UNIVERSITY OF CALIFORNIA
Santa Barbara

Parameter Estimation for Stable Distributions: Spacings-based and Indirect Inference

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Statistics and Applied Probability

by

Gaoyuan Tian

Committee in Charge:

Professor S.Rao Jammalamadaka, Chair

Professor John Hsu

Professor Tomoyuki Ichiba

March 2016

The Dissertation of
Gaoyuan Tian is approved:

Professor John Hsu

Professor Tomoyuki Ichiba

Professor S.Rao Jammalamadaka, Committee Chairperson

March 2016

Parameter Estimation for Stable Distributions: Spacings-based and Indirect Inference

by

Gaoyuan Tian

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisor, Professor S. Rao Jammalamadaka. Only his continuous guidance, support, and care made the research presented in this dissertation possible. I am very grateful to my committee members Professor Tomoyuki Ichiba and Professor John Hsu for their friendly advice and kindly help while serving on my committee. I thank Dr. Stéphane Guerrier, who as a visiting faculty member in our department, introduced me to indirect inference and made many helpful suggestions. During the final phase of my work on indirect inference for stable distributions, I discovered this overlapped with some of the work of Professor David Veredas. I am grateful to him for pointing this out and for his support and suggestions.

I would never have been able to finish my dissertation without the guidance of many faculty members, help from friends, as well as support from my family. I wish to thank Professors David Hinkley, Andrew Carter, Jean-Pierre Fouque, and Michael Ludkovski for their wonderful teaching and instruction during my studies here at UCSB. Also my time at UCSB has been made most enjoyable in large part due to the pleasure I had studying and working with many great graduate students, who have become my lifetime friends. I also want to thank them for their help and useful discussions in my study.

Finally, I would like to thank my wonderful parents Jun Tian and Haifeng Yuan for all their love and support during all these years.

Curriculum Vitæ

Gaoyuan Tian

Education

2015	Doctor of Philosophy in Statistics and Applied Probability, Department of Statistics and Applied Probability, University of California, Santa Barbara
2011	Master of Art in Applied Statistics, University of California, Santa Barbara
2009	Master of Art in Economics,University of California, Santa Barbara
2008	Bachelor of Economics in Finance, Renmin University of China, China.

Experience

2010-2015	Teaching Assistant, Department of Statistics and Applied Probability, University of California, Santa Barbara
2009-2010	Readership Assistant, Department of Statistics and Applied Probability, University of California, Santa Barbara.

Abstract

Parameter Estimation for Stable Distributions: Spacings-based and Indirect Inference

Gaoyuan Tian

Stable distributions are important family of parametric distributions widely used in signal processing as well as in mathematical finance. Estimation of the parameters of this model, is not quite straightforward due to the fact that there is no closed-form expression for their probability density function. Besides the computationally intensive maximum likelihood method where the density has to be evaluated numerically, there are some existing adhoc methods such as the quantile method, and a regression based method. These are introduced in Chapter 2. In this thesis, we introduce two new approaches: One, a spacing based estimation method introduced in Chapter 3 and two, an indirect inference method considered in Chapter 4. Simulation studies show that both these methods are very robust and efficient and do as well or better than the existing methods in most cases. Finally in Chapter 5, we use indirect inference approach to estimate the best fitting income distribution based on limited information that is often available.

Contents

Acknowledgements **iv**

Curriculum Vitæ **v**

Abstract **vi**

List of Figures **x**

List of Tables **xi**

1 Introduction to Stable Distributions **1**
1.1 Definitions 2
1.1.1 A Different Parametrization of Stable Laws 3
1.2 Basic Properties 5
1.2.1 Densities and Distribution Functions 5
1.2.2 Tail Probabilities and Moments 7
1.3 Simulation Methods 9

2 Existing Estimation Methods **10**
2.1 Maximum Likelihood Estimation 11
2.1.1 The Integral Representations of Zolotarev 12
2.1.2 Fast Fourier Transformation 13
2.2 Quantile Based Estimation 15
2.3 Characteristic Function based Estimation 16

3 Spacing Based Estimation for Stable Distributions **18**
3.1 Introduction of Spacings Based Estimation 19
3.1.1 Definition 19
3.1.2 Properties 20
3.1.3 Examples 20
3.2 GSE Applied in Stable Distributions 22

3.2.1 Estimating Tail and Skewness Parameters 22
3.2.2 Monto Carlo Studies 25
3.3 Comparison Between Different Methods 26
3.4 Conclusion 27

4 Indirect Inference Method Applied to Stable Distributions 28
4.1 Introduction 28
4.2 Existing Estimation Methods 30
4.2.1 Introduction of the Stable Distribution 30
4.2.2 Available Estimation Methods 31
4.3 Indirect Inference in Stable Distributions 33
4.3.1 Indirect Inference Method 33
4.3.2 Quanitle-based Indirect Inference 35
4.3.3 Theoretical Properties 37
4.4 Simulation Study 41
4.4.1 Choice of the number of quantiles, m 41
4.4.2 Weight Matrix 45
4.4.3 Comparison Between different Methods 47
4.5 Case Study 47
4.6 Conclusion 52
4.7 Appendix: Influence Function and Robust Property of Quantiles . 53

5 Indirect Inference Applied to Income Distributions 55
5.1 Introduction to Some Inequality Measures 56
5.1.1 Lorenz curve 56
5.1.2 Gini Index and Other Inequality Measures 57
5.1.3 Some Popular Parametric Income Distributions 59
5.2 Indirect Inference Method 60
5.2.1 Indirect inference framework 60
5.2.2 Theoretical Properties 60
5.2.3 Goodness of Fit Analysis 61
5.2.4 Data 61
5.3 Simulation Study 63
5.3.1 Numerical Optimization 63
5.3.2 Monte Carlo Study 65
5.4 Case Study 66
5.4.1 Data 68
5.4.2 Result 69
5.4.3 Conclusions 72

6 Conclusions and Discussion 73

A Code **75**
A.1 R code of Estimating income distribution by indirect inference . . 75
A.2 Matlab code for spacing based estimation of stable distribution . . 77

Bibliography **80**

List of Figures

1.1 Pdf of stable distribution of different parameters 6
1.2 Stable noise with different tail parameters 8

3.1 Greenwood statistic vs.α 23
3.2 Greenwood statistic vs.(α, β) 24

4.1 Estimation Algorithm 37
4.2 MSE of $\hat{\alpha}$ by different q 43
4.3 MSE of $\hat{\beta}$ by different q 43
4.4 MSE of scale $\hat{\sigma}$ by different q 44
4.5 MSE of location $\hat{u}$ by different q 44
4.6 Sum of MSE 45
4.7 Plot of index and return 49
4.8 histogram 50
4.9 QQ plot 50

5.1 Lorenz Curve of lognormal and exponential 58
5.2 Objective function vs (θ_1, θ_2) 64
5.3 Objective function vs θ_2 64
5.4 Boxplot of $\hat{\mu}$ 65
5.5 Boxplot of $\hat{\sigma}$ 66
5.6 GDP growth rate 67
5.7 Lorenz curve 2010 USA vs China 69
5.8 Lorenz Curve of USA 1981 vs 2010 70
5.9 Lorenz Curve of China 1981 vs 2010 70
5.10 Income distribution of China 1981 vs 2010 71
5.11 Income distribution of USA 1981 vs 2010 71

List of Tables

3.1 Mean Square Error: MLE vs GSE 22
3.2 Mean square error and bias of $\hat{\alpha}$ with various sample size 25
3.3 Mean square error and bias of $(\hat{\alpha}, \hat{\beta})$ with various sample size . . . 26
3.4 Monte Carlo mean and mean square error of $(\hat{\alpha}, \hat{\beta})$ with different methods 27

4.1 MSE comparison: Identity matrix vs Two-step weight matrix . . . 46
4.2 MSE of different methods 48
4.3 Stable vs Skewed-t 52

5.1 Lorenz Curve for some distributions 57
5.2 Original Data 62
5.3 Transformed Data 62
5.4 Goodness of Fit Assessment 63
5.5 Income inequality of USA: 1981 v.s 2010 68
5.6 Income inequality of China: 1981 v.s 2010 68

Chapter 1

Introduction to Stable Distributions

Stable distributions are a rich class of probability distributions that allow high skewness and heavy tails, compared to the most commonly used Normal distributions, and enjoy many interesting and useful properties. They are introduced by Lévy (1924) in his study of sums of independently identically distributed random variables. The lack of closed-form expression for their densities and distribution functions for all but some special cases viz. the Gaussian, Cauchy and Levy, has been a major drawback for applications. Stable distributions are found to be useful for many reasons. First, there are some theoretical reasons for using stable distributions, e.g. hitting times for a Brownian motion yield a Levy distribution. Secondly, stable distributions turn out to be only possible non-trivial limits of normalized sums of independently and identically distributed (i.i.d.) random variables – a property that is considered as one of the main reasons that these distributions are viewed as suitable for describing stock-returns since a stock price may be considered the result of random instantaneous arrival of information. Mandelbrot (1963) was among the first to apply the stable laws to stock-return data. Thirdly, stable distributions have four parameters instead of two as in Gaussian, which makes them much more flexible to adapt to empirical data for calibration and model testing. Finally, many practical data sets exhibit heavy tails and skewness which stable distributions are able to capture.

1.1 Definitions

An important property of normal random variables is that the sum of any two of them is itself a normal random variable. This property nearly characterizes a stable distribution.

Definition 1. *A random variable X is said to have a stable distribution if for any $n \geq 2$ and independent copies $X_1, \cdots, X_n$ of X, there is a positive real number C_n and a real number D_n, such that $X_1 + X_2 + \cdots + X_n \stackrel{D}{=} C_n X + D_n$, where $\stackrel{D}{=}$ denotes distributional equivalence.*

The word "stable" is used since the type of distribution is unchanged under sums of independent copies. Two random variables X and Y are said to be of the same type if there are constants $a > 0$ and $b \in \mathbb{R}$ with $X \stackrel{\mathrm{D}}{=} aY + b$. Here stable stands for "sum stable", and there are similar notions of max-stable, min-stable, multiplication stable and geometric stable, etc (Kozubowski et al. 2005).

The following definition states that stable distributions are the only possible non-trivial limits of normalized sums of i.i.d. random variables. This result is sometimes called Generalized Central Limit Theorem (Gnedenko and Kolmogorov 1954).

Definition 2. *A random variable X is said to have a stable distribution if it has a domain of attraction, i.e. if there is a sequence of i.i.d. random variables Y_1, Y_2,$\cdots$,Y_n and sequences of positive numbers of d_n and real number a_n, such that $\frac{Y_1+Y_2+\ldots+Y_n}{d_n} + a_n \xrightarrow{D} X$, where $\xrightarrow{D}$ denotes convergence in distribution.*

While the above definition is quite interesting, yet it does not give a concrete way of describing a stable distribution. The most concrete way to describe a stable distribution is through its characteristic function.

Definition 3. *A random variable X follows stable distribution $S(\alpha, \beta, \sigma, \mu_0)$ if its characteristic function $\varphi_0(t) = Ee^{itX}$ has the following form, where $0 < \alpha \leq 2$ measures the tail thickness, $-1 \leq \beta \leq 1$ determines skewness, and $\mu_0 \in \mathbb{R}$, $\sigma > 0$ are location and scale parameters in the sense that $\frac{X-\mu_0}{\sigma} \sim S(\alpha, \beta, 1, 0)$. $\varphi_0(t) =$*

$$\begin{cases} \exp(-\sigma^\alpha |t|^\alpha (1 - i\beta \frac{t}{|t|} \tan(\frac{\pi\alpha}{2})) + i\mu_0 t), \alpha \neq 1 \\ \exp(-\sigma|t|^\alpha (1 + i\beta \frac{2}{|\pi|} \ln(t)) + i\mu_0 t), \alpha = 1 \end{cases} \tag{1.1}$$

1.1.1 A Different Parametrization of Stable Laws

The parametrization in Definition 3 has the advantage that the parameters are easy to interpret in terms of location and scale. But there is a disadvantage, namely that when it comes to numerical or statistical work, it is discontinuous at $\alpha = 1$ and $\beta \neq 0$. An alternative parametric representation $S(\alpha, \beta, \sigma, \mu_1)$ (denoted as P_1) with the following characteristic function overcomes this problem:

$$\varphi_1(t) =$$

$$\begin{cases} \exp(-|\sigma t|^\alpha + i\sigma t\beta(|\sigma t|^{\alpha-1} - 1)\tan(\frac{\pi\alpha}{2}) + i\mu_1 t), \alpha \neq 1 \\ \exp(-|\sigma t| + i\sigma t\beta\frac{2}{\pi}\ln|\sigma t| + i\mu_1 t), \alpha = 1 \end{cases} \tag{1.2}$$

where $0 < \alpha \leq 2$,$-1 \leq \beta \leq 1$,$\sigma > 0$ and $\mu_1 \in \mathbb{R}$.

The relationship between P_0 and P_1 is given by,

$$\mu_1 =$$

$$\begin{cases} \mu_0 + \beta\sigma\tan(\frac{\pi\alpha}{2}), \alpha \neq 1 \\ \mu_0 + \beta\sigma\frac{2}{\pi}\ln\sigma, \alpha = 1 \end{cases} \tag{1.3}$$

Another parametrization $S(\alpha, \beta_2, \sigma_2, \mu)$ (denoted as P_2) proposed by Zolotarev (1986) appears to be more suitable in the derivation of some analytic properties of stable law

$$\varphi_2(t) =$$

$$\begin{cases} \exp(i\mu t - \sigma_2^\alpha|t|^\alpha + exp(-i\frac{\pi\beta_2}{2}sign(t)\min(\alpha, 2-\alpha))), \alpha \neq 1 \\ \exp(i\mu t - \sigma_2|t|(1 + i\beta_2\frac{2}{\pi}sign(t)\ln(\sigma_2|t|))), \alpha = 1 \end{cases} \tag{1.4}$$

The relationship between P_0 and P_2 is given by,

$$\begin{cases} \beta = \cot\frac{\pi\alpha}{2}\tan(\frac{\pi\beta_2}{2}\min(\alpha, 2-\alpha)) \\ \sigma = \sigma_2(\cos(\frac{\pi\beta_2}{2}\min(\alpha, 2-\alpha)))^{\frac{1}{\alpha}} \end{cases} \tag{1.5}$$

and α and μ remain unchanged. Unless it is specifically mentioned otherwise, the default parameter set will be assumed to be in the form of P_0.

1.2 Basic Properties

1.2.1 Densities and Distribution Functions

Except for some special cases, say, Normal ($\alpha = 2$), Cauchy ($\alpha = 1, \beta = 0$) and Levy ($\alpha = 1/2, \beta = 1$), the density function and distribution function of α-stable distributions cannot be written analytically. However, the most basic fact is the following.

Theorem 1. *(Nolan 2005) All (non-degenerate) stable distributions are continuous unimodal distributions with an infinitely differentiable density.*

Since all stable distributions are shifts and scales of standard stable $S(\alpha, \beta)$ where $\sigma = 1, \mu = 0$, we will focus on these distributions for simplicity. The following fact is about the reflection property.

Proposition 1. *For any α and β, $X \sim S(\alpha, \beta)$, the distribution function F satisfies $F(x|\alpha, \beta) = 1 - F(-x|\alpha, -\beta)$.*

First consider the symmetric case when $\beta = 0$. In this case, the reflection property simply says the density and distribution function are symmetric around 0. Also notice as α increases, the tails get heavier and the peak gets higher. If $\beta > 0$, then the distribution is skewed with the right tail heavier than the left tail which means $P(X > x) > P(X < -x)$ for large $x > 0$. When $\beta = 1$, the stable distribution is totally skewed to the right. By the reflection property, the

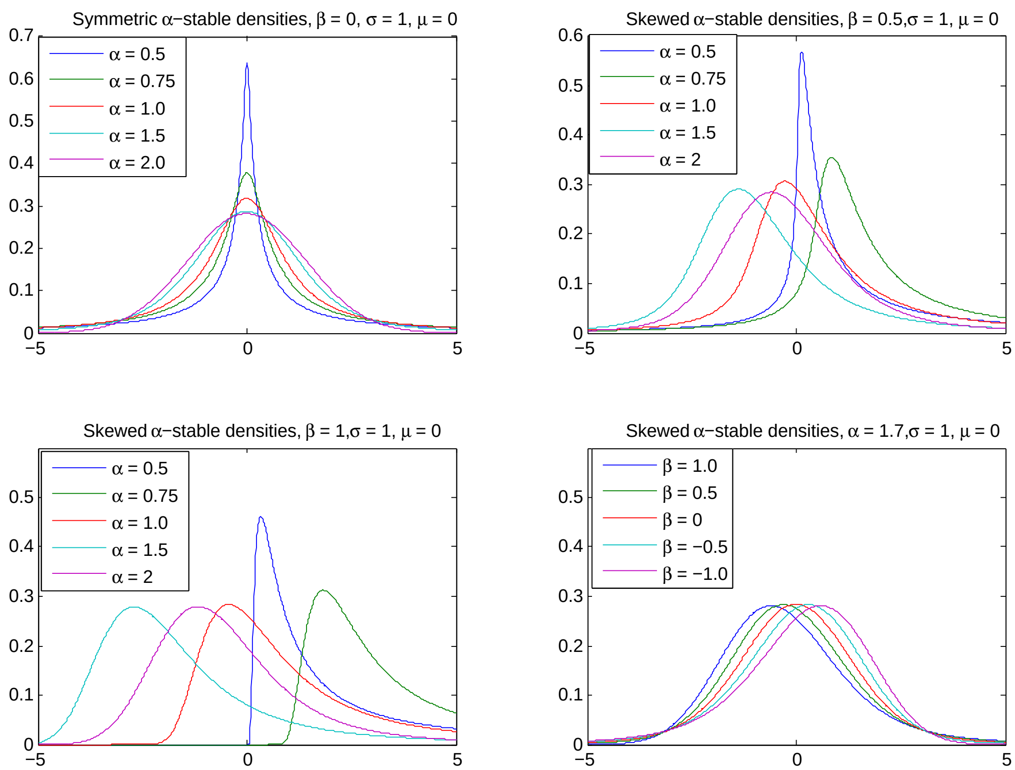

Figure 1.1: Pdf of stable distribution of different parameters

behavior of the $\beta < 0$ cases are reflections of the $\beta > 0$ ones, with the left tail being heavier (see Figure 1.1).

1.2.2 Tail Probabilities and Moments

The tail of the stable distribution behaves similarly to the tail of the Pareto distribution. Thus the stable distribution is also called the stable Pareto distribution. This stable Paretian law (Mandelbrot 1961) is used to distinguish between the fast decay of Gaussian law and the Pareto like tail behavior in $\alpha < 2$ case.

Theorem 2. *If $X \sim S(\alpha, \beta, \sigma, \mu)$, with $0 < \alpha \leq 2$,$-1 \leq \beta \leq 1$. As $x \to \infty$,*

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

where $C_{\alpha} = \sin(\frac{\pi\alpha}{2})\Gamma(\alpha)/\pi$.

By the reflection property, $P(X < -x) \sim \sigma^{\alpha} C_{\alpha}(1-\beta)(-x)^{-\alpha}$ for large x. For all $\alpha < 2$ and $-1 < \beta < 1$, the tails are asymptotically power laws. When $\beta = -1$, the right tail of the distribution is not asymptotically a power law. When $\beta = 1$, the left tail of the distribution is not asymptotically a power law.

One consequence of heavy tails is that not all moments exist. The fractional absolute moment of X, $E|X|^p < \infty$ for $0 < p < \alpha$ and $E|X|^p = \infty$ for $p \geq \alpha$. Thus, α-stable random variable does not have finite mean and variance for $0 < \alpha < 1$. It has finite first mean but infinite variance for $1 < \alpha < 2$.

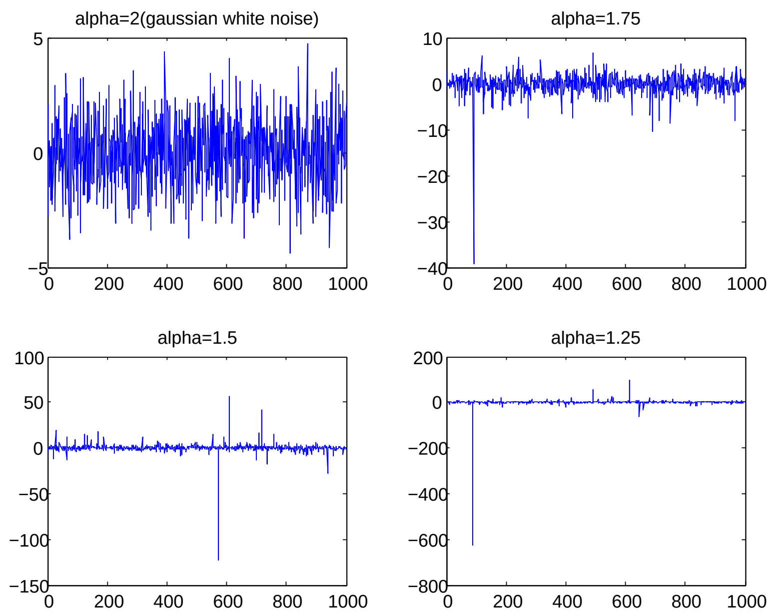

Figure 1.2: Stable noise with different tail parameters

1.3 Simulation Methods

The simulation method of stable random variable Y is given by Chambers et al. (1976).

Step 1 Generate a random variable U uniformly distributed on $(-\frac{\pi}{2}, \frac{\pi}{2})$ and an independent exponential random variable E with mean 1.

Step 2 For $\alpha \neq 1$, compute $X = S_{\alpha,\beta}\frac{\sin(\alpha(U+B_{\alpha,\beta}))}{(\cos U)^{1/\alpha}}(\frac{\cos(U-\alpha(U+B_{\alpha,\beta}))}{E})^{(1-\alpha)/\alpha}$, Where $B_{\alpha,\beta} = \frac{\arctan(\beta\tan(\pi\alpha/2))}{\alpha}$, $S_{\alpha,\beta} = [1+\beta^2\tan^2\frac{\pi\alpha}{2}]^{1/(2\alpha)}$.

Step 3 for $\alpha = 1$, compute $X = \frac{2}{\pi}[(\pi/2+\beta U) - \beta\log(\frac{E(\cos U)}{\pi/2+\beta U})]$.

Step 4 Set $Y =$

$$\begin{cases} \sigma X + \mu, \alpha \neq 1 \\ \sigma X + \frac{2}{\pi}\beta\sigma\log(\sigma) + \mu, \alpha = 1 \end{cases} \tag{1.6}$$

is $S(\alpha, \beta, \sigma, \mu)$.

Chapter 2

Existing Estimation Methods

The popular parameter estimation techniques for stable distributions fall into three categories: quantile methods, characteristic function based methods and maximum likelihood method (MLE). The quantile method of McCulloch (1986) gives a simple and consistent estimation for all four parameters in stable distribution. However, the quantile method requires a large amount of computation in the form of some precisely tabulated values. MLE has theoretically the smallest variance for large samples but at a high computational cost. Careful numerical implementation is needed for the density function and the searching procedure for the maximum (See e.g. Nolan (2002)). Parameter estimation based on characteristic function was originally proposed by Press (1972). Later the iterative weighted regression method of Koustrouvellis (1980) was shown to have somewhat better performance. Characteristic function based methods avoid the inversion procedure for evaluating the density function. Nevertheless, no single method is efficient and/or simple.

2.1 Maximum Likelihood Estimation

The method of maximum likelihood is very attractive because of the good asymptotic properties of the estimates, provided that the likelihood function obeys certain general conditions. The likelihood function is

$$L(x_1, x_2, \cdots, x_n|\boldsymbol{\theta}) = \prod_{k=1}^{n} f(x_k|\boldsymbol{\theta}),$$

where $x_1, x_2, ..., x_n$ is a sample of i.i.d. observations of a random variable X, $f(x|\boldsymbol{\theta})$ is the pdf of X and $\boldsymbol{\theta}$ is a vector of parameters. In the case of stable distributions, $\boldsymbol{\theta} = (\alpha, \beta, \sigma, \mu)$. Maximum likelihood estimates are found by searching for those parameter values which maximize the likelihood function, or equivalently, the log-likelihood function $l(\boldsymbol{\theta}) = \log(L(x_1, x_2, ..., x_n|\boldsymbol{\theta}))$. Maximum likelihood estimation is theoretically the most efficient estimating method when the sample size is big enough. But it is computationally intensive, the density function and the maximum searching procedure have to be both carefully numerically evaluated. Zolotarev (1986) gives computational formulae for the density and distribution function. These formulae are carefully used in a software called STABLE by Nolan (1997). Another more general method to evaluate the density is by using Fast fourier transform(FFT), described below.

2.1.1 The Integral Representations of Zolotarev

The density and the distribution function of stable laws can be very accurately evaluated with the help of integral representations derived by Zolotarev in parametrization P_1. The density and distribution function of stable laws can be expressed as, for $x > \xi$,

$$f(x; \alpha, \beta, P_1) = c_2(x; \alpha, \beta) \int_{\theta_0}^{\pi/2} g(\theta; x, \alpha, \beta) exp(-g(\theta; x, \alpha, \beta)) d\theta \tag{2.1}$$

and

$$F(x; \alpha, \beta, P_1) = c_1(\alpha, \beta) + c_3(\alpha) \int_{\theta_0}^{\pi/2} exp(-g(\theta; x, \alpha, \beta)) d\theta \tag{2.2}$$

where for $\alpha \neq 1$,

- $c_1(\alpha, \beta) = \frac{1}{\pi}(\frac{\pi}{2} - \theta_0)$ for $\alpha < 1$, and 1 for $\alpha > 1$
- $c_2(x; \alpha, \beta) = \frac{\alpha}{\pi|\alpha-1|(x-\xi)}$
- $c_3(\alpha) = \frac{sign(1-\alpha)}{\pi}$
- $g(\theta; x, \alpha, \beta) = (x - \xi)^{\frac{\alpha}{\alpha-1}} V(\theta; \alpha, \beta)$
- $\xi = \xi(\alpha, \beta) = -\beta \tan(\frac{\pi\alpha}{2})$
- $\theta_0 = \theta_0(\alpha, \beta) = \frac{1}{\alpha} arctan(\beta \tan(\frac{\pi\alpha}{2}))$
- $V(\theta; \alpha, \beta) = (\cos(\alpha\theta_0))^{\frac{1}{1-\alpha}} (\frac{\cos\theta}{\sin\alpha(\theta_0+\theta)})^{\frac{\alpha}{\alpha-1}} \frac{\cos(\alpha\theta_0+(\alpha-1)\theta)}{\cos\theta}$;

for $\alpha = 1$,

- $c_1(\alpha, \beta) = 0$
- $c_2(x; \alpha, \beta) = \frac{1}{2|\beta|}$
- $c_3(\alpha) = \frac{1}{\pi}$
- $g(\theta; x, \alpha, \beta) = exp(\frac{\pi x}{2\beta})V(\theta; \alpha, \beta)$
- $\xi = 0$
- $\theta_0 = \frac{\pi}{2}$
- $V(\theta; \alpha, \beta) = \frac{2}{\pi}(\frac{\frac{\pi}{2}+\beta\theta}{\cos\theta})exp(\frac{1}{\beta}(\frac{\pi}{2} + \beta\theta)\tan\theta)$.

The case $x < \xi$ can be treated by taking advantage of the relationship

$$f(x; \alpha, \beta; P_1) = f(-x; \alpha, -\beta; P_1) \tag{2.3}$$

and

$$F(x; \alpha, \beta; P_1) = 1 - F(-x; \alpha, -\beta; P_1). \tag{2.4}$$

2.1.2 Fast Fourier Transformation

Mittnik et al. (1999) carefully presented and implemented the Fast Fourier Transform (FFT) algorithm for calculating the density function of stable distribution. A brief introduction is given below. Recall the inversion formula:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)dt. \tag{2.5}$$

For grids of equally spaced x values with $x_k = (k-1-\frac{N}{2})h$, where $k = 1, 2, \cdots, N$.

$$f(x_k) = \int_{-\infty}^{\infty} e^{-i2\pi w(k-1-\frac{N}{2})h}\phi(2\pi w)dw, t = 2\pi w. \tag{2.6}$$

Since this integral is convergent, it can be approximated by Riemann sum for N points with spacing s, where $w = s(n-1-\frac{N}{2})$:

$$f(x_k) \approx s\sum_{n=1}^{N}\phi(2\pi s(n-1-\frac{N}{2}))e^{-i2\pi(k-1-\frac{N}{2})(n-1-\frac{N}{2})sh}, k = 1, ..., N. \tag{2.7}$$

By setting $s = (hN)^{-1}$, for $k = 1, ..., N$, we have

$$f(x_k) \approx \frac{1}{hN}\sum_{n=1}^{N}\phi(2\pi\frac{1}{hN}(n-1-\frac{N}{2}))e^{-i2\pi(k-1-\frac{N}{2})(n-1-\frac{N}{2})\frac{1}{N}}. \tag{2.8}$$

Having rearranged the terms in the exponent, finally, for $k = 1, ..., N$ we arrive at

$$f(x_k) \approx \frac{(-1)^{k-1+\frac{N}{2}}}{hN}\sum_{n=1}^{N}(-1)^{n-1}\phi(\frac{2\pi}{hN}(n-1-\frac{N}{2}))e^{\frac{-i2\pi(n-1)(k-1)}{N}}. \tag{2.9}$$

The discrete FFT is a numerical method developed for calculation of sequences such as $f(x_k)$ in (2.9) given the sequence $(-1)^{n-1}\phi(\frac{2\pi}{hN}(n-1-\frac{N}{2}))$.

It should be noted that the approximation errors may arise from the interchange of the infinite integral bounds (2.6) with finite ones, or from the approximation of (2.7) with the Riemann sum. Also, the FFT method does not have a good performance around the tail of the distribution. Thus density functions (2.5) are evaluated on an equally spaced grid over a certain interval. For the points outside the interval, we need to employ integral representations of Zolotarev for good tail approximation.

2.2 Quantile Based Estimation

McCulloch (1986) obtained consistent estimators for four parameters in stable distributions based on five sample quantiles. The main estimating algorithm is described as follows:

Step 1 Estimating α and β. X_p is the $p-th$ quantile if $F(x_p)=p$, where $F(x)$ is the distribution function. $\hat{x_p}$ is the sample quantile if $F_n(\hat{x_p})=p$, where $F_n(x)$ is empirical distribution function. Define two functions of theoretical quantiles:

$$\begin{cases} v_\alpha = \dfrac{x_{0.95}-x_{0.05}}{x_{0.75}-x_{0.25}} = \phi_1(\alpha,\beta). \\ v_\beta = \dfrac{(x_{0.95}-x_{0.5})-(x_{0.5}-x_{0.05})}{x_{0.95}-x_{0.05}} = \phi_2(\alpha,\beta). \end{cases} \tag{2.10}$$

Replace v_α and v_β with their sample counterparts $\hat{v_\alpha}$ and $\hat{v_\beta}$, define φ as the solution to the equation (2.10), we get estimators

$$\begin{cases} \hat{\alpha} = \varphi_1(\hat{v_\alpha}, \hat{v_\beta}) \\ \hat{\beta} = \varphi_2(\hat{v_\alpha}, \hat{v_\beta}) \end{cases} \tag{2.11}$$

with $\hat{v}_\alpha = \frac{\hat{x}_{0.95}-\hat{x}_{0.05}}{\hat{x}_{0.75}-\hat{x}_{0.25}}$ and $\hat{v}_\beta = \frac{(\hat{x}_{0.95}-\hat{x}_{0.5})-(\hat{x}_{0.5}-\hat{x}_{0.05})}{\hat{x}_{0.95}-\hat{x}_{0.05}}$

Step 2 Estimating scale parameter σ. Let us first define v_σ as $v_\sigma = \frac{x_{0.75}-x_{0.25}}{\sigma} = \phi_3(\alpha,\beta)$. The estimator $\hat{\sigma}$ is obtained by replacing (α,β) with $(\hat{\alpha},\hat{\beta})$, thus $\hat{\sigma} = \frac{\hat{x}_{0.75}-\hat{x}_{0.25}}{\phi_3(\hat{\alpha},\hat{\beta})}$.

Step 3 Estimating location parameter μ. Let us define $\frac{x_{0.5}-\mu}{\sigma} = \phi_4(\alpha,\beta)$. The estimator $\hat{\mu}$ is obtained by replacing (α,β,σ) with $(\hat{\alpha},\hat{\beta},\hat{\sigma})$, thus $\hat{\mu} = \phi_4(\hat{\alpha},\hat{\beta})\hat{\sigma}+\hat{x}_{0.5}$.

The main idea is to use quantile-differences in order to get rid of the dependence on the location parameter, and then take ratios of these to remove the scale parameter. Then, two functions on α and β are numerically calculated from sample quantiles and inverted to get the corresponding parameter estimates. A tabulated table is needed for equations of (2.11), $\phi_3(\alpha,\beta)$ and $\phi_4(\alpha,\beta)$.

2.3 Characteristic Function based Estimation

Since there is a closed form of characteristic function, the estimator based on empirical characteristic function can be developed. The regression -type estimation of Koustrouvellis (1980) starts with an initial estimate (in practice, we usually choose the quantile estimate) and proceeds iteratively until some convergence criterion is satisfied.

Directly from the convenient form of the logarithm of the CF, we have the following linear equations:

$$\ln(-\Re(\ln\phi_0(t))) = \alpha\ln\sigma + \alpha\ln|t| \tag{2.12}$$

and

$$\Im(\ln\phi_0(t)) = \mu_1 t + \beta\sigma t(|\sigma t|^{\alpha-1} - 1)\tan(\frac{\pi\alpha}{2}) \tag{2.13}$$

The estimation algorithm is as follows,

Step 1 Given a sample of i.i.d observations $x_1, x_2, ..., x_n$ first we find preliminary estimates σ_0 and μ_{01} by the quantile method of McCulloch and we normalize the observations as $x'_j = \frac{x_j - \hat{\mu}_{01}}{\hat{\sigma}_0}$ for $j = 1, 2, ..., n$.

Step 2 Consider the regression equation constructed above $y_k = b + \alpha\omega_k + \epsilon_k, k = 0, 1, 2, ..., 9$, where $y_k = \ln(-\Re(\ln\hat{\phi}_0(t)))$, $\omega_k = \ln|t_k|$, $t_k = 0.1 + 0.1k$ and ϵ_k denotes the error term. The empirical CF $\hat{\phi}_0(t)$ is defined as

$$\hat{\phi}_0(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itx_j} = (\frac{1}{n}\sum_{j=1}^{n}\cos tx_j) + i(\frac{1}{n}\sum_{j=1}^{n}\sin tx_j), t \in \mathbb{R} \tag{2.14}$$

We find $\hat{\alpha}$ and $\hat{b}$ according to the method of least squares using the normalized sample $x'_1, x'_2, ...x'_n$. The estimator $\hat{\sigma_1}$ of the scale parameter of the normalized sample is $\hat{\sigma} = exp(\frac{\hat{b}}{\hat{\alpha}})$.

Step 3 Estimators $\hat{\beta}$ and $\hat{\mu_{11}}$ of the skewness parameter and the modified location parameter respectively are derived from the second regression equation based on (2.13): $z_k = \mu_{11}t_k + \beta\nu_k + \eta_k$, where $z_k = \Im(\ln\hat{\phi}(t))$, $\nu_k = \hat{\sigma}_1 t_k(|\hat{\sigma}t|^{\hat{\alpha}-1} - 1)\tan(\frac{\pi\hat{\alpha}}{2})$, $t_k = 0.1 + 0.1k$ and η_k is the error term.

Step 4 Compute the final estimates $\hat{\sigma} = \hat{\sigma}_0\hat{\sigma}_1$ and $\hat{\mu}_1 = \hat{\mu}_{01}\hat{\sigma}_0 + \hat{\mu}_{11}$. If we aim to estimate the location parameter μ, we need to take advantage of the connection between the two parametric forms P_0 and P_1:

$$\hat{\mu} = \hat{\mu}_1 - \hat{\beta}\hat{\sigma}\tan\frac{\pi\hat{\sigma}}{2} \tag{2.15}$$

Repeat Step 1 to Step 4 until the estimator fulfills some convergence criterion.

Chapter 3

Spacing Based Estimation for Stable Distributions

The idea of spacing is introduced by Cheng and Amin (1983) and independently by Ranneby (1984) to estimate finite dimensional parameters in continuous univariate distributions. This idea is adopted to estimate parameters in stable distributions in this chapter which is organized as the following manner. In Section 1, we briefly introduce Generalized Spacing Estimator, about its flexibility of choosing different measures of information and asymptotic normal property. Also, we give some cases where this method is better than MLE. In Section 2, spacing based estimation is applied to estimate (α, β). As an M-estimator, its optimization algorithm is verified to have a local minimum for certain selected point at stable distribution. Also, A Monto Carlo study to compare the mean square errors of this method under different measures of information is illustrated. Section 3 compares this method with others. Section 4 concludes.

3.1 Introduction of Spacings Based Estimation

3.1.1 Definition

Given an i.i.d random sample, $x_1, \ldots, x_n$ from a univariate distribution with distribution function $F(x; \boldsymbol{\theta})$. Let $x_{(1)}, \ldots, x_{(n)}$ be the corresponding order statistics. Define spacings as the gaps between the values of the distribution function at adjacent ordered points, $D_i(\boldsymbol{\theta}) = F(x_{(i)}; \boldsymbol{\theta}) - F(x_{(i-1)}; \boldsymbol{\theta}), i = 1, \ldots, n+1$, and we denote $F(x_{(0)}; \boldsymbol{\theta}) = 0, F(x_{(n+1)}; \boldsymbol{\theta}) = 1$. Then, for any convex function $h : (0, \infty) \to \Re$, minimize the quantity $T_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} h(nD_i(\boldsymbol{\theta}))$. The resulting minimizer $\hat{\boldsymbol{\theta}}$ is called the Generalized Spacing Estimator (GSE) of $\boldsymbol{\theta}$.

The choice of different $h(x)$ yields different criterions of spacing estimation. If $h(x) = -\log(x)$ is chosen, $T_n(\boldsymbol{\theta}) = \sum_{i=1}^{n+1} \log D_i(\boldsymbol{\theta})$, which is called maximum product of spacing (criterion 1). If $h(x) = (x-1)^2$ is chosen, $T_n(\boldsymbol{\theta}) = G_n(\boldsymbol{\theta}) = \sum_{i=1}^{n+1} (D_i(\boldsymbol{\theta}) - \frac{1}{n})^2$, which is called Greenwood statistics (criterion 2). If $h(x) = |x-1|$ is chosen, $T_n(\boldsymbol{\theta}) = \sum_{i=1}^{n+1} |D_i(\boldsymbol{\theta}) - \frac{1}{n}|$, which is called Rao-statistic (criterion 3). Each criterion represents a different measure of information called entropy in information theory. Criterion 1, the Kullback-Leibler divergence, is the most popular measure. Unless specifically pointed out, it is the default one used in this chapter.

3.1.2 Properties

Similar to MLE, one advantage of spacing based estimator is the asymptotic normality of the estimator. Ghosh and Jammalamadaka (2001) show, under some regularity conditions on the density and $h(\cdot)$,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\text{D}} N(0, \sigma_h^2 / I(\theta_0)) \tag{3.1}$$

where $I(\theta_0)$ is the Fisher Information in one observation from the true distribution and

$$\sigma_h^2 = \frac{E(Wh'(W))^2 - 2EWh'(W)Cov\{Wh^{'}(W), W\}}{[EW^2 h''(W)]^2} \tag{3.2}$$

with $W \sim Exp(1)$. Also, they show that the Cramer-Rao lower bound is only reached for the GSE under criterion 1.

3.1.3 Examples

The purpose of this section is to give some simple cases where GSE beats MLE. As mentioned below, GSE has the asymptotic properties closely parallel to ML estimators. GSE may perform better than MLE in the following cases. First, there are certain cases where the ML method breaks down, e.g. for three parameter Weibull distribution or mixtures of continuous distributions (see e.g. Hinkley (1974)). Second, when the end points of density are not known, the log-likelihood is unbounded (see Example 1). Finally, from the robustness perspective, GSE may be more efficient than MLE in small sample when the distribution is skewed or heavy tailed (see Example 2).

Example 1. *Suppose $x_{(1)}, \cdots, x_{(n)}$ is the ordered sample from a uniform distribution $U(a,b)$ with unknown endpoints a and b. The cumulative distribution function is $F(x) = \dfrac{x-a}{b-a}$ when $x \sim [a,b]$. Therefore spacings are given by*

$$D_1 = \frac{x_1 - a}{b-a}, D_i = \frac{x_i - x_{i-1}}{b-a}, i = 2, \cdots, n, D_{n+1} = \frac{b - x_n}{b-a} \tag{3.3}$$

Then, the GSE estimator maximizes the logarithm of the geometric mean of sample spacings:

$$S_n = \log[(D_1 D_2 \cdots D_{n+1})^{\frac{1}{n+1}}] = \frac{1}{n+1}\sum_{i=1}^{n+1} \log D_i \tag{3.4}$$

Differentiating with respect to parameters a and b and solving the resulting linear system, the maximum spacing estimators will be

$$\hat{a} = \frac{nx_{(1)} - x_n}{n-1}, \hat{b} = \frac{nx_{(n)} - x_{(1)}}{n-1} \tag{3.5}$$

These are known to be the uniformly minimum variance unbiased estimators for this continuous uniform distribution. In comparison, the maximum likelihood estimates $\hat{a} = x_{(1)}$ and $\hat{b} = x_{(n)}$ are biased and have a higher mean-squared error. In this case, it is possible the log-likelihood is unbounded in MLE but S_n is always bounded (see Cheng and Amin (1983)).

Example 2. *Consider the exponential distribution $f(x,\lambda) = \lambda \exp(-\lambda x)$ for $x \geq 1$. Suppose the true value of λ equals 1. Here is the result of MSE of both methods obtained from 10,000 simulations by different sample size N.*

	N=6	*N=10*	*N=20*
MLE	*0.3852*	*0.1636*	*0.0640*
GSE	*0.2519*	*0.1200*	*0.0550*

Table 3.1: Mean Square Error: MLE vs GSE

3.2 GSE Applied in Stable Distributions

Here we are interested in estimating tail and skewness parameters (α, β) by assuming $\alpha > 1$, which means the distribution has a finite mean. The location and scale parameters (α, β) are known. And the distribution function of stable laws will be evaluated by the integral representations derived by Zolotarev (1995).

3.2.1 Estimating Tail and Skewness Parameters

In some cases, the practitioner strongly believes the distribution is symmetric and thus, only α has to be estimated. Firstly we consider estimating α when $\beta = 0$ is known. Sample $(x_1, \ldots, x_{1000})$ is simulated from $S(\alpha = 1.5, \beta = 0, \sigma = 1, \mu = 0)$. The information measure we used here is criterion 2. As we could see in Figure 3.1, the estimated value of spacing estimator that minimizes Greenwood-statistic is $\hat{\alpha} = 1.5126$.

Then, consider the case where α and β are both unknown. Sample$(x_1, \ldots, x_{1000})$ is simulated from $S(\alpha = 1.5, \beta = 0, \sigma = 1, \mu = 0)$. Following the similar estimating procedure as above, we get this two dimensional estimator $(\hat{\alpha}, \hat{\beta}) = (1.5158, 0.0572)$. Figure 3.2 shows that this estimator is the local minimal point.

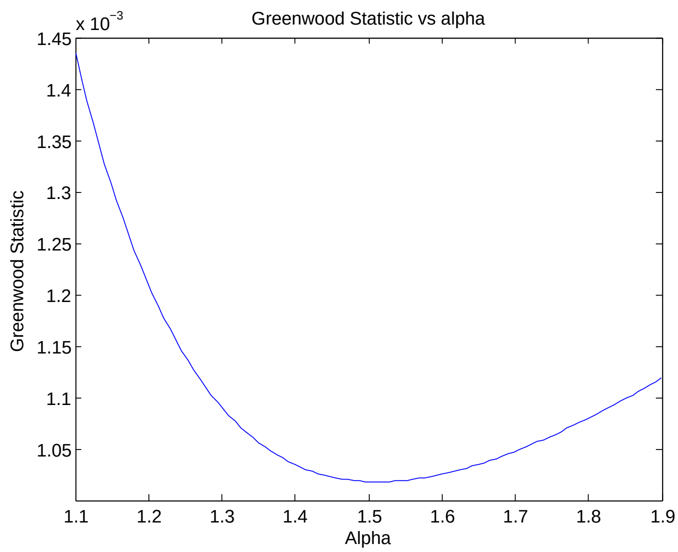

Figure 3.1: Greenwood statistic vs.α

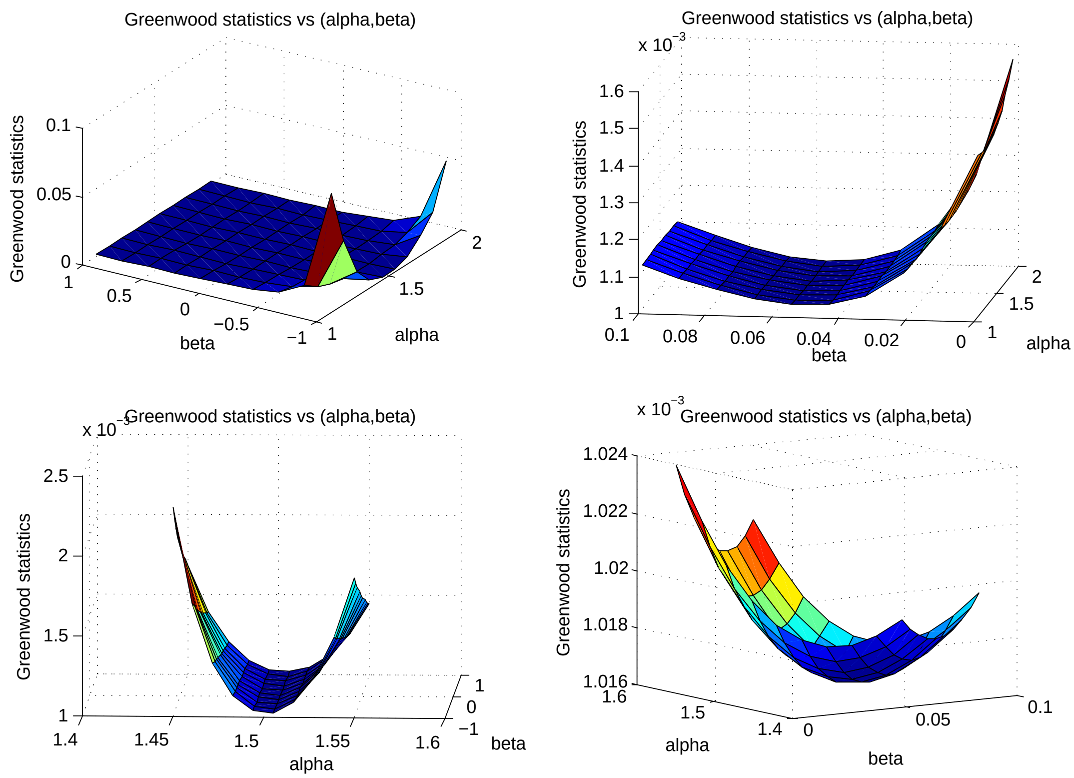

Figure 3.2: Greenwood statistic vs.(α, β)

N=200, M=500			
	Criterion1	Criterion2	Criterion3
$E(\hat{\alpha})$	1.4639	1.4397	1.4649
MSE of $\hat{\alpha}$	0.0110	0.0214	0.0162

N=500, M=500			
	Criterion1	Criterion2	Criterion3
$E(\hat{\alpha})$	1.4786	1.4677	1.4783
MSE of $\hat{\alpha}$	0.0051	0.0088	0.0072

N=1000, M=500			
	Criterion1	Criterion2	Criterion3
$E(\hat{\alpha})$	1.4892	1.4824	1.4938
MSE of $\hat{\alpha}$	0.0023	0.0038	0.0036

Table 3.2: Mean square error and bias of $\hat{\alpha}$ with various sample size

3.2.2 Monto Carlo Studies

In this section, we will compare these three information measures of their performance in estimating parameters in stable distributions by Monto Carlo.

Suppose we have M samples based on the data $(x_1, \ldots, x_N)$ generated from $S(\alpha = 1.5, 0, 1, 0)$. In each sample i, we have spacing estimator $\hat{\alpha}_i$, $i = 1, \ldots, M$. The mean and mean square error of the estimator could be approximated for large M: $E(\hat{\alpha}) = \frac{1}{M}\sum_{i=1}^{M}\hat{\alpha}_i$, $MSE(\hat{\alpha}) = \frac{1}{M}\sum_{i=1}^{M}(\hat{\alpha}_i - 1.5)^2$.

As expected, Criterion 1 has the smallest MSE in Table 3.2. Also as sample size increases, the estimator will converge to the true value with smaller MSE. A similar conclusion could be drawn for estimating (α, β) from Table 3.3.

N=1000, M=500			
	Criterion1	Criterion2	Criterion3
$E(\hat{\alpha}, \hat{\beta})$	(1.4861,0.1940)	(1.4823,0.1909)	(1.4874,0.1989)
MSE of $(\hat{\alpha}, \hat{\beta})$	(0.0025,0.0029)	(0.0051,0.0066)	(0.0037,0.0064)

N=500, M=500			
	Criterion1	Criterion2	Criterion3
$E(\hat{\alpha}, \hat{\beta})$	(1.4816,0.1941)	(1.4675,0.1824)	(1.4848,0.2071)
MSE of $(\hat{\alpha}, \hat{\beta})$	(0.0039,0.0046)	(0.0086,0.0117)	(0.0062,0.0125)

Table 3.3: Mean square error and bias of $(\hat{\alpha}, \hat{\beta})$ with various sample size

3.3 Comparison Between Different Methods

A Monte Carlo evaluation to compare different methods at the point $S(\alpha = 1.5, \beta = 0.2, \sigma = 1, \mu = 0)$ with α, β unknown is listed in the following Table 3.4. For large sample size, GSE and MLE are equivalently good as expected, they have smaller MSE than the other two methods (especially for β) even if the numerical error of evaluating the density and distribution function occurs.

N=1000 M=1000				
	MLE method	Quantile method	Regression method	Spacing method
$(\hat{\alpha}, \hat{\beta})$	(1.5017,0.2045)	(1.5049,0.2085)	(1.5042,0.2010)	(1.4983,0.1998)
MSE	(0.0025,0.0092)	(0.0045,0.0126)	(0.0034,0.0143)	(0.0035,0.0058)
N=500 M=1000				
	MLE method	Quantile method	Regression method	Spacing method
$(\hat{\alpha}, \hat{\beta})$	(1.5016,0.2159)	(1.5039,0.2242)	(1.5031,0.2058)	(1.4951,0.1944)
MSE	(0.0138,0.0567)	(0.0250,0.0822)	(0.0179,0.0833)	(0.0173,0.0741)

Table 3.4: Monte Carlo mean and mean square error of $(\hat{\alpha}, \hat{\beta})$ with different methods

3.4 Conclusion

Spacing-based estimation has several advantages compared with other methods in terms of estimating stable distribution. First, this estimator is equivalently good as MLE in the large sample estimation. However, it provides more robustness. Second, spacing-based estimation is considerably flexible, namely different information measures could be selected in specific cases. Finally, the spacing idea could be used in goodness of fit test and model selection (Jammalamadaka and Goria 2004).

Chapter 4

Indirect Inference Method Applied to Stable Distributions

4.1 Introduction

Stable distributions comprise an entire class of distributions and was first described by Lévy (1924) in a study of normalized sums of independently and identically distributed (i.i.d.) random variables. The Gaussian and Cauchy distributions are important special cases of stable distributions. This family of distributions delivers an extensive class of distributions that provide a flexible framework to consider various features such as skewness and heavy tails. These features make stable distribution useful under many situations. First, this distribution has four parameters compared to the two for Gaussian and enable the distribution to be considerably more flexible when adapting to empirical data for calibration and model testing. Second, the stable distribution comes from the limit of normalized sums of iid random variables that constitute one of the main reasons it is suitable for describing stock-returns. Stock price can be considered the result of the random and instantaneous arrival of information corresponding to the hitting times for a Brownian motion. Accordingly, Mandelbrot (1963) was among the first to apply stable distribution to stock-return data where heavy tails and skewness are frequent and complicated to capture.

Despite the aforementioned flexibility, a stable distribution lacks a closed form expression for its density, with the exception of the few cases where it takes the parametric form of the Gaussian, Cauchy and Lévy distributions. This can represent a major drawback in practice in terms of the estimation of its parameters. Thus, several methods from different perspectives have been developed. This chapter discusses the major drawbacks of currently existing methods, and proposes a quantile-based indirect inference method. The rest of this chapter is organized as follows. Section 1 describes stable distributions in considerable detail and then discusses the existing estimation methods to underscore their limitations and motivate the proposal of the new estimation recommended in this chapter. Section 2 provides an overview of indirect inference estimation and describes how this method is used using quantiles as the auxiliary parameter to deliver robust, efficient and easy-to-compute estimations. Section 3 presents a simulation study that considers different parametric settings, thereby showing that, this method has the smallest mean square error (MSE), particularly when in the presence of heavy tail parametrizations. Section 4 describes an application to the S&P 500 index returns. Section 5 provides some conclusions.

4.2 Existing Estimation Methods

4.2.1 Introduction of the Stable Distribution

As mentioned above, stable distribution is the only possible limit distribution of sum of normalized iid random variables. This property is also known as generalized central limit theorem which uniquely define stable distribution (Gnedenko and Kolmogorov 1954).

Definition 4. *A random variable X is said to have a stable distribution if it has a domain of attraction, i.e. if there is a sequence of i.i.d. random variables Y_1, Y_2,$\cdots$,Y_n and sequences of positive numbers of d_n and real numbera_n, such that $\frac{Y_1+Y_2+\ldots+Y_n}{d_n}+a_n \xrightarrow{D} X$, where $\xrightarrow{D}$ denotes convergence in distribution.*

Although this definition is quite interesting, it does not include the parameters of interest. Another definition to overcome this problem is through its characteristic function:

Definition 5. *A random variable $X \sim S(\alpha, \beta, \sigma, \mu)$, if its characteristic function has the following form, where $0 < \alpha \leq 2$ measuring the tail thickness,$-1 \leq \beta \leq 1$ determining skewness. And $\mu \in \mathbb{R}$ and $\sigma > 0$ are location and scale parameters in the sense that $\frac{X-\mu}{\sigma} \sim S(\alpha, \beta, 1, 0)$.*

$$\varphi(t) = Ee^{itX} = \begin{cases} exp(-\sigma^{\alpha}|t|^{\alpha}(1 - i\beta\frac{t}{|t|}tan(\frac{\pi\alpha}{2})) + i\mu t), \alpha \neq 1 \\ exp(-\sigma|t|^{\alpha}(1 + i\beta\frac{2}{|\pi|}\ln(t)) + i\mu t), \alpha = 1 \end{cases} \tag{4.1}$$

The non-Gaussian stable distribution is also called stable Pareto since the asymptotic tail behavior of stable laws is Pareto, namely, for sufficient large x, we have

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

where C_{α} is a function of α. More properties of stable distribution can be found in Samorodnitsky and Mittnik (1994).

4.2.2 Available Estimation Methods

The currently popular parameter estimation techniques are divided into three categories, viz. maximum likelihood estimation (MLE), characteristic function estimation (CFE) and quantile methods (QM). Theoretically, MLE is the most efficient estimator for large samples but is obtained at a high computational cost and is quite unstable. MLE requires careful numerical implementation for the density function and the maximum searching procedure (e.g., Nolan (2002)). The indirect inference method considered here is based on simulation that would avoid this problem. CFE was originally proposed by Press (1972) to avoid the inversion procedure for evaluating the density function. Other methods such as the iterative weighted regression method of Koustrouvellis (1980) were later proposed, thereby enabling simpler and more efficient estimation. However, a problem of these methods is the choice of the grids to evaluate the characteristic function in the regression. The ideal selection of grid is on a case-by-case basis. QM proposed by McCulloch (1986) gives simple estimators for four parameters in stable distributions based on five sample quantiles. The main idea is to match the functions of the sample quantiles with their theoretical counterparts using a table. This method is easy and convenient and avoids optimization. However, theoretical properties of these estimator are unclear. Moreover, interpolation is necessary when the sample values are not precisely equal to the tabulated theoretical values.

C.Gourierou and Renault (1993) first introduced indirect inference as a simulation based method for estimating the parameters of an extensive class of models. It was first applied in this context by Garcia et al. (2011). They suggest using the skewed-t distribution as an auxiliary model which has four parameters, each of which plays the same role as one of the parameters in the stable distribution. However, this setting is limited to the number of parameters in the auxiliary model, therefore is not flexible. Moreover, this method may not be as robust because the parameters of skewed-t are estimated by MLE. By setting the auxiliary parameters to be quantiles, the proposed method could guarantee robustness and be flexible. This latter idea of using quantiles as auxiliary parameters was first adopted by Dominicy and Veredas (2012). In their paper, method of simulated quantile (MSQ) was proposed. By setting the auxiliary parameters to be functions of quantiles (same as those used by McCulloch (1986)), MSQ extended the idea of QM. While QM is based on the tabulated tables, MSQ utilizes simulation. Such simulation based methods are flexible, and enable one to adjust the functions of quantiles or add further information to the the auxiliary parameters.

4.3 Indirect Inference in Stable Distributions

4.3.1 Indirect Inference Method

As mentioned earlier C.Gourierou and Renault (1993) first introduced indirect inference as a simulation based method for estimating the parameters of an extensive class of models. This method is particularly important when, the likelihood function is analytically intractable or considerably difficult to evaluate. The density of stable distribution does not have a closed form expression, thus has to be evaluated through characteristic function by Fourier inversion. This difficulty can be overcome through indirect inference, that greatly simplifies the estimation problem by only requiring that points could be simulated from the model.

The auxiliary parameter vector, which is denoted as $\boldsymbol{\pi}(\theta)$, is a function of $\boldsymbol{\theta}$, and has an easy-to-compute empirical estimator $\hat{\boldsymbol{\pi}}$. The relationship between $\hat{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}(\boldsymbol{\theta})$ is not required to be explicit, compared with the generalized method of moment (GMM) proposed by Hansen (1982). In general, an estimator of $\boldsymbol{\theta}$ could be defined by the solution of the following optimization problem:

$$\underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\ (\hat{\boldsymbol{\pi}}-\boldsymbol{\pi}(\boldsymbol{\theta}))^T\boldsymbol{\Omega}(\hat{\boldsymbol{\pi}}-\boldsymbol{\pi}(\boldsymbol{\theta}))$$

where $\boldsymbol{\Theta}$ is the parameter space, and $\boldsymbol{\Omega}$ is a positive definite weight matrix. The idea here is to find the parameter vector $\boldsymbol{\theta}$ such that $\hat{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}^*(\boldsymbol{\theta})$ are as close as possible. If $\boldsymbol{\pi}(\boldsymbol{\theta})$ could be calculated given $\boldsymbol{\theta}$, either by an explicit relationship or a function in a software, then the estimator could be obtained by the standard optimization algorithm. Otherwise, the estimator could be approximated by parametric bootstrap as follows:

Step 1 H samples of sample size N is simulated from $F_{\boldsymbol{\theta}}$.

Step 2 For each sample h,$h=1,2,\ldots,H$, its $\boldsymbol{\pi}^{*h}(\boldsymbol{\theta})$ is calculated based on its empirical distribution function.

Step 3 $\boldsymbol{\pi}(\boldsymbol{\theta})$ could be approximated by $\boldsymbol{\pi}^*(\boldsymbol{\theta})=\dfrac{1}{H}\sum_{h=1}^{H}\boldsymbol{\pi}^{*h}(\boldsymbol{\theta})$

Thus the indirect inference estimator $\hat{\boldsymbol{\theta}}$ is defined as follows:

$$\hat{\boldsymbol{\theta}}=\underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\ (\hat{\boldsymbol{\pi}}-\boldsymbol{\pi}^*(\boldsymbol{\theta}))^T\boldsymbol{\Omega}(\hat{\boldsymbol{\pi}}-\boldsymbol{\pi}^*(\boldsymbol{\theta}))$$

As a simulation based method, the auxiliary parameter $\boldsymbol{\pi}^*(\boldsymbol{\theta})$ is computed from the simulated sample. Bootstrap methods (see e.g. Efron (1979)) have been shown to operate a bias correction, as also in indirect inference (Gourierou et al. 1995). This feature gives indirect inference an advantage in finite sample estimation.

4.3.2 Quanitle-based Indirect Inference

Given the non-existence of moments in stable distribution, quantiles are a natural option for auxiliary parameters for the following reasons. First, the function of quantiles can be informative of the tail and skewness parameters in the stable distribution (McCulloch 1986). Moreover, the properties of quantile are necessary to derive the theoretical properties of the estimator. The asymptotic property in Lemma 4 enables us to derive the consistency and asymptotical efficiency of the estimator. Quantiles have a bounded influence function to ensure that the estimator is robust (Theorem 6).

For i.i.d. observations $\boldsymbol{X} = (X_1, X_2, \cdots, X_N)^T$ from the distribution function $F(\cdot)$, $\boldsymbol{\pi}(\boldsymbol{\theta})$ can be represented by $\boldsymbol{q}(\boldsymbol{\theta}) = (X_{p_1}, X_{p_2}, \cdots, X_{p_m})^T$, where $F(X_{p_i}) = p_i, 0 < p_i < 1, i = 1, 2, \cdots, m$. Thus, $\hat{\boldsymbol{q}}$ and $\boldsymbol{q}(\boldsymbol{\theta})$ represent the $m \times 1$ vectors of estimated and theoretical quantiles respectively. Given that the stable distribution is $F_{\boldsymbol{\theta}}$, with parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^4$, the proposed estimator is defined as follows:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmin}} \;\; (\hat{\boldsymbol{q}} - \boldsymbol{q}^*(\boldsymbol{\theta}))^T \, \boldsymbol{\Omega} \, (\hat{\boldsymbol{q}} - \boldsymbol{q}^*(\boldsymbol{\theta}))$$

The optimal choice of the weight matrix is given by $\boldsymbol{\Omega} = (Var(\hat{\boldsymbol{q}}))^{-1}$, thereby ensuring that the estimator $\hat{\boldsymbol{\theta}}$ is asymptotically efficient. However, $(Var(\hat{\boldsymbol{q}}))^{-1}$ is a function of $\boldsymbol{\theta}$ and a consequent manner to obtain an estimate for this matrix is through a two-step procedure similar to the two-step GMM as follows:

Step 1 $\boldsymbol{\Omega} = \boldsymbol{I}$ is used with $\boldsymbol{I}$ denoting the identity matrix, to solve the optimization problem and obtain the initial estimate $\boldsymbol{\theta}_1$.

Step 2 The weighting matrix is estimated with $\hat{\boldsymbol{\Omega}} = (Var(\hat{\boldsymbol{q}}(\boldsymbol{\theta}_1)))^{-1}$.

The expression for $Var(\hat{\boldsymbol{q}}(\boldsymbol{\theta}_1))$ can either be based on the asymptotic form of the variance of $\hat{\boldsymbol{q}}$ evaluated at $\boldsymbol{\theta}_1$ or can be obtained through parametric bootstrap with simulations using $\boldsymbol{\theta}_1$. This chapter adopts the latter approach.

The optimization algorithm used in this case is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm which is an iterative method that solves non-linear optimization problems. The quantile based indirect inference could be described by the genetic algorithm shown in Figure 4.1. $\boldsymbol{\pi}(\boldsymbol{\theta})(\boldsymbol{q}^*(\boldsymbol{\theta})$ in this case) could be calculated given a reasonable initial value of $\boldsymbol{\theta}$, which is obtained by QM. Thereafter an iterative process is triggered to search the optimal θ until some convergence criterions are satisfied.

Having defined our proposed estimator and described the procedure to obtain the estimator in practice, the next section studies the asymptotic and robustness properties of this estimator.

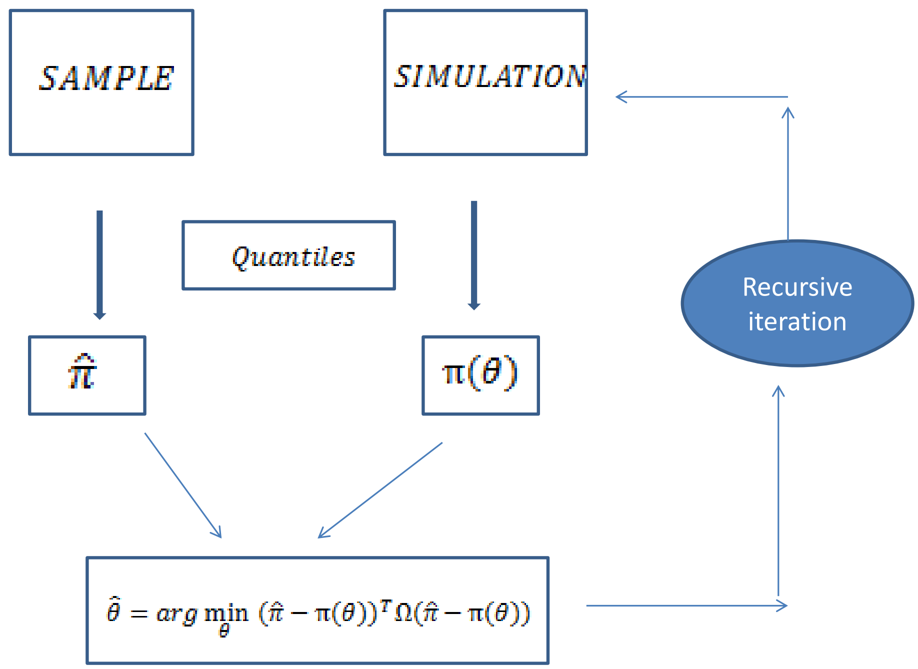

Figure 4.1: Estimation Algorithm

4.3.3 Theoretical Properties

Asymptotic Properties

To study the asymptotic properties we make use of the existing results and conditions for indirect estimators given in C.Gourierou and Renault (1993). Denoting $\boldsymbol{\theta}_0$ as the true parameter vector, let us first investigate the conditions which ensure the consistency and asymptotic normality of the proposed estimator in our case:

(A1) $\boldsymbol{\xi}_n = \sqrt{n}(\hat{\boldsymbol{q}} - \boldsymbol{q}(\boldsymbol{\theta}_0)) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{V})$ where $\boldsymbol{V} = \lim_{n\to\infty} Var(\boldsymbol{\xi}_n)$

(A2) There is a unique $\boldsymbol{\theta}_0$ such that sample quantiles equal the theoretical ones: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ if and only if $\hat{\boldsymbol{q}} = \boldsymbol{q}(\boldsymbol{\theta}_{\mathbf{0}})$.

(A3) If $\boldsymbol{\Omega}$ is estimated by $\hat{\boldsymbol{\Omega}}$, then $\hat{\boldsymbol{\Omega}} \xrightarrow{P} \boldsymbol{\Omega}$, where $\boldsymbol{\Omega} > 0$

(A4) $\boldsymbol{q}(\boldsymbol{\theta})$ is a differentiable function with $\mathbf{D}(\boldsymbol{\theta}) = \partial \boldsymbol{q}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^T$.

(A5) The matrix $\mathbf{D}^T(\boldsymbol{\theta})\,\boldsymbol{\Omega}\mathbf{D}(\boldsymbol{\theta})$ is full rank.

(A6) $\boldsymbol{\Theta}$ is compact.

(A7) The choice of the initial value of $\boldsymbol{\theta}$ is independent of the estimation algorithm.

Theorem 3. *(C.Gourierou and Renault 1993) and (Dominicy and Veredas 2012) Under the conditions of (A1)-(A7) and the other usual regularity conditions, our indirect estimator is asymptotically normal, when H is fixed and n goes to infinity:*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xmapsto{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda})$$

with $\boldsymbol{\Lambda} = (1 + \frac{1}{H})\boldsymbol{\Gamma}\mathbf{V}\boldsymbol{\Gamma}^T$ *where* $\boldsymbol{\Gamma} = \left(\mathbf{D}^T(\boldsymbol{\theta}_0)\,\boldsymbol{\Omega}\mathbf{D}(\boldsymbol{\theta}_0)\right)^{-1}\mathbf{D}^T(\boldsymbol{\theta}_0)\,\boldsymbol{\Omega}$.

This theorem provides the asymptotic normality of the estimator $\hat{\boldsymbol{\theta}}$ by that of auxiliary statistics $\hat{\boldsymbol{q}}$. Since the asymptotic normality is obtained, the consistency property follows. Notice, the factor $(1 + \frac{1}{H})$ distinguish the asymptotic variance of indirect inference with that of GMM: when H goes to infinity, they have the same expression. H is set to be 100 in this chapter. Now let us explain these conditions (A1) to (A7).

Condition (A1) is satisfied because of the following Lemma 4 and Lemma 5.

Lemma 4. *(Cramer 1946, page 369) Let* $0 < p_1 < \cdots < p_m < 1$. *Suppose that cumulative distribution function* F *has a density* f *in neighborhoods of quantiles* $\boldsymbol{q} = (X_{p_1}, \cdots, X_{p_m})^T$ *and that* f *is positive and continuous at* $\boldsymbol{q}$. *Then the empirical quantiles* $\hat{\boldsymbol{q}} = (\hat{X}_{p_1}, \cdots, \hat{X}_{p_m})^T$ *has asymptotically normal distribution:* $\sqrt{n}\,(\hat{\boldsymbol{q}} - \boldsymbol{q}) \xmapsto{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{V})$. *where the* (i, j)*-th element of covariance matrix* $\mathbf{V}$ *is* $\mathbf{V}_{ij} = \dfrac{p_i(1-p_j)}{f(X_{p_i})f(X_{p_j})} = \dfrac{p_i(1-p_j)}{f(F^{-1}(p_i))f(F^{-1}(p_j))}$, *for* $1 \le i \le j \le m$.

Lemma 5. *(Nolan 2015, page 12) All (non-degenerate) stable distributions are continuous unimodal distributions with an infinitely differentiable distribution function.*

Condition (A2) is often called the "global identifiability" problem in econometrics and is often hard to prove and such, is assumed in many cases. In indirect inference framework, the auxiliary parameters $\boldsymbol{\pi}(\boldsymbol{\theta})$ usually does not have an explicit expression which makes it even harder to verify. In condition (A3), our 2-step matrix $\hat{\boldsymbol{\Omega}}$ is estimated through the 2-step GMM procedure described above and thus is consistent (Hansen 1982). The rest of the conditions are standard conditions for indirect estimators such as the one put forward in this chapter. We therefore have that the estimator $\hat{\boldsymbol{\theta}}$ proposed here is consistent and asymptotically normally distributed.

Robustness Property

The use of quantiles as auxiliary parameters for estimation not only provides a wide range of auxiliary parameters which can make $\hat{\boldsymbol{\theta}}$ efficient but can also allow this estimator to be robust. Indeed, Genton and Ronchetti (2003) showed that if the auxiliary parameter $\boldsymbol{\pi}(\boldsymbol{\theta})$ in the indirect inference approach has a bounded influence function, then so does the indirect estimator $\hat{\boldsymbol{\theta}}$. The influence function is a tool used in robust statistics to study the impact of an infinitesimal contamination on a statistical functional (i.e. a test-statistic or estimator). If the latter is bounded, then the statistical functional is robust. Considering these results, we have the following theorem.

Theorem 6. *Our estimator $\hat{\boldsymbol{\theta}}$ has a bounded influence function, and thus is a robust estimator.*

The proof of this theorem together with an introduction of influence function can be found in Section 4.7. This result allows the proposed estimator to be robust implying that its bias will be bounded if the sample suffers from a small degree of contamination. This is especially important when choosing the quantiles to be used in the proposed indirect inference procedure.

4.4 Simulation Study

The estimator is approximated by Monte Carlo with B replications. For each parameter θ in $\boldsymbol{\theta}$, $E(\hat{\boldsymbol{\theta}}) \approx \dfrac{1}{B}\sum_{i=1}^{B}\hat{\boldsymbol{\theta}}_{\boldsymbol{i}}$. where each $\hat{\boldsymbol{\theta}}_{\boldsymbol{i}}$ is estimated by its individual sample with sample size N. The mean square error(MSE) is approximated by $MSE(\hat{\boldsymbol{\theta}}) = E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^2 \approx \dfrac{1}{B}\sum_{i=1}^{B}(\hat{\boldsymbol{\theta}}_{\boldsymbol{i}} - \boldsymbol{\theta})^2$. Since the MSE is estimated by simulation, some simulation bias correction techniques may apply (James and Anthony 1998) when the sample size N is small. However, N is set to be 1000 in this chapter.

We are interested in estimating parameters in the stable distribution with finite mean ($\alpha > 1$). This method is flexible because the auxiliary parameters could be adjusted on a case by case basis. If $\alpha > 1$ is known, then the mean could be added to the auxiliary parameters which may increase estimation efficiency although a few instances of robustness are lost. Thus for iid observations $\boldsymbol{X}$ from distribution function $F(\cdot)$, the auxiliary estimator $\hat{\boldsymbol{\pi}}$ is set to be quantiles $\hat{\boldsymbol{q}} = (X_{p_1}, X_{p_2}, \cdots, X_{p_m})^T$ plus sample mean $\bar{X}$: $\hat{\boldsymbol{\pi}} = (X_{p_1}, X_{p_2}, \cdots, X_{p_m}, \bar{X})^T$, where $F(X_{p_i}) = p_i, 0 < p_i < 1, i = 1, 2, \cdots, m$. The selected quantiles have equal space, that is, $p_{i+1} - p_i = p_i - p_{i-1}$, for $1 < i < m-1, p_0 = 0$.

4.4.1 Choice of the number of quantiles, m

The number of quantiles m is above or equal to 3. The best choice of m depends on sample size, parameters of interest and true parameter value. m

increases the dimension of auxiliary parameters, thereby due to the problem of collinearity, increasing m may negatively affect the estimation when m is already above a certain value. In other words, if the auxiliary parameters are already "sufficient statistics" for $\boldsymbol{\theta}$, adding more information will hurt the estimation. For select interesting case we could evaluate the best m by Monte Carlo studies. m is set as odd because the median could be included in the auxiliary parameters. The weight matrix used is the aforementioned two-step weight matrix.

The MSE of B=1000 Monte Carlo estimate is compared by selecting different number of quantiles of sample size N=1000 realization of iid random variables from $S(1.5, -0.2, 1, 0)$. Figure 4.1 to Figure 4.5 show the MSE of different m. Selecting $m = 3$ is ideal for location parameter μ, when only the first quantile, median and third quantile are adopted. If one pays considerable attention to the tail and skewness parameters, then $m = 9, 11$ minimize the MSE of $\hat{\alpha}$, and $m = 9$ minimizes the MSE of $\hat{\beta}$. $m = 9$ also has smallest sum of MSE, as shown in Figure 4.5. Thus, the auxiliary estimator $\hat{\boldsymbol{\pi}} = (X_{0.1}, X_{0.2}, \cdots, X_{0.9}, \bar{X})^T$. Selecting $m = 9$ may not be the best in every case, but is adopted here for simplicity in the rest of this chapter without seriously compromising the spirit of indirect inference.

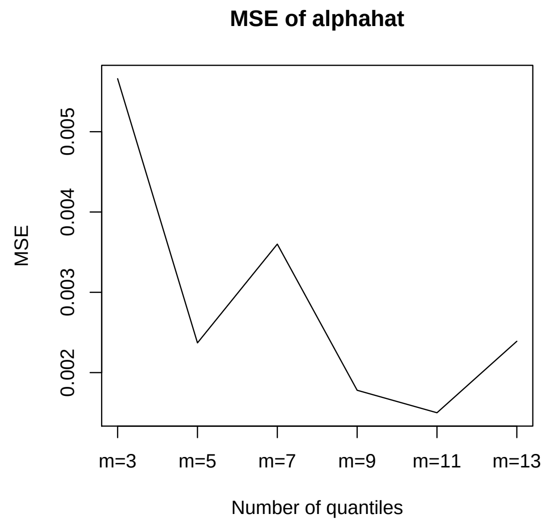

Figure 4.2: MSE of $\hat{\alpha}$ by different q

$MSE(\hat{\alpha}) = \dfrac{1}{B}\sum_{i=1}^{B}(\hat{\alpha}_i - \alpha)^2$, where the true parameter $\alpha = 1.5$. B=1000.

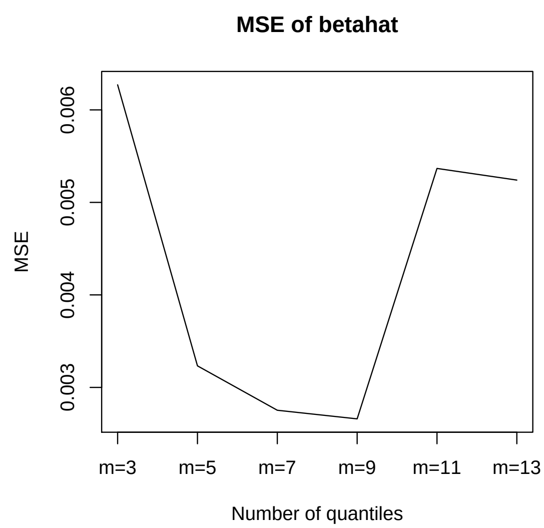

Figure 4.3: MSE of $\hat{\beta}$ by different q

$MSE(\hat{\beta}) = \dfrac{1}{B}\sum_{i=1}^{B}(\hat{\beta}_i - \beta)^2$, where the true parameter $\beta = -0.2$. B=1000.

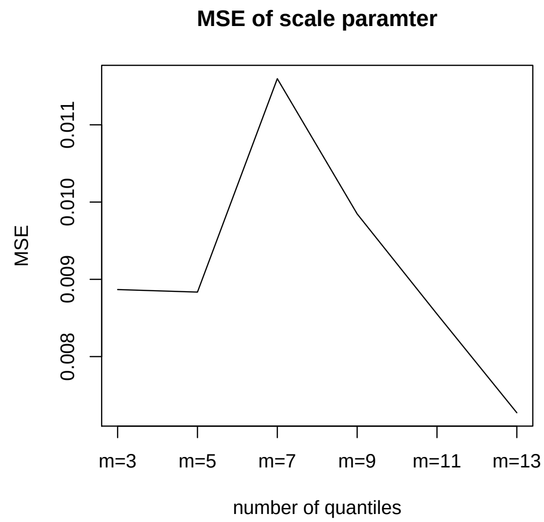

Figure 4.4: MSE of scale $\hat{\sigma}$ by different q

$MSE(\hat{\sigma}) = \frac{1}{B}\sum_{i=1}^{B}(\hat{\sigma}_i - \sigma)^2$, where the true parameter $\sigma = 1$. B=1000.

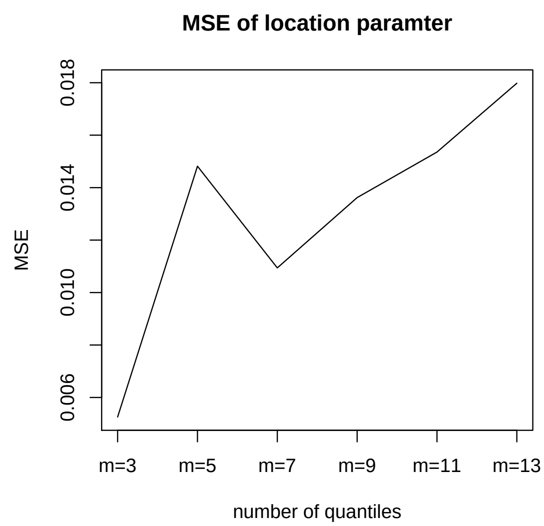

Figure 4.5: MSE of location $\hat{u}$ by different q

$MSE(\hat{\mu}) = \frac{1}{B}\sum_{i=1}^{B}(\hat{\mu}_i - \mu)^2$, where the true parameter $\mu = 0$. B=1000.

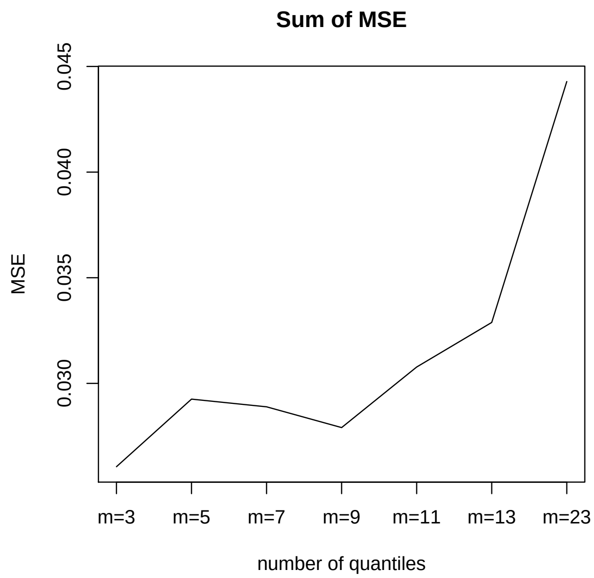

Figure 4.6: Sum of MSE

$$\text{Sum of MSE} = MSE(\hat{\alpha}) + MSE(\hat{\beta}) + MSE(\hat{\sigma}) + MSE(\hat{\mu})$$

4.4.2 Weight Matrix

The previously proposed identity matrix and the two-step weight matrix were compared at certain points of their MSE using Monte Carlo. Table 4.1 shows the MSE of α at certain points when other parameters are fixed at particular values. When α is close to 1, the two-step weight matrix performs better than the identity matrix. By contrast, the identity matrix is good when α is close to 2.

A trade-off is determined between the benefit of using a weight matrix and the estimation error of the weight matrix. When α is close to 1, less weight given on the tail makes the two-step weight matrix have considerably small MSE. When α is close to 2, the estimation error of the weight matrix makes an identity matrix

$N = 1000, B = 1000$ $\beta = -0.2, \sigma = 1, \mu = 0$		
	Identity matrix	Two-step weight matrix
$\alpha = 1.2$	0.0092	0.0005
$\alpha = 1.3$	0.0067	0.0005
$\alpha = 1.4$	0.0052	0.0009
$\alpha = 1.5$	0.0047	0.0013
$\alpha = 1.6$	0.0040	0.0023
$\alpha = 1.7$	0.0033	0.0037
$\alpha = 1.8$	0.0023	0.0030
$\alpha = 1.9$	0.0014	0.0025

$N = 1000, B = 1000$ $\alpha = 1.5, \sigma = 1, \mu = 0$		
	Identity matrix	Two-step weight matrix
$\beta = 0$	0.0045	0.0057
$\beta = -0.1$	0.0050	0.0051
$\beta = -0.2$	0.0082	0.0069
$\beta = -0.3$	0.0121	0.0103
$\beta = -0.4$	0.0235	0.0177
$\beta = -0.5$	0.0247	0.0169

Table 4.1: MSE comparison: Identity matrix vs Two-step weight matrix

All parameters are assumed to be unknown and have to be estimated. The true parameter value is assumed to be known when evaluating the MSE.

better. Overall, identity matrix performs better when the distribution is close to Gaussian where β is close to 0 and α is close to 2. Two-step weight matrix performs well when the distribution is heavy-tailed and skewed.

4.4.3 Comparison Between different Methods

The MSE of B=1000 Monte Carlo estimate of N=1000 realizations of iid random variables from $S(\alpha, -0.2, 1, 0)$ and $S(1.5, \beta, 1, 0)$ by different methods. In indirect inference method, two-step weight matrix is selected for $\alpha \leq 1.6$, $\beta \leq -0.2$, otherwise, the identity matrix is chosen. Table 4.2 shows that indirect inference method has considerably small mean square error compared with other methods when the distribution is heavy tailed (i.e., α is close to 1) and close to symmetric (i.e., β is close to 0).

4.5 Case Study

Mandelbrot (1963) and Fama (1963) proposed that the stable distribution could be a candidate model to characterize asset returns. A few opinions criticize the stable distributions without bounded variation. Moreover, the iid assumption seems naive that it could not model the volatility clustering phenomena of asset return. The stable distribution remains a robust model that identifies heavy tail and skewness. McCulloch (1997) analyzed 40 years of monthly stock price data from the Center for Research in Security Prices and concluded a good fit. Nolan (2005) analyzed 16 years of monthly return of exchange of British Pound vs. German Mark and calculated the Value at risk based on the stable distribution.

$\beta = -0.2$ $\sigma = 1$ $\mu = 0$				
	Indirect inference	MLE	Quantile method	Regression method
$\alpha = 1.2$	0.0005	0.0018	0.0027	0.0029
$\alpha = 1.3$	0.0005	0.0020	0.0033	0.0031
$\alpha = 1.4$	0.0009	0.0025	0.0036	0.0034
$\alpha = 1.5$	0.0013	0.0026	0.0046	0.0035
$\alpha = 1.6$	0.0023	0.0026	0.0057	0.0036
$\alpha = 1.7$	0.0033	0.0024	0.0075	0.0030
$\alpha = 1.8$	0.0023	0.0019	0.0091	0.0027
$\alpha = 1.9$	0.0014	0.0013	0.0080	0.0018

$\alpha = 1.5$ $\sigma = 1$ $\mu = 0$				
	Indirect inference	MLE	Quantile method	Regression method
$\beta = 0$	0.0045	0.0096	0.0126	0.0161
$\beta = -0.1$	0.0050	0.0099	0.0137	0.0147
$\beta = -0.2$	0.0069	0.0095	0.0117	0.0158
$\beta = -0.3$	0.0103	0.0097	0.0127	0.0144
$\beta = -0.4$	0.0177	0.0084	0.0136	0.0153
$\beta = -0.5$	0.0169	0.0078	0.0156	0.0144

Table 4.2: MSE of different methods

This table evaluates the MSEs of $\hat{\alpha}$ and $\hat{\beta}$ at different points using different methods. Although the other parameters are known, they are assumed to be unknown. Thus, all methods will estimate the four parameters. MLE refers to Nolan (2002), Quantile method refers to McCulloch (1986) and regression method refers to Koustrouvellis (1980). The 3 methods are carefully implemented by a program called STABLE on Nolan's personal website: http://academic2.american.edu/ jpnolan/stable/stable.html.

The Data used in the current study are the daily return of S&P 500 from January 1, 2008 to January 1, 2011 (757 trading days). We let $S_i, i = 1, 2, \cdots, 757$ be the closing price (index) on that day. The daily return is defined as $R_i =$

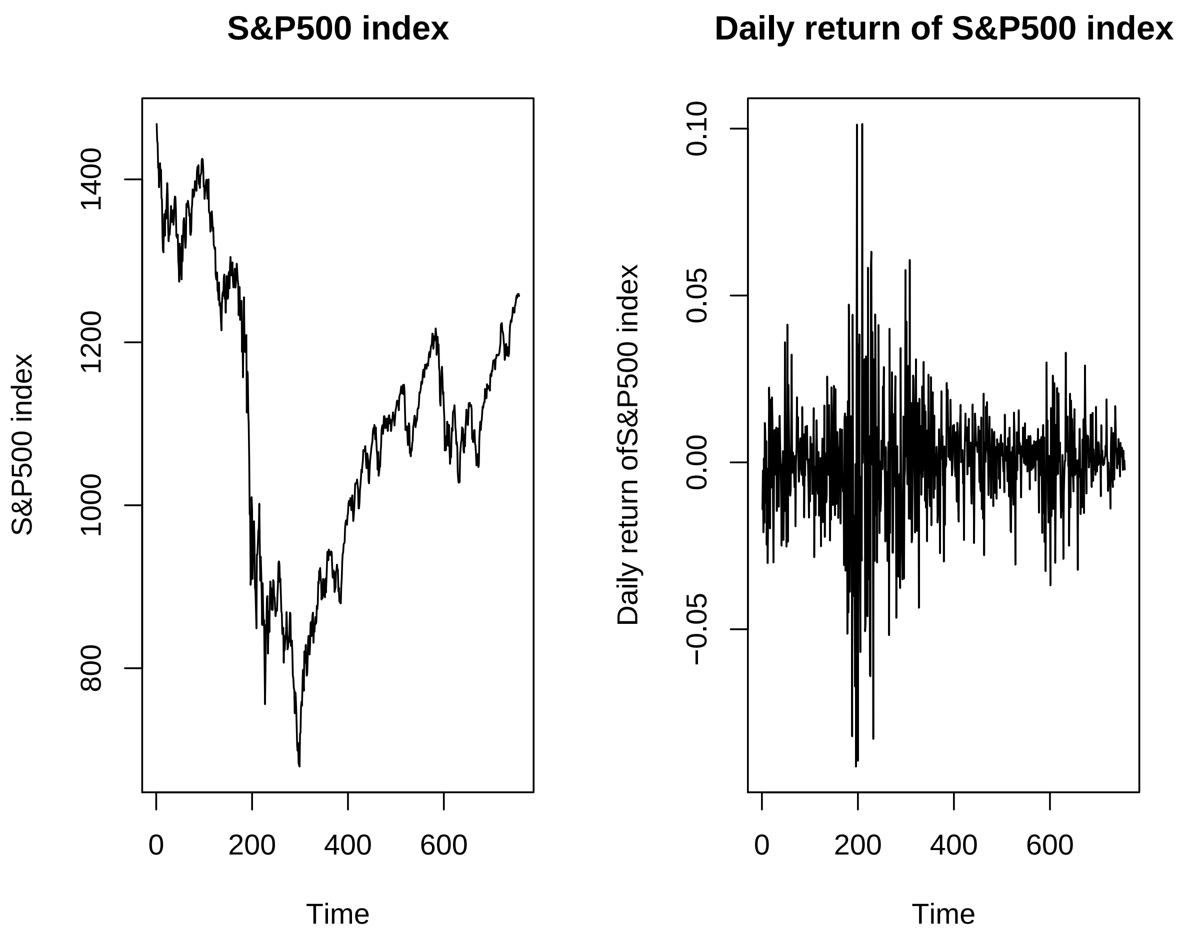

Figure 4.7: Plot of index and return

A total of 757 trading days, and thus 756 daily returns. In the x-axis of the index, 0 represents the January 1, 2008, which is the starting day. In the x-axis of the daily return, 0 represents January 2, 2008.

$\log \frac{S_i}{S_{i-1}}$. The jump of index shown in Figure 4.6 is due to the financial crisis fueled by the collapse of subprime mortgage-backed securities. The histogram and QQ normal plot shows that the data has a serious heavy tail and a possible negative skewness.

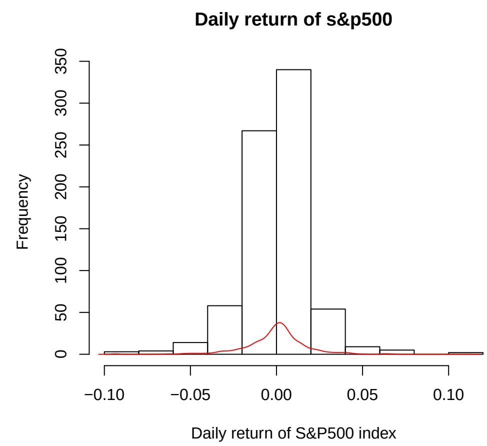

Figure 4.8: histogram

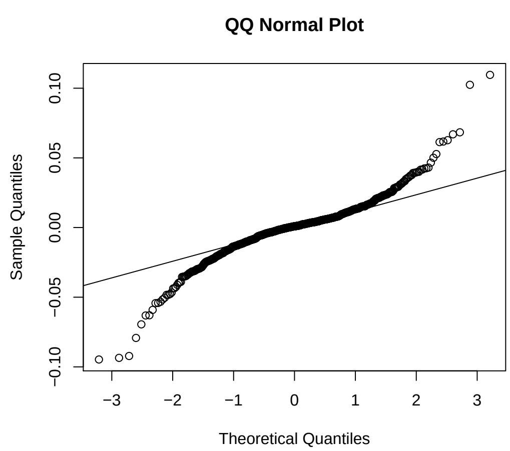

Figure 4.9: QQ plot

A central issue in this study is the test of skewness. If no skewness of data is determined, then the data could be modeled by t-distribution or symmetric stable distribution which has less parameters. A nonparametric asymptotic test could be developed based on the following statistic:

$$S_3 = \frac{(X_{0.75} - X_{0.5}) - (X_{0.5} - X_{0.25})}{X_{0.75} - X_{0.25}} \tag{4.2}$$

This statistics was first proposed by Bowely (1920) and was previously used in QM. This statistics converges to symmetric normal distribution where the variance could be quantified under the null. Ekström and Jammalamadaka (2012) extended this test by adding additional quantiles. They conclude that, a reasonably good test may rely on five quantiles as follows:

$$S_5 = \frac{(X_{0.9} - X_{0.5}) + (X_{0.8} - X_{0.5}) - (X_{0.5} - X_{0.2}) - (X_{0.5} - X_{0.1})}{(X_{0.9} - X_{0.1}) + (X_{0.8} - X_{0.2})} \tag{4.3}$$

Under the null which says the distribution is symmetric, S_5 will converge to $N(0, V)$ where V is a function of the density function f (Ekström and Jammalamadaka 2012). f would be approximated by the kernel density estimator with normal kernel. If S_3 is applied, then the p-value of this test is 0.0161. If S_5 is applied, then the p-value is 0.0060. The distribution is slightly negatively skewed. Hence the candidate model would be asymmetric stable distribution or skewed-t distribution. Skewed-t distribution is introduced by Fernandez and Steel (1998). It has four parameters, each parameter plays the same role as the one in stable distribution. Table 4.3 shows the estimated value of these two models. The quantile-based indi-

Parameters	Alpha stable	Skewed-t
Tail thickness	1.3500	2.233
Skewness	-0.1490	0.9121
Location	0.0077	-0.0005
Scale	0.0109	0.0318

Table 4.3: Stable vs Skewed-t

rect inference is applied in the stable distribution and MLE is applied in skewed-t distribution.

4.6 Conclusion

Quantile-based indirect inference for the stable distributions is studied in this chapter. Asymptotic and robust properties of these estimators have been shown when quantiles are chosen as the auxiliary parameters. Quantile-based indirect inference has several advantages compared with other methods in terms of estimating stable distribution. First, it only requires that distribution can be simulated, and thus avoids numerical evaluation of density and/or distribution function. Secondly, the simulation study shows that this method has a considerably small mean square error at heavy tailed points compared with other methods. Third, the method is robust because the quantiles are adapted. Finally, this method is considerably flexible, i.e. the auxiliary parameter could be adjusted, the candidate distribution could be changed and certain parameters could be fixed easily. This feature is beneficial for the goodness of fit test and model selection. As a final comment, it is acknowledged that parts of the work in this chapter overlap with that of Dominicy and Veredas (2012), and this was discovered only after all the work in this chapter was completed.

4.7 Appendix: Influence Function and Robust Property of Quantiles

Let $\boldsymbol{A}$ be a convex subset of the set of all finite signed measures on $\boldsymbol{\Sigma}$. We want to estimate the parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ of a distribution F in $\boldsymbol{A}$. Let the functional $\boldsymbol{T} : \boldsymbol{A} \to \Gamma$ be the asymptotic value of some estimator sequence $(\boldsymbol{T}_n)_{n \in \mathbb{N}}$. We will suppose that this functional is Fisher consistent, i.e. $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{T}(F_{\boldsymbol{\theta}}) = \boldsymbol{\theta}$. This means that at the model F, the estimator sequence asymptotically measures the correct quantity. Let $x \in \boldsymbol{\chi}$, Δ_x is the probability measure which gives mass 1 to x. The influence function is then defined by

$$IF(x; \boldsymbol{T}; F) := \lim_{\varepsilon \to 0} [\frac{\boldsymbol{T}((1-\varepsilon)F + \varepsilon \Delta_x) - \boldsymbol{T}(F)}{\varepsilon}]. \tag{4.4}$$

The influence function describes the effect of an infinitesimal contamination at the point x on the estimate we are seeking. For a robust estimator, we want a bounded influence function, that is, one which does not go to infinity as x becomes arbitrage large.

Let F be strictly increasing with positive density f, $\phi = T(F) = F^{-1}(p)$ be the p^{th} quantile. The influence function of quantile could be obtained (Hinkley 1974):

$$IF(x) = \begin{cases} \dfrac{p-1}{f(\phi)}, x < \phi \\ \\ \dfrac{p}{f(\phi)}, x > \phi \end{cases} \tag{4.5}$$

As x goes to infinity, $IF(x)$ is bounded by $\frac{p}{f(\phi)}$. Then the influence function of our auxiliary parameter $\boldsymbol{\pi}(\theta) = (F^{-1}(p_1), F^{-1}(p_2), \cdots, F^{-1}(p_m))^T$ is therefore bounded by the chain rule as described by Lemma 7.

Lemma 7. *(Hinkley 1974) Suppose statistical functionals take the form* $\boldsymbol{T}(F) = \boldsymbol{a}(T_1(F), \cdots, T_m(F)) = \boldsymbol{a}(t_1, \cdots, t_m)$. $IF_i(x)$ *is the influence function of* $T_i(F)$*, for* $i = 1, 2, \cdots, m$*. By the chain rule,the influence function of* $\boldsymbol{T}(F)$ *is*

$$IF(x) = \sum_{i=1}^{m} \frac{\partial \boldsymbol{a}}{\partial t_i} IF_i(x) \tag{4.6}$$

Chapter 5

Indirect Inference Applied to Income Distributions

The distribution of income and wealth play an important role in the measurement of inequality and poverty among people as well as nations. Various methods and different models for income distribution are developed in a number of articles by many economists— see e.g. Chotikapanich et al. (2007), McDonald and Xu (1995). This chapter provides an extension of the work in Hajargasht et al. (2012) and suggests a general method of fitting income distributions. In their paper, Generalized Method of Moments (GMM) method is applied to estimate the income distribution which may take several parametric forms. For each parametric form, the explicit expressions of the moment conditions are needed. In this chapter, the indirect inference method allows us to estimate income distribution without specifying the explicit expression for the moments.

This chapter is organized as follows. In Section 2, we give a brief introduction to some measures of inequality including the Gini index and the Lorenz Curve

(LC). Also, some popular parametric income distributions are introduced. In Section 3, we point out that indirect inference method is a suitable approach for these types of data sets. Theoretical properties of this estimator and a goodness-of-fit test are provided. In Section 4, we test the optimization algorithm used in our method. Also a Monte Carlo study is conducted to compare and evaluate these estimators. In Section 5, we illustrate our method by comparing the income distributions and inequality indices for both China and USA over the past 30 years.

5.1 Introduction to Some Inequality Measures

5.1.1 Lorenz curve

Let $x_1 \leq x_2 \leq \cdots \leq x_n$ be ordered data, say on incomes. The empirical Lorenz Curve is defined as

$$L(i/n) = s_i/s_n \tag{5.1}$$

where $s_i = x_1 + x_2 + \cdots + x_i, L(0) = 0, i = 0, \cdots, n$.

Let x_i denote data drawn from the distribution function $F(x)$ with mean μ. Let z_p denote the quantile corresponding to a proportion $0 \leq p \leq 1$ i.e.

$$p = F(z_p) = \int_0^{z_p} f(t)\, dt \tag{5.2}$$

Distribution	CDF	lorenz curve
Exponential	$F(x) = 1 - \exp^{-\lambda x}, x > 0$	$p + (1-p)\log(1-p)$
General Uniform	$F(x) = \dfrac{x-a}{\theta}, a < x < a+\theta$	$\dfrac{ap + \theta p^2/2}{a + \theta/2}$
Pareto	$F(x) = 1 - (a/x)^a, x > a, a > 1$	$1 - (1-p)^{(a-1)/a}$
lognormal	$F(x) = 1/2 + 1/2\,\mathrm{erf}[\dfrac{\log x - \mu}{\sqrt{2}\sigma}]$	$\Phi(\Phi^{-1}(p) - \sigma)$

Table 5.1: Lorenz Curve for some distributions

and then the theoretical Lorenz Curve is defined

$$L(p) = \mu^{-1} \int_0^z t f(t)\, dt = \frac{\int_0^z t f(t)\, dt}{\int_0^\infty t f(t)\, dt} \tag{5.3}$$

The numerator sums the incomes of the bottom p proportion of the population, while the denominator sums the incomes of all the population.

Assuming that F is continuous, one may write $z = F^{-1}(p)$ and a change of variable to write the LC in a direct way:

$$L(p) = \mu^{-1} \int_0^p F^{-1}(t)\, dt \tag{5.4}$$

Table 5.1 shows LC expression for some common distributions. Notice that, for exponential distributions, LC does not depend on the scale-parameter. This property could be used for goodness of fit tests (see Gail and Gastwirth (1978)). Figure 5.1 compares LC for lognormal and exponential.

5.1.2 Gini Index and Other Inequality Measures

Gini index is a number between 0 and 1 which gives information about the income inequality of a country, and is the most commonly used measure of inequal-

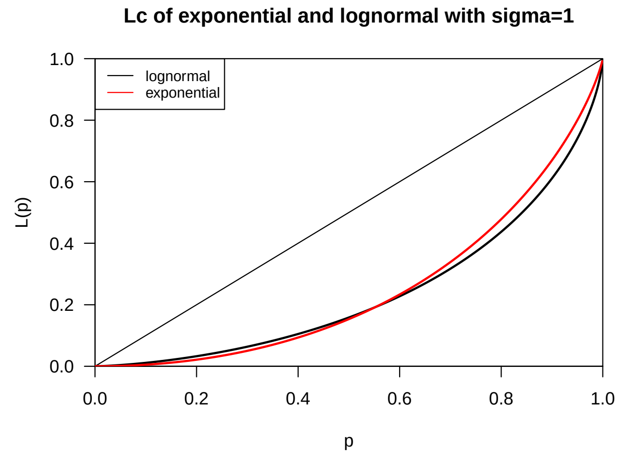

Figure 5.1: Lorenz Curve of lognormal and exponential

ity. It is also a U-statistic widely used in goodness of fit tests. Jammalamadaka and Goria (2004) introduced a test of goodness of fit based on Gini index of spacings. Recently, Noughabi (2014) introduced a general test of goodness of fit based on the Gini index of data. One way to define Gini index is through expected mean difference.

Definition 6. *Gini* $:= \dfrac{E|X-Y|}{2 \cdot E(X)}$ *where* X, Y *are two random points drawn independently from the distribution* F.

The sample version could be written in the following way:

$$Gini(S) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2(n-1) \sum_{i=1}^{n} x_i} \tag{5.5}$$

It could also be calculated via LC (Gastwirth 1972):

$$G(t) = 2 \cdot \int_0^1 (t - L(t))\, dt \tag{5.6}$$

5.1.3 Some Popular Parametric Income Distributions

The income distribution is heavily positively skewed and has a long right tail. The popular income distribution models include Generalized Beta-2 distribution, Gamma distribution and the lognormal distribution.

Generalized Beta-2 distribution (5.7) is widely used for modeling income distribution. Beta-2 ($a = 1$), Singh-Maddala ($p = 1$), Dagum ($q = 1$) and Generalized gamma ($q \to \infty$) are special cases of Generalized beta-2 distribution (see McDonald and Xu (1995)).

$$f(x; a, b, p, q) = \frac{ax^{ap-1}}{b^{ap}B(p, q)(1 + (x/b)^a)^{p+q}}, x > 0 \tag{5.7}$$

Lognormal distribution (5.8) is another popular income distribution model, its pdf could be derived from $log(X) = Y$ which has a normal distribution.

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(log(x) - \mu)^2}{2\sigma^2}}, x > 0, \sigma > 0 \tag{5.8}$$

Many alternate models exist, but as Cowell (1995) says, the more complicated four parameters densities are not particularly good choices. Their parameters are hard to interpret and may have an over-fitting problem. He is more in favor of lognormal and gamma density which has two parameters. Among the distribution with two parameters, the Pareto density is nice for modeling high incomes while gamma and lognormal are nice for modeling middle range incomes. In this chapter, lognormal distribution is chosen for illustrative purposes.

5.2 Indirect Inference Method

We have described the general methodology of indirect inference and properties of resulting estimators, in Section 4.3

5.2.1 Indirect inference framework

Remember in Chapter 4, the indirect inference estimator for $\boldsymbol{\theta}$ is defined as

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \quad (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^*(\boldsymbol{\theta}))^T \, \boldsymbol{\Omega} \, (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^*(\boldsymbol{\theta})) \tag{5.9}$$

The auxiliary estimator $\hat{\boldsymbol{\pi}}$ is set to be the sample mean and 9 points on empirical LC in Table 5.3: $\hat{\boldsymbol{\pi}} = (\bar{X}, \hat{L}(0.1), \cdots, \hat{L}(0.9))$. The auxiliary parameters corresponds to the theoretical mean and 9 points on theoretical LC implied by lognormal distribution. As opposed to the GMM, $\boldsymbol{\pi}^*(\boldsymbol{\theta})$ will be calculated by parametric bootstrap. $\boldsymbol{\Omega}$ is estimated by 2-step weight matrix. The details of this estimation algorithm is already described in Chapter 4.

5.2.2 Theoretical Properties

Compare with the auxiliary parameters in Chapter 4, here we replace quantile $F^{-1}(\boldsymbol{p})$ with LC $L(\boldsymbol{p})$. $L(\boldsymbol{p})$ and $F^{-1}(\boldsymbol{p})$ share the same properties: under some mild conditions, Goldie (1977) proved that the empirical LC $L_n(\boldsymbol{p})$ converges, uniformly to the theoretical LC $L(\boldsymbol{p})$. Also, he derived the weak convergence of the Lorenz process $l_n(\boldsymbol{p}) = \sqrt{n}[L_n(\boldsymbol{p}) - L(\boldsymbol{p})], 0 \leq \boldsymbol{p} \leq 1$, to a Gaussian process if $L(\boldsymbol{p})$ is continuous at the empirical points. Thus the asymptotical property of our auxiliary parameters $L(\boldsymbol{p}, \boldsymbol{\theta})$ is established. The consistency and asymptotic normality of our estimator $\hat{\boldsymbol{\theta}}$ could be obtained by Theorem 3 if Conditions (A2)-(A7) hold.

5.2.3 Goodness of Fit Analysis

Since $l_n(\boldsymbol{p}) = \sqrt{n}[L_n(\boldsymbol{p}) - L(\boldsymbol{p})], 0 \leq \boldsymbol{p} \leq 1$ converges to the Gaussian process. The J-test (Hansen 1982) could be developed through the following theorem:

Theorem 8. *(Hayashi 2009) If $\boldsymbol{y} \sim N(0, \boldsymbol{I}_p)$ and $\boldsymbol{A}$ is an idempotent matrix with rank R, then $\boldsymbol{y}^T \boldsymbol{A} \boldsymbol{y} \sim \chi^2_R$.*

Here the Test statistics J_n can be used to test the validity of the assumed income distribution.

$$J_n = n \left(\hat{L}(\boldsymbol{p}) - L(\boldsymbol{p}, \hat{\boldsymbol{\theta}}) \right)^T \hat{\boldsymbol{\Omega}} \left(\hat{L}(\boldsymbol{p}) - L(\boldsymbol{p}, \hat{\boldsymbol{\theta}}) \right) \xrightarrow{D} \chi^2_{M-K} \tag{5.10}$$

In this case, the dimension of auxiliary parameters $M = 10$, the number of parameter in lognormal $K = 2$. The sample size n is the number of the surveyed citizens which is unknown. The test results varies for different choices of n. Hajargasht et al. (2012) assume $n = 10000$ in their paper.

5.2.4 Data

The data comes from the Website of the World Bank, it takes the form of summary statistics including mean income, measures of inequality and 9 points

USA 's Income share by deciles(%)										
Year	lowest	2nd	3rd	4th	5th	6th	7th	8th	9th	highest
2010	1.70	3.40	4.56	5.73	7.00	8.44	10.19	12.52	16.25	30.19

USA's poverty index				
Year	mean($/month)	pov.line	headcount(%)	Gini index(%)
2010	1917.38	1.90	1.00	41.06

Table 5.2: Original Data

By deciles(%)									
p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{L}(p)$	1.70	5.10	9.66	15.39	22.39	30.83	41.02	53.54	69.77

Table 5.3: Transformed Data

on the empirical LC. In Table 5.2, the poverty line is the minimum level of income deemed adequate in a particular country. The head-count ratio is the proportion of a population lives below the poverty line. The first part of Table 5.2 shows the data in the following way: the first 10% of the population owns 1.7% of the total income, the second 10% of the population owns 3.4% of the total income, etc. Since the sum of these 10 numbers equals 1, only the numbers of the first 9 groups need to be included in the moment conditions. The cumulation of these 9 numbers yields the 9 points on the empirical LC $\hat{L}(\boldsymbol{p})$ in Table 5.3.

With our indirect inference estimator $\hat{\boldsymbol{\theta}}$, $\hat{L}(\boldsymbol{p})$ and $L(\boldsymbol{p}, \hat{\boldsymbol{\theta}})$ are compared as shown at Table 5.4 . This table could be extended for different models to assess the goodness of fit.

USA 2010									
p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{L}(\boldsymbol{p})$	1.70	5.10	9.66	15.39	22.39	30.83	41.02	53.54	69.77
$L(\boldsymbol{p}, \hat{\boldsymbol{\theta}})$	2.15	5.63	10.14	15.63	22.31	30.59	40.66	52.98	69.17

Table 5.4: Goodness of Fit Assessment

5.3 Simulation Study

5.3.1 Numerical Optimization

The default optimization algorithm used in R is Broyden-Fletcher-Goldfarb-Shanno(BFGS) algorithm. Similar to Newton's method, it is a iterative method solving non-linear optimization problems. In this case, the parameter space of σ is $(0, \infty)$. Since it has a lower bound, sometimes this optimization algorithm breaks down when searching the nearby points slightly bigger than 0. Instead, we would estimate the parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)$, where $(\mu, \sigma) = (\theta_1, \exp(\theta_2))$. The estimated parameter $\hat{\sigma}$ approximately equals to $\log(\hat{\theta_2})$.

Here we want to verify that the estimated point is the local minimum. The true parameters $\boldsymbol{\theta} = (4.8276, -0.4963)$ is obtained from the estimate value of data in Table 5.2. The data (9 points on lorenz curve and mean) is simulated from lognormal distribution with above parameters with sample size $N = 1000$. The estimated value $\hat{\boldsymbol{\theta}} = (4.8381, -0.4515)$. It has a local minimum as we could see from Figure 5.2 and Figure 5.3.

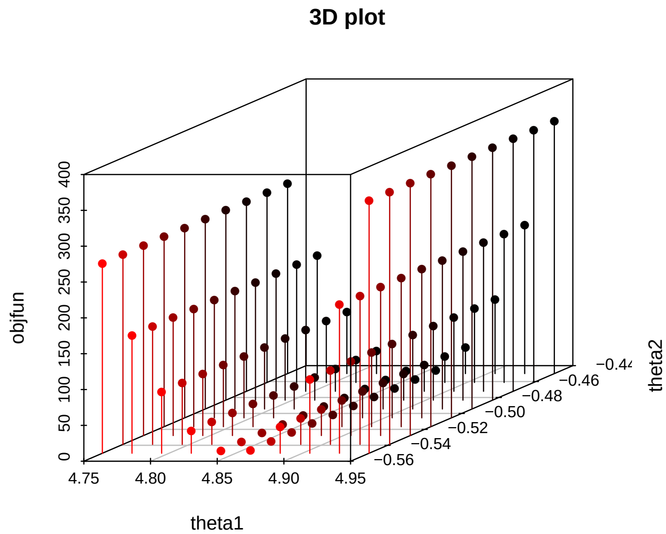

Figure 5.2: Objective function vs (θ_1, θ_2)

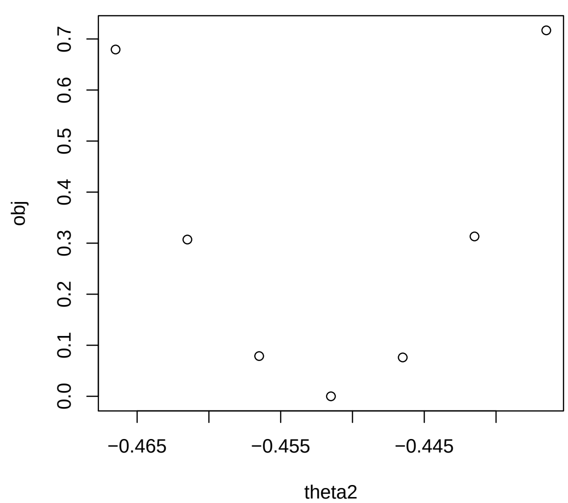

Figure 5.3: Objective function vs θ_2

5.3.2 Monte Carlo Study

Suppose we only have the 9 points on the LC, sample median and sample mean. For lognormal distribution, the mean $EX = \exp(\mu + \sigma^2/2)$, Median $m = \exp(\mu)$. By setting these equal to their empirical parts, a method of moment estimator has obtained:

$$\hat{\mu} = \log(m), \hat{\sigma} = \sqrt{2(\log(\bar{x}) - \log(m))} \tag{5.11}$$

Suppose the true parameters $(\mu, \sigma) = (4.8276, \exp(-0.4963))$. Box-plots to compare these two estimators are obtained by Monte Carlo study with sample size $N = 1000$ and Monte Carlo replication $B = 1000$ in Figure 5.4 and Figure 5.5. Our indirect inference method has smaller variance especially for σ.

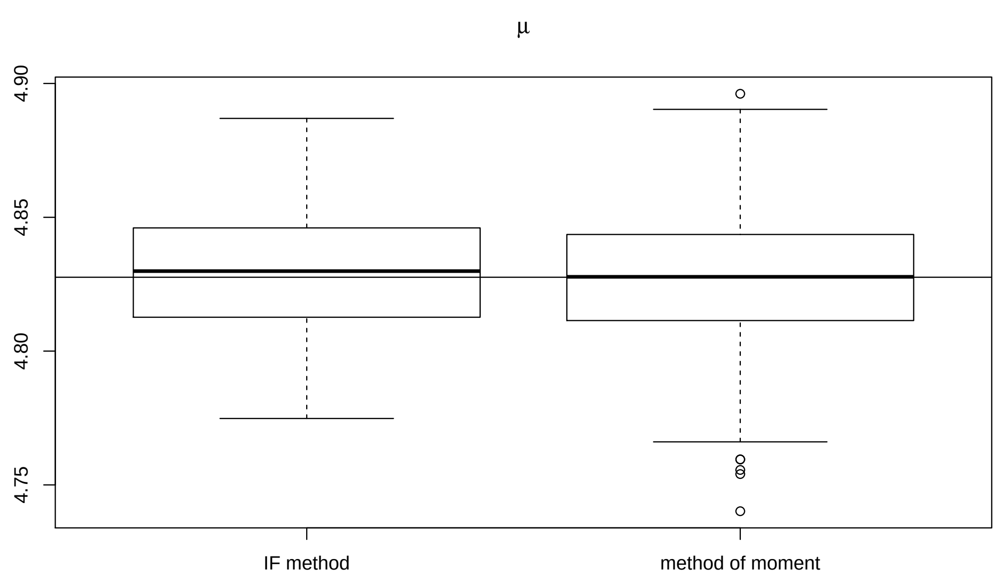

Figure 5.4: Boxplot of $\hat{\mu}$

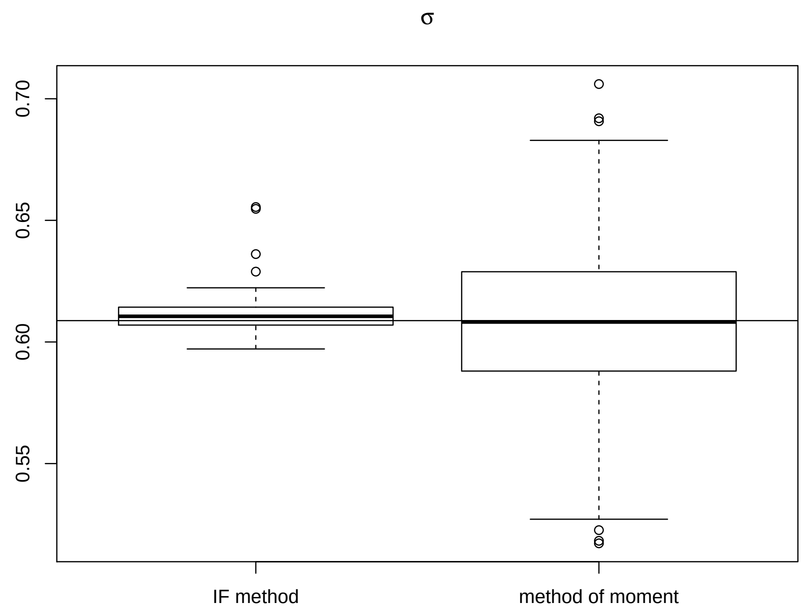

Figure 5.5: Boxplot of $\hat{\sigma}$

5.4 Case Study

USA and China are currently the largest two economies in the world. In 2015, the nominal GDP of USA is \$18,287 billion while the nominal GDP of China is \$11,285 billion. It is known that China keeps a high growing rate in the last 35 years as we could see in Figure 5.6.

Greenwood and Jovanovic (1990) found a positive correlation between growth and income inequality in a cross-section of international data. Here we are interested to see whether economic growth brings more income inequality in China and USA. In this section, a comparison of USA and China's income distribution and inequality in the last 30 years is illustrated.

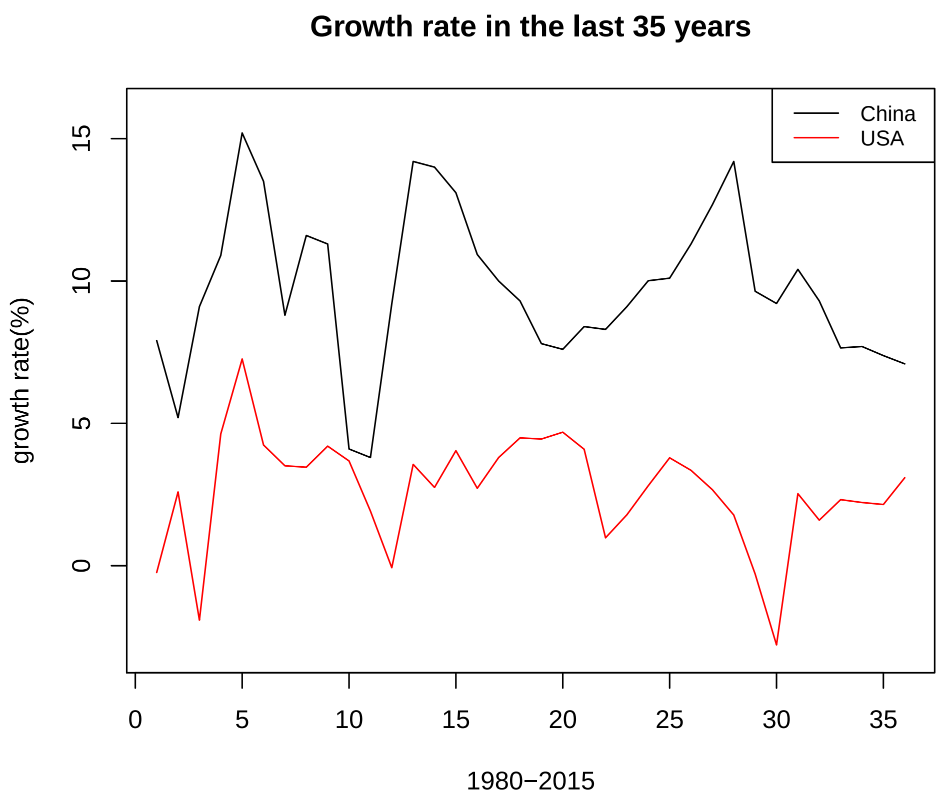

Figure 5.6: GDP growth rate

1 represents year 1980, 36 represents year 2015.

5.4.1 Data

Data is collected every 3 years by the World Bank. It takes the form of summary statistics as shown at Table 5.5 and Table 5.6.

USA 's Income share by deciles(%)										
Year	lowest	2nd	3rd	4th	5th	6th	7th	8th	9th	highest
2010	1.70	3.40	4.56	5.73	7.00	8.44	10.19	12.52	16.25	30.19
1981	1.81	3.59	5.00	6.23	7.51	8.95	10.69	12.96	16.40	26.86

Poverty Index				
Year	mean($/month)	pov.line($/day)	headcount(%)	Gini index(%)
2010	1917.38	1.9	1	41.06
1981	1581.81	1	0.67	37.73

Table 5.5: Income inequality of USA: 1981 v.s 2010

China 's Income share by deciles(%)										
Year	lowest	2nd	3rd	4th	5th	6th	7th	8th	9th	highest
2010	1.69	2.98	4.23	5.51	6.88	8.43	10.31	12.88	17.11	29.98
1981	3.72	4.96	6.05	7.08	8.12	9.25	10.58	12.31	15.08	22.86

Poverty Index				
Year	mean($/month)	pov.line($/day)	headcount(%)	Gini index(%)
2010	218.54	1.9	11.18	42.06
1981	34.64	1	88.32	18.46

Table 5.6: Income inequality of China: 1981 v.s 2010

5.4.2 Result

With the 9 points on the empirical LC, a smooth empirical LC is estimated by the non-parametric spline technique in R. The income distributions are assumed to be lognormal and are estimated by above indirect inference method. The Results are illustrated from Figure 5.7 to Figure 5.11.

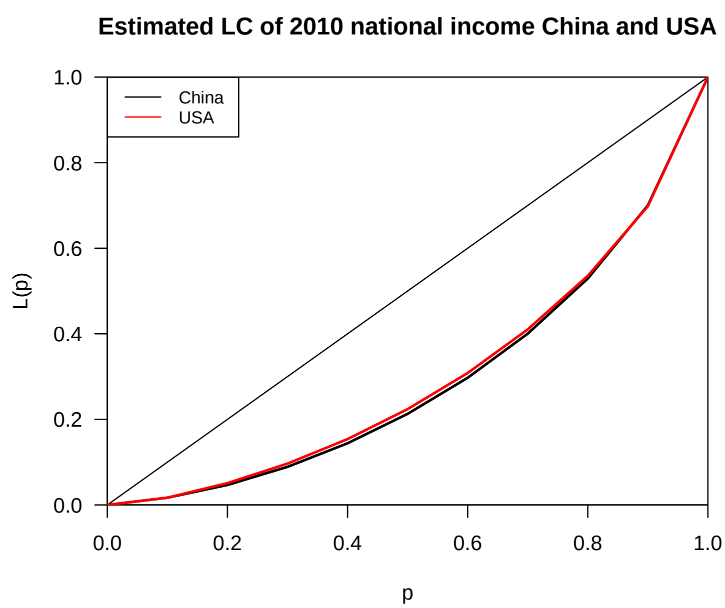

Figure 5.7: Lorenz curve 2010 USA vs China

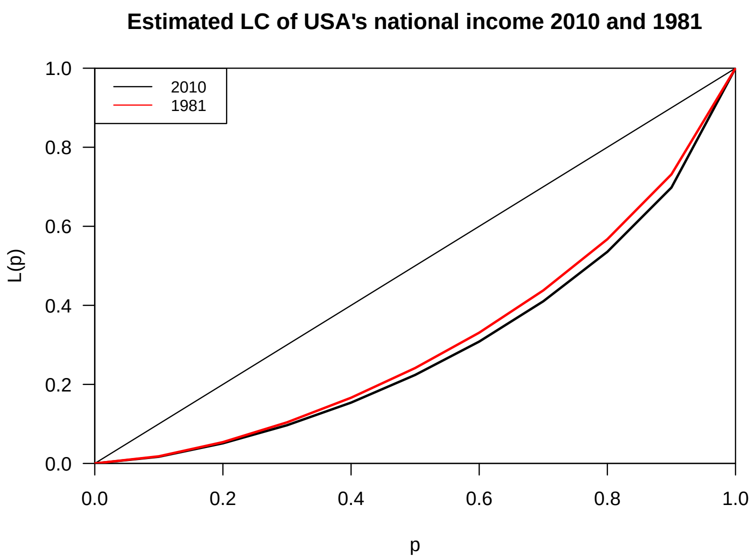

Figure 5.8: Lorenz Curve of USA 1981 vs 2010

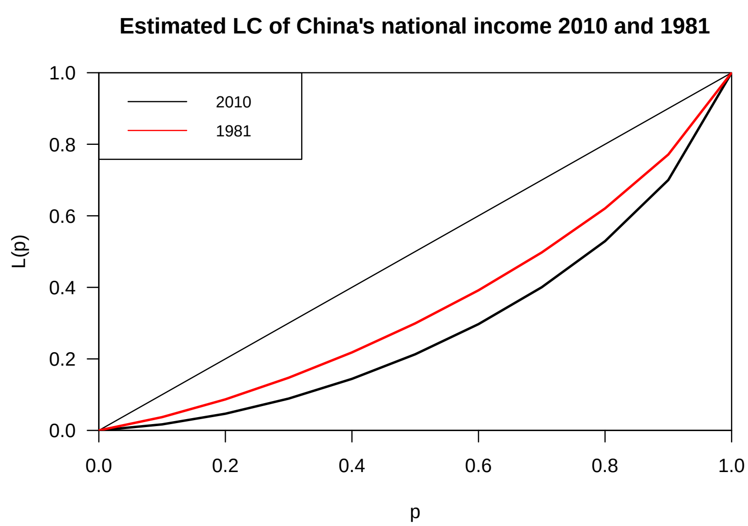

Figure 5.9: Lorenz Curve of China 1981 vs 2010

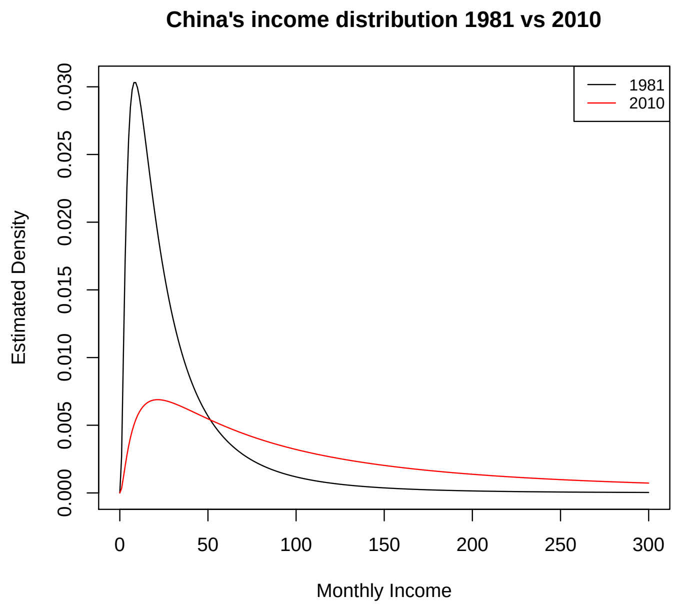

Figure 5.10: Income distribution of China 1981 vs 2010

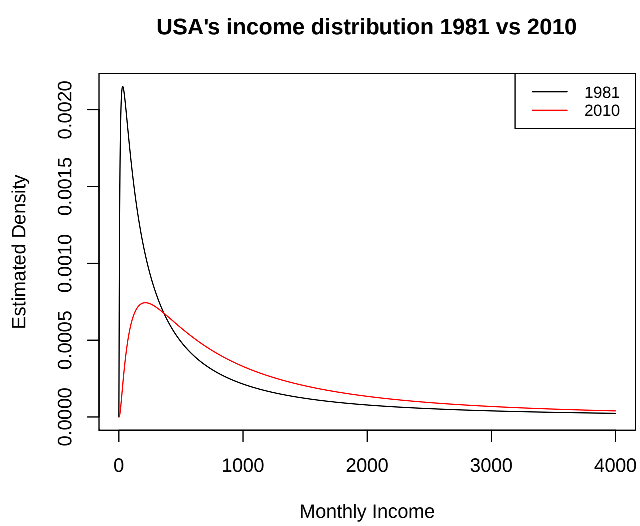

Figure 5.11: Income distribution of USA 1981 vs 2010

5.4.3 Conclusions

Kuznets (1995) has advanced the conjecture that evolution of income distribution follows an inverted U-shaped curve: growth results in relatively more inequality in the initial stage of economic development, and greater equality at advanced stages. But this statement is controversial: M.Ravallion (1995) among others, showed that there is no empirical support for this conjecture.

Our analysis of the data we looked at, seems to partially support this conjecture. As we could see from our example, the Gini index of China increased from 18.46% to 42.06%, however the poverty is significantly improved due to the overall income increase. Compared with China's big change, inequality indices and income distribution of USA are stable over the last 30 years. Interestingly, USA and China's LC are close in 2010.

Chapter 6

Conclusions and Discussion

Conclusions

Two new estimation methods are introduced in this thesis in connection with estimating the parameters of a stable distribution: spacing based estimation and indirect inference. For spacing based estimation in stable distributions in chapter 3, we showed that it performs as good as the MLE for large samples. Also we concluded that it is a flexible method as one has the choice of distance measures that could be used. As for the indirect inference, we developed a general framework for estimating stable distribution as well as income distribution with limited data. This simulation based method is very flexible, namely that the parametric model and/or auxiliary parameters could be adjusted. In Chapter 4, we showed that this method has the smallest mean square error among the existing popular methods of estimating stable distribution parameters, at most parameter values. In Chapter 5, we developed a practical estimation framework of analyzing income distribution and income equality given the limited data. This analytical tool helps develop some interesting practical conclusions as we showed in that chapter.

Future Work

Our work could be extended in several directions. For the linear stable distribution we studied, it could be transformed to wrapped stable distribution. After wrapping, the trigonometric moments and likelihood start to exist. Method of trigonometric methods is applied by Gatto and Jammalamadaka (2003). Spacing-based idea could also be applied for wrapped distribution, either for inference or goodness of fit testing.

Since stable distributions do not have finite variance which is a major drawback for their application in finance. Different schemes for the truncation were proposed. Tempered stable distributions and process proposed by Rosiński (2002) is a popular one, and has been widely applied in finance (for example, see Kim and Rachev (2009)). For the calibration, our quantile-based indirect inference method is applicable if one wants to circumvent this standard, but complicated analytical methods.

Appendix A

Code

A.1 R code of Estimating income distribution by indirect inference

```
* Emperical function: 9 points on Emperical lorenz curve and mean
emp.fun = function(x){
mean = mean(x)
x = sort(x)
N = length(x)
q010a=sum(x[1:round(N*0.10)])
q020a=sum(x[1:round(N*0.20)])
q030a=sum(x[1:round(N*0.30)])
q040a=sum(x[1:round(N*0.40)])
q050a=sum(x[1:round(N*0.50)])
q060a=sum(x[1:round(N*0.60)])
q070a=sum(x[1:round(N*0.70)])
q080a=sum(x[1:round(N*0.80)])
q090a=sum(x[1:round(N*0.90)])

q010p=q010a/sum(x)
q020p=q020a/sum(x)
q030p=q030a/sum(x)
q040p=q040a/sum(x)
q050p=q050a/sum(x)
q060p=q060a/sum(x)
q070p=q070a/sum(x)
```

```
q080p=q080a/sum(x)
q090p=q090a/sum(x)

g.x = c(mean,q010p,q020p,q030p,q040p
,q050p,q060p,q070p,q080p,q090p)
return(g.x)
}

***Theoretical function
theo.fun = function(theta){
mu = theta[1]
sigma = exp(theta[2])

g.theta = matrix(NA,H,10)
for (j in 1:H){
set.seed(j + 13212341)
x.star = rlnorm(n, meanlog=mu,sdlog=sigma)
g.theta[j,] = emp.fun(x.star)
}

g.theta = apply(g.theta,2,mean)
return(g.theta)
}

***Objective function
obj.fun = function(theta){
theo = theo.fun(theta)
dif = theo - emp.estim
obj = (t(dif))\%*\%Omega\%*\%dif
return(obj)
}

*** 2-step procedue weight matrix.

Omega=I   *set Omega to be identity matrix
thetahat1=optim(theta.start,obj.fun)\$par   * initial estimate

boot.Var = function(theta, B = 1000){

emp.boot = matrix(NA,B,10)
for (i in 1:B){
set.seed(i)
```

```
x.star = rlnorm(n,meanlog=theta[1],sdlog=exp(theta[2]))
emp.boot[i,]= emp.fun(x.star)
}
return(cov(emp.boot))
}                  *Covariance matrix approximation

V=boot.Var(thetahat1)
Omegahat=solve(V)       *2-step weight matrix is obtained

***Obtain the final estimate thetahat2 by optimization
Omega=Omegahat   *Set Omega to be 2-step weight matrix
thetahat2=optim(theta.start,obj.fun)\$par   *final estimate

*** Gini index function
Gini.fun = function(x){
inter = 0
N = length(x)
  for (i in 2:N){
  for (j in 1:(i-1)){
inter = inter + abs(x[i]-x[j])
  }
}
  Gini = (1/(N*(N-1)))*inter/mean(x)
  return(Gini)
}
*** Headcount ratio caculation
n.pov=which.min(abs(x-pl))     *pl represents poverty line
HC=n.pov/N                            *HC is headcount ratio
```

A.2 Matlab code for spacing based estimation of stable distribution

```
*Stable random number generator stblrnd(alpha,beta,gamma,delta,N)
if alpha == 2                 * Gaussian distribution
    r = sqrt(2) * randn(N);
elseif alpha==1 \&\& beta == 0       * Cauchy distribution
    r = tan( pi/2 * (2*rand(N) - 1) );
```

```
elseif alpha == .5 \&\& abs(beta) == 1   * Levy distribution
    r = beta ./ randn(N).\^2;

elseif beta == 0                * Symmetric alpha-stable
    V = pi/2 * (2*rand(N) - 1);
    W = -log(rand(N));
    r = sin(alpha * V) ./ ( cos(V).\^(1/alpha) ) .* ...
        ( cos( V.*(1-alpha) ) ./ W ).\^( (1-alpha)/alpha );

elseif alpha ~= 1               * General case, alpha not 1
    V = pi/2 * (2*rand(N) - 1);
    W = - log( rand(N) );
    const = beta * tan(pi*alpha/2);
    B = atan( const );
    S = (1 + const * const).\^(1/(2*alpha));
    r = S * sin( alpha*V + B ) ./ ( cos(V) ).\^(1/alpha) .* ...
       ( cos( (1-alpha) * V - B ) ./ W ).\^((1-alpha)/alpha);

else                            * General case, alpha = 1
    V = pi/2 * (2*rand(N) - 1);
    W = - log( rand(N) );
    piover2 = pi/2;
    sclshftV =  piover2 + beta * V ;
    r = 1/piover2 * ( sclshftV .* tan(V) - ...
        beta * log( (piover2 * W .* cos(V) ) ./ sclshftV ) );

end

*  Scale and shift
if alpha ~= 1
   r = gamma * r + delta;
else
   r = gamma * r + (2/pi) * beta * gamma * log(gamma) + delta;
end
end

 *  Spacing estimation for alpha and beta
   X = stblrnd(1.5,0,1,0,1000);
   gam=1;
   delta=0;
   y=sort(X);
```

```
 F=@(theta)[0,(stblcdf(y,theta(1),theta(2),gam,delta))'];
 G=@(theta)[(stblcdf(y,theta(1),theta(2),gam,delta))',1];
 H=@(theta)-(sum(log(G(theta)-F(theta))));


[theta,fval]=fminsearch(H,[1.4,0.1]);

*two dimensional  graph
  alpha=linspace(1.1,1.9,10);
  beta=linspace(-1,1,10);
  z=zeros(10,10);
  for i=1:10
      for j=1:10
          z(i,j)=H([alpha(i),beta(j)]);
      end
  end
 subplot(2,2,1)
 surf(alpha,beta,z);
 title('Greenwood statistics vs (alpha,beta)');
```

Bibliography

Bowely, A. (1920). *Elements of statistics*, Charles Scribner's Sons.

C.Gourierou, A. M. and Renault, E. (1993). Indirect inference, *Journal of Applied Econometrics* **8**: 1–10.

Chambers, J. et al. (1976). A method for simulating stable random variables, *JASA* **71**: 340–344.

Cheng, R. and Amin, N. (1983). Estimating parameters in continuous univariate distributions with a shifted origin, *J.Roy. Statist.Soc.Ser.B* **45**: 394–403.

Chotikapanich, D. et al. (2007). Estimating and combining national income distribution using limited data, *Journal of Business & Economic statistics* **25**: 97–109.

Cowell, F. (1995). *Measuring Inequality*, Harvester Wheatsheaf, Hemel Hempstead.

Cramer, H. (1946). *Mathematical Method of Statistics*, Princenton University Press, Princeton NJ.

Dominicy, Y. and Veredas, D. (2012). The method of simulated quantiles, *Journal of Econometrics* **172**: 235–247.

Efron, B. (1979). Boostrap methods: Another look at jackknife, *Journal of Annual Statistics* **7**: 1–26.

Ekström, M. and Jammalamadaka, S. R. (2012). An asymptotically distribution-free test of symmetry, *Journal of Statistical Planning and Inference* **137**: 799–810.

Fama, E. F. (1963). Mandelbrot and the stable paretian hypothesis, *The Journal of Business* **36**: 420–429.

Fernandez, C. and Steel, M. (1998). On beyesian modelling of fat tails and skewness, *Journal of the American Statistical Association* **93**: 359–371.

Gail, M. H. and Gastwirth, J. L. (1978). A scale-free goodness of fit test for the exponential distribution based on the lorenz curve, *Journal of the American Statistical Association* **73**: 787–793.

Garcia, R. et al. (2011). Estimation of stable distribution by indirect inference, *Journal of Econometrics* **161**: 325–337.

Gastwirth, J. L. (1972). The estimation of the lorenz curve and gini index, *The Review of Economics and statistics* **54**: 306–316.

Gatto, R. and Jammalamadaka, S. (2003). Inference for wrapped symmetric alpha stable circular models, *The Indian Journal of Statistics* **65**: 333–355.

Genton, M. G. and Ronchetti, E. (2003). Robust indirect inference, *Journal of the American Statistical Association* **98**: 67–76.

Ghosh, K. and Jammalamadaka, S. (2001). A general estimation method using spacing, *Journal of Statistical Planning and Inference* **93**: 71–82.

Gnedenko, B. and Kolmogorov, A. (1954). *Limit distributions for sums of independent random variables*, Cambridge,Addison-Wesley.

Goldie, C. M. (1977). Convergence theorems for empirical lorenz curve and their inverses, *Advances in Applied Probability* **9**: 765–791.

Gourierou, C. et al. (1995). *Calibration by Simulation for Small Sample Bias Correction*, INSEE, Paris.

Greenwood, J. and Jovanovic, B. (1990). Financial development, growth, and the distribution of income, *The Journal of Political Economy* **98**: 1076–1107.

Hajargasht, G. et al. (2012). Inference for income distribution using grouped data, *Journal of Business & Economic statistics* **30:4**: 563–575.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica* **50**: 1029–1054.

Hayashi, F. (2009). *Econometrics*, Princeton University Press.

Hinkley, D. C. . D. (1974). *Theoretical statistics*, Chapman & Hall, London.

James, G. and Anthony, A. (1998). Approximate bias correction in econometrics, *Journal of Econometrics* **85**: 205–230.

Jammalamadaka, S. R. and Goria, M. (2004). A test of goodness of fit based on gini's index of spacing, *Statistics & probability letters* **68**: 177–187.

Kim, Y. and Rachev, S. (2009). *Risk Assessment*, Phyica-Verlag Heidelberg.

Koustrouvellis, I. (1980). Regression type estimation of the parameters of stable laws, *American Statistical Association* **75**: 918–928.

Kozubowski, T. J. et al. (2005). Operator geometric stable laws, *Journal of Multivariate Analysis* **92**: 298–323.

Kuznets, S. (1995). Economic growth and income inequality, *American Economic Review*.

Lévy, P. (1924). Théorie des erreurs la loi de gauss et les lois exceptionelles, *bulletin de la Socété de France* **52**: 49–85.

Mandelbrot, B. (1961). Stable paretian random functions and the multiplicative variation of income, *Econometrica* **29**: 517–543.

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

McCulloch, J. (1986). Simple consistent estimator of stable distribution parameters, *Communication in Statistics* **15**: 1109–1136.

McCulloch, J. H. (1997). Measuring tail thickness to estimate stable index: A critique, *Journal of Business and Economic statistics* **15**: 74–81.

McDonald, J. and Xu, Y. (1995). A generalization of the beta distribution with applications, *Journal of Econometrics* **66**: 133–152.

Mittnik, S. et al. (1999). Computing the probability density function of stable paretian distribution, *Mathematical and Computer Modelling* **29**: 235–240.

M.Ravallion (1995). Growth and poverty:evidence for developing countries in the 1980s, *Economics letters* **48**: 411–417.

Nolan, J. (1997). Numerical computation of stable densities and distribution functions, *Stochastic models* **13**: 759–774.

Nolan, J. (2002). Maximum likelihood estimation and diagnostic for stable distributions.

Nolan, J. (2005). Modeling financial data with stable distributions.

Nolan, J. (2015). *Stable Distributions-Models for Heavy Tailed Data*, Birkhauser, Boston.

Noughabi, H. (2014). A test of goodness of fit based on gini index, *Journal of the turkish statistical association* **7**: 23–32.

Press, S. (1972). Estimation in univariate and multivariate stable distribution, *JASA* pp. 842–846.

Ranneby, B. (1984). The maximum spacing method, an estimation method related to maximum likelihood method, *Scandinavian Journal of Statistics* **11**: 93–112.

Rosiński, J. (2002). Tempering stable processes, *Stochastic Processes and their Applications* **117**: 677–707.

Samorodnitsky, G. and Mittnik, S. (1994). *Stable Non-Gaussian Random Processes*, Chapman & Hall.

Zolotarev, V. (1986). One-dimensional stable distributions, *American mathematical Society* **65**: 284–293.

Zolotarev, V. (1995). On the representations of stable laws by special functions, *Theory Probab. Appl.* **39**: 354–362.