (where the tail starts) and the tail index $\hat{\alpha}$ are derived accordingly. Then calculate the corresponding empirical distance D_{Emp}.

2. Generate a number (e.g., 1000) of comparable synthetic data sets with a Pareto tail based on the estimates $\hat{\alpha}$ and $\hat{x}_{min}$ derived from step 1. Then repeat step 1 to each synthetic data set individually to calculate the synthetic distance D_{syn} of each data set with respect to its own fitted distribution.

3. Count the times that the synthetic distance D_{syn} is larger than the empirical one D_{emp}. This constructs the p-value of the test.

$$p = \frac{no.\ of\ \{D_{syn} > D_{emp}\}}{1000}. \tag{4.10}$$

There are two important issues to note. The first is how to obtain a relatively reliable estimator of the tail index α. The best known estimator of the tail index is the Hill estimator based on k upper order statistics. Since the Hill estimator is known to be sensitive to the choice of k, the so-called sequential testing method, a reasonably quick and reliable method introduced by Nguyen & Samorodnitsky (2012), is applied here to decide "where the tail begins".

Remark 4.1. *The reason for not applying the estimation method proposed in Chapter 2 is that the starting point of the tail is not estimated using this method. Without an estimate of the starting point, the synthetic data cannot be generated. However, as shown by the results reported in Section 4.3 below, the estimates obtained from the RA are more informative than those of the Hill estimation with sequential testing.*

This sequential testing method is based on the idea that upper order statistics in a tail, which is regularly varying at infinity, "behave like points of a Poisson random measure with a power intensity" (Nguyen & Samorodnitsky, 2012). Suppose $X_{1,n} \leq X_{2,n} \leq \ldots X_{n,n}$ are the order statistics of a positive sample $X_1, \ldots, X_n$. By sequentially testing the quantities

$$\{\ln \frac{X_{n-i,n}}{X_{n-k,n}} : i = 0, 1 \ldots k-1\},$$

with k increasing, for the null hypothesis of exponentiality, the number of upper order statistics used in the Hill estimation, namely N_{n^*}, is the smallest k rejecting exponentiality in the test. The statistic of the test is the moment statistic

$$Q_{k,n} = \frac{\sqrt{k}}{2} \left\{ \frac{\frac{1}{k}\sum_{i=0}^{k-1} (\ln \frac{X_{n-i,n}}{X_{n-k,n}})^2}{(\frac{1}{k}\sum_{i=0}^{k-1} \ln \frac{X_{n-i,n}}{X_{n-k,n}})^2} - 2 \right\}.$$

$Q_{k,n}$ converges to the standard normal distribution under exponentiality (Dahiya & Gurland, 1972). Then N_n^* is given by

$$N_n^* := \inf\{k : 1 \leq k \leq n, |Q_{k,n}| \geq \omega \sqrt{\frac{\theta_n}{k}}\},$$

where $\omega > 0$ is decided by the chosen quantile of the standard normal distribution and $\theta_n = (\ln n)^2$ which makes the critical value of the test increase with sample size n in order to take account of more order statistics. Then the Hill estimator based on N_n^* upper order statistics is consistent, i.e.,

$$\hat{\alpha}^{Hill}_{N_{n^*},n} = \left(\frac{1}{N_n^*} \sum_{i=0}^{N_n^*-1} \ln \frac{X_{n-i,n}}{X_{n-N_n^*,n}} \right)^{-1} \xrightarrow{p} \alpha.$$

The second important issue is how to generate a synthetic data set based on empirical data. In order to get an accurate estimator of the p-value, the synthetic data set should be comparable to the empirical one. This means that the synthetic data set should have the same non-Pareto distribution below $\hat{x}_{min} = X_{N_n^*}$ and follow the same Pareto tail above $\hat{x}_{min}$. Suppose n_{tail}

observations in the empirical data set with size n belong to the tail region, i.e., above $\hat{x}_{min}$. When generating a synthetic data set, with probability n_{tail}/n, a random number is drawn from the fitted Pareto tail with parameters $\hat{\alpha}$ and $\hat{x}_{min}$. With probability $1 - n_{tail}/n$, one random element is drawn uniformly from observations below $\hat{x}_{min}$ in the empirical data set. This process is repeated until all n observations in the synthetic data set are generated. This approach to generate a complete synthetic data set is similar to that applied in Clauset et al. (2009).

Finally, a decision rule about when to reject the null hypothesis is needed. In this chapter, the null hypothesis of the power law is rejected if $p \leq 0.05$.

4.3 Empirical results on the distribution of city sizes

In this section, the RA and GA are applied to different data sets of city sizes. Data descriptions are displayed first and the results are reported subsequently.

4.3.1 Data

Different definitions of cities have been proposed in the literature. One definition uses the concept of the MSA, which contains entities with a population of at least 50,000 in the census. Only the largest cities are included, therefore data based on the MSAs include just the upper tail of the distribution. Data of this type will not be considered in this chapter, despite its popularity in the literature, because the main advantage of the RA and GA is to use the entire data set to make inference for the tail and not only the extreme values. Two alternative definitions are used. One is the so-called "places" which refer to administrative or legal bounded units. The other constructs "cities" from micro data. Data based on these two definitions cover the entire range of sizes with population size ranging from 1 to several million. Details of the data sets are provided in the following.

US Census 2000 places Places are Census Designated Places (CDP), consolidated cities, or incorporated places under their respective states' laws as defined by the US Census Bureau[4]. Since Census 2000, CDPs are included without applying any minimum population restriction. Therefore, the data on all populated "places" are included in US Census 2000. The data set covers 25,358 places (cities, towns, and villages) ranging in size from 1 to 8,008,278.

Places derived from US Census 2010 Bee et al. (2013) generated a data set of "places" based on disaggregated data from the US census 2010. This data set contains 28,916 observations aggregated from 6,127,259 blocks in Census 2010.

Area clusters Rozenfeld et al. (2011) introduce a "bottom-up" approach to define a city based on a geographical criteria, rather than arbitrary "legal" boundaries. A City Clustering Algorithm (CCA) was introduced to define a "city" as an agglomeration of maximally connected populated cells with high spatial resolution. An agglomeration of population "is made of contiguous populated sites within a prescribed distance that cannot be expanded: all sites immediately outside the cluster have a population density below a cutoff threshold". Cities with different degrees of aggregation can be obtained by setting various "coarse-graining" level, denoted l.

Two area-cluster data sets are used in this section. Both of them are used as benchmarks by Rozenfeld et al. (2011). One is the population of CCA clusters in GB with 10,157 observations for $l = 1km$. The raw data for GB consist of 5.75 million cells with a total population of around 55 million in 1991 obtained from the Economic and Social Research Council

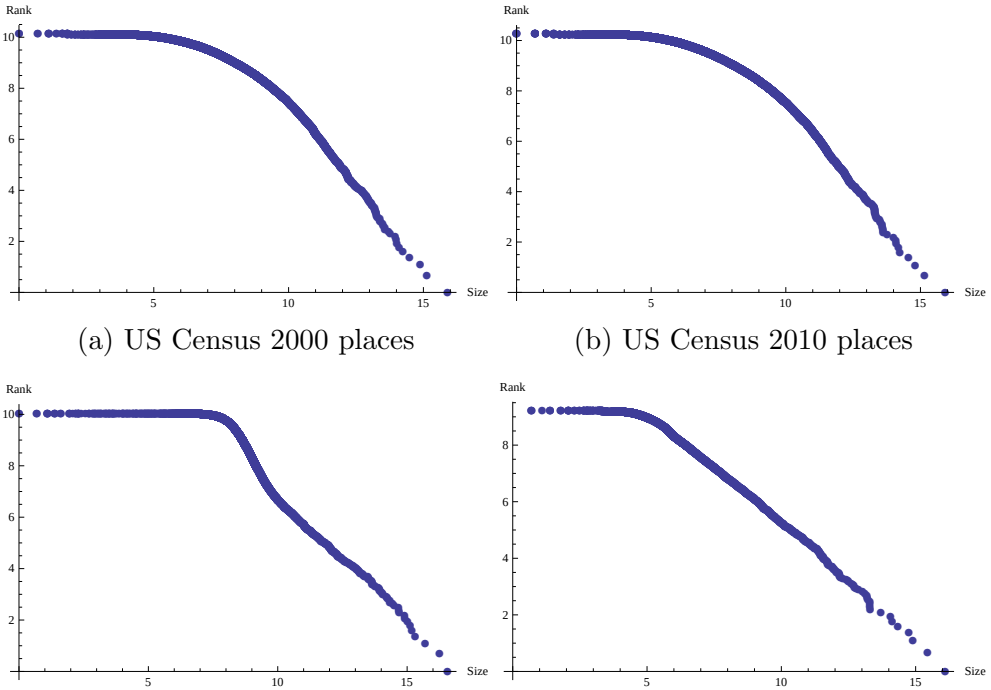
[4]Incorporated places may have legal descriptions of "city, town (except in the New England states, New York, and Wisconsin), borough (except in Alaska and New York), or village and having legally prescribed limits, powers, and functions". A CDP is a "statistical entity that serves as a statistical counterpart of an incorporated place for the purpose of representing census data for a concentration of population, housing, and commercial structures that is identifiable by name, but is not with in an incorporated place". A consolidated city refers to the primary incorporated place governed by a consolidated government. Source: http://www.census.gov/prod/cen2000/doc/sf1.pdf

(ESRC). The second data set is from the US with 23,499 observations for $l = 3km$. The raw data for the US are composed of 61,224 points, defined by Federal Information Processing Standards (FIPS), with population size varying between 1500 to 8000 in 2000.[5]

4.3.2 Empirical results

The plots of the population (ln scale) against the rank (ln scale) for each data set are shown in Figure 4.4. When the population is above some threshold, all plots are approximately linear. It seems reasonable to suspect that the distributions follow a Pareto tail, with index α given by the absolute slope of the straight line.

Figure 4.4: Distributions of city sizes



(a) US Census 2000 places

(b) US Census 2010 places

(c) Area clusters of the US 2000

(d) Area clusters of GB 1991

[5]Data on area clusters are available at
http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.5.2205

Results from the regression approach

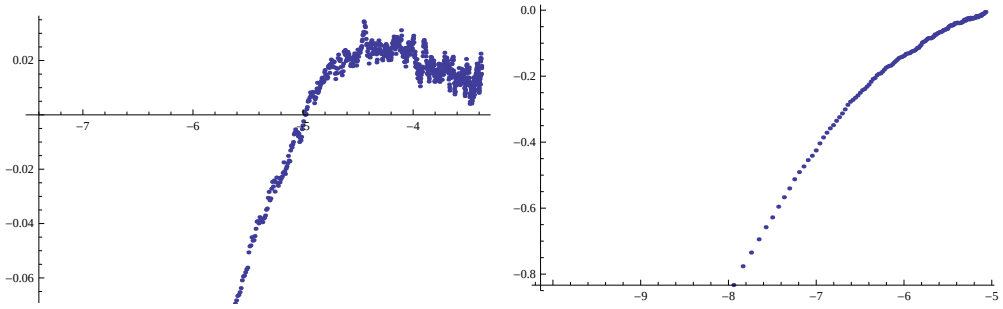
First, the RA is applied to test whether the distribution of city sizes follows a Pareto law or not. To choose point t_j in the regression, the rule

$$t_j = j/n,\ \ j = 1, 2, \ldots, m = n^{\delta},\ \ where\ 0 < \delta < 1, \tag{4.11}$$

is applied. We ran regressions with $\delta = 2/3, 1/2, 1/3, 1/4$ in each case. The plots of the regression points, the estimates of α, and the numbers of points, denoted by N_{t_j}, in the regression for different δ's and data sets are shown in Figures 4.5 to 4.8.

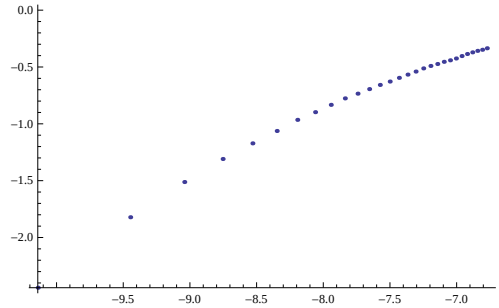
Remark 4.2. *The entire sample is used to calculate each point in the regression and N_{t_j} is not directly linked to the sample size. This is the key difference between the RA and the tail estimation based on extreme order statistics.*

Figure 4.5: Regression results for US Census 2000 places

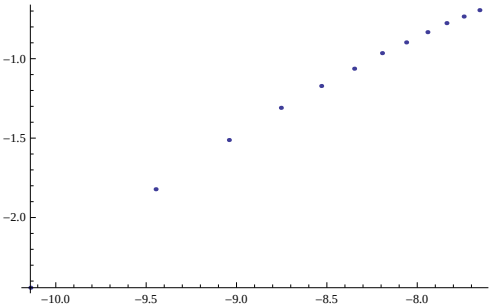


(a) $\hat{\alpha} = 0.13$, $\delta = 2/3$, $N_{t_j} = 863$

(b) $\hat{\alpha} = 0.34$, $\delta = 1/2$, $N_{t_j} = 159$

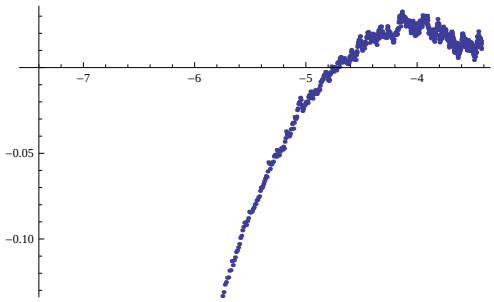


(c) $\hat{\alpha} = 0.57$, $\delta = 1/3$, $N_{t_j} = 29$

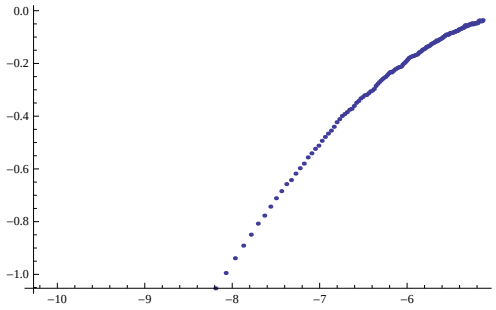


(d) $\hat{\alpha} = 0.69$, $\delta = 1/4$, $N_{t_j} = 12$

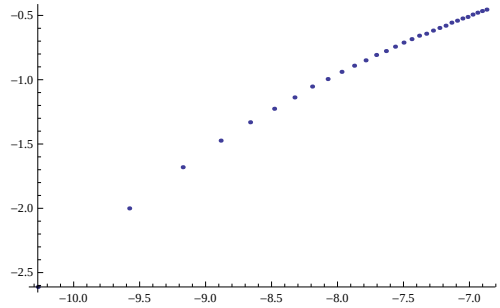
Figure 4.6: Regression results for places derived from US Census 2010



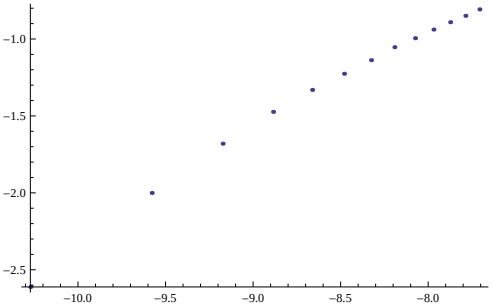
(a) $\hat{\alpha} = 0.16$, $\delta = 2/3$, $N_{t_j} = 942$



(b) $\hat{\alpha} = 0.37$, $\delta = 1/2$, $N_{t_j} = 170$

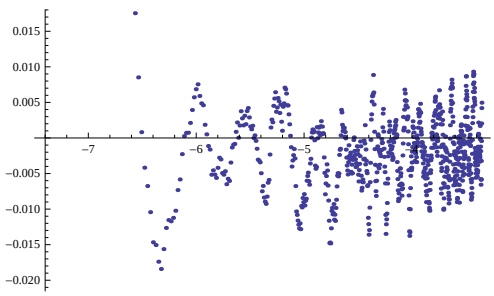


(c) $\hat{\alpha} = 0.58$, $\delta = 1/3$, $N_{t_j} = 30$

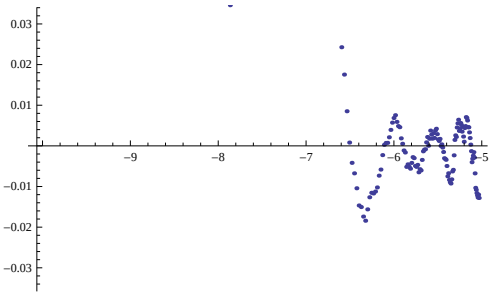


(d) $\hat{\alpha} = 0.68$, $\delta = 1/4$, $N_{t_j} = 13$

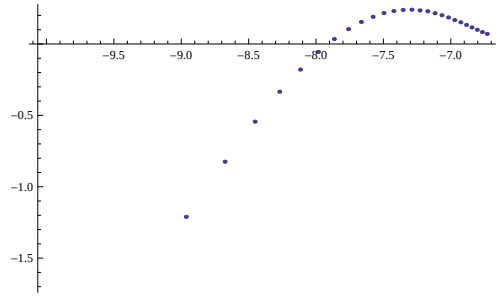
Figure 4.7: Regression results for area clusters of the US 2000



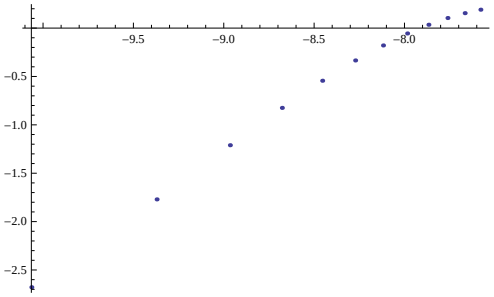
(a) $\hat{\alpha} = 0.04$, $\delta = 2/3$, $N_{t_j} = 820$



(b) $\hat{\alpha} = 0.15$, $\delta = 1/2$, $N_{t_j} = 153$

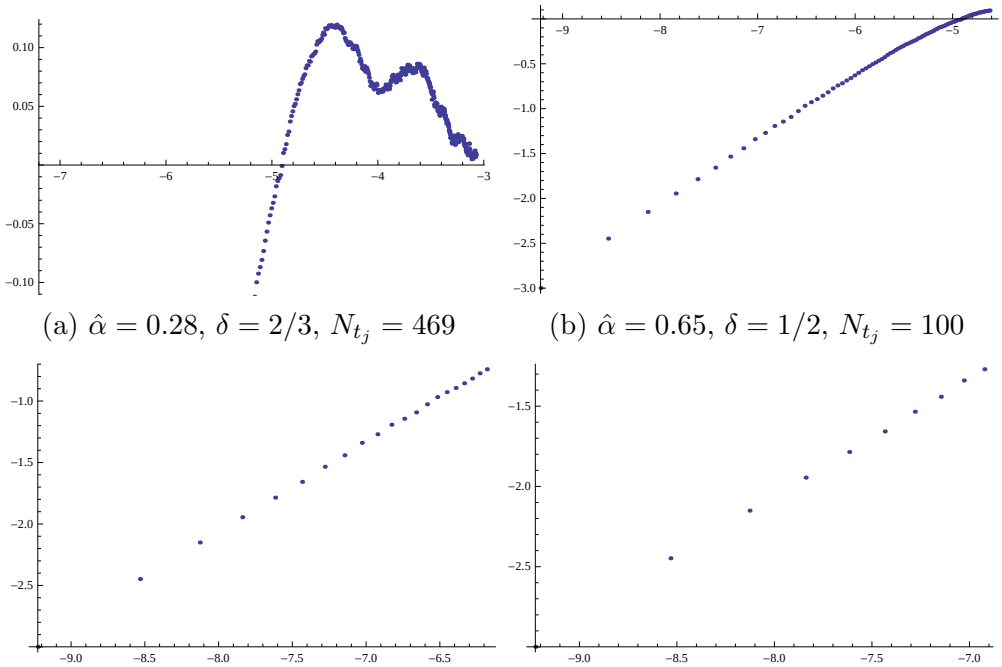


(c) $\hat{\alpha} = 0.74$, $\delta = 1/3$, $N_{t_j} = 28$



(d) $\hat{\alpha} = 1.17$, $\delta = 1/4$, $N_{t_j} = 12$

Figure 4.8: Regression results for area clusters of GB 1991



(a) $\hat{\alpha} = 0.28$, $\delta = 2/3$, $N_{t_j} = 469$

(b) $\hat{\alpha} = 0.65$, $\delta = 1/2$, $N_{t_j} = 100$

(c) $\hat{\alpha} = 0.74$, $\delta = 1/3$, $N_{t_j} = 21$

(d) $\hat{\alpha} = 0.75$, $\delta = 1/4$, $N_{t_j} = 10$

Note that each t_j in the regression is calculated using the entire sample. N_{t_j} reports the number of points in each regression.

In Figures 4.5 to 4.8, as δ decreases, t_j is closer to zero. This means that the regression reflects more accurately the behaviour of the distribution function at infinity. When applying the RA, we can check the plot of the regression points to obtain an accurate estimate. In order to illustrate how the RA works in practice, we use the data on US census 2000 places as an example. When $\delta = 2/3$, the plot in Figure 4.5a is not on a line but can be roughly divided into two pieces. The left piece is approximately a straight line, which indicates that point t in the regression (4.3) is not close enough to zero. Note that relation (4.3) holds only when t is quite near the origin. Hence, some points should be excluded in the regression which means that δ should decrease. As δ decreases, the plot of the regression function is closer to linearity as shown in Figures 4.5b to 4.5d, suggesting that the estimate

obtained is more reliable.

Similar patterns are found in the cases of US census 2010 and GB area clusters as shown in Figures 4.6 and 4.8, respectively. However, the tail behaviour is not obvious for area clusters of the US 2000 as shown in Figure 4.7. Indeed, the plots have almost no regularity for all but very small δ (1/4 or less). The reason might be that too few points fall in the Pareto tail region in this case. Figure 4.4c supports this explanation: the points in the right part of the plot do not lie on the line as presented in Figures 4.4a, 4.4b, and 4.4d but fluctuate around the fitted straight line.

In short, all the plots of the regression points with $\delta = 1/4$ are approximately a straight line. The estimates of the tail index are not far away from one as expected by Zipf's law for small δ. According to the RA, the distribution of city sizes has a Pareto tail regardless of the country and time factors. It also means that consistent regularities are exhibited for data sets based on different definitions of "cities".

Results from the graphical approach

Since a power transformation to the original data is needed for this approach, estimates of the parameters needed for the transformation of these data sets are provided first. Table 4.2 reports: 1. the estimated parameters if the underlying distribution is log-normal; 2. the sequential testing Hill estimate of the tail index and the estimated number of observations in the tail if the underlying distribution has a Pareto tail. Because the GA needs the power-transformed tail index to be less than two, only the two data sets on area clusters are used.

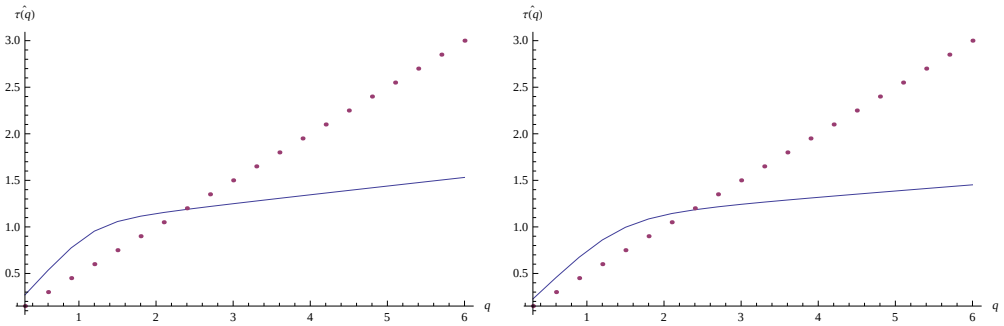
Table 4.2: Estimates for each data set

Data set	Places: US Census		Area Clusters	
	year 2000	year 2010	US year 2000	GB year 1991
$LN(\hat{\mu}, \hat{\sigma})$	(7.28,1.75)	(7.10,1.82)	(8.43,0.91)	(6.04,1.44)
$\hat{\alpha}$	1.45	1.39	0.87	0.81
$\hat{n}_{tail}$	616	781	206	164

The plots of the empirical scaling functions of the two

power-transformed data sets are given in Figure 4.9. For the Monte Carlo procedure, the number of synthetic data sets is set to 1000. Synthetic data are generated with the method described in Section 4.2.2 if the hypothesised distribution has a Pareto tail. If the hypothesised distribution is a log-normal distribution, synthetic data are randomly drawn from the estimated log-normal distribution. The estimates of the parameters of the log-normal distribution are obtained by maximum likelihood estimation.

Figure 4.9: Plots of empirical scaling functions



(a) Area clusters of the US 2000 (b) Area clusters of GB 1991

The empirical scaling function and the baseline $q/2$ is drawn by a solid line and a dotted line, respectively.

For the data set on area clusters of the US in 2000, the null hypothesise of a Pareto tail is rejected at the significance level of 0.05: $D_{Emp} = 0.53$ and the p-value is equal to 0.04. Then testing the null hypothesis of a log-normal distribution gives $D_{Emp} = 1.47$ and the p-value is equal to 0.14. For the data set on area clusters of GB in 1991, the null hypothesise of a Pareto tail cannot be rejected: $D_{Emp} = 0.45$ and the p-value is equal to 0.24.

According to the GA, the distribution of city sizes for the data set on area clusters of the US in 2000 is unlikely to be drawn from a distribution with a Pareto tail. It is more likely drawn from a log-normal distribution. The different conclusion of the GA and RA may be the result of having too few points included in the tail region in the GA. Although the two plots in Figure 4.9 are very similar, according to sequential testing, the fraction of observations in the tail region, i.e., n_{tail}/n, for GB is double that for the US.

The relatively small proportion of sample in the tail for the US clusters is also reflected in Figure 4.7 as discussed in the analysis of the RA results. In this sense, the conclusions from these two approaches are consistent. Anyhow, since the theoretical plot of the scaling function holds only asymptotically, the GA is expected to work better for large samples. Hence, larger empirical data sets should be found to further investigate this approach.

4.4 Applications to data from other fields

In this section, the RA is applied to other six empirical data sets in which power laws are suspected.[6] These data sets are drawn from a variety of fields, including the social sciences, ecology, computer and information sciences, and earth sciences. The following data sets are included in the analysis[7]:

a Word frequency: the frequency of word occurrence in the text of Herman Melville's novel *Moby Dick* (Newman, 2005).

b Severe terrorist event frequency: the severity of worldwide terrorist attacks measured by the number of deaths from 1968 to 2006[8] (Clauset et al., 2007).

c Wild fire size: the sizes of all wild fires (in acres) occurring in the US from 1986 to 1992 inclusive[9] (Newman, 2005).

d Solar flare intensity: the peak gamma-ray intensity of solar flares measured as scintillation counts per second from 1980 to 1989

[6]Only the RA is applied to detect the power law and obtain an estimator of the tail index if the power-law assumption is not ruled out. The reason is that the GA is mainly used as an exploratory method to detect the power law and is based on an asymptotic result which needs more strict conditions on empirical data.

[7]All data sets used in this section, together with detailed description, are available from the website http://tuvalu.santafe.edu/ aaronc/powerlaws/data.htm

[8]Data source: the National Memorial Institute for the Prevention of Terrorism (MIPT; 2006) database

[9]National Fire Occurrence Database, United States Department of Agriculture (USDA) Forest Service and Department of the Interior

inclusive[10] (Newman, 2005).

e Earthquake intensity: the maximum amplitude of motion, transformed from the Richter magnitude, occurring in California between 1910 and 1992[11] (Newman, 2005).

f Web link: the number of links to web sites on a Web crawl containing approximately 200 million pages in 1997 (Broder et al., 2000).

All data sets used in this section are analysed in Clauset et al. (2009). They apply the maximum likelihood (ML) estimation combined with the Kolmogorov-Smirnov (KS) test to identify power-law distributions. The ML estimator of the tail index is given by

$$\hat{\alpha} = n \left[\sum_{i}^{n} \ln \frac{x_i}{x_{min}} \right]^{-1},$$

where x_{min} is the smallest value above which the power law holds and $x_i,\ i = 1, 2, \ldots, n$, are the observed values of random variable X such that $x_i \geq x_{min}$. This estimator is equivalent to the Hill estimator. However, the estimator of the lower bound x_{min} is derived by the KS test rather than the sequential testing method. In Table 4.3, the estimates obtained by the ML estimation combined with the KS test (denoted by ML-KS) and the sequential testing method (denoted by ST) are reported respectively, along with descriptive statistics including mean, standard deviation, and maximum. Figure 4.10 shows the log-log plots of these six data sets.

[10]Data source: the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center, http://umbra.nascom.nasa.gov/index.html

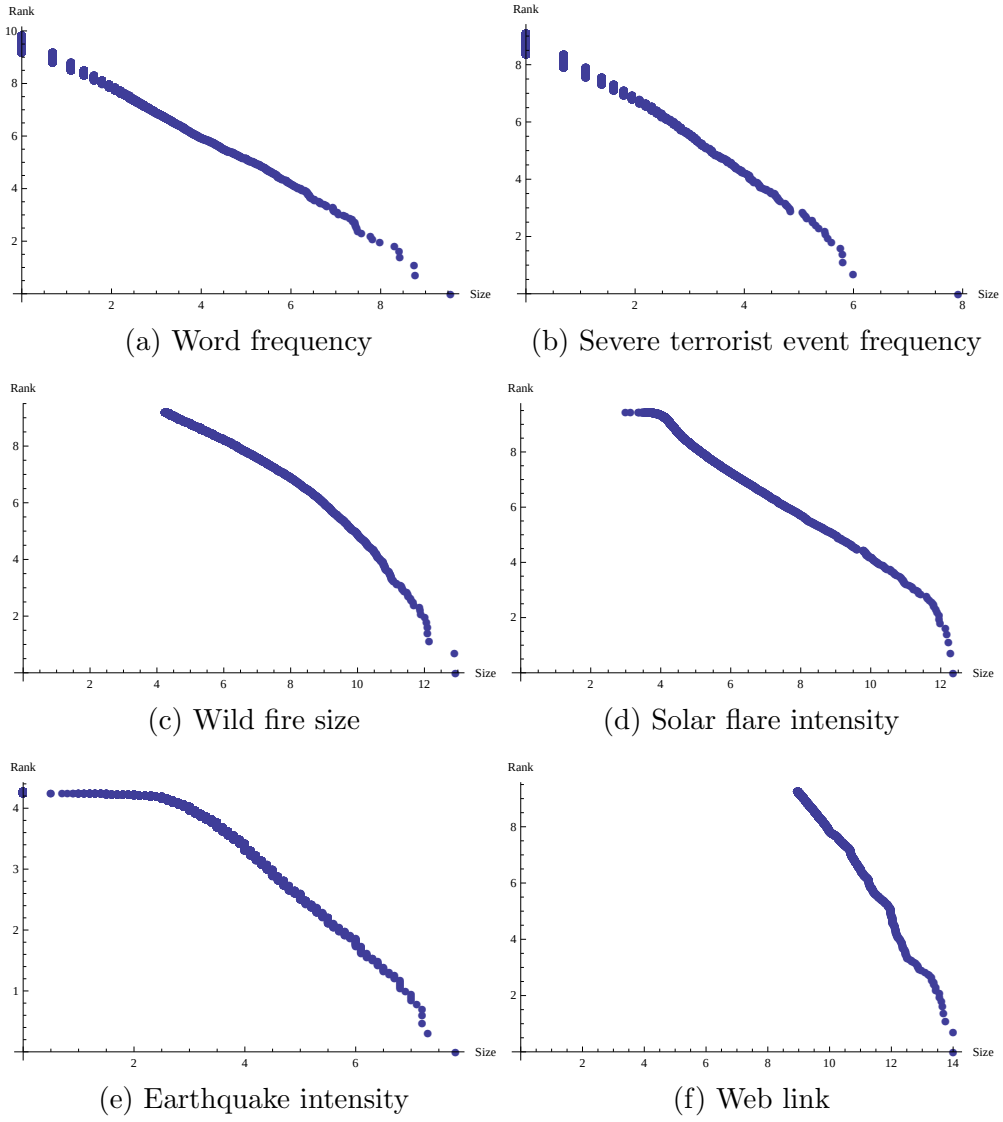
[11]Data source: the National Geophysical Data Center, http://www.ngdc.noaa.gov

Table 4.3: Description and estimated results of the six data sets

Data set	n	μ	σ	x_{max}	$\hat{\alpha}_{ML-KS}$	$n_{tail(ML-KS)}$	$\hat{\alpha}_{ST}$	$n_{tail(ST)}$
a: Word frequency	18855	11.14	148.33	14086	0.95	2958	1.00	154
b: Severe terrorist event frequency	9101	4.35	31.58	2749	1.4	547	1.31	306
c: Wild fire size	203785	89.56	2098.73	412050	1.2	521	1.34	156
d: Solar flare intensity	12773	689.41	6520.59	231300	0.79	1711	0.95	86
e: Earthquake intensity	19302	24537.20	563831	63095700	0.64	11697	0.79	95
f: Web link	241428853	9.15	106871.65	1199466	1.336	28986	2.58	136

Note: The estimates derived by the ML-KS method is the same as those displayed in Clauset et al. (2009), since it is not necessary to redo all the calculations they have done for the same data sets.

Figure 4.10: Log-log plots of the six data sets reputed to follow power laws



(a) Word frequency

(b) Severe terrorist event frequency

(c) Wild fire size

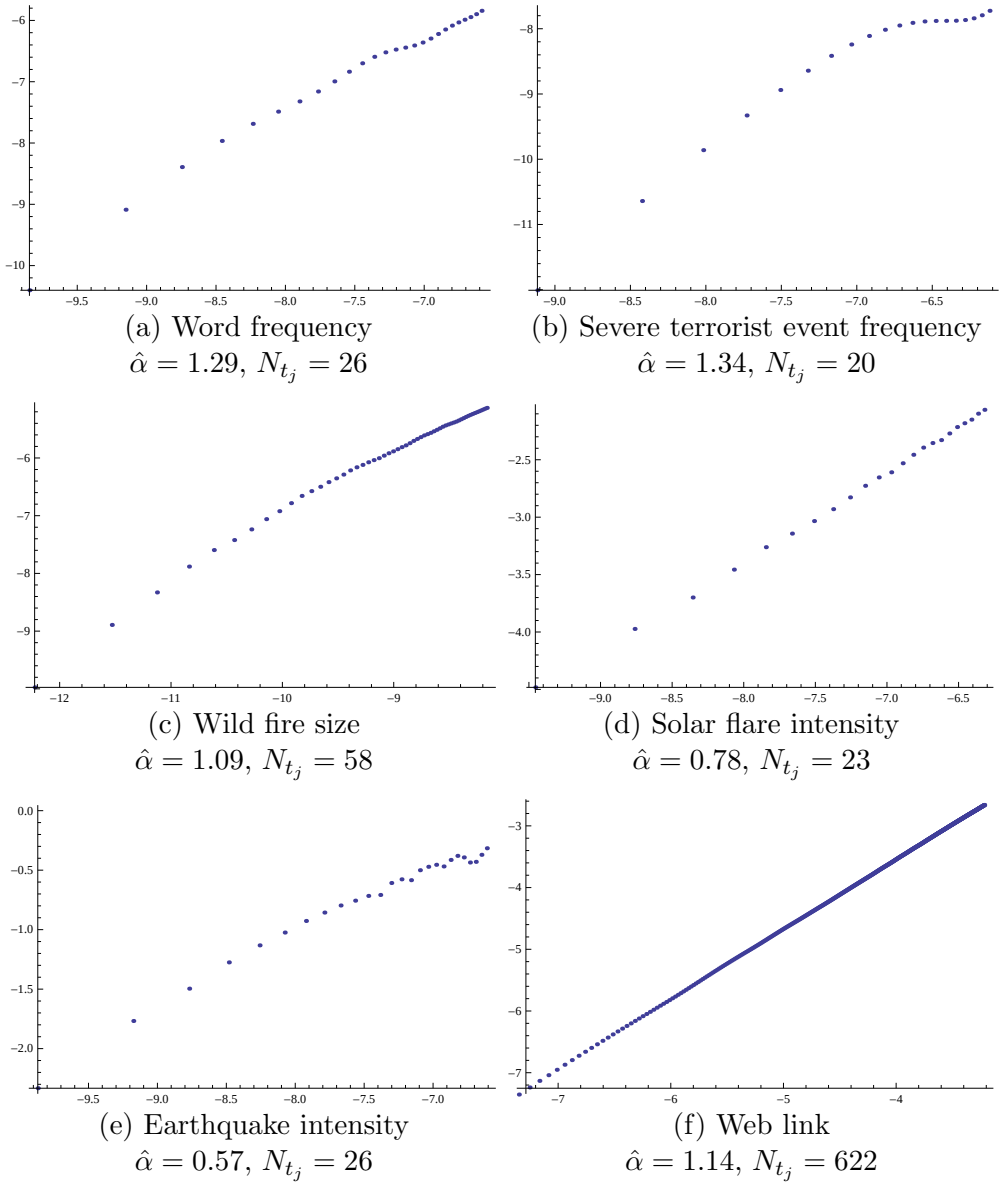
(d) Solar flare intensity

(e) Earthquake intensity

(f) Web link

Note: For wild fire and web link data (Figures 4.10c and 4.10f, respectively), only the upper tails are plotted because of the large size of the datasets. For the earthquake data set, the base of the logarithm is 10, because the Richter magnitude is proportional to the logarithm with base 10 of the maximum amplitude of motion detected in the earthquake. Otherwise, base e is used.

Figure 4.11: Regression results on six quantities reputed to follow power laws



(a) Word frequency
$\hat{\alpha} = 1.29$, $N_{t_j} = 26$

(b) Severe terrorist event frequency
$\hat{\alpha} = 1.34$, $N_{t_j} = 20$

(c) Wild fire size
$\hat{\alpha} = 1.09$, $N_{t_j} = 58$

(d) Solar flare intensity
$\hat{\alpha} = 0.78$, $N_{t_j} = 23$

(e) Earthquake intensity
$\hat{\alpha} = 0.57$, $N_{t_j} = 26$

(f) Web link
$\hat{\alpha} = 1.14$, $N_{t_j} = 622$

Note that each t_j in the regression is calculated using the entire sample. N_{t_j} reports the number of points in each regression.

As shown in Table 4.3, some estimates obtained by the ST method differ to some degree from those by the ML-KS method, despite the fact that

the only difference between these two methods is the determination of the tail. For instance, the ST estimate for the web link case is greater than 2, which indicates the variance is finite. By contrast, the ML-KS estimate is around 1.3, which means the tail decreases slowly and only the mean exists. Indeed, the numbers of observations in the tail region derived by these two methods differ substantially, especially for large sample sizes. Even when the estimates of α are similar, the tail fractions differ a lot. For example, in the case of earthquake intensity, both estimates of α are around 0.7, however, the number of observations in the tail for the ML-KS method is almost 120 times that of the ST method.

Figure 4.11 shows the RA results. Again, each point in the regression is calculated using the entire sample. With the exception of the data on web links, all the points near the origin in the regression are selected according to (4.11) with $\delta = 1/3$. In the special case of web link,

$$t_j = j/\sqrt{n},\ j = 1, 2, \ldots, m = n^{1/3}$$

is applied. The reason is that the large sample size will result in high volatility in the estimation. More precisely, t approaches zero as sample size n increases. However, if t is extremely small, the variance of the error term in the regression will not converge and the estimator is not robust. The sample size for web links is over 200 million and the plot in Figure 4.11f is nearly straight for $t_j = j/\sqrt{n}$. There is no need to sacrifice accuracy to get a straighter line in the regression (more details see Section 2.3.2).

As shown in Figure 4.11, all the plots are nearly straight. Estimates of α from the RA are shown in Figures 4.11a to 4.11e. Note that, similar conclusions to those by the ML-KS and ST methods are obtained by the RA even without having to answer the question "where the tail starts". For web links, compared with the ST method, the result obtained by the RA is more similar to that from the ML-KS method. In short, combining the plot of the points in the regression and the corresponding estimate obtained, the RA is more informative on the tail behaviour.

4.5 Conclusion

Power laws are found in many fields, such as economics, finance, biology, computer science, earth science and so on. However, there is a lack of robust methods to detect a power law. In this chapter, we question the widespread use of the log-log plot. Instead, we present two different approaches using the entire sample to identify a power law. The first approach (RA) exploits the fact that the behaviour of the characteristic function near the origin reflects the behaviour of the distribution function at infinity. The second approach (GA) is based on the asymptotic properties of the so-called partition function.

Both of these two approaches are applied to describe the distribution of city sizes where the debate about whether it fits a power-law or log-normal distribution is still active. The RA generally works well and suggests that the power law holds remarkably well for data on the entire range of all sizes for two different countries and different definitions of cities. By contrast, the results obtained from the GA are inconclusive. The drawback of the GA is exposed in the study of city sizes. The GA is based on asymptotic properties and thus requires large sample sizes. Moreover, in order to generate synthetic data and get a p-value of the test, the GA must apply other techniques to estimate the tail index and the minimum value above which the power law holds. This makes the results of the GA heavily reliant on the chosen estimation method. Further research could be done to develop and improve the GA. Finally the RA is applied to other six data sets from various fields. Generally, the power-law hypothesis cannot be rejected with given data sets.

4.A Appendix

Proofs of (4.7) *and* (4.8). Suppose $X_i \in (0, \infty)$ are i.i.d. random variables drawn from a log-normal distribution with location parameter μ and scaling parameter σ, denoted by $LN(\mu, \sigma^2)$. Then

$$E(X) = e^{\mu+\frac{1}{2}\sigma^2}, \quad E(X^2) = e^{2\mu+2\sigma^2}, \quad \textit{and} \quad Var(X) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}.$$

Moreover, since $X \sim LN(\mu, \sigma^2)$ and let $Y = \ln(x)$, then $Y \sim \mathcal{N}(\mu, \sigma^2)$. The moment generating function of a normal distribution is

$$M_Y(t) = E(e^{tY}) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}.$$

It is easy to obtain that

$$E(|X|^q) = E(X^q) = E(e^{qY}) = e^{\mu q + \frac{1}{2}\sigma^2 q^2},$$

for any real number q.

Let $Z = X - E(X)$. Thus

$$E(Z^2) = Var(X) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}.$$

For $q \geq 2$, since $|a+b|^q \leq 2^{q-1}(|a|^q + |b|^q)$ (this inequality holds for all $q > 1$),

$$\begin{aligned} E(|Z|^q) &= E(|X - E(X)|^q) \\ &\leq 2^{q-1}(E(|X|^q) + |E(X)|^q) \\ &= 2^{q-1}(e^{\mu q + \frac{1}{2}\sigma^2 q^2} + e^{\mu q + \frac{1}{2}\sigma^2 q}) \\ &\leq C_1 e^{\mu q + \frac{1}{2}\sigma^2 q^2}. \end{aligned}$$

Now using Rosenthal's inequality (Merlevède & Peligrad, 2013),

$$\begin{aligned} E\Big(\Big|\sum_{j=1}^{\lfloor n^s \rfloor} Z_j\Big|^q\Big) &\leq C(q)\Big(\sum_{j=1}^{\lfloor n^s \rfloor} E(|Z_j|^q) + \big(\sum_{j=1}^{\lfloor n^s \rfloor} E(Z_j^2)\big)^{q/2}\Big) \\ &\leq C(q)\Big(n^s C_1 e^{\mu q + \frac{1}{2}\sigma^2 q^2} + n^{\frac{sq}{2}}\big((e^{\sigma^2} - 1)e^{2\mu + \sigma^2}\big)^{q/2}\Big) \\ &\leq C(q)\big(n^s e^{\mu q + \frac{1}{2}\sigma^2 q^2} + n^{\frac{sq}{2}} e^{\mu q + \sigma^2 q}\big) \\ &\leq C(q)\big(\max\{n^s e^{\mu q + \frac{1}{2}\sigma^2 q^2}, n^{\frac{sq}{2}} e^{\mu q + \sigma^2 q}\}\big). \end{aligned} \tag{4.12}$$

Note: The terms with factors of *exp* are treated as constants since they are finite, i.e., the term on the left-hand side of (4.12) is less than or equal to $C_2 n^{\frac{sq}{2}}$ for a given q.

Notice that

$$n^{\frac{\ln S_q(n,n^s)}{\ln n}} = S_q(n,n^s) = \frac{1}{\lfloor n^{1-s} \rfloor} \sum_{i=1}^{\lfloor n^{1-s} \rfloor} \Big| \sum_{j=1}^{\lfloor n^s \rfloor} Z_{\lfloor n^s \rfloor (i-1)+j} \Big|^q.$$

For $\delta > 0$,

$$\begin{aligned} P\Big(\frac{\ln S_q(n,n^s)}{\ln n} > \frac{sq}{2} + \delta\Big) &= P\Big(S_q(n,n^s) > n^{\frac{sq}{2}+\delta}\Big) \\ &= P\Big(\frac{1}{\lfloor n^{1-s} \rfloor} \sum_{i=1}^{\lfloor n^{1-s} \rfloor} \Big| \sum_{j=1}^{\lfloor n^s \rfloor} Z_{\lfloor n^s \rfloor (i-1)+j} \Big|^q > n^{\frac{sq}{2}+\delta}\Big) \\ &\leq \frac{E\Big(\Big| \sum_{j=1}^{\lfloor n^s \rfloor} Z_{\lfloor n^s \rfloor (i-1)+j} \Big|^q\Big)}{n^{\frac{sq}{2}+\delta}} \to 0. \end{aligned}$$

It follows that

$$\operatorname*{plim}_{n\to\infty} \frac{\ln S_q(n,n^s)}{\ln n} \leq \frac{sq}{2}, \tag{4.13}$$

which establishes the upper bound for $q \geq 2$.

Now let's consider the case $q < 2$. Combining Jensen's Inequality and Rosenthal's Inequality,

$$\begin{aligned} E\Big(\Big| \sum_{j=1}^{\lfloor n^s \rfloor} Z_j \Big|^q\Big) &\leq \left(E\Big(\Big| \sum_{j=1}^{\lfloor n^s \rfloor} Z_j \Big|^2\Big)\right)^{\frac{q}{2}} \\ &\leq C_3 n^{\frac{sq}{2}} e^{\mu q + \sigma^2 q}. \end{aligned} \tag{4.14}$$

Again, because $e^{\mu q+\sigma^2 q}$ is treated as a constant, the same upper bound is obtained for the case $q < 2$.

To prove the lower bound, by the central limit theorem for large n, $P(|N(0,1|) > 1) > 1/4$, we obtain

$$P\Big(\Big| \sum_{j=1}^{\lfloor n^s \rfloor} Z_{n^s(i-1)+j} \Big| > n^{\frac{s}{2}} \sigma_Z\Big) > 1/4,$$

where $\sigma_Z^2 = \lim_{n\to\infty} \frac{E\left(\sum_{j=1}^n Z_j\right)^2}{n}$.

Then for $\epsilon > 0$,

$$\begin{aligned}
P\Big(\frac{\ln S_q(n,n^s)}{\ln n} < \frac{sq}{2} - \epsilon\Big) &= P\Big(S_q(n,n^s) < n^{\frac{sq}{2}-\epsilon}\Big) \\
&\le P\Big(\sum_{i=1}^{\lfloor n^{1-s}\rfloor} \Big|\sum_{j=1}^{\lfloor n^s\rfloor} Z_{n^s(i-1)+j}\Big|^q < n^{\frac{sq}{2}-\epsilon+1-s}\Big) \\
&\le P\Big(\sum_{i=1}^{\lfloor n^{1-s}\rfloor} \mathbf{1}\Big(\Big|\sum_{j=1}^{\lfloor n^s\rfloor} Z_{n^s(i-1)+j}\Big| > n^{\frac{s}{2}}\sigma_Z\Big) < \frac{n^{\frac{sq}{2}-\epsilon+1-s}}{n^{\frac{sq}{2}}\sigma_Z^q}\Big) \\
&= P\Big(\sum_{i=1}^{\lfloor n^{1-s}\rfloor} \mathbf{1}\Big(\Big|\sum_{j=1}^{\lfloor n^s\rfloor} Z_{n^s(i-1)+j}\Big| > n^{\frac{s}{2}}\sigma_Z\Big) < \frac{n^{1-s-\epsilon}}{\sigma_Z^q}\Big) \\
&\le P\Big(\mathcal{B}(\lfloor n^{1-s}\rfloor, 1/4) < \frac{n^{1-s-\epsilon}}{\sigma_Z^q}\Big) \to 0,
\end{aligned}$$

where $\mathcal{B}(\lfloor n^{1-s}\rfloor, 1/4)$ is the binomial distribution. Hence

$$\operatorname*{plim}_{n\to\infty} \frac{\ln S_q(n,n^s)}{\ln n} \ge \frac{sq}{2}. \tag{4.15}$$

Combining (4.13), (4.14), and (4.15), this completes the proof of (4.7).

When n is large, the (empirical) scaling function can be obtained as

$$\tau(q) = \frac{\int_0^1 s\frac{\ln S_q(n,n^s)}{\ln n}\,\mathrm{d}s - \int_0^1 s\,\mathrm{d}s\int_0^1 \frac{\ln S_q(n,n^s)}{\ln n}\,\mathrm{d}s}{\int_0^1 s^2\,\mathrm{d}s - (\int_0^1 s\,\mathrm{d}s)^2}.$$

As $n \to \infty$, according to (4.7),

$$\tau(q) = 12\int_0^1 \frac{s^2 q}{2}\,\mathrm{d}s - 6\int_0^1 \frac{sq}{2}\,\mathrm{d}s = \frac{q}{2}.$$

This completes the proof of (4.8). □

Chapter 5

Conclusion

This thesis suggests that it may be inefficient to use subsamples containing only extreme values for making inference of the tail index of a heavy-tailed distribution. Therefore, new approaches that are based on the entire sample to analyse the tail behaviour are proposed. A regression method and a graphical method are introduced individually. The former is a regression technique based on the characteristic function, whose behaviour near the origin reflects the behaviour of the distribution function at infinity. The latter utilises the scaling function based on the asymptotic properties of the partition function, a function usually employed in the analysis of multifractality.

These two methods were applied to detect power laws and to distinguish them from log-normal distributions in empirical data sets. In most cases, the regression method shows excellent performance. For instance, in the analysis of the city-size distribution, a power law is detected for all data sets that include the full range of city sizes. The plot of the regression points indicates whether the selected point t in the regression is close enough to the origin or not. This helps to solve the practical problem regarding the choice of appropriate values of δ. In this thesis, this method is developed for i.i.d. random samples; extending it beyond the i.i.d. variables would be an interesting further research direction. With regard to the graphical method, the results are inconclusive in some of the cases

studied. The main reason for the inconclusive results is that the graphical method relies on other estimation techniques for a robust test to detect a power law. Visual inspection and subjective judgement are not sufficient for a reliable conclusion. Therefore, the graphical method should be regarded as an exploratory tool, rather than a testing tool.

Finally, the drawback of the proposed methods should be stressed because as in other cases "every coin has two sides". The main advantage of these two methods is that they circumvent the problem of estimating the starting point of the tail of the underlying distribution. However, this turns into a disadvantage if one wants to separate the tail from the entire distribution. There is no single estimation method that outperforms the others in all situations.

Bibliography

Abramowitz, M., & Stegun, I. A. (1965). *Handbook of Mathematical Functions*, chap. Numerical Interpolation, Differentiation, and Integration, (p. 886). Dover Books on Mathematics. Dover, New York.

Bee, M., Riccaboni, M., & Schiavo, S. (2013). The size distribution of US cities: Not Pareto, even in the tail. *Economics Letters*, *120*(2), 232–237.

Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester.

Beirlant, J., Vynckier, P., & Teugels, J. L. (1996). Tail index estimation, Pareto quantile plots regression diagnostics. *Journal of the American Statistical Association*, *91*(436), 1659–1667.

Bibby, B. M., Skovgaard, I. M., & Sørensen, M. (2005). Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli*, *11*(2), 191–220.

Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (2007). *Discrete Multivariate Analysis: Theory and Practice*. Springer, New York.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., & Wiener, J. (2000). Graph structure in the web. *Computer Networks*, *33*(1), 309–320.

Chan, G., Hall, P., & Poskitt, D. S. (1995). Periodogram-based estimators of fractal properties. *The Annals of Statistics*, *23*(5), 1684–1711.

Chechkina, A. V., & Gonchara, V. Y. (2000). Self and spurious multi-affinity of ordinary Levy motion, and pseudo-Gaussian relations. *Chaos, Solitons & Fractals*, *11*(14), 2379–2390.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703.

Clauset, A., Young, M., & Gleditsch, K. S. (2007). On the frequency of severe terrorist events. *Journal of Conflict Resolution*, *51*(1), 58–87.

Cooray, K., & Ananda, M. M. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal*, *2005*(5), 321–334.

Dahiya, R. C., & Gurland, J. (1972). Goodness of fit tests for the gamma and exponential distributions. *Technometrics*, *14*(3), 791–801.

Danielsson, J., Jansen, D. W., & De vries, C. G. (1996). The method of moments ratio estimator for the tail shape parameter. *Communications in Statistics-Theory and Methods*, *25*(4), 711–720.

de Haan, L., & Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer, New York.

de Haan, L., & Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica*, *52*(1), 60–70.

de Haan, L., & Resnick, S. (1998). On asymptotic normality of the Hill estimator. *Communications in Statistics. Stochastic Models*, *14*(4), 849–866.

Dekkers, A. L. M., Einmahl, J. H. J., & de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, *17*(4), 1833–1855.

del Castillo, J., & Puig, P. (1999). The best test of exponentiality against singly truncated normal alternatives. *Journal of the American Statistical Association*, *94*(446), 529–532.

Donatos, G. S., & Meintanis, S. G. (1996). On robustness and efficiency of certain statistics involving the empirical characteristic function. *Journal of the Italian statistical society*, *5*(1), 149–161.

Doob, J. L. (1942). The Brownian movement and stochastic equations. *Annals of Mathematics*, *43*(2), 351–369.

DuMouchel, W. H. (1983). Estimating the stable index α in order to measure tail thickness: a critique. *The Annals of Statistics*, *11*(4), 1019–1031.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *The American Economic Review*, *94*(5), 1429–1451.

Eeckhout, J. (2009). Gibrat's law for (all) cities: reply. *The American Economic Review*, *99*(4), 1676–1683.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events*. Springer, Berlin.

Falk, M. (1994). Efficiency of convex combinations of Pickands estimator of the extreme value index. *Journal of Nonparametric Statistics*, *4*(2), 133–147.

Feller, W. (1967). On regular variation and local limit theorems. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2: Contributions to Probability Theory, Part 1, (pp. 373–388). The Regents of the University of California.

Fofack, H., & Nolan, J. P. (1999). Tail behavior, modes and other characteristics of stable distributions. *Extremes*, *2*(1), 39–58.

Frisch, U. (1980). Fully developed turbulence and intermittency. *Annals of the New York Academy of Sciences*, *357*(1), 359–367.

Gabaix, X. (1999). Zipf's law for cities: an explanation. *The Quarterly Journal of Economics*, *114*(3), 739–767.

Gabaix, X., & Ioannides, Y. M. (2004). The evolution of city size distributions. *Handbook of Regional and Urban Economics*, *4*, 2341–2378.

Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, *4*(4), 221–238.

Gibrat, R. (1931). *Les inégalités économiques*. Librarie du Recueil Sirey, Paris.

Grahovac, D., Jia, M., Leonenko, N. N., & Taufer, E. (2014). Asymptotic properties of the partition function and applications in tail index inference of heavy-tailed data. *Statistics: A Journal of Theoretical and Applied Statistics*. DOI: 10.1080/02331888.2014.969267.

Haeusler, E., & Teugels, J. L. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *The Annals of Statistics*, *13*(2), 743–756.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society. Series B (Methodological)*, *44*(1), 37–42.

Hall, P., & Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, *13*(1), 331–341.

Heyde, C. C. (2009). Scaling issues for risky asset modelling. *Mathematical Methods of Operations Research*, *69*(3), 593–603.

Heyde, C. C., & Kou, S. G. (2004). On the controversy over tailweight of distributions. *Operations Research Letters*, *32*(5), 399–408.

Heyde, C. C., & Leonenko, N. N. (2005). Student processes. *Advances in Applied Probability*, *37*(2), 342–365.

Heyde, C. C., & Liu, S. (2001). Empirical realities for a minimal description risky asset model. The need for fractal features. *Journal of the Korean Mathematical Society*, *38*(5), 1047–1059.

Heyde, C. C., & Sly, A. (2008). A Cautionary note on modeling with fractional Lévy flights. *Physica A: Statistical Mechanics and its Applications*, *387*(21), 5024–5032.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annuals of Statistics*, *3*(5), 1163–1174.

Hill, J. B. (2010). On tail index estimation for dependent, heterogeneous data. *Econometric Theory*, *26*(05), 1398–1436.

Hsing, T. (1991). On tail index estimation using dependent data. *The Annals of Statistics*, *19*(3), 1547–1569.

Hurst, S. R., & Platen, E. (1997). The marginal distributions of returns and volatility. *Lecture Notes-Monograph Series*, *31*, 301–314.

Iacus, S. M. (2008). *Simulation and Inference for Stochastic Differential Equations: With R Examples*, vol. 1 of *Springer Series in Statistics*. Springer, New York.

Ioannides, Y., & Skouras, S. (2013). US city size distribution: Robustly Pareto, but only in the tail. *Journal of Urban Economics*, *73*(1), 18–29.

Janicki, A., & Weron, A. (1994). *Simulation and Chaotic Behavior of Alpha-stable Stochastic Processes*, vol. 178 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker, New York.

Kapteyn, J. C. (1903). *Skew Frequency Curves in Biology and Statistics*. Astronomical Laboratory, Groningen: Noordhoff.

Karr, A. F. (1993). *Probability*. Springer Texts in Statistics. Springer, New York.

Koutrouvelis, I. A. (1980). Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, *75*(372), 918–928.

Kratz, M., & Resnick, S. I. (1996). The qq-estimator and heavy tails. *Communications in Statistics. Stochastic Models*, *12*(4), 699–724.

Leonenko, N. N., & Šuvak, N. (2010). Statistical inference for Student diffusion process. *Stochastic Analysis and Applications*, *28*(6), 972–1002.

Levy, M. (2009). Gibrat's law for (all) cities: comment. *The American Economic Review*, *99*(4), 1672–1675.

Malevergne, Y., Pisarenko, V., & Sornette, D. (2011). Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Physical Review E*, *83*(3), 036111.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394–419.

Mandelbrot, B. B., Fisher, A. J., & Calvet, L. E. (1997). A multifractal model of asset returns. Cowles foundation discussion papers no. 1164, Sauder School of Business Working Paper.

Mason, D. M. (1982). Laws of large numbers for sums of extreme values. *The Annals of Probability*, *10*(3), 754–764.

Masuda, H. (2004). On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. *Bernoulli*, *10*(1), 97–120.

McElroy, T., & Politis, D. N. (2007). Moment-based tail index estimation. *Journal of Statistical Planning and Inference*, *137*(4), 1389–1406.

McNeil, A. J. (1996). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, *27*(1), 117–138.

Meerschaert, M. M., & Scheffler, H. P. (1998). A simple robust estimator for the thickness of heavy tails. *Journal of Statistical Planning and Inference*, *71*(1), 19–34.

Meintanis, S., & Koutrouvelis, I. A. (1990). Adaptive estimation of the center of symmetry based on the empirical characteristic function. *Communication in Statistics-Theory and Methods*, *19*(1), 381–394.

Merlevède, F., & Peligrad, M. (2013). Rosenthal-type inequalities for the maximum of partial sums of stationary processes and examples. *The Annals of Probability*, *41*(2), 914–960.

Mina, J., & Xiao, J. Y. (2001). Return to Riskmetrics: the evolution of a standard. Tech. rep., RiskMetrics Group.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323–351.

Nguyen, T., & Samorodnitsky, G. (2012). Tail inference: where does the tail begin? *Extremes*, *15*(4), 437–461.

Pareto, V. (1896). *Cours d'économie politique: professé à l'Université de Lausanne*, vol. 1. F. Rouge.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, *3*(1), 119–131.

Pitman, E. J. G. (1968). On the behaviour of the characteristic function of a probability distribution in the neighbourhood of the origin. *Journal of the Australian Mathematical Society*, *8*(03), 423–443.

Politis, D. N. (2002). A new approach on estimation of the tail index. *Comptes Rendus Mathematique*, *335*(3), 279–282.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, New York.

Rachev, S. T. (Ed.) (2003). *Handbook of Heavy Tailed Distributions in Finance: Handbooks in Finance*, vol. 1. Elsevier, Amsterdam.

Resnick, S., & Stărică, C. (1995). Consistency of Hill's estimator for dependent data. *Journal of Applied Probability*, *32*(1), 139–167.

Resnick, S., & Stărică, C. (1997). Smoothing the Hill estimator. *Advances in Applied Probability*, *29*(1), 271–293.

Resnick, S. I. (2006). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering. Springer, New York.

Rozenfeld, H. D., Rybski, D., Gabaix, X., & Makse, H. A. (2011). The area and population of cities: New insights from a different perspective on cities. *The American Economic Review*, *101*(5), 2205–2225.

Segers, J. (2005). Generalized Pickands estimators for the extreme value index. *Journal of Statistical Planning and Inference*, *128*(2), 381–396.

Sly, A. (2005). Self-similarity, multifractionality and miltifractality. *MPhil Thesis, Australian National University*.

Smith, R. L. (1987). Estimating tails of probability distributions. *The Annals of Statistics*, *15*(3), 1174–1207.

Smyth, G. K. (2002). Nonlinear regression. *Encyclopedia of Environmetrics Vol. 3*, (pp. 1405–1411).

Taufer, E., & Leonenko, N. N. (2009). Simulation of Lévy-driven Ornstein-Uhlenbeck processes with given marginal distribution. *Computational Statistics & Data Analysis*, *53*(6), 2427–2437.

Wang, H., & Tsai, C. L. (2009). Tail index regression. *Journal of the American Statistical Association*, *104*(487), 1233–1240.

Welsh, A. H. (1986). On the use of the empirical distribution and characteristic function to estimate parameters of regular variation. *Australian Journal of Statistics*, *28*(2), 173–181.