On Robust Asymptotic Theory of Unstable AR(p) Processes with Infinite Variance

Maryam Sohrabi

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in
Mathematics [1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1] The Ph.D. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

In this thesis, we explore some asymptotic results in heavy-tailed theory. There are many empirical and compelling evidence in statistics that require modeling with heavy tailed observations.

This thesis is divided into three parts. First, we consider a robust estimation of the mean vector for a sequence of independent and identically distributed observations in the domain of attraction of a stable law with possibly different indices of stability, $DS(\alpha_1, \ldots, \alpha_p)$, such that $1 < \alpha_i \leq 2$, $i = 1, \ldots, p$. The suggested estimator is asymptotically normal with unknown parameters. We apply an asymptotically valid bootstrap to construct a confidence region for the mean vector. Furthermore, a simulation study is performed to show that the estimation method is efficient for conducting inference about the mean vector for multivariate heavy-tailed observations.

In the second part, we present the asymptotic distribution of M-estimators for parameters in an unstable AR(p) process. The innovations are assumed to be in the domain of attraction of a stable law with index $0 < \alpha \leq 2$. In particular, when the model involves repeated unit roots or conjugate complex unit roots, M-estimators have a higher asymptotic rate of convergence compared to the least square estimators. Moreover, we show that the asymptotic results can be written as Itô stochastic integrals.

Finally, the preceding methodologies lead to develop the asymptotic theory of M-estimators for parameters in unstable AR(p) processes with nonzero location parameter. Similar to the preceding cases, we assume that the process is driven by innovations in the domain of attraction of a stable law with index $0 < \alpha \leq 2$. In this thesis, for all models, we also cover the finite variance case ($\alpha = 2$).

Acknowledgements

I would not have been able to write this thesis without the invaluable help and the continuous support that I received from my supervisor, Dr. Mahmoud Zarepour. I offer my sincerest gratitude to Dr. Mahmoud Zarepour, who has supported me with his patience, assistance, motivation, immense knowledge and enthusiasm throughout my Ph.D. study and research.

My sincere thanks also go to my thesis examining committee. I would like to thank professor Andrei Igorevich Volodin, professor Chul Park, professor Mayer Alvo, and professor Rafal Kulik for their brilliant comments, encouragement, insightful questions, and thoughtful suggestions. In addition, I thank with deep appreciation professor Mohamedou Ould Haye and professor Raluca Balan for their contribution to my learning and ultimate success in the Ph.D. process.

I take this opportunity to express gratitude to the Department of Mathematics and Statistics of University of Ottawa and Carleton and Faculty of Graduate and Postdoctoral Studies which provided me the excellent academic environment. I also appreciate the financial support from University of Ottawa Admission Scholarship and Ontario Government Scholarship that funded parts of the research discussed in this dissertation.

Most importantly, none of this would have been possible without the love and patience of my family. I would like to express my special thanks to my husband, Sadegh, for his endless love and support. A heartfelt thank to my mother, father, Mohammad, Forough, Marjan, and Hasti for their support, inspiration, motivation, and endless patience at all times.

Last but not the least, I would like to thank my officemates and friends in the Department of Mathematics and Statistics for all times we were working together and for all the fun we have had in the last four years.

Dedication

To my parents Fereshteh and Shaban.

All I have and will accomplish are only possible due to their love and sacrifices.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Historical background

Heavy tails are characteristic of phenomena where the probability of observing values with very large magnitude is relatively big. A distribution is said to be heavy-tailed if its tails behave asymptotically like a power function. The standard Pareto distributions, the family of stable distributions, and the extreme-value Frechet distributions are some examples of heavy-tailed distributions.

The application of heavy-tailed distributions have been considered from the 1960s by papers of Mandelbrot [32, 33] and Fama [17]. They show that the distributions of commodity and financial returns are often heavy-tailed with possible infinite variance, and thus the corresponding risk management often relies on heavy-tail analysis.

Heavy-tailed distributions also have diverse applications in data networks. For instance, data sets coming from the teletraffic industry such as file lengths, CPU time to complete a job, call holding times seem to have heavy-tailed distributions, for example see [54]. Resnick [47] reviews various estimation and model identification methods for the time series coming from teletraffic industry which exhibit many non-standard characteristics such as heavy-tailed distributions. For a comprehensive review on heavy-tailed distributions and their applications, we refer to [1], [35], and [49].

On the other hand, most statistical experiments are multivariate by nature and explosive advances in computer technology make it feasible to work with large scale multivariate data sets. Several authors contribute to identify the multivariate heavy-tailed distributions and their domains of attraction. For example, see [21], [44], and [48].

Due to the complexity of multivariate stable densities, most researchers assume that the indices of stability for all components are the same. However, in many real life examples, some coordinates may have light tails while other coordinates may have heavier tails. Resnick and Greenwood [44] consider the limiting distribution of observations in the domain of a bivariate stable law with possibly different indices of stability. A similar idea is also considered by Zarepour and Roknossadati [58] for a multivariate random-walk model, when the innovations are in the domain of attraction of a multivariate stable law with different indices of stability in their coordinates.

1.2 Thesis context

In this thesis, we are interested in models involving univariate and multivariate heavy-tailed distributions. The asymptotic theory in modeling such phenomena is tail dependent and different from classical modeling and regular statistical analysis. We study the limit theory for the M-estimators of parameters of the corresponding models. Our approach also covers innovations with finite variance.

This thesis is divided into three parts. First, we develop a large sample inference procedure for the population mean vector for a sequence of random vectors with multivariate heavy-tailed distributions.

The sample mean is the most common statistic to make inference for the mean when observations have finite and infinite variance. Although the sample mean is a consistent estimate for the population mean, the convergence of the sample mean is very slow when the observation are in the domain of attraction of a stable law with tail index close to 1.

On the other hand, the asymptotic distribution of the sample mean depends on the unknown parameters and thus cannot be used for inference. A bootstrap procedure can rectify this difficulty. It has been shown that the regular bootstrap is not consistent for estimating the distribution of the mean. Athreya [4] shows that the bootstrap distribution of the sample mean of infinite variance observations does not converge weakly to a fixed distribution. Instead the convergence occurs in distribution to a random probability distribution. Several investigations have been considered to enforce the bootstrap to work. For example, a bootstrap with m out of n subsampling procedure when $m/n \to 0$ is asymptotically valid. Some of these investigations can be found in [3], [20], and [26].

In this thesis, we propose a robust estimate method (M-estimate method) to make inference about the mean vector when the observations belong to the domain of attraction of a multivariate stable law with possibly different indices of stability. We will show that the regular bootstrap will remain valid if we make inference about the mean vector using a certain objective function.

For the second part of this thesis, we study the limiting behaviour of an autoregressive process driven by infinite variance innovations in the domain of attraction of a stable law with index $0 < \alpha \leq 2$. Consider the autoregressive process of order p (AR(p))

$$\phi(B)X_t = \epsilon_t, \tag{1.2.1}$$

where B is the backward operator and

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p.$$

The errors $\{\epsilon_t\}$ in (1.2.1) form a sequence of independent and identically distributed (i.i.d.) random variables in the domain of attraction of a symmetric stable law with index $0 < \alpha \leq 2$. The model (1.2.1) is referred to as non-stationary autoregressive time series, if the characteristic polynomial $\phi(\cdot)$ has at least one root on the boundary of the unit circle. It is well known that the unit root tests are particulary a descriptive tool performed to classify time series as stationary and non-stationary.

It has always been an interesting and challenging task to conduct asymptotic inference for time series with unit roots. In the case where the errors (innovations) have finite variance, Anderson [2], Dickey and Fuller [15], and Philips and Perron [43] provide the asymptotic theory for the least squares (LS) estimators in an AR(1) process with one unit root. Chan and Wei [11] provide the fundamental results on non-stationary autoregressive AR(p) model, when the innovations form a sequence of martingale differences with respect to an increasing sequence of σ-fields $\{\mathcal{F}_n\}$. They derive the limiting distribution of the LS estimates of the parameters.

With infinite variance innovations, Chan and Tran [10] consider the Dickey-Fuller test when the errors are in the domain of attraction of a stable law. Phillips [42] extends the results of Chan and Tran [10] to find the limit theory of the parameters in an AR(1) process with weakly dependent errors in the domain of attraction of a stable law. Caner [9] generalizes the univariate results of Chan and Tran [10] and Phillips [42] to multivariate time series. Since both the Dickey-Fuller [15] and Phillips-Perron [43] statistics are based on LS estimation, they do not take advantage of the heavy tails of the innovations. Thus, it is important to consider estimation and inference procedures which are robust to departures from finite variance condition. One way to achieve robustness is the use of M-estimate method where, with an appropriate chosen loss function, it has a number of desirable properties when the errors are heavy tailed. Knight [27] considers the asymptotic behaviour of the LS estimates and M-estimates for the random-walk model. The results establish that M-estimates are asymptotically normal and their rate of convergence is higher than the LS estimates. Davis, Knight, and Liu [14] mention that M-estimates are more appropriate when the distribution of innovations belongs to the class of heavy-tailed distributions since M-estimates give less weight to the outliers. Zarepour and Roknossadati [58] consider the asymptotic behavior of a multivariate random-walk model of Caner [9], when the innovations are in the domain of attraction of a multivariate stable law with different indices of stability in their coordinates. Tanaka [53] considers a non-stationary AR model with complex unit roots, when the innovations are stationary linear processes. The asymptotic properties of the LS estimates of the parameters are derived. Samarakoon and Knight [51] consider Dickey-Fuller-type tests for at least one unit root in an AR process driven by infinite variance innovations.

Recently, Chan and Zhang [12] derive the limiting distribution for the LS estimates of the parameters for unstable AR(p) processes, with i.i.d. innovations in the domain of attraction of a stable law. They show that the limiting distribution of the LS estimate is a function of integrated stable processes.

The goal of the second part is to derive the asymptotic distribution of M-estimators for the parameters in an unstable AR(p) process, when the innovations are in the domain of attraction of a stable law with index $0 < \alpha \leq 2$. Our results show that, similar to the previous cases, M-estimators have a higher asymptotic rate of convergence than LS estimators. However, similar asymptotic results with the same rate of consistency hold for LAD estimates if we impose further assumptions on innovations. For example, the innovations need to have 0 median and a density with respect to Lebesgue measure; see Knight (1989).

However, in the infinite variance case, the time series analysis usually relies on the assumption that the location parameter is zero, which often does not necessarily happen in practice. Finally, we extend the robust estimation method to the case where the autoregressive model has location parameter μ. When $1 < \alpha \leq 2$, μ is the mean of our time series.

This thesis is structured as follows. In Chapter 2, we introduce some preliminary concepts of heavy tailed phenomena, with emphasis on regular variation, point processes, Lévy processes, and stable processes.

In chapter 3, we consider a robust estimation of the mean vector for a sequence of i.i.d. observations in the domain of attraction of a stable law with different indices of stability, $DS(\alpha_1, \ldots, \alpha_p)$, such that $1 < \alpha_i \leq 2$, $i = 1, \ldots, p$. We apply an asymptotically valid bootstrap to construct a confidence region for the mean vector. A simulation study is performed to show that the estimation method is efficient for making inference about the mean vector for multivariate heavy-tailed observations.

In chapter 4, we present the asymptotic distribution of M-estimators for parameters in non-stationary AR(2) processes. The innovations are assumed to be in the domain of attraction of a stable law with index $0 < \alpha \leq 2$. To investigate the results given in this chapter, we also present a simulation study. Due to complex limiting distributions for the estimates, a brief discussion and simulation results of the bootstrap scheme are presented in this chapter.

In chapter 5, we develop the asymptotic theory for M-estimators for parameters in non-stationary AR(p) processes. Similar to Chapter 4, we assume that the process is driven by infinite variance innovations and derive their asymptotic properties.

In chapter 6, we generalize the results of preceding chapters to a non-stationary AR(p) process with non-zero location parameter μ. In this chapter, we estimate simultaneously both μ and the AR parameters using M-estimation method. We also provide the simulation results to study the estimation behaviour of M-estimates of the parameters.

Finally, we propose some future research in Chapter 6. Some discussion about the multivariate stable characterization and the domain of attraction and space $D[0, 1]$ are presented in appendixes A and B, respectively.

Chapter 2

Definitions and Preliminaries

The main objective of this chapter is to review some properties of heavy tailed phenomena, with special emphasis on regular variation, point processes, Lévy processes, and stable processes. The notations and the results introduced in this chapter will be employed in this thesis.

2.1 Functions of regular variation

The theory of regularly varying functions plays an essential role in dealing with heavy tails and domains of attraction.

Definition 2.1.1 *A positive measurable function U is called regularly varying at infinity with index $\alpha \in \mathbb{R}$ ($U \in RV_\alpha$) if for all $x > 0$*

$$\lim_{x\to\infty} \frac{U(tx)}{U(x)} = t^\alpha \quad \textit{for all} \;\; t > 0,$$

where α is called the exponent of variation.

If $\alpha = 0$, U is said to be slowly varying at infinity, denoted by $U \in \mathrm{SV}$.

Corollary 2.1.1 *Let $\alpha \neq 0$ and $U \in RV_\alpha$. Then, there exists a slowly varying function L such that*

$$U(x) = x^\alpha L(x).$$

Examples of slowly varying functions are positive constants or functions converging to a positive constant, logarithms ($\ln x$) and iterated logarithms ($\ln(\ln x)$). The properties of regularly varying functions are thus deduced from those of slowly varying functions. Few key properties of the slowly varying functions are presented in the following theorems.

Theorem 2.1.1 *(Karamata Representation Theorem [49]) A positive measurable function L on $[0,\infty)$ is slowly varying if and only if L can be represented as*

$$L(x) = c(x) \exp\left(\int_0^x \frac{\varepsilon(t)}{t} dt\right),$$

where $c(\cdot)$ is a measurable non-negative function such that $\lim_{x\to\infty} c(x) = c_0 \in (0,\infty)$ and $\lim_{x\to\infty} \varepsilon(x) = 0$.

The following result of Karamata theorem mentions that the integrals of regularly varying functions are again regularly varying. In other words, the slowly varying function can be taken out of the integral.

Theorem 2.1.2 *(Karamata's Theorem [49]) Let L be slowly varying and locally bounded in $[0,\infty)$. Then*

(a) for $\alpha > -1$,

$$\int_0^x t^\alpha L(t) dt \sim \frac{x^{\alpha+1}}{\alpha+1} L(x) \quad \text{as } x \to \infty,$$

(b) for $\alpha < -1$,

$$\int_x^\infty t^\alpha L(t) dt \sim -\frac{x^{\alpha+1}}{\alpha+1} L(x) \quad \text{as } x \to \infty.$$

Note that the notation $f(x) \sim g(x)$ as $x \to \infty$ means their ratio approaches 1 as $x \to \infty$ for two real functions f and g. For $\alpha = -1$, the result of Theorem 2.1.2 is still valid. Then

$$\frac{1}{L(x)}\int_0^x \frac{L(t)}{t}dt \to \infty \quad \text{as } x \to \infty$$

and $\int_0^x \frac{L(t)}{t}dt$ is slowly varying. If $\int_0^x \frac{L(t)}{t}dt < \infty$, then

$$\frac{1}{L(x)}\int_x^\infty \frac{L(t)}{t}dt \to \infty \quad \text{as } x \to \infty$$

and $\int_x^\infty \frac{L(t)}{t}dt$ is slowly varying.

Random variables and more generally random vectors with regularly varying tails have wide variety of applications in various fields of applied probability. In what follows, some properties of regularly varying functions in terms of regularly varying distributions are presented. Denote the distribution function of any random variable X as

$$F(x) = P(X \leq x) \quad \text{for } x \in \mathbb{R},$$

and its right tail as

$$\bar{F}(x) = 1 - F(x) \quad \text{for } x \in \mathbb{R}.$$

Proposition 2.1.1 *(Regularly varying distributions [35]) Suppose F is a distribution function with $F(x) < 1$ for all $x > 0$.*
(a) If for some real function h and all x from a dense subset of $(0, \infty)$, we have

$$\lim_{n\to\infty} a_n \bar{F}(b_n x) = h(x) \in (0, \infty),$$

where two sequences $\{a_n\}$ and $\{b_n\}$ are satisfying $a_n/a_{n+1} \to 1$ and $b_n \to \infty$ for all $x > 0$ as $n \to \infty$, then $h(x) = x^{-\alpha}$ for some $\alpha \geq 0$ and $\bar{F}$ is regularly varying.
(b) Let F be absolutely continuous with density f such that for some $\alpha > 0$,

$$\lim_{x\to\infty} \frac{x f(x)}{\bar{F}(x)} = \alpha.$$

Then f is regularly varying with index $-(1+\alpha)$ and, consequently, F is regularly varying with index $-\alpha$.

(c) Suppose $\bar{F}(x) \sim x^{-\alpha}L(x)$ as $x \to \infty$ for some $\alpha > 0$ and $\beta \geq \alpha$. Then

$$\lim_{x\to\infty} \frac{x^\beta \bar{F}(x)}{\int_0^x u^\beta dF(u)} = \frac{\beta - \alpha}{\alpha}.$$

The converse also holds in the case that $\beta > \alpha$. For $\beta = \alpha$, we have $\bar{F}(x) = o(x^{-\alpha}L(x))$ for some slowly varying L.
(d) The following are equivalent:

- *$\int_0^x u^2 dF(u)$ is slowly varying,*
- *$\bar{F}(x) = o(x^{-2}\int_0^x u^2 dF(u))$ as $x \to \infty$.*

The following corollary implies that regularly varying distributions are closed under convolution.

Lemma 2.1.1 *([35]) Let X and Y be two non-negative, regularly varying, and independent random variables with index $\alpha > 0$. Then $X + Y$ is regularly varying with index α and*

$$P(X + Y > x) \sim P(X > x) + P(Y > x) \quad \text{as } x \to \infty.$$

An immediate consequence of Lemma 2.1.1 is the following result.

Corollary 2.1.2 *Let $X, X_1, \ldots, X_n$ be i.i.d. non-negative regularly varying random variables and $S_n = X_1 + \cdots + X_n$ for $n \geq 1$. Then*

$$P(X_1 + \cdots + X_n > x) \sim nP(X > x) \quad \text{as } x \to \infty. \tag{2.1.1}$$

Equation (2.1.1) has an intuitive interpretation which is another closure property of the class of regularly varying random variables. Let $M_n = \max(X_1, \ldots, X_n)$ for $n \geq 1$. Then it is easily seen that for every $n \geq 1$

$$P(S_n > x) \sim nP(X > x) \sim P(M_n > x) \quad \text{as } x \to \infty.$$

The relation $P(M_n > x) \sim nP(X > x)$ for large x implies that M_n is regularly varying with the same index as X.

In the following, we shall define the concept of multivariate regularly varying functions.

Definition 2.1.2 *([48]) Suppose that $h(\cdot) : \mathbb{R}^d \to (0, \infty)$ is measurable function. We say that h is multivariate regularly varying with limit function $\lambda \neq 0$ if for any $\mathbf{x} \in \mathbb{R}^d$,*

$$\lim_{t\to\infty} \frac{h(t\mathbf{x})}{h(t\mathbf{1})} = \lambda(\mathbf{x}),$$

where $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^d$.

2.2 Point processes

Point processes are used as a tool to study asymptotic behaviour of statistics constructed from observations with heavy-tailed distributions. Let $(\Omega, \mathcal{F}, P)$ be a probability space and assume that $\mathbb{E}$ is a locally compact state space with countable base. A measure $\mu : \mathbb{E} \to [0, \infty]$ is called Radon if $\mu(K) < \infty$ for any compact subset K in $\mathbb{E}$. Define

$$M_+(\mathbb{E}) = \{\mu : \mu \text{ is a nonnegative measure on } \mathbb{E} \text{ and } \mu \text{ is Radon}\}.$$

The space $M_+(\mathbb{E})$ can be made into a complete separable metric space under the vague metric. To describe the notion of convergence in $M_+(\mathbb{E})$ consistent with the metric, we define

$$C_K^+(\mathbb{E}) = \{f : \mathbb{E} \to \mathbb{R}_+ : f \text{ is continuous with compact support}\}.$$

If $\mu_n \in M_+(\mathbb{E})$ for $n \geq 0$, then μ_n converges vaguely to μ, denoted by $\mu_n \xrightarrow{v} \mu$, if

$$\mu_n(f) := \int_{\mathbb{E}} f(x)\mu_n(dx) \to \mu(f) := \int_{\mathbb{E}} f(x)\mu(dx),$$

for all $f \in C_K^+(\mathbb{E})$, as $n \to \infty$. In $M_+(\mathbb{E})$, the random measure μ_n converges weakly to the random measure μ if weak convergence occurs with respect to the vague topology. If μ_n and μ are probability measures, it is not hard to see that vague convergence is equivalent to weak convergence.

Example 2.2.1 *([7]) Suppose μ_n and μ are finite measures on $(\mathbb{R}, \mathcal{R})$. Let μ_n be a unit mass at n and $\mu(\mathbb{R}) = 0$. Notice that $\int f d\mu_n = f(n) = 0$ since f vanishes at infinity for all $f \in C_K^+(\mathbb{E})$, as $n \to \infty$. Thus, $\mu_n \xrightarrow{v} \mu$ but μ_n does not converge to μ weakly since μ is not a probability measure.*

Let ζ be the Borel σ-algebra of subsets of $\mathbb{E}$, and

$$\varepsilon_x(F) = \begin{cases} 1 & \text{if } x \in F, \\ 0 & \text{otherwise}, \end{cases} \tag{2.2.1}$$

for $x \in \mathbb{E}$, and $F \in \zeta$. A point measure m is a finite measure on relatively compact subsets of $\mathbb{E}$ defined to be a measure of the form $\sum_{i=1}^{\infty} \varepsilon_{x_i}$. Let $M_P(\mathbb{E})$ be the class of such point measures which is the closed subset of $M_+(\mathbb{E})$ and $\mathcal{M}_P(\mathbb{E})$ be the smallest Borel σ-algebra of subsets of $M_P(\mathbb{E})$ generated by open sets in the vague topology. A point process on $\mathbb{E}$ is a measurable map from a probability space $(\Omega, \mathcal{F}, P)$ to $(M_p(\mathbb{E}), \mathcal{M}_p(\mathbb{E}))$.

Definition 2.2.1 *([45]) A point process ξ is a Poisson process defined on $(\mathbb{E}, \zeta)$ with mean measure μ if it satisfies in the following properties.*
(i) For each $A \in \zeta$,

$$P(\xi(A) = x) = \begin{cases} \frac{e^{-\mu(A)}(\mu(A))^x}{x!} & \text{if } \mu(A) < \infty, \\ 0 & \text{if } \mu(A) = \infty. \end{cases}$$

(ii) Whenever $A_1, \ldots, A_k \in \zeta$ are disjoint, the random variables $\xi(A_1), \ldots, \xi(A_k)$ are independent.

We shall call ξ a Poisson random measure with mean measure μ (PRM(μ)) on $(\mathbb{E}, \zeta)$. The Poisson process ξ can be identified by the characteristic form of its Laplace functional

$$\Psi_\xi(f) = \mathrm{E}\exp\{-\xi(f)\} = \exp\{-\int(1-e^{-f})d\mu\}.$$

A process whose mean measure is a multiple of Lebesgue measure is called a homogeneous process. The following proposition derives a representation for PRM(λ) where λ is the Lebesgue mean measure. The PRM(λ) can be used to generate a general PRM with any mean measure by transforming the points of a PRM(λ).

Proposition 2.2.1 *([49]) Let $\Gamma_n = \sum_{i=1}^n E_i$ where $E_j, j \geq 1$ are i.i.d. random variables with a standard exponential distribution. Set $\xi = \sum_{n=1}^\infty \varepsilon_{\Gamma_n}$. Then ξ is a homogeneous Poisson process on $[0,\infty)$ with unit rate $\lambda = 1$; that is, ξ satisfies Definition 2.2.1, and the mean measure is Lebesgue.*

Theorem 2.2.1 *([49]) Suppose that $\{X_{n,j}, j \geq 1\}$, $n \geq 1$, is a sequence of i.i.d. random elements of $(\mathbb{E}, \zeta)$, and ξ is a PRM(μ) on $M_p(\mathbb{E})$. Then*

$$\sum_{j=1}^n \varepsilon_{X_{n,j}} \xrightarrow{d} \xi$$

on $M_p(\mathbb{E})$ if and only if (iff) we have

$$nP[X_{n,1} \in \cdot] = \mathrm{E}\left(\sum_{j=1}^n \varepsilon_{X_{n,j}}(\cdot)\right) \xrightarrow{v} \mu$$

in $M_+(\mathbb{E})$, where $\xrightarrow{d}$ denotes the convergence in distribution.

2.3 Multivariate regularly varying tail probabilities

In this section, we review some concepts related to multivariate probabilities. To state the equivalent forms of multivariate regular variation for the distribution of a random vector, we need the polar coordinate transformation. Suppose that $\|\cdot\| : \mathbb{R}^d \to [0,\infty)$ is a norm on $\mathbb{R}^d$. The most useful norms for us are the usual Euclidean L_2 norm, the L_p norm for $p \geq 0$ and the L_∞ norm: $\|x\| = \vee_{i=1}^n |x^{(i)}|$. Given a chosen norm $\|\cdot\|$, the unit sphere is

$$\mathcal{N} := \{x : \|x\| = 1\}.$$

Definition 2.3.1 *The polar coordinate transformation $T : \mathbb{R}^d \setminus \{\mathbf{0}\} \to (0,\infty) \times \mathcal{N}$ is defined by:*

$$T(x) = \left(\|x\|, \frac{x}{\|x\|} \right) =: (r, a),$$

and the inverse transformation $T^{-1} : (0,\infty) \times \mathcal{N} \to \mathbb{R}^d \setminus \{\mathbf{0}\}$ is given by

$$T^{-1}(r, a) = ra.$$

Theorem 2.3.1 *(Multivariate regularly varying tail probabilities [48]) Suppose that $\{\mathbf{X}_n, n \geq 1\}$ are i.i.d. random vectors that take values in $[\mathbf{0},\infty) = [0,\infty)^d$ with common distribution $F(\cdot)$ and let $\mathbb{E} = [\mathbf{0},\infty)\backslash\{\mathbf{0}\}$. We also define the measure on $[0,\infty)$:*

$$\nu_\alpha(x,\infty] = x^{-\alpha}, \quad x > 0, \; \alpha > 0,$$

and $\mathcal{N}_+ = \mathcal{N} \cap \mathbb{E}$. The following statements are equivalent.

(i) There exists a Radon measure ν on $\mathbb{E}$ such that

$$\lim_{t\to\infty} \frac{1 - F(t\mathbf{x})}{1 - F(t\mathbf{1})} = \lim_{t\to\infty} \frac{P\left(\mathbf{X}_1/t \in [\mathbf{0},\mathbf{x}]^c\right)}{P\left(\mathbf{X}_1/t \in [\mathbf{0},\mathbf{1}]^c\right)} = c\nu([\mathbf{0},\mathbf{x}]^c)$$

for all points $\mathbf{x} \in [\mathbf{0},\infty)\backslash\{\mathbf{0}\}$ which are continuity points of $\nu([\mathbf{0},\mathbf{x}]^c)$.

(ii) There exists a function $b(\cdot)$ which approaches to ∞ and a Radon measure ν on $\mathbb{E}$, such that

$$tP\left(\frac{\mathbf{X}_1}{b(t)} \in \cdot\right) \xrightarrow{v} \nu \quad \text{as } t \to \infty$$

in $M_+(\mathbb{E})$.

(iii) There exists a sequence b_n which approaches to ∞ and a Radon measure ν on $\mathbb{E}$, such that

$$nP\left(\frac{\mathbf{X}_1}{b_n} \in \cdot\right) \xrightarrow{v} \nu \quad \text{as } n \to \infty$$

in $M_+(\mathbb{E})$.

(iv) There exists a probability measure $S(\cdot)$ on $\mathcal{N}_+$ and a function $b(\cdot)$ which tends to ∞ such that, for $(R_1, \Theta_1) = (\|X_1\|, X_1/\|X_1\|)$,

$$tP\left[\left(\frac{R_1}{b(t)}, \Theta_1\right) \in \cdot\right] \xrightarrow{v} c\nu_\alpha \times S$$

in $M_+(\mathbb{E} \times \mathcal{N}_+)$ for some $c > 0$.

(v) There exists a probability measure $S(\cdot)$ on $\mathcal{N}_+$ and a sequence b_n which approaches to ∞ such that, for $(R_1, \Theta_1) = (\|X_1\|, X_1/\|X_1\|)$,

$$nP\left[\left(\frac{R_1}{b_n}, \Theta_1\right) \in \cdot\right] \xrightarrow{v} c\nu_\alpha \times S$$

in $M_+(\mathbb{E} \times \mathcal{N}_+)$ for some $c > 0$.

(vi) There exists a sequence b_n which approaches to ∞ such that

$$\sum_{i=1}^{n} \varepsilon_{\mathbf{X}_i/b_n}$$

converges weakly to a PRM(ν).

Theorem 2.3.1 (ii)-(iv) can be extended to the general case where $\{\mathbf{X}_n, n \geq 1\}$ are i.i.d. random vectors with values in $\mathbb{R}^d$. Part (i) of Theorem 2.3.1 does not have an easy analogue in terms of multivariate regular variation of the distribution function tail, except when $d = 1$.

If $d = 1$, let $\mathbb{E} = [-\infty, \infty]$, then (iii) of Theorem 2.3.1 implies that

$$nP\left(\frac{X_1}{b_n} \in \cdot\right) \xrightarrow{v} \nu$$

on $M_+([-\infty,\infty]\setminus\{o\})$ is the basic condition. This means that, for $x>0$

$$nP\left(\frac{X_1}{b_n} > x\right) \xrightarrow{v} \nu(x,\infty] \quad \text{as } n\to\infty.$$

So we have $\nu(x,\infty] = c_+x^{-\alpha}$, $c_+ \geq 0$. Similarly, for $x>0$

$$nP\left(\frac{X_1}{b_n} < -x\right) \xrightarrow{v} \nu[-\infty,-x) \quad \text{as } n\to\infty.$$

Thus we have $\nu[-\infty,-x) = c_-x^{-\alpha}$, $c_- \geq 0$. The parameter α for both right and left tails must be the same since the same b_n works for both tails and b_n relates to α through the fact that $b_n = b(n)$, where $b(\cdot)\in RV_{1/\alpha}$. Thus, we have

$$\nu(dx) = \left(c_+\alpha x^{-\alpha-1}I_{(0,\infty)}(x) + c_-\alpha|x|^{-\alpha-1}I_{(-\infty,0)}(x)\right)dx,$$

where $0<\alpha<2$ and for $d=1$ case, $\mathcal{N}=\{-1,1\}$ and

$$S(\{1\}) = \frac{c_+}{c_+ + c_-},\quad S(\{-1\}) = \frac{c_-}{c_+ + c_-}.$$

The following theorem shows that multivariate regular variation also has an exact probabilistic equivalence in terms of convergence of empirical measures to a limiting PRM.

Theorem 2.3.2 *([45]) Suppose $\{\mathbf{X}_1,\mathbf{X}_2,\ldots\}$ are i.i.d. random vectors in $\mathbb{R}^d$ with common distribution $F(\cdot)$. Then*

$$nP\left(\frac{\mathbf{X}_1}{b_n}\in\cdot\right) \xrightarrow{v} \nu,$$

iff

$$\sum_i \varepsilon_{(i/n,\mathbf{X}_i/b_n)} \xrightarrow{d} \xi,$$

where ξ ia a PRM on $(0,\infty)\times\mathbb{R}^d$ with mean measure $dt\times d\nu$.

This property is also used to prove weak convergence of partial sum processes or maximal processes in the space $D=[0,\infty)$, the space of real-valued, right-continuous functions on $[0,\infty)$ with finite left limits existing on $(0,\infty)$. If $d=1$, i.e., $\mathbb{E} = [0,\infty)\backslash\{0\}$, then ξ in Theorem 2.3.2 is a PRM$(\lambda\times\nu)$, where λ denotes the Lebesgue measure.

2.4 Lévy Processes

Definition 2.4.1 *(Lévy Process [49]) A càdlàg (right-continuous with left limits a.s.) stochastic process* $\{X(t)\}_{t\geq 0}$ *defined on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is called a Lévy process if it satisfies the following conditions:*

1. $X(0) = 0$ *a.s.*

2. $X(t)$ *has independent increments, i.e. for* $0 < s \leq t$, $X(t) - X(s)$ *is independent of the* σ*-algebra generated by* $\{X(r), r \leq s\}$.

3. $X(t)$ *has stationary increments, i.e., for* $s > 0$, $X(t+s) - X(t) \stackrel{d}{=} X(s)$ *for all* $t \geq 0$, *where* $\stackrel{d}{=}$ *means equallity of the finite dimensional distributions.*

4. $X(t)$ *is stochastically continuous, i.e. for all* $\epsilon > 0$ *and for all* $s \geq 0$,

$$\lim_{t \to s} P\left(|X(t) - X(s)| > \epsilon\right) = 0. \tag{2.4.1}$$

The Poisson process, Brownian motion, and gamma process are some examples of Lévy processes. A Lévy process $X(t)$ is called a Poisson process of intensity $\lambda > 0$ if $X(t) - X(s) \sim Poisson(\lambda(t-s))$. The Lévy process $X(t)$ is called a Brownian motion (Wiener process) if $X(t) - X(s) \sim N(0, t-s)$ for $t > s > 0$. The Lévy process $X(t)$ is a gamma process with parameters $a, b > 0$ if $X(t) - X(s) \sim gamma(a(t-s), b)$. Note that (2.4.1) implies only continuity in probability and not all Lévy processes have continuous sample paths. For example, the gamma process is a pure jump process while almost all sample paths of Brownian motion are continuous (a.s. continuous).

Definition 2.4.2 *([49]) A measure* ν *is called the Lévy measure on* $\mathbb{R}\backslash\{0\}$ *if it satisfies in the following property:*

(i) $\nu(\{0\}) = 0$, *and for every* $x > 0$, $\nu\{u \in \mathbb{R}\backslash\{0\} : |u| > x\} < \infty$.

(ii) $\int_0^1 (x^2 \wedge 1)\nu(dx) < \infty$, *where the notation* $x^2 \wedge 1$ *denotes the minimum of* x^2 *and 1.*

The Lévy-Khintchine representation theorem states that a Lévy process is the sum of a deterministic function, a Brownian motion, and a pure jump part. The Lévy process $\{X(t), t \geq 0\}$ in $\mathbb{R}$ has the following characteristic function

$$\begin{aligned}\mathrm{E}\left(e^{isX(t)}\right) = &\exp\left\{t\left[i\gamma s - \frac{cs^2}{2} + \int_{|x|>1}\left(e^{isx} - 1\right)\nu(dx)\right.\right.\\ &\left.\left.+ \int_{0<|x|\leq 1}\left(e^{isx} - 1 - isx\right)\nu(dx)\right]\right\},\end{aligned}$$

where $s \in \mathbb{R}$, $\gamma \in \mathbb{R}$, $c > 0$, and ν is the Lévy measure; for more details see [45].

Let $\xi = \sum_{k=1}^{\infty} \varepsilon_{(t_k, J_k)}$ be a PRM on $\mathbb{R}^+ \times \mathbb{R}\backslash\{0\}$ with mean measure $dt \times d\nu$. A (one-dimensional) Lévy process can be represented in the following way:

$$\begin{aligned}X_t = &\gamma t + \sqrt{c}W_t + \sum_{t_k \leq t} j_k I_{[|j_k|>1]}\\ &+ \lim_{\delta \downarrow 0}\left[\sum_{t_k \leq t} j_k I_{[|j_k| \in (\delta,1]]} - t\int_{|x|\in(\delta,1]} x\nu(dx)\right],\end{aligned}$$

where $(W_t)_{t\geq 0}$ is a standard Brownian motion; see [24]. When $c = 0$, i.e., if the continuous Gaussian part vanishes, the process X_t is a pure jump Lévy process. Therefore, the Lévy measure ν satisfies

$$\int_0^{\infty} (x \wedge 1)\nu(dx) < \infty.$$

Thus, it is convenient to express the Itô representation as

$$X_t = \sum_{t_k \leq t} j_k$$

and so we get

$$\mathrm{E}\left(e^{iuX(t)}\right) = \exp\left\{t\int_0^{\infty}\left(e^{iux} - 1\right)\nu(dx)\right\}.$$

For more details see [45].

2.5 Stable processes

Stable distributions are a rich class of probability distributions that describe heavy tails and perhaps asymmetric behavior. The theory of stable distributions was characterized in the 1920s by Paul Lévy in his study of sums of i.i.d. terms. This class does not have simple closed formulas for densities and distribution functions for all but a few stable distributions (Gaussian, Cauchy and Lévy). We start with the definition of a stable distribution.

Definition 2.5.1 *Let $X, X_1, X_2, \ldots, X_n$ be mutually independent random variables with a common distribution S. Then the distribution S is stable if for any constants $a_n > 0$ and b_n we have*

$$X_1 + X_2 + \cdots + X_n \stackrel{d}{=} a_n X + b_n, \tag{2.5.1}$$

where $\stackrel{d}{=}$ denotes equality in distribution.

The distribution S is said to be *strictly* stable if (2.5.1) holds with $b_n = 0$. If $X \stackrel{d}{=} -X$, then X is called *symmetric*. A symmetric stable distribution is obviously strictly stable. Feller [19] shows that in (2.5.1) it is necessary to have $a_n = n^{1/\alpha}$ for some $0 < \alpha \leq 2$. The following definition specifies the characteristic function of a stable random variable.

Definition 2.5.2 *([50]) A random variable X is said to have a stable distribution if there are parameters $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\sigma \geq 0$, and $\mu \in \mathbb{R}$ such that its characteristic function has the form*

$$\mathrm{E}\exp(i\theta X) = \begin{cases} \exp\{-\sigma^\alpha|\theta|^\alpha(1 - i\beta(\operatorname{sign}\theta)\tan\frac{\pi\alpha}{2}) + \mathrm{i}\mu\theta\} & \text{if } \alpha \neq 1, \\ \exp\{-\sigma|\theta|(1 + i\beta\frac{2}{\pi}(\operatorname{sign}\theta)\ln|\theta| + \mathrm{i}\mu\theta\} & \text{if } \alpha = 1, \end{cases}$$

where θ is a real number and

$$\text{sign } \theta = \begin{cases} 1 & \text{if } \theta > 0, \\ 0 & \text{if } \theta = 0, \\ -1 & \text{if } \theta < 0. \end{cases}$$

The parameter α is called the index of stability. The other parameters include the shift parameter μ, the skewness parameter β (a measure of asymmetry), and the scale parameter σ. The shift parameter equals to the mean when $1 < \alpha \leq 2$. Since the univariate stable distribution is characterised by these four parameters, its distribution is denoted by $S_\alpha(\sigma, \beta, \mu)$. The probability density of X exists and is continuous but it cannot be written in a simple closed form except for a few cases. The Gaussian distribution $S_2(\sigma, 0, \mu) = N(\mu, 2\sigma^2)$, the Cauchy distribution $S_1(\sigma, 0, \mu)$, and the Lévy distribution $S_{1/2}(\sigma, 1, \mu)$ are the exceptions. Therefore, we mainly use their characteristic functions.

A random variable $X \sim S_\alpha(\sigma, \beta, \mu)$ is symmetric ($S\alpha S$) iff $X \sim S_\alpha(\sigma, 0, 0)$, with characteristic function $\mathrm{E}\left(e^{i\theta X}\right) = \exp\{-\sigma^\alpha |\theta|^\alpha\}$. When $\alpha \neq 1$, X is strictly stable iff $\mu = 0$. In this case,

$$X \sim S_\alpha(\sigma, \beta, 0) \Leftrightarrow -X \sim S_\alpha(\sigma, -\beta, 0)$$

and $X \sim S_1(\sigma, \beta, \mu)$ is strictly stable iff $\beta = 0$.

Remark 2.5.1 *If $X \sim S_\alpha(\sigma, \beta, \mu)$, then $\mathrm{E}\left(|X|^\gamma\right) = \infty$ for any $\gamma > \alpha$, $0 < \alpha < 2$. It means that the second moment does not exist when $\alpha < 2$ and $\mathrm{E}|X| = \infty$, when $\alpha \leq 1$. Thus many of the suitable techniques applicable for the Gaussian case are not valid.*

The following definition is the general case of the Central Limit Theorem when the random variables have infinite variance (or even $\mathrm{E}(|X_1|) = \infty$). This definition states that stable distributions can be obtained as limits of centered normalized sum of i.i.d. random variables.

Definition 2.5.3 *(equivalent to Definition 2.5.1 [50]) 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 $X_1, X_2, \ldots$ and sequences of positive numbers $\{a_n\}$ and real numbers $\{b_n\}$, such that*

$$a_n^{-1}\left(\sum_{i=1}^{n} X_i - b_n\right) \xrightarrow{d} X. \tag{2.5.2}$$

When $\alpha = 2$, X has a normal distribution with mean zero and variance $\sigma^2 < \infty$. In (2.5.2), $\{a_n\}$ is a sequence of positive constants such that

$$a_n = \inf\{x : P[|X_1| > x] \leq n^{-1}\} \tag{2.5.3}$$

and $b_n = \mathrm{E}(X_1 I(|X_1| < a_n))$. It can be proved that $a_n = n^{1/a}L(n)$ for some slowly varying function $L(\cdot)$. When $a_n = n^{1/\alpha}$, the X_j's are said to belong to the normal domain of attraction of X.

Proposition 2.5.1 *[19] The random variable X is in the domain of attraction of a stable law with index $0 < \alpha < 2$, denoted by DS(α), if and only if*

(i) $P(|X| > x) = x^{-\alpha}L(x)$, where L is slowly varying at ∞,

(ii) $\frac{P(X>x)}{P(|X|>x)} \to p$ and $\frac{P(X\leq -x)}{P(|X|>x)} \to q$, as $x \to \infty$, $0 \leq p \leq 1$, and $q = 1 - p$.

Condition (i) describes the tails of the distribution function of X which vanishes smoothly while condition (ii) points out that the left and right tails are balanced in some way.

The following theorem presents the characterisation of the domain of attraction of a normal law for the case $\alpha = 2$.

Theorem 2.5.1 *([19]) The distribution F belongs to the domain of attraction of a normal law if and only if*

$$\int_{|y|<x} y^2 dF(y)$$

is slowly varying.

We denote the domain of attraction of a normal law by D(2). Note that Proposition 2.1.1 implies that slow variation of $\int_{|y|<x} y^2 dF(y)$ is equivalent to the tail condition

$$G(x) = P(|X| > x) = o\left(x^{-2} \int_{|y|<x} y^2 dF(y)\right). \tag{2.5.4}$$

Moreover, if $\mathrm{E}(X^2) < \infty$ then

$$\int_{|y|<x} y^2 dF(y) \to \mathrm{E}\left(X^2\right) \quad \text{as } x \to \infty,$$

hence $X \in D(2)$. Consequently, we have the following result for the domain of attraction of a normal distribution.

Corollary 2.5.1 *([35]) A random variable $X \in D(2)$ if and only if one of the following conditions holds:*

- $\mathrm{E}(X^2) < \infty$,
- $\mathrm{E}(X^2) = \infty$ *and* (2.5.4) *is satisfied.*

The multivariate stable distribution, which is the distribution of a stable random vector, is defined in the following definition by simply extending to $\mathbb{R}^d$ the definition of a stable random variable.

Definition 2.5.4 *A random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ is said to be a stable random vector in $\mathbb{R}^d$ if for any positive numbers A and B there is a positive number C and a vector $\mathbf{D} \in \mathbb{R}^d$ such that*

$$A\mathbf{X}^{(1)} + B\mathbf{X}^{(2)} \stackrel{d}{=} C\mathbf{X} + \mathbf{D}, \tag{2.5.5}$$

where $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are independent copies of $\mathbf{X}$.

The vector $\mathbf{X}$ is called strictly stable if (2.5.5) holds with $\mathbf{D} = 0$ for any $A > 0$ and $B > 0$. The vector $\mathbf{X}$ is called symmetric stable if it is stable and satisfies in addition the relation

$$P\{\mathbf{X} \in A\} = P\{-\mathbf{X} \in A\}$$

for any Borel set A of $\mathbb{R}^d$. As in $\mathbb{R}^1$, a symmetric stable vector is strictly stable.

Definition 2.5.4 imposes conditions on the joint distribution of $(X_1, X_2, \ldots, X_d)$. What do they imply about the individual components $X_1, X_2, \ldots, X_d$? Gaussian vectors are defined in the following way: a random vector is Gaussian if and only if any linear combination of its components is a Gaussian random variable. If $\mathbf{X}$ is a stable random vector, are its components stable random variables? Are linear combinations of its components stable random variables? The following theorem provides an answer.

Theorem 2.5.2 *([50]) Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a stable (respectively, strictly stable, symmetric stable) vector in $\mathbb{R}^d$. Then there is a constant $\alpha \in (0, 2]$ such that, in (2.5.5), $C = (A^\alpha + B^\alpha)^{1/\alpha}$. Moreover, any linear combination of the components of $\mathbf{X}$ of the type $Y = \sum_{i=1}^{k} b_k X_k$ is an α-stable (respectively, strictly α-stable, symmetric α-stable) random variable.*

As in the Gaussian case, any linear combination of the components of a stable random vector is a stable random variable, but the converse is not always true. The converse holds when the linear combinations are either strictly stable or when $\alpha \geq 1$. For more details see [50].

Stable processes is a subclass of Lévy processes corresponding to $c = 0$ (no Wiener component). The finite dimensional distribution of a stochastic process $\{X(t)\}$ are the distributions of the vectors $(X(t_1), \ldots, X(t_d))$, for $d \geq 1$.

Definition 2.5.5 *A stochastic process $\{X(t)\}$ is (strictly/symmetric) stable if all its finite dimensional distributions are (strictly/symmetric) stable.*

Definition 2.5.6 *[46] Suppose $d = 1$, $0 < \alpha < 2$, $0 \leq p \leq 1$, and $q = 1 - p$, and define the Lévy measure*

$$\nu_\alpha(dx) = \left(p\alpha x^{-\alpha-1} I_{(0,\infty)}(x) + q\alpha |x|^{-\alpha-1} I_{(-\infty,0)}(x)\right) dx, \tag{2.5.6}$$

The Lévy process with Lévy measure ν_α is called the stable Lévy motion and is denoted by $X_\alpha(\cdot)$.

The role that α-stable Lévy motions play among α-stable processes is similar to the role of Browian motions among Gaussian processes.

Corollary 2.5.2 *[45] Let $\{X_n, n \geq 1\}$ be i.i.d. random variables on $\mathbb{R}$ and set $X_{n,i} = X_i/a_n$, where a_n is defined as in (2.5.3). Define ν_α to be the Lévy measure in (2.5.6) with $0 < \alpha < 2$. Then*

$$a_n^{-1}\sum_{i=1}^{[n\cdot]} X_i - [n\cdot]\mathrm{E}\left(a_n^{-1}X_1 I(a_n^{-1}X_1 \leq 1)\right) \xrightarrow{d} X_\alpha(\cdot)$$

in $D[0,\infty)$, where the limit is the α-stable Lévy motion with Lévy measure ν_α, iff

$$nP\left(a_n^{-1}X_1 \in \cdot\right) \xrightarrow{v} \nu_\alpha(\cdot)$$

in $M_+([-\infty,\infty] \setminus \{0\})$.

Chapter 3

Bootstrapping the Mean Vector for the Observations in the Domain of Attraction of a Multivariate Stable Law

We consider a robust estimator of the mean vector for a sequence of i.i.d. observations in the domain of attraction of a stable law with different indices of stability, $DS(\alpha_1, \ldots, \alpha_p)$, such that $1 < \alpha_i \leq 2$, $i = 1, \ldots, p$. The suggested estimator is asymptotically Gaussian with unknown parameters. We apply an asymptotically valid bootstrap to construct a confidence region for the mean vector. A simulation study is performed to show that the estimation method is efficient for conducting inference about the mean vector for multivariate heavy-tailed distributions.

3.1 Introduction

Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables from some distribution F with mean μ. Traditionally, studentization has been considered for inference about the mean for relatively light-tailed distributions. This approach demands only a second moment assumption. Moreover, bootstrap inference is arguably accurate and is a simple approach to inference for the univariate mean of finite variance observations; see, for example, Diciccio and Efron (1996) and Singh (1981).

Now suppose that $\{X_k\}$ are in the domain of attraction of a stable law with infinite second moment. In other words, there exist constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$S_n = a_n^{-1} \sum_{k=1}^{n} (X_k - b_n) \xrightarrow{d} S_\alpha,$$

where S_α is a stable random variable with index $0 < \alpha \leq 2$. It is known that $a_n = n^{1/\alpha} L(n)$, where L is a slowly varying functions at ∞; see Feller (1971) for more details. Throughout this paper, we assume that $1 < \alpha \leq 2$ and $\mathrm{E}(X_1) < \infty$, so we can take $b_n = \mu$. Since the sample mean $\bar{X}_n$ is the usual estimator of the mean, it is natural to base inference about μ on $\bar{X}_n$. Despite the fact that the sample mean is an intuitive estimate for the population mean, the rate of convergence of the sample mean is na_n^{-1} which approaches zero very slowly when α is close to 1.

Properties of the various bootstrap procedures for the mean of heavy-tailed distributions have been considered extensively in the statistical literature; see Hall (1990) and Knight (1989a). It has been shown that the regular bootstrap is not consistent for estimating the distribution of the mean. For finite variance observations, the bootstrap distribution of the sample mean converges almost surely to a fixed distribution. On the other hand, Athreya (1987) shows that the bootstrap distribution of the sample mean of infinite variance observations converges in distribution to a random probability distribution. Athreya, Lahiri, and Wu (1998) demonstrate that bootstrapping based on m out of n resampling, such that $m/n \to 0$, rectifies the asymptotic failure of the regular bootstrap for heavy-tailed distributions. They also consider bootstrap methods for conducting inference about the mean of a sequence of i.i.d. random variables in the domain of attraction of a stable law whose index exceeds 1. Arcones and Giné (1989) discuss almost sure and in probability bootstrap central limit theorem when the random variable X is in the domain of attraction of a stable law with infinite second moment. Hall and LePage (1996) propose a bootstrap method for estimating the distribution of the studentized mean under more general conditions on the tails of the sampling distributions. They also show that this method holds even when the sampling distribution is not in the domain of attraction of any limit law. Zarepour and Knight (1999b) consider the weak limit behavior of a point process obtained by replacing the original observations by the bootstrap sample.

In this chapter, we wish to make inference about the mean vector $\boldsymbol{\mu}$ of a multivariate heavy-tailed distribution. Consider the model

$$\mathbf{X}_i = \boldsymbol{\mu} + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \ldots, n, \tag{3.1.1}$$

where $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})$, $i = 1, \ldots, n$, are $\mathbb{R}^p$-valued random vectors, and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$ is an unknown fixed parameter vector. Let $\{\boldsymbol{\epsilon}_i\} = \{(\epsilon_{i1}, \ldots, \epsilon_{ip})\}$ form a sequence of i.i.d. random vectors with zero mean in the domain of attraction of a multivariate stable law. The following generalizes the definition of the domain of attraction of a bivariate stable law in Resnick and Greenwood (1979) to the multivariate stable law.

Definition 3.1.1 *Given $\{\mathbf{X}_n = (X_{n1}, X_{n2}, \ldots, X_{np})\}$ i.i.d. random vectors on $\mathbb{R}^p$ with distribution* F*, let $\mathbf{S}_n = \sum_{i=1}^n \mathbf{X}_i$ and $S_n^{(j)} = \sum_{i=1}^n X_{ij}$ for $j = 1, \ldots, p$. Then,* $\mathrm{F} \in DS(\alpha_1, \ldots, \alpha_p)$, $\alpha_j \in (0, 2]$, *if there exist sequences $\mathbf{a}_n = (a_n^{(1)}, \ldots, a_n^{(p)}), \mathbf{b}_n \in \mathbb{R}^p$ with $a_n^{(j)} > 0$ such that*

$$\left(S_n^{(1)}/a_n^{(1)}, \ldots, S_n^{(p)}/a_n^{(p)}\right) - \mathbf{b}_n \xrightarrow{d} \mathbf{Y}, \tag{3.1.2}$$

where $\mathbf{Y}$ is a random vector on $\mathbb{R}^p$ with stable distribution.

For more discussion about the class of all possible limits in (3.1.2) see Resnick and Greenwood (1979). For the multivariate case, observations can be in the domain of attraction of a stable law with different indices of stability. In many real life examples, some coordinates may have light tails while other coordinates may have heavier tails. In this paper, we assume that errors are in $DS(\alpha_1, \ldots, \alpha_p)$ with possibly different values of $\alpha_j \in (1, 2]$ for $j = 1, \ldots, p$.

It is obvious that the limiting distribution of the sample mean depends on the tail indices when the errors are in the domain of attraction of a stable law. Thus, it is hard to derive any inference for the mean vector $\boldsymbol{\mu}$ based on the limit, especially when the limiting distributions of the coordinates may have different indices of stability. A bootstrap procedure may circumvent this difficulty but, as mentioned before, the ordinary bootstrap fails in this case. Using a m out of n bootstrap, when $m/n \to 0$, typically resolves the problem. Note that the choice of m is a key issue and controversial. See Bickel and Sakov (2008) for more details.

These difficulties prompted us to look at a robust estimator based on an M-estimate method for constructing any inference about the mean vector such as confidence regions when the errors are in the domain of attraction of a multivariate stable law with possibly different indices of stability in $(1, 2]$. Moreover, in our approach, the proposed robust estimation method for the mean vector has higher rate of convergence compared to the sample mean. We also show that the regular bootstrap is applicable since the limiting distribution is a multivariate normal distribution.

Section 3.2 presents our main theorem, robust estimation of the mean vector for a sequence of i.i.d. observations in the domain of attraction of a stable law with different indices of stability, $DS(\alpha_1, \ldots, \alpha_p)$, such that $1 < \alpha_i \leq 2$, $i = 1, \ldots, p$. The bootstrap procedure is discussed in Section 3.3. Section 3.4 presents some simulations supporting the results of this paper.

3.2 M-estimates of the mean vector

Let $\boldsymbol{\mu}$ be the parameter vector of interest and let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random samples satisfying (3.1.1). The classical M-estimate for $\boldsymbol{\mu}$, denoted by $\hat{\boldsymbol{\mu}}_M$, is defined as the minimizer of the function

$$\sum_{i=1}^{n} \left(\rho\left(\mathbf{X}_i - \boldsymbol{\beta}\right) - \rho\left(\boldsymbol{\epsilon}_i\right)\right)$$

with respect to $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$, where ρ is an almost everywhere differentiable convex function. This guarantees the uniqueness of the solution. For more details see Davis, Knight, and Liu (1992). For convenience, similar to Zarepour and Roknossadati (2008), we consider the multivariate loss function as

$$\rho\left(x_1, \ldots, x_p\right) = \rho_1\left(x_1\right) + \cdots + \rho_p\left(x_p\right), \tag{3.2.1}$$

where ρ_j, $j = 1, \ldots, p$, are univariate loss functions. A good justification for using the objective function of the form (3.2.1) is the ability to calibrate with respect to the thickness of the tails for each coordinate to derive more precise estimates in practice.

Here, we impose the following assumptions on the functions ρ_j, for $j = 1, \ldots, p$.

Assumption 1 *(A1) $\rho_j : \mathbb{R} \to \mathbb{R}$ is a convex and twice differentiable function, and take $\psi_j = \rho_j'$, and $\psi_j' = \rho_j''$.*

Assumption 2 *(A2) $\mathrm{E}(\psi_j(\cdot)) = 0$, $\mathrm{E}(\psi_j^2(\cdot)) < \infty$, and $0 < |\mathrm{E}(\psi_j'(\cdot))| < \infty$.*

Assumption 3 *(A3) ψ_j has Lipschitz-continuous derivative ψ_j'; i.e., there exists a real constant $k \geqslant 0$ such that for all x and y,*

$$|\psi_j^{'}(x) - \psi_j^{'}(y)| \leq k|x - y|.$$

Definition 3.2.1 *Let $(\Omega, \mathcal{A}, P)$ be a probability space and $(a, b) \in \mathbb{R}$ be an interval. We say that a stochastic process $X : (a, b) \times \Omega \to \mathbb{R}$ is convex if*

$$X(\lambda s + (1 - \lambda)t, \cdot) \leq \lambda X(s, \cdot) + (1 - \lambda)X(t, \cdot)$$

almost everywhere for all $s, t \in (a, b)$ and $\lambda \in [0, 1]$.

The following lemma is used to prove our main results.

Lemma 3.2.1 *Suppose that $\{S_n(\cdot)\}$ is a sequence of convex stochastic processes on $\mathbb{R}$ and suppose that*

$$S_n(\cdot) \xrightarrow{d} S(\cdot).$$

Then $\{S_n(\cdot)\}$ has a unique minimum $\boldsymbol{\kappa}_n$. If $\boldsymbol{\kappa}$ minimizes $S(\cdot)$, then

$$\boldsymbol{\kappa}_n \xrightarrow{d} \boldsymbol{\kappa}.$$

Proof. The proof is given in Lemma 2.2 of Davis et al. (1992). See also page 276 from Knight (1989b).

□

Theorem 3.2.1 *Suppose* (3.1.1) *holds. With the loss function* (3.2.1)*, let $\hat{\boldsymbol{\mu}}_M$ be the M-estimator of the mean vector for a sequence of i.i.d. observations in the domain of attraction of a stable law with indices of stability $(\alpha_1, \ldots, \alpha_p)$ such that $1 < \alpha_j \leq 2$, $j = 1, \ldots, p$. Then, we have*

$$\mathbf{W}_n = \sqrt{n}(\hat{\boldsymbol{\mu}}_M - \boldsymbol{\mu}) \xrightarrow{d} \mathbf{W}, \tag{3.2.2}$$

where $\mathbf{W}$ has a multivariate normal distribution with mean zero and covariance matrix $\Sigma = \operatorname{diag}\left(\frac{\mathrm{E}[(\psi_1(\epsilon_{11}))^2]}{\mathrm{E}^2(\psi_1'(\epsilon_{11}))}, \ldots, \frac{\mathrm{E}[(\psi_\mathrm{p}(\epsilon_{1\mathrm{p}}))^2]}{\mathrm{E}^2(\psi_\mathrm{p}'(\epsilon_{1\mathrm{p}}))}\right)$.

Proof: Under conditions A1-A3, define the convex process

$$\begin{aligned} A_n(u_1, \ldots, u_p) &= \sum_{j=1}^{p} \sum_{i=1}^{n} \left(\rho_j\left(\epsilon_{ij} - n^{-1/2} u_j\right) - \rho_j(\epsilon_{ij})\right) \\ &= \frac{-1}{\sqrt{n}} \sum_{j=1}^{p} u_j \sum_{i=1}^{n} \psi_j(\epsilon_{ij}) + \frac{1}{2n} \sum_{j=1}^{p} u_j^2 \sum_{i=1}^{n} \psi_j'(c_{ij}), \end{aligned} \tag{3.2.3}$$

where $u_j = n^{1/2}(\hat{\mu}_{Mj} - \mu_j)$, for $j = 1, \ldots, p$, and c_{ij} is between ϵ_{ij} and $\epsilon_{ij} - n^{-1/2}u_j$. Asymptotically, $\psi_j'(c_{ij})$ can be replaced by $\psi_j'(\epsilon_{ij})$ in (3.2.3) since

$$n^{-1}\sum_{j=1}^{p}|\psi_j'(\epsilon_{ij}) - \psi_j'(c_{ij})| \le kn^{-1}\sum_{j=1}^{p}|n^{-1/2}u_j| \xrightarrow{P} 0.$$

It is well known that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_j(\epsilon_{ij}) \xrightarrow{d} \mathrm{E}[(\psi_j(\epsilon_{1j}))^2]^{1/2}Z_j,$$

where Z_j, $j = 1, \ldots, p$, have standard normal distributions. Therefore,

$$A_n(u_1, \ldots, u_p) \xrightarrow{d} A(u_1, \ldots, u_p) = -\sum_{j=1}^{p}u_j\mathrm{E}[(\psi_j(\epsilon_{1j}))^2]^{1/2}Z_j + \frac{1}{2}\sum_{j=1}^{p}u_j^2\mathrm{E}(\psi_j'(\epsilon_{1j})).$$

From Lemma 3.2.1, the minimizer of A_n converges to the minimizer of A. Thus, (3.2.2) follows by setting the derivative of $A(u_1, \ldots, u_p)$ to 0 and solving for $u_1, u_2, \ldots,$ and u_p. Note that $\mathbf{W} = (W_1, \ldots, W_p)$ in (3.2.2) has a multivariate normal distribution and

$$W_j = \frac{\mathrm{E}[(\psi_j(\epsilon_{1j}))^2]^{1/2}}{\mathrm{E}(\psi_j'(\epsilon_{1j}))}Z_j,$$

$j = 1, \ldots, p$, are independent.

□

Based on Theorem 3.2.1, a simple approach to construct a $100(1-\alpha)\%$ confidence region for the mean of a $p-$dimensional random vector in the domain of attraction of a multivariate stable law with large sample size is the ellipsoid determined by all $\boldsymbol{\mu}$ such that

$$CR_{1-\alpha} = \{\boldsymbol{\mu} : n(\hat{\boldsymbol{\mu}}_M - \boldsymbol{\mu})^T\mathbf{S}^{-1}(\hat{\boldsymbol{\mu}}_M - \boldsymbol{\mu}) \le \tau_{1-\alpha}\}. \qquad (3.2.4)$$

Here, $\tau_{1-\alpha}$ is the $1-\alpha$ quantile of a $\chi^2(p)$ distribution and $\mathbf{S}$ is the estimated value of Σ using residuals, $\boldsymbol{e}_i = \mathbf{X}_i - \hat{\boldsymbol{\mu}}_M$ for $i = 1, 2, \ldots, n$.

The confidence region in (3.2.4) gives the joint knowledge concerning reasonable values of $\boldsymbol{\mu}$ when the correlation between the measured variables is taken into account. Typically, any summary of conclusions includes confidence statements about the individual component means especially when the covariance matrix is diagonal, similar to our case. Let (3.1.1) hold for i.i.d. random vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$. Consider the following linear combination

$$\mathbf{a}^T\mathbf{X} = a_1\mathbf{X}_1 + a_2\mathbf{X}_2 + \cdots + a_p\mathbf{X}_p.$$

Simultaneously for all $\mathbf{a}$, the interval

$$\left(\mathbf{a}^T\hat{\boldsymbol{\mu}}_M - \sqrt{\tau_{1-\alpha}\mathbf{a}^T\mathbf{S}\mathbf{a}}\ ,\ \mathbf{a}^T\hat{\boldsymbol{\mu}}_M - \sqrt{\tau_{1-\alpha}\mathbf{a}^T\mathbf{S}\mathbf{a}}\right)$$

contains $\mathbf{a}^T\hat{\boldsymbol{\mu}}_M$ with probability $1-\alpha$. The consecutive choices $\mathbf{a}^T = (1, 0, \ldots, 0)$, $\mathbf{a}^T = (0, 1, \ldots, 0)$, and so on through $\mathbf{a}^T = (0, 0, \ldots, 1)$ for the χ^2-intervals allow us to conclude that

$$\begin{aligned}
\hat{\mu}_{M1} - \sqrt{\tau_{1-\alpha}s_{11}} \leq &\mu_1 \leq \hat{\mu}_{M1} + \sqrt{\tau_{1-\alpha}s_{11}} \\
\hat{\mu}_{M2} - \sqrt{\tau_{1-\alpha}s_{22}} \leq &\mu_2 \leq \hat{\mu}_{M2} + \sqrt{\tau_{1-\alpha}s_{22}} \\
&\vdots \\
\hat{\mu}_{Mp} - \sqrt{\tau_{1-\alpha}s_{pp}} \leq &\mu_p \leq \hat{\mu}_{Mp} + \sqrt{\tau_{1-\alpha}s_{pp}}
\end{aligned}$$

all hold simultaneously with confidence coefficient $1-\alpha$.

3.3 Bootstrapping the Mean Vector

It has been pointed out that the regular bootstrap fails to estimate the distribution of the sample mean of heavy-tailed observations. The main reason for the failure of the regular bootstrap comes from the fact that rare events occur when we resample the data. This means the resampling procedure will remember the magnitude of the observations in the resampled data. This fact is reflected in point process theory which is used as a tool for the asymptotic theory of heavy-tailed observations. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be a sequence of i.i.d. random vectors in $DS(\alpha_1, \ldots, \alpha_p)$. Given $\mathbf{X}_1, \ldots, \mathbf{X}_n$, we draw an i.i.d. sequence of observations $\mathbf{X}_1^*, \ldots, \mathbf{X}_n^*$ from the empirical distribution

$$F_n(\cdot) = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{\mathbf{X}_i}(\cdot),$$

where ε_x is defined as (2.2.1). Define $M_{i,n}^* = \sum_{k=1}^{n} I(\mathbf{X}_k^* = \mathbf{X}_i)$. By Lemma 3.1 of Zarepour and Knight (1989a), we have

$$\left(M_{1,n}^*, \ldots, M_{n,n}^*, 0, 0, \ldots\right) \xrightarrow{d} \left(M_1^*, \ldots, M_n^*, \ldots\right),$$

where $M_1^*, M_2^*, \ldots$ are i.i.d. Poisson(1) random variables. Let $\mathbf{X}_i^* = \left(X_{i1}^*, \ldots, X_{ip}^*\right)$. Now, for the corresponding sequence of point processes, we have

$$\sum_{i=1}^{n} \varepsilon_{\left((a_n^{(1)})^{-1} X_{i1}^*, \ldots, (a_n^{(p)})^{-1} X_{ip}^*\right)} = \sum_{i=1}^{n} M_{i,n}^* \varepsilon_{\left((a_n^{(1)})^{-1} X_{i1}, \ldots, (a_n^{(p)})^{-1} X_{ip}\right)}.$$

By Theorem 4 of Resnick and Greenwood (1979) and Resnick (2004) and Zarepour and Knight (1999b) (Appendix A), it can be shown that

$$\sum_{i=1}^{n} \varepsilon_{\left((a_n^{(1)})^{-1} X_{i1}^*, \ldots, (a_n^{(p)})^{-1} X_{ip}^*\right)} \xrightarrow{d} \sum_{i=1}^{\infty} M_i^* \varepsilon_{\left(\mathrm{sign}(\gamma_{i1})|\gamma_{i1}|^{1/\alpha_1} \Gamma_i^{-1/\alpha_1}, \ldots, \mathrm{sign}(\gamma_{ip})|\gamma_{ip}|^{1/\alpha_p} \Gamma_i^{-1/\alpha_p}\right)}, \tag{3.3.1}$$

in distribution. Here, $\{\Gamma_1, \Gamma_2, \ldots\}$ is a sequence of arrival times of a Poisson process with unit arrival rate and $\boldsymbol{\gamma}_i = (\gamma_{i1}, \ldots, \gamma_{ip}) \sim G$ and G is a distribution on the boundary of unit sphere. To have a valid bootstrap, we expect to have $M_i^* = 1$ which is not the case here.

The limiting distribution for the bootstrap sample mean, $\bar{\mathbf{X}}^*$, can be derived from (3.3.1) and the continuous mapping theorem along with extra mathematical steps. Similar to the univariate case, this result shows that the regular bootstrap fails asymptotically. As discussed in the introduction, a subsampling scheme (m out of n bootstrap such that $m/n \to 0$) is an appropriate approach to achieve asymptotic validity of a bootstrap procedure for constructing a confidence region for the mean vector of i.i.d. heavy-tailed data. However, choosing the proper subsample size m is of great concern to many authors.

On the other hand, Theorem 3.2.1 shows that the weak limit behavior of $\sqrt{n}(\hat{\boldsymbol{\mu}}_M - \boldsymbol{\mu})$ is a multivariate normal distribution. Thus, the regular bootstrap works if we use the robust estimates (M-estimates) for the mean vector. Our approach in this section is to consider a bootstrap approach to estimate the confidence region for $\boldsymbol{\mu}$. Given $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, find M-estimates of $\boldsymbol{\mu}$ in model (3.1.1) using the objective function in (3.2.1). Then calculate the residuals, where

$$\boldsymbol{e}_i = \mathbf{X}_i - \hat{\boldsymbol{\mu}},\ i = 1, 2, \ldots, n. \tag{3.3.2}$$

Let $\boldsymbol{e}_1^*, \ldots, \boldsymbol{e}_n^*$ be a sample of size n from the centered residuals in (3.3.2). These assumptions imply the following lemma.

Lemma 3.3.1 *Let $\{\boldsymbol{e}_1^*, \ldots, \boldsymbol{e}_n^*\}$ be an i.i.d. sample from $F_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_{(\boldsymbol{e}_i - \bar{\boldsymbol{e}} \leq \cdot)}$ where $\boldsymbol{e}_i^* = \left(e_{i1}^*, \ldots, e_{ip}^*\right)$, $i = 1, \ldots, n$, and E^* denotes the expectation under F_n. Also, let ψ_j, $j = 1, \ldots, p$, satisfy conditions A2-A3. Then, for $j = 1, \ldots, p$, we have*

(i) $\mathrm{E}^*(\psi_j(e_{1j}^*)) = 0$.

(ii) $n^{-1/2}\hat{\sigma}_j^{-1} \sum_{i=1}^{n} \psi_j(e_{ij}^*) \xrightarrow{d} Z_j$, *in probability, where Z_j has a standard normal distribution and* $\hat{\sigma}_j^2 = \mathrm{E}^*(\psi_j^2(e_{1j}^*)) = \frac{1}{n}\sum_{i=1}^{n} \psi_j^2(e_{ij}) \xrightarrow{p} \mathrm{E}(\psi_j^2(\epsilon_{1j}))$.

(iii) $\mathrm{E}^*(\psi_j'(e_{1j}^*)) = \frac{1}{n}\sum_{i=1}^{n} \psi_j'(e_{ij}) \xrightarrow{p} \mathrm{E}(\psi_j'(\epsilon_{1j}))$.

Proof: The proofs are straightforward under the condition that the bootstrap errors are symmetric. We omit the proofs here; for more details see Singh (1981) and Moreno and Romo (2012).

□

Now, we find $\{\mathbf{X}_i^*\}_{i=1}^n$ from the model $\mathbf{X}_i^* = \hat{\boldsymbol{\mu}} + \boldsymbol{e}_i^*$. Then, we have

$$\begin{aligned} A_n^*(u_1^*, \ldots, u_p^*) &= \sum_{i=1}^{n} \left(\rho\left(\mathbf{X}_i^* - \hat{\boldsymbol{\mu}}\right) - \rho\left(\mathbf{e}_i^*\right)\right) \\ &= \frac{-1}{\sqrt{n}} \sum_{j=1}^{p} u_j^* \sum_{i=1}^{n} \psi_j(e_{ij}^*) + \frac{1}{2n} \sum_{j=1}^{p} u_j^{*2} \sum_{i=1}^{n} \psi_j^{'}(c_{ij}^*), \end{aligned}$$

where $u_j^* = n^{1/2}(\hat{\mu}_j^* - \hat{\mu}_j)$ for $j = 1, \ldots, p$ and c_{ij}^* is between e_{ij}^* and $e_{ij}^* - n^{-1/2}u_j^*$. Then by Lemma 3.3.1, we have

$$\mathbf{W}^* = \sqrt{n}(\hat{\boldsymbol{\mu}}^* - \hat{\boldsymbol{\mu}}) \xrightarrow{d} \mathbf{W}, \tag{3.3.3}$$

in probability, where $\mathbf{W}$ is defined in (3.2.2). We apply (3.3.3) to approximate the critical points of $\mathbf{W}$. To do so, we carry out a large number, say B, of the bootstrap replicates of size n from

$$C^* = n(\hat{\boldsymbol{\mu}}_M^* - \hat{\boldsymbol{\mu}})^T {\mathbf{S}^*}^{-1}(\hat{\boldsymbol{\mu}}_M^* - \hat{\boldsymbol{\mu}}).$$

Set $\hat{\tau}_\alpha$ to be the 100α-th percentile value of $\{C^*(b), b = 1, 2, \ldots, B\}$. Thus an approximate confidence region for $\boldsymbol{\mu}$ at level $100(1-\alpha)\%$ will be

$$n(\hat{\boldsymbol{\mu}}_M - \boldsymbol{\mu})^T \mathbf{S}^{-1}(\hat{\boldsymbol{\mu}}_M - \boldsymbol{\mu}) \leq \hat{\tau}_{1-\alpha}. \tag{3.3.4}$$

3.4 Simulation

To illustrate the preceding results, some simulation studies are performed. The first step in our simulation study is choosing the loss function. An example for the univariate $\rho_j(\cdot)$ is the Huber loss function given by

$$\rho_j\left(x\right) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq c_j, \\ c_j\left|x\right| - \frac{1}{2}c_j^2 & \text{if } |x| > c_j, \end{cases} \tag{3.4.1}$$

for a known constant c_j; see Huber (1981). Then, $\psi_j(x) = \max[\min(x, c_j), -c_j]$. The Huber loss function satisfies conditions A1-A3 except ψ_j might not exist everywhere.

In this case, although ψ_j is not differentiable at a countable number of points, the results will usually hold with some additional complexity in the proofs.

The choice of a truncation value c_j is of practical interest especially when we have different indices of stability. The following univariate simulation study is undertaken in order to explore whether there is a relationship between values of c_j in the Huber loss function and the index of stability α.

Consider the univariate model

$$X_i = \mu + \epsilon_i,\ i = 1, 2, \ldots, n, \tag{3.4.2}$$

where $\{\epsilon_i\} \in DS(1 < \alpha \leq 2)$. We generate the random samples $\{X_i\}_{i=1}^{n}$ in model (3.4.2) for $\mu = 3$ and $n = 100$ with different values of $1 < \alpha \leq 2$. Then, the M-estimates of μ in (3.4.2) are calculated from the generated random samples using the Huber loss function given in (3.4.1). To seek a more efficient c in the Huber loss function, the estimation is repeated for different values of c between 0.5 and 4.5 for each choice of α. We find the average deviation by calculating the absolute deviation $|\hat{\mu}_M - \mu|$ and then carrying out 10,000 replications. The numbers in Table 3.1 are the averages of the replications. Meanwhile, the scatterplot of the average deviations is presented in Figure 3.1. For each level of α and c, the minimum of the average deviations appear in boldface in Table 3.1. This table shows that, for instance, for $1.1 \leq \alpha \leq 1.4$ if we choose $c = 1$, we get the minimum of the error estimation. Table 3.1 and Figure 3.1 also show that there is a positive relationship between the truncation value c and the index of stability α. In fact, to have less estimation error, we must choose a larger value of c as α gets larger and this conclusion is not surprising.

Now consider the bivariate model

$$\begin{pmatrix} X_{i1} \\ X_{i2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix},\ i = 1, 2, \ldots, n. \tag{3.4.3}$$

Set $(\mu_1, \mu_2) = (1, 14)$ and $\{\boldsymbol{\epsilon}_i\}$ are in a domain of attraction of a symmetric bivariate stable laws with indices of stability $1 < \alpha_1 \leq 2$ and $1 < \alpha_2 \leq 2$. To generate

Table 3.1: Average values of $|\hat{\mu}_M - \mu|$ for different values of α and truncation values of c in the Huber loss function with the replication size of 10,000.

	c								
α	0.5	1	1.5	2	2.5	3	3.5	4	4.5
1.1	**0.123**	**0.126**	0.134	0.143	0.153	0.162	0.171	0.180	0.188
1.2	**0.125**	**0.125**	0.131	0.138	0.146	0.154	0.161	0.169	0.176
1.3	0.128	**0.126**	**0.127**	0.136	0.142	0.149	0.156	0.162	0.168
1.4	0.128	**0.125**	**0.127**	0.131	0.137	0.142	0.147	0.153	0.158
1.5	0.129	**0.125**	**0.125**	**0.128**	0.132	0.136	0.140	0.145	0.148
1.6	0.130	0.128	**0.123**	**0.125**	**0.126**	0.131	0.134	0.137	0.141
1.7	0.130	0.124	**0.122**	**0.122**	**0.123**	0.125	0.127	0.130	0.132
1.8	0.129	0.122	0.120	**0.118**	**0.118**	**0.119**	0.121	0.122	0.124
1.9	0.131	0.124	0.120	**0.118**	**0.117**	**0.118**	**0.118**	0.119	0.120
2.0	0.131	0.123	0.119	**0.116**	**0.115**	**0.114**	**0.114**	**0.114**	**0.114**

$\{\boldsymbol{\epsilon}_1\} = \{(\epsilon_{11}, \epsilon_{12})\}$ in (3.4.3) with the preceding indices of stabilities, consider the set of $K = 10,000$ points $\{\boldsymbol{\gamma}_i = (\cos\theta_i, \sin\theta_i) : \theta_i \in [0, 2\pi],\ i = 1, \ldots, 10000\}$ on the boundary of the unit circle. By (3.3.1), we draw the error $\{\boldsymbol{\epsilon}_1\}$ from

$$\boldsymbol{\epsilon}_1 = (\epsilon_{11}, \epsilon_{12}) = \left(\sum_{i=1}^{K} \mathrm{sign}(\boldsymbol{\gamma}_{\mathrm{i}1}) |\boldsymbol{\gamma}_{\mathrm{i}1}|^{1/\alpha_1} \Gamma_{\mathrm{i}}^{-1/\alpha_1}, \sum_{\mathrm{i}=1}^{\mathrm{K}} \mathrm{sign}(\boldsymbol{\gamma}_{\mathrm{i}2}) |\boldsymbol{\gamma}_{\mathrm{i}2}|^{1/\alpha_2} \Gamma_{\mathrm{i}}^{-1/\alpha_2} \right), \quad (3.4.4)$$

where $\Gamma_i = E_1 + \cdots + E_i$ and $E_j, j \geq 1$ are i.i.d. random variables with a standard exponential distribution. To get the exact value of the innovations, K must tend to ∞. Perform this procedure again n times independently to generate random numbers $\{\boldsymbol{\epsilon}_i\}$, $i = 1, \ldots, n$.

To acquire an intuitive feel for the bivariate observations with different indices of stability, we generate errors from (3.4.4) with indices of stability $(\alpha_1, \alpha_2) = (1.3, 1.8)$ and $(2.0, 1.2)$ and sample size $n = 1,000$. The observations (X_{i1}, X_{i2}), $i = 0, 1, \ldots, n$,

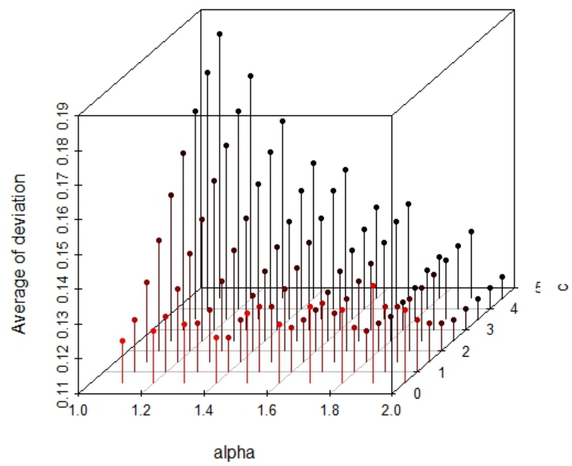


Figure 3.1: Average values of $|\hat{\mu}_M - \mu|$ for different values of α and truncation values of c in the Huber loss function with the replication size of 10,000.

are simulated from (3.4.3), and based on these observations, we plot the joint density of (X_{i1}, X_{i2}), $i = 1, \ldots, n$. Figures 3.2 and 3.3 present the joint density of (X_{i1}, X_{i2}) when $(\alpha_1, \alpha_2) = (1.3, 1.8)$ and $(\alpha_1, \alpha_2) = (2.0, 1.2)$, respectively.

To illustrate the results of Theorem 3.2.1, we perform the following simulation study to construct $100(1-\alpha)\%$ confidence region for the mean vector in model (3.4.3). All the corresponding distributions of the innovations come from symmetric bivariate stable laws with indices of stability $(\alpha_1, \alpha_2) = (1.2, 1.1)$, $(1.5, 1.5)$, $(1.5, 1.9)$, $(1.3, 1.8)$, and $(2.0, 1.2)$ with sample sizes n = 100, 200, and 500. The simulation scheme for each choice of n and (α_1, α_2) is as follows:

(i) Generate $\{\boldsymbol{\epsilon}_i\}$ in (3.4.3) with the preceding indices of stability.

(ii) Find $\{\boldsymbol{X}_i\}_{i=1}^n$ from (3.4.3). Then estimate $\boldsymbol{\mu}$ by $\hat{\boldsymbol{\mu}}_M$ using the bivariate convex function ρ given in (3.2.1) and apply the Huber loss function in (3.4.1) for ρ_1 and ρ_2. Note that, according to the values of α_1 and α_2, the values of c_1 and c_2

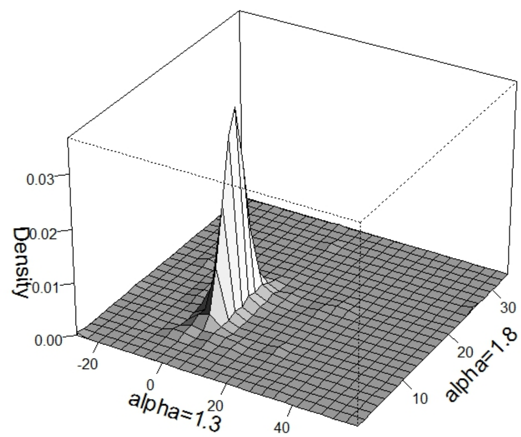


Figure 3.2: Density plot for the bivariate observations (X_{i1}, X_{i2}) given in (3.4.3) where $\alpha_1 = 1.3$ and $\alpha_2 = 1.8$.

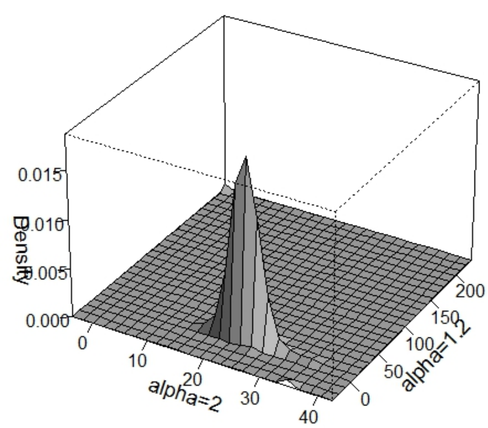


Figure 3.3: Density plot for the bivariate observations (X_{i1}, X_{i2}) given in (3.4.3) where $\alpha_1 = 2$ and $\alpha_2 = 1.2$.

are chosen from Table 3.1 and plugging them in the result in Theorem 3.2.1.

(iii) Estimate Σ by using the residuals, $\boldsymbol{e}_i = \mathbf{X}_i - \hat{\boldsymbol{\mu}}_M,\ i = 1, 2, \ldots, n$.

(iv) Estimate the naive $(1-\alpha)$-percentiles from 3,000 bootstrap replications of $C^* = n(\hat{\boldsymbol{\mu}}_M^* - \hat{\boldsymbol{\mu}})^T \mathbf{S}^{*^{-1}} (\hat{\boldsymbol{\mu}}_M^* - \hat{\boldsymbol{\mu}})$. To do so, draw a sample of size n from centered residuals denoted by $\hat{\boldsymbol{\epsilon}}_1^*, \ldots, \hat{\boldsymbol{\epsilon}}_n^*$, and find $\{\boldsymbol{X}_i^*\}_{i=1}^n$ from (3.4.3). Then estimate the parameters (μ_1, μ_2) using the bootstrap observations by the same minimization of the objective function used in step (ii). Out of 3,000 bootstrap replications, compute the $(1-\alpha)$ quantile of $C^*(b), b = 1, 2, \ldots, 3000$.

(v) Compute the confidence region from (3.3.4).

This experiment was repeated $8,000$ times to estimate the coverage probabilities by checking whether $(\mu_1, \mu_2) = (1, 14)$ are located within the estimated confidence intervals or not. Table 3.2 presents the coverage probabilities for different combinations of α_1 and α_2 when we set confidence levels equal to 0.90, 0.95 and 0.99. As seen in Table 3.2, the M-estimates provide significantly precise estimation such that the coverage probabilities for their related confidence region are fairly close to the nominal confidence levels.

Table 3.2: Estimated coverage probabilities by employing different choices of sample size, indices of stability, and confidence levels

	$n = 100$			$n = 200$			$n = 500$		
(α_1, α_2)	$I_{0.90}$	$I_{0.95}$	$I_{0.99}$	$I_{0.90}$	$I_{0.95}$	$I_{0.99}$	$I_{0.90}$	$I_{0.95}$	$I_{0.99}$
(1.2,1.1)	0.886	0.944	0.987	0.889	0.948	0.987	0.898	0.948	0.988
(1.5,1.5)	0.890	0.945	0.986	0.903	0.948	0.989	0.902	0.946	0.986
(1.5,1.9)	0.888	0.939	0.983	0.900	0.947	0.991	0.891	0.947	0.990
(1.3,1.8)	0.885	0.943	0.984	0.900	0.950	0.988	0.889	0.949	0.986
(2.0,1.2)	0.892	0.942	0.987	0.893	0.945	0.988	0.900	0.943	0.990

3.5 Some notes and remarks

In this chapter,robust estimation of the mean vector is considered when the observations are a sequence of i.i.d. random vectors in the domain of attraction of a stable law with different indices of stability, $DS(\alpha_1, \ldots, \alpha_p)$, such that $1 < \alpha_i \leq 2$, $i = 1, \ldots, p$.

Notice that it is not necessary to enforce the condition that $1 < \alpha \leq 2$. For $0 < \alpha \leq 1$, $E(\mathbf{X}_i)$, $i = 1, \ldots, n$, does not exist. Therefore, $\boldsymbol{\mu}$ is not the mean for the observations $\{\mathbf{X}_i\}$ in (3.1.1) but we can still consider $\boldsymbol{\mu}$ as the location parameter. The estimation procedure and the asymptotic behaviour of our M-estimate remains valid and bootstrapping still works when we wish to estimate the location parameter $\boldsymbol{\mu}$.

The model (3.1.1) can be modified to a more complex form when $\boldsymbol{\mu}$ plays the role of the location parameter in other statistical methods. As an example, in Chapter 6 we will consider the time series models with the location parameter, $X_t = \mu + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t$. Thus, such analysis can help to characterize the asymptotic distribution for the location parameter in more complex statistical models.

Chapter 4

Asymptotic Theory for M-Estimates in Unstable AR(2) Processes with Infinite Variance Innovations

In this chapter, for clarity and simplicity, we present the asymptotic distribution of M-estimators for parameters in non-stationary AR(2) processes. The innovations are assumed to be in the domain of attraction of a stable law with index $0 < \alpha \leq 2$. In particular, when the model involves repeated unit roots or conjugate complex unit roots, M-estimators have a higher asymptotic rate of convergence compared to the LS estimators and the asymptotic results can be written as Itô stochastic integrals. The generalization will be considered in Chapter 5.

4.1 Introduction

Consider the AR(p) model

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t, \quad t = 1, 2, \ldots, n, \tag{4.1.1}$$

where ϵ_t, $t = 1, 2, \ldots, n$, are i.i.d. random variables from some distribution F. Let

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \tag{4.1.2}$$

denote the characteristic polynomial of the model (4.1.1). When all roots of $\phi(z)$ are outside the unit circle, the model is stationary. A commonly used estimate of the parameter vector $\Phi = (\phi_1, \ldots, \phi_p)^T$ is the LS estimate

$$\hat{\Phi}_{LS} = \left(\sum_{t=0}^{n-1} \mathbf{X}_t \mathbf{X}_t^T \right)^{-1} \sum_{t=1}^{n} \mathbf{X}_{t-1} X_t, \quad n > p,$$

where (1.4)

$$\mathbf{X}_t = (X_t, \ldots, X_{t-p+1})^T,$$

and $X_0 = \cdots = X_{1-p} = 0$. The statistical properties of $\hat{\Phi}_{LS}$ have been considered in the literature for the stationary case. Mann and Wald [34] derive the asymptotic distribution of $\hat{\Phi}_{LS}$ when X_n is stationary. They show that if $\{\epsilon_t\}$ are i.i.d. with all moments, we have

$$\sqrt{n}(\hat{\Phi}_{LS} - \Phi) \xrightarrow{d} N(0, \Sigma),$$

where Σ is a positive definite matrix. Anderson [2] obtains the asymptotic distribution of $\hat{\Phi}_{LS}$ by removing the higher moment assumption.

For the unstable case, White [55] considers the AR(1) model when $\phi_1 = 1$ and errors are i.i.d. random variables from $N(0, \sigma^2)$. It can be proved that

$$n(\hat{\phi}_{LS} - 1) \xrightarrow{d} \tau = \frac{W^2(1) - 1}{2 \int_0^1 W^2(s) ds}, \tag{4.1.3}$$

where $W(\cdot)$ is a standard Brownian-motion process. Notice that $\hat{\phi}_{LS} \xrightarrow{P} \phi$ at the unusual rate n instead of $\sqrt{n}$. This rate of convergency only holds for the unit root models ($\phi = 1$). Dickey and Fuller [15] prove that

$$n(\hat{\phi}_{LS} - 1) \xrightarrow{d} \frac{T^2 - 1}{2\Gamma}, \tag{4.1.4}$$

where

$$T = \sqrt{2}\sum_{i=1}^{\infty} \gamma_i Z_i, \qquad \Gamma = \sum_{i=1}^{\infty} \gamma_i^2 Z_i^2, \qquad \gamma_i = 2(-1)^{i+1}/[(2i-1)\pi], \quad \text{for } i = 1, 2, \ldots$$

and the $\{Z_i\}$ are i.i.d. N(0,1) random variables. They also tabulate the percentiles of τ in (4.1.3) by the Monte Carlo method using (4.1.4).

For the infinite variance case, when $p = 1$, Chan and Tran [10] consider the limiting distribution of $\hat{\phi}_{LS}$ for a random-walk process. When the errors are in the domain of attraction of a stable law, the limiting distribution of $\hat{\phi}_{LS}$ is as follows:

$$n(\hat{\phi}_{LS} - 1) \xrightarrow{d} \frac{S^2(1) - R(1)}{2\int_0^1 S^2(s)ds}, \tag{4.1.5}$$

where $S(\cdot)$ and $R(\cdot)$ are stable processes and from Resnick and Greenwood [44] and Knight [27], their representations are as follows

$$S(t) = \begin{cases} \sum_{k=1}^{\infty} \delta_k \Gamma_k^{-1/\alpha} I(U_k \le t) & \text{if } 0 < \alpha < 2, \\ \text{standard Brownian motion} & \text{if } \alpha = 2, \end{cases} \tag{4.1.6}$$

$$R(t) = \begin{cases} \sum_{k=1}^{\infty} \Gamma_k^{-2/\alpha} I(U_k \le t) & \text{if } 0 < \alpha < 2, \\ t & \text{if } \alpha = 2. \end{cases}$$

Here and throughout this chapter, $\{U_k\}$ is a sequence of i.i.d. $U[0,1]$ random variables and $\{\delta_k\}$ is a sequence of i.i.d. random variables such that $P(\delta_k = 1) = p$, $P(\delta_k = -1) = q$, and $p + q = 1$. Also, $\Gamma_1, \Gamma_2, \ldots$ are the arrival times of a Poisson process with Lebesgue mean measure and $\{U_k, \Gamma_k, \delta_k\}$ are mutually independent. Note that, when $\alpha = 2$, $R(1) = 1$ and so the limiting random variable is the same as in the finite-variance case.

On the other hand, there are important time series models, for instance seasonal models, where $\phi(z)$ may have several roots on the unit circle. It has always been an interesting and challenging task to conduct asymptotic inference for time series with multiple unit roots.

For a general unit root AR(p) model like (4.1.1), we not only have to pay attention to the multiplicity of each root but also the relationship between different roots has to be taken into account. This general case was first considered by Chan and Wei [11] when $\{\epsilon_t\}$ is a martingale difference sequence. It was generalized by Chan and Terrin [13] for the case that $\{\epsilon_t\}$ is long-memory and Jeganathan [25] to the nearly non-stationary case when $\{\epsilon_t\}$ is a sequence i.i.d. random variables with finite variance.

Chan and Zhang [12] obtain the limiting distribution for the LS estimates of the parameters for non-stationary AR(p) processes, with i.i.d. innovations in the domain of attraction of a stable law. They show that the limiting distribution of the LS estimate is a function of integrated stable processes. However, for model (4.1.1) with unit roots when $\{\epsilon_t\}$ is a sequence of random variables with infinite variance, a complete theory on more efficient estimating technique is still missing in the literature.

In the forthcoming chapters of this thesis, we will consider a robust estimate of Φ, denoted by $\hat{\Phi}_M = (\hat{\phi}_{1M}, \ldots, \hat{\phi}_{pM})^T$, using the objective function

$$\sum_{t=p+1}^{n} \rho(X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p})$$

for some properly chosen convex function ρ. These estimates are commonly called M-estimates and are usually used as a robust alternative to LS estimates. The asymptotic behaviour of the robust estimates for stationary AR(p) processes was considered by Davis et al. [14]. They show that with $\{\epsilon_t\} \in DS(\alpha)$, we have

$$a_n(\hat{\Phi}_M - \Phi) \xrightarrow{d} \text{to some random vector } \boldsymbol{\xi}_1.$$

Unfortunately, due to complexity, there is no tractable representation for the limiting distribution. Knight [27] considers the limit theory for AR parameter estimates in an infinite variance random walk. By imposing some conditions on the errors, ρ, and taking $\psi = \rho'$, Knight [27] for AR(1) proves that

$$n^{1/2} a_n(\hat{\phi}_M - 1) \xrightarrow{d} \frac{E^{\frac{1}{2}}\left(\psi^2(\epsilon_1)\right)}{E\left(\psi'(\epsilon_1)\right)} \frac{\int_0^1 S(t)dW(t)}{\int_0^1 S^2(t)dt}$$

and

$$\left(\sum_{t=2}^{n} X_{t-1}^2\right)^{1/2} (\hat{\phi}_M - 1) \xrightarrow{d} \frac{E^{\frac{1}{2}}\left(\psi^2(\epsilon_1)\right)}{E\left(\psi'(\epsilon_1)\right)} W(1), \tag{4.1.7}$$

where $S(\cdot)$ is a stable process and $W(\cdot)$ is a standard Brownian motion independent of $S(\cdot)$. The self normalized asymptotic result in (4.1.7) avoids using any Monte Carlo method as the limiting distribution is normally distributed.

In this chapter, we derive the asymptotic distribution of M-estimators for the parameters in an unstable AR(2) process, where the innovations are in the domain of attraction of a stable law with index $0 < \alpha \leq 2$. Our results show that, similar to the previous cases, M-estimators have a higher asymptotic rate of convergence than LS estimators. Notice that our analysis covers the finite variance case ($\alpha = 2$).

Section 4.2 contains some necessary preliminary concepts. In section 4.3, the limiting distribution of M-estimates in an AR(2) model is derived. In section 4.4, we carry out a simulation study to demonstrate the practical precision. Due to complexity of the limiting distributions for the estimates, a brief discussion and a bootstrap simulation scheme are presented in section 4.5. Note that, since we only consider M-estimators, the subscript M is dropped in all equations in the forthcoming chapters.

4.2 Preliminaries

Consider the autoregressive process of order 2 (AR(2))

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t, \tag{4.2.1}$$

where the errors $\{\epsilon_t\}$ in (4.2.1) form a sequence of i.i.d. random variables in the domain of attraction of a symmetric stable law with index $0 < \alpha \leq 2$. The classical M-estimator $\hat{\Phi} = (\hat{\phi}_1, \hat{\phi}_2)^T$ of $\Phi = (\phi_1, \phi_2)^T$ minimizes

$$\sum_{t=3}^{n} \rho(X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2}), \tag{4.2.2}$$

where ρ is an almost everywhere differentiable convex function. This guarantees the uniqueness of the solution. Usually, $\rho(x)$ grows at a slower rate than x^2, as $|x|$ gets large. For more details see [14]. Throughout this chapter we impose assumptions A1-A3, defined in Chapter 3, on the function $\rho(\cdot)$.

An example for $\rho(\cdot)$ is the Huber loss function defined in (3.4.1). We can also consider the loss function $\rho^*(x) = \ln(b + x^2)$, for some $b > 0$. The function $(d/dx)\rho^*(x) = 2x/(b + x^2)$ is the score function of a t-distribution with b degrees of freedom. The sufficiently smooth loss function ρ^* does not satisfy A3 and is just locally convex around its global minimum. Our simulations in Section 4.4 show that the performance of ρ^* is as good as that of the Huber loss function.

Moreover, we assume that the innovations $\{\epsilon_t\}$ satisfy:

Assumption 4 *(A4) The innovations* $\{\epsilon_t\}$ *are i.i.d. random variables in the domain of attraction of a stable law with index* $0 < \alpha \leq 2$ *(see Proposition 2.5.1).*

It is common to assume symmetry for the innovations. Note that for $0 < \alpha < 1$, symmetry is not required. Therefore, for $1 < \alpha \leq 2$ we impose symmetry on the innovations; i.e., $p = q = 1/2$. However, for $1 < \alpha \leq 2$ symmetry is not required, if $\mathrm{E}(\epsilon_1) = 0$.

Assumption A4 implies that:

$$S_n(t) = a_n^{-1} \sum_{k=1}^{[nt]} \epsilon_k \ \xrightarrow{d}\ S(t) \quad \text{in} \quad D[0,1], \tag{4.2.3}$$

where $\{a_n\}$ is a sequence of positive constants defined as in (2.5.3) and $\xrightarrow{d}$ denotes here convergence in distribution with respect to the Skorohod topology. The Skorohod space of the càdlàg functions defined on $[0,1]$, equipped with the Skorohod topology, is denoted by $D = D[0,1]$. Moreover, $S(\cdot)$ is a stable process and defined in (4.1.6). The series in (4.1.6) is convergent if either $0 < \alpha < 1$ or $p = q = 1/2$. For more details, see LePage, Woodroofe, and Zinn [29].

To derive the main result of this chapter, we assume that conditions A1-A4 hold and we define the following processes on the Skorohod space $D[0,1]$:

$$\begin{aligned}
S_n^{(1)}(t) &= a_n^{-1} \sum_{k=1}^{[nt]} (-1)^k \epsilon_k, \\
\mathbf{T}_n(t) &= \begin{pmatrix} T_{1,n} \\ T_{2,n} \end{pmatrix} = a_n^{-1} \sum_{k=1}^{[nt]} \begin{pmatrix} \cos(k\theta) \\ \sin(k\theta) \end{pmatrix} \epsilon_k, \\
W_n(t) &= n^{-1/2} \sum_{k=1}^{[nt]} \psi(\epsilon_k), \\
\mathbf{R}_n(t) &= \begin{pmatrix} R_{1,n} \\ R_{2,n} \end{pmatrix} = n^{-1/2} \sum_{k=1}^{[nt]} \begin{pmatrix} \sin\left((k-1)\theta\right) \\ \cos\left((k-1)\theta\right) \end{pmatrix} \psi(\epsilon_k), \\
V_n(t) &= n^{-1/2} \sum_{k=1}^{[nt]} \left[\psi'(\epsilon_k) - \mathrm{E}\left(\psi'(\epsilon_k)\right) \right],
\end{aligned} \tag{4.2.4}$$

where $[x]$ stands for integer part of x.

From Resnick [46], we have

$$\sum_{k=1}^{n} \varepsilon_{\left(k/n, a_n^{-1}\epsilon_k\right)} \ \xrightarrow{d}\ \sum_{k=1}^{\infty} \varepsilon_{\left(U_k, \delta_k \Gamma_k^{-1/\alpha}\right)}, \tag{4.2.5}$$

where $U_k \overset{i.i.d.}{\sim} U[0,1]$ is independent from $\{\delta_k, \Gamma_k\}$, which are defined in (4.1.6). By applying a technique similar to Proposition 3.4 of Resnick [45] and the continuous mapping theorem on (4.2.5), we have

$$\begin{pmatrix} S_n \\ S_n^{(1)} \\ \mathbf{T}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} S \\ S^{(1)} \\ \mathbf{T} \end{pmatrix},$$

where S is defined in (4.1.6) and $\mathbf{T}$ is a bivariate stable process which is defined in Lemma 4.3.2; see [46] for more details. Furthermore

$$S^{(1)} = \sum_{k=1}^{\infty} \delta'_k \Gamma_k^{-1/\alpha} I(U_k \le t),$$

where $\{\delta'_k\}$ are independent from $\{\delta_k\}$ and

$$P(\delta_k = 1) = P(\delta'_k = 1) = P(\delta_k = -1) = P(\delta'_k = -1) = 1/2.$$

We also have

$$\begin{pmatrix} W_n \\ \mathbf{R}_n \\ V_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} W \\ \mathbf{R} \\ V \end{pmatrix},$$

where $W(\cdot)$ and $V(\cdot)$ are standard Brownian-motion processes with $\mathrm{E}(W^2(t)) = t\mathrm{E}\left(\psi^2(\epsilon_1)\right)$ and $\mathrm{E}(V^2(t)) = t\mathrm{Var}\left(\psi'(\epsilon_1)\right)$. From the fact that $\sum_{k=1}^{n} \sin^2((k-1)\theta)$ and $\sum_{k=1}^{n} \cos^2\left((k-1)\theta\right)$ are $O_p(n)$ for large values of n, the Lindeberg Feller central limit theorem and the tightness of the partial sum process $\mathbf{R}_n$ give

$$\mathbf{R}(t) = \begin{pmatrix} R_1(t) \\ R_2(t) \end{pmatrix} = \frac{\mathrm{E}\left(\psi'^2(\epsilon_1)\right)}{2} \begin{pmatrix} \mathcal{W}_1(t) \\ \mathcal{W}_2(t) \end{pmatrix}. \tag{4.2.6}$$

Here, $\mathcal{W}_1(\cdot)$, and $\mathcal{W}_2(\cdot)$ are independent standard Brownian motion processes. Therefore, similar to Theorem 4 of Resnick and Greenwood [44], we can show that

$$\left(S_n, S_n^{(1)}, \mathbf{T}_n, W_n, \mathbf{R}_n, V_n\right)^T \xrightarrow{d} \left(S, S^{(1)}, \mathbf{T}, W, \mathbf{R}, V\right)^T. \tag{4.2.7}$$

For $0 < \alpha < 2$, $\left(S(\cdot), S^{(1)}(\cdot), \mathbf{T}(\cdot)\right)^T$ is independent of $\left(W(\cdot), \mathbf{R}(\cdot), V(\cdot)\right)^T$. For more details see [44].

4.3 The limiting distribution for the parameters of unstable AR(2) processes

The limiting distribution for the M-estimates of the parameters for an infinite-variance random-walk processes is obtained in Knight [27]. To generalize, we extend the results of Knight [27] to the AR(2) process when characteristic roots may have different multiplicities and lie on the unit circle. Theorem 4.3.1 presents the weak limit behaviour of M-estimates for the AR(2) processes defined in (4.2.1).

Theorem 4.3.1 *Suppose $\{X_t\}$ is an AR(2) process. Consider the following cases.*

(i) When $\phi(z) = 1 - 2z\cos\theta + z^2$ for $\theta \in (0, \pi)$,

$$n^{1/2}a_n \begin{pmatrix} \hat{\phi}_1 - 2\cos\theta \\ \\ \hat{\phi}_2 + 1 \end{pmatrix} \xrightarrow{d} \Gamma_1^{-1} \begin{pmatrix} \frac{2\mathrm{E}^{1/2}\left(\psi^2(\varepsilon_1)\right)\digamma_1 \sin\theta}{\mathrm{E}(\psi'(\varepsilon_1))\left(\int_0^1 T_1^2(t)d(t) + \int_0^1 T_2^2(t)d(t)\right)} \\ \\ \frac{2\mathrm{E}^{1/2}\left(\psi^2(\varepsilon_1)\right)\digamma_2 \sin\theta}{\mathrm{E}(\psi'(\varepsilon_1))\left(\int_0^1 T_1^2(t)d(t) + \int_0^1 T_2^2(t)d(t)\right)} \end{pmatrix},$$

where

$$\Gamma_1 = \begin{pmatrix} 1 & \cos\theta \\ \cos\theta & 1 \end{pmatrix},$$

$$\begin{aligned} \digamma_1 = &\cos\theta \left(\int_0^1 T_1(t)dR_1(t) - \int_0^1 T_2(t)dR_2(t) \right) \\ &+ \sin\theta \left(\int_0^1 T_1(t)dR_2(t) + \int_0^1 T_2(t)dR_1(t) \right), \end{aligned}$$

and

$$\digamma_2 = \int_0^1 T_1(t)dR_1(t) - \int_0^1 T_2(t)dR_2(t).$$

(ii) When $\phi(z) = 1 - 2z + z^2$,

$$\begin{pmatrix} n^{1/2}a_n(\hat{\phi}_1 - 2) \\ \\ n^{3/2}a_n(\hat{\phi}_1 - 2) + n^{3/2}a_n(\hat{\phi}_2 + 1) \end{pmatrix} \xrightarrow{d} \Gamma_2^{-1} \begin{pmatrix} \frac{\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right) \int_0^1 S(t)\, dW(t)}{\mathrm{E}(\psi'(\epsilon_1))} \\ \\ \frac{\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right) \int_0^1 \int_0^t S(s)\, ds\, dW(t)}{\mathrm{E}(\psi'(\epsilon_1))} \end{pmatrix},$$

where

$$\Gamma_2 = \begin{pmatrix} \int_0^1 S^2(t)\ dt & \int_0^1 S(t) \int_0^t S(s)\ ds\, dt \\ \int_0^1 S(t) \int_0^t S(s)\ ds\, dt & \int_0^1 \left(\int_0^t S(s)\ ds\right)^2 dt \end{pmatrix}.$$

(iii) When $\phi(z) = 1 + 2z + z^2$,

$$\begin{pmatrix} n^{3/2} a_n(\hat{\phi}_1 + 2) - n^{3/2} a_n(\hat{\phi}_2 + 1) \\ \\ n^{1/2} a_n(\hat{\phi}_2 + 1) \end{pmatrix} \xrightarrow{d} -\Gamma_3^{-1} \begin{pmatrix} \frac{\mathrm{E}^{1/2}\left(\psi^2(\varepsilon_1)\right) \int_0^1 \int_0^t S^{(1)}(s)\, ds\, dW(t)}{\mathrm{E}(\psi'(\varepsilon_1))} \\ \\ \frac{\mathrm{E}^{1/2}\left(\psi^2(\varepsilon_1)\right) \int_0^1 S^{(1)}(t)\, dW(t)}{\mathrm{E}(\psi'(\varepsilon_1))} \end{pmatrix},$$

where

$$\Gamma_3 = \begin{pmatrix} \int_0^1 \left(\int_0^t S^{(1)}(s)\, ds\right)^2 dt & \int_0^1 S^{(1)}(t) \int_0^t S^{(1)}(s)\, ds\, dt \\ \int_0^1 S^{(1)}(t) \int_0^t S^{(1)}(s)\ ds\, dt & \int_0^1 \left(S^{(1)}(t)\right)^2 dt \end{pmatrix}.$$

(iv) When $\phi(z) = 1 - z^2$,

$$\begin{pmatrix} n^{1/2} a_n \hat{\phi}_1/2 + n^{1/2} a_n(\hat{\phi}_2 - 1)/2 \\ \\ n^{1/2} a_n \hat{\phi}_1/2 - n^{1/2} a_n(\hat{\phi}_2 - 1)/2 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{\mathrm{E}^{1/2}\left(\psi^2(\varepsilon_1)\right) \int_0^1 S(t)\, dW(t)}{\mathrm{E}(\psi'(\varepsilon_1)) \int_0^1 S^2(t)\, dt} \\ \\ -\frac{\mathrm{E}^{1/2}\left(\psi^2(\varepsilon_1)\right) \int_0^1 S^{(1)}(t)\, dW(t)}{\mathrm{E}(\psi'(\varepsilon_1)) \int_0^1 \left(S^{(1)}(t)\right)^2 dt} \end{pmatrix},$$

where $S(\cdot)$, $S^{(1)}(\cdot)$, $\mathbf{T}(\cdot)$, $W(\cdot)$, *and* $\mathbf{R}(\cdot)$ *are defined in* (4.2.7).

Remark:

- The results in Theorem 4.3.1, cases (ii) and (iii), show different convergence rates for the estimators. Broadly speaking, the linear combination of $\hat{\phi}_1$ and $\hat{\phi}_2$ converges faster than each of them individually.
- When $\alpha = 2$, T_1 and T_2 are Gaussian processes with certain covariance functions.

To facilitate the proof of Theorem 4.3.1, we first present the following lemmas and a proposition.

Lemma 4.3.1 *Let* $(Z_{1,k}, Z_{2,k})^T$ *be a sequence of symmetric i.i.d. random vectors on* $\mathcal{Z}_2 = \{(z_1, z_2)^T : z_1^2 + z_2^2 = 1\}$ *with a probability distribution* Q *on the boundary of the unit circle, and*

$$\mathcal{X} = \left(\sum\nolimits_{k=1}^{\infty} Z_{1,k}\Gamma_k^{-1/\alpha}, \sum\nolimits_{k=1}^{\infty} Z_{2,k}\Gamma_k^{-1/\alpha}\right)^T.$$

Then the characteristic function for $\mathcal{X}$ *is* $\phi(\mathbf{s}) = \exp\left[-KE\left(|s_1 Z_{1,1} + s_2 Z_{2,1}|^\alpha\right)\right]$, *where* $\mathbf{s} = (s_1, s_2)$, *and* K *is given by*

$$K = \begin{cases} \cos(\pi\alpha/2)\Gamma(1-\alpha) & \text{if } 0 < \alpha < 1, \\ \pi(2-\alpha)/(2\alpha) & \text{if } \alpha = 1, \\ \cos(\pi\alpha/2)\Gamma(3-\alpha)/(\alpha^2-\alpha) & \text{if } 1 \le \alpha < 2. \end{cases}$$

Here, Γ_k *is defined in* (4.1.6) *and* $\{Z_k\}$ *is independent from* $\{\Gamma_k\}$.

Proof. The proof is similar to the proof given in Theorem 1 of Banjevic, Ishwaran, and Zarepour [8]. For $u \ge 0$ define

$$\mathcal{X}(u) = \left(\sum\nolimits_{k=1}^{\infty} Z_{1,k}(\Gamma_k + u)^{-1/\alpha}, \sum\nolimits_{k=1}^{\infty} Z_{2,k}(\Gamma_k + u)^{-1/\alpha}\right)^T.$$

By conditioning on Γ_1 and $(Z_{1,1}, Z_{2,1})^T$, we can prove that the characteristic function for $\mathcal{X}(u)$ is

$$\phi(\mathbf{s}, u) = \exp\left[-\int_u^\infty \int_{\mathcal{Z}_2} \left(1 - \cos(s_1 w^{-1/\alpha} z_1 + s_1 w^{-1/\alpha} z_2)\right)\right) dQ((z_1, z_2))\, dw\right].$$

For $u = 0$, we get

$$\begin{aligned}\phi(\mathbf{s}, 0) &= \exp\left[-K\int_{\mathcal{Z}_2} |s_1 z_1 + s_1 z_2|^\alpha dQ(z_1, z_2)\right] \\ &= \exp\left(-K\mathrm{E}|s_1 Z_{1,1} + s_1 Z_{2,1}|^\alpha\right).\end{aligned}$$

For more details see also LePage et al. [29], and Samorodnitsky and Taqqu [50].

$\square$

Lemma 4.3.2 *Let $\mathbf{T}_n$ be defined by (4.2.4), then*

$$\mathbf{T}_n \xrightarrow{d} \mathbf{T},$$

where $\mathbf{T}(t) = (T_1(t), T_2(t))^T$ is a bivariate stable process with index α.

Proof. By applying the continuous mapping theorem on (4.2.5), for any $\theta \in (0, 2\pi)$, we have

$$\sum_{k=1}^{n} \varepsilon_{\left(\cos(\theta k/n), \sin(\theta k/n), a_n^{-1}\epsilon_k\right)} \xrightarrow{d} \sum_{k=1}^{\infty} \varepsilon_{\left(\cos(\theta U_k), \sin(\theta U_k), \delta_k \Gamma_k^{-1/\alpha}\right)}.$$

Similar to Resnick's Proposition 3.4 [45], it can be shown that

$$a_n^{-1} \sum_{k=1}^{[nt]} \left(\cos\left(\theta k/n\right), \sin\left(\theta k/n\right)\right) \epsilon_k \xrightarrow{d} \sum_{k=1}^{\infty} \left(\cos\left(\theta U_k\right), \sin\left(\theta U_k\right)\right) \delta_k \Gamma_k^{-1/\alpha} I(U_k \leq t). \tag{4.3.1}$$

Notice that we can easily find the finite-dimensional distribution for the limiting process in (4.3.1). For instance, for $t = 1$ and by applying Lemma 4.3.1 where $(Z_{1,k}, Z_{2,k}) = (\cos(\theta U_k)\delta_k, \sin(\theta U_k)\delta_k)$, and $p = q = 1/2$, we have

$$\phi(\mathbf{s}) = \exp\left[-K \int_0^1 |s_1 \cos(\theta u) + s_2 \sin(\theta u)|^\alpha du\right].$$

□

Lemma 4.3.3 *Suppose that $S_n(t)$, $n \geq 1$, is a sequence of processes such that*

$$S_n(t) \xrightarrow{d} S(t),$$

then

$$\sup_{0<j\leq n} \left| \sum_{k=1}^{j} e^{ik\theta} S_n\left(\frac{k}{n}\right) \right| = o_p(n). \tag{4.3.2}$$

Proof. See Proposition 8 of Jeganathan [25]. □

Proposition 4.3.1 *Suppose that* (4.2.3) *holds and let* $M_n(s) = a_n^{-1}\sum_{k=1}^{[ns]}\cos(k\theta)\epsilon_k$ *such that*

$$\begin{pmatrix} S_n(s) \\ M_n(s) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} S(s) \\ M(s) \end{pmatrix}.$$

Then, as $n \to \infty$*, for* $0 \le t \le 1$*, we have*

$$\begin{aligned} &\left(S_n(s), M_n(s), \int_0^t S_n(s)dS_n(s), \int_0^t S_n(s)dM_n(s)\right)^T \\ &\xrightarrow{d} \left(S(s), M(s), \int_0^t S(s)dS(s), \int_0^t S(s)dM(s)\right)^T. \end{aligned}$$

Proof. The proof is given in Proposition 2 of Paulauskas and Rachev [40]. See also [12], [28], and [36]. □

Proof of Theorem 4.3.1 Case (i). Consider the model

$$[1 - 2\cos\theta B + B^2]X_t = \epsilon_t, \quad t = 1, 2, \dots, n.$$

We can write

$$\begin{aligned} X_t &= [(1 - e^{i\theta}B)(1 - e^{-i\theta}B)]^{-1}\epsilon_t \\ &= \frac{1}{B\sin\theta}\sum_{k=0}^{\infty} B^k \sin(k\theta)\epsilon_t \\ &= \frac{1}{\sin\theta}\sum_{k=0}^{\infty} B^{k-1}\sin(k\theta)\epsilon_t \\ &= \frac{1}{\sin\theta}\left[\sin\theta + B\sin 2\theta + \cdots\right]\epsilon_t \\ &= \frac{1}{\sin\theta}\sum_{k=1}^{t}\sin\left(\theta(t-k+1)\right)\epsilon_k. \end{aligned}$$

By applying trigonometric identities, we have

$$\sin\theta X_t = a_n \sin\left((t+1)\theta\right) T_{1,n}\left(\frac{t}{n}\right) - a_n \cos\left((t+1)\theta\right) T_{2,n}\left(\frac{t}{n}\right), \tag{4.3.3}$$

where

$$T_{1,n}(\cdot) = a_n^{-1}\sum_{k=1}^{[n\cdot]}\cos(k\theta)\epsilon_k$$

and

$$T_{2,n}(\cdot) = a_n^{-1}\sum_{k=1}^{[n\cdot]}\sin(k\theta)\epsilon_k.$$

Now by (4.2.2), define the following two-parameter stochastic process

$$A_n(u,v) = \sum_{t=3}^{n}\left[\rho(\epsilon_t - n^{-1/2}a_n^{-1}uX_{t-1} - n^{-1/2}a_n^{-1}vX_{t-2}) - \rho(\epsilon_t)\right].$$

Here, ρ is a convex and differentiable function. Note that the minimizer of $A_n(u,v)$ is $(u,v) = \left(n^{1/2}a_n(\hat\phi_1 - 2\cos\theta), n^{1/2}a_n(\hat\phi_2+1)\right)^T$. Using a Taylor series expansion of each summand of A_n around $u=0$ and $v=0$, we get

$$\begin{aligned} A_n(u,v) &= -un^{-1/2}a_n^{-1}\sum_{t=3}^{n}X_{t-1}\psi(\epsilon_t) - vn^{-1/2}a_n^{-1}\sum_{t=3}^{n}X_{t-2}\psi(\epsilon_t) \\ &\quad + \frac{1}{2}u^2n^{-1}a_n^{-2}\sum_{t=3}^{n}X_{t-1}^2\psi'(c_t^n) + \frac{1}{2}v^2n^{-1}a_n^{-2}\sum_{t=3}^{n}X_{t-2}^2\psi'(c_t^n) \\ &\quad + uvn^{-1}a_n^{-2}\sum_{t=3}^{n}X_{t-1}X_{t-2}\psi'(c_t^n), \end{aligned} \tag{4.3.4}$$

where $c_t^n = \epsilon_t - un^{-1/2}a_n^{-1}X_{t-1} - vn^{-1/2}a_n^{-1}X_{t-2}$, and

$$|\psi'(\epsilon_t) - \psi'(c_t^n)| \le \lambda\left|un^{-1/2}a_n^{-1}X_{t-1} + vn^{-1/2}a_n^{-1}X_{t-2}\right|.$$

Asymptotically, $\psi'(c_t^n)$ can be replaced by $\psi'(\epsilon_t)$ in (4.3.4). For simplicity we only consider the third term of A_n (the proof is similar for the other terms in (4.3.4)). We have

$$u^2n^{-1}a_n^{-2}\sum_{t=3}^{n}X_{t-1}^2|\psi'(\epsilon_t) - \psi'(c_t^n)|$$

$$
\begin{aligned}
&\leq ku^2 n^{-1} a_n^{-2} \sum_{t=3}^{n} X_{t-1}^2 |un^{-1/2} a_n^{-1} X_{t-1} + vn^{-1/2} a_n^{-1} X_{t-2}| \\
&\leq ku^3 n^{-1/2} n^{-1} a_n^{-3} \sum_{t=3}^{n} |X_{t-1}|^3 \\
&+ ku^2 v n^{-1/2} n^{-1} a_n^{-3} \sum_{t=3}^{n} |X_{t-1}|^2 |X_{t-2}| \xrightarrow{P} 0.
\end{aligned}
$$

In the following equations, we show that (4.3.3) implies all terms of A_n can be expressed as functionals of $T_{1,n}(\cdot)$ and $T_{2,n}(\cdot)$.

$$
\begin{aligned}
a_n^{-1} \sin\theta \sum_{t=3}^{n} X_{t-1}\epsilon_t &= \sum_{t=3}^{n} T_{1,n}\left(\frac{t-1}{n}\right) \sin(t\theta)\epsilon_t \\
&- \sum_{t=3}^{n} T_{2,n}\left(\frac{t-1}{n}\right) \cos(t\theta)\epsilon_t,
\end{aligned}
$$

$$
\begin{aligned}
n^{-1/2} a_n^{-1} \sin\theta \sum_{t=3}^{n} X_{t-1}\psi(\epsilon_t) &= n^{-1/2} \sum_{t=3}^{n} T_{1,n}\left(\frac{t-1}{n}\right) \sin(t\theta)\psi(\epsilon_t) \\
&- n^{-1/2} \sum_{t=3}^{n} T_{2,n}\left(\frac{t-1}{n}\right) \cos(t\theta)\psi(\epsilon_t) \\
&= n^{-1/2} \cos\theta \sum_{t=3}^{n} T_{1,n}\left(\frac{t-1}{n}\right) \sin((t-1)\theta)\psi(\epsilon_t) \\
&+ n^{-1/2} \sin\theta \sum_{t=3}^{n} T_{1,n}\left(\frac{t-1}{n}\right) \cos((t-1)\theta)\psi(\epsilon_t) \\
&- n^{-1/2} \cos\theta \sum_{t=3}^{n} T_{2,n}\left(\frac{t-1}{n}\right) \cos((t-1)\theta)\psi(\epsilon_t) \\
&+ n^{-1/2} \sin\theta \sum_{t=3}^{n} T_{1,n}\left(\frac{t-1}{n}\right) \sin((t-1)\theta)\psi(\epsilon_t),
\end{aligned}
$$

$$
\begin{aligned}
n^{-1/2} a_n^{-1} \sin\theta \sum_{t=3}^{n} X_{t-2}\psi(\epsilon_t) &= n^{-1/2} \sum_{t=3}^{n} T_{1,n}\left(\frac{t-2}{n}\right) \sin((t-1)\theta)\psi(\epsilon_t) \\
&- n^{-1/2} \sum_{t=3}^{n} T_{2,n}\left(\frac{t-2}{n}\right) \cos((t-1)\theta)\psi(\epsilon_t),
\end{aligned}
$$

$$
\begin{aligned}
2n^{-1} a_n^{-2} \sin^2\theta \sum_{t=3}^{n} X_{t-1}^2 \psi'(\epsilon_t) &= n^{-1} \sum_{t=3}^{n} T_{1,n}^2\left(\frac{t-1}{n}\right) \psi'(\epsilon_t) \\
&+ n^{-1} \sum_{t=3}^{n} T_{2,n}^2\left(\frac{t-1}{n}\right) \psi'(\epsilon_t) \\
&- n^{-1} \sum_{t=3}^{n} \cos(2t\theta) \\
&\times \left(T_{1,n}^2\left(\frac{t-1}{n}\right) - T_{2,n}^2\left(\frac{t-1}{n}\right)\right) \psi'(\epsilon_t) \\
&- 2n^{-1} \sum_{t=3}^{n} \sin(2t\theta) \\
&\times T_{1,n}\left(\frac{t-1}{n}\right) T_{2,n}\left(\frac{t-1}{n}\right) \psi'(\epsilon_t),
\end{aligned}
$$

$$
\begin{aligned}
2n^{-1} a_n^{-2} \sin^2\theta \sum_{t=3}^{n} X_{2-1}^2 \psi'(\epsilon_t) &= n^{-1} \sum_{t=3}^{n} T_{1,n}^2\left(\frac{t-2}{n}\right) \psi'(\epsilon_t) \\
&+ n^{-1} \sum_{t=3}^{n} T_{2,n}^2\left(\frac{t-2}{n}\right) \psi'(\epsilon_t) \\
&- n^{-1} \sum_{t=3}^{n} \cos(2(t-1)\theta) \\
&\times \left(T_{1,n}^2\left(\frac{t-2}{n}\right) - T_{2,n}^2\left(\frac{t-2}{n}\right)\right) \psi'(\epsilon_t) \\
&- 2n^{-1} \sum_{t=3}^{n} \sin(2(t-1)\theta) \\
&\times T_{1,n}\left(\frac{t-2}{n}\right) T_{2,n}\left(\frac{t-2}{n}\right) \psi'(\epsilon_t),
\end{aligned}
$$

$$
\begin{aligned}
n^{-1}a_n^{-2}\sin^2\theta\sum_{t=3}^{n}X_{t-1}X_{t-2}\psi'(\epsilon_t) &= \frac{1}{2}n^{-1}\sum_{t=3}^{n}T_{1,n}\left(\frac{t-2}{n}\right)\sin(t-1)\epsilon_{t-1}\psi'(\epsilon_t)\\
&- \frac{1}{2}n^{-1}\sum_{t=3}^{n}T_{2,n}\left(\frac{t-2}{n}\right)\cos(t-1)\epsilon_{t-1}\psi'(\epsilon_t)\\
&+ n^{-1}\sum_{t=3}^{n}T_{1,n}^2\left(\frac{t-2}{n}\right)\psi'(\epsilon_t)\\
&+ n^{-1}\sum_{t=3}^{n}T_{2,n}^2\left(\frac{t-2}{n}\right)\psi'(\epsilon_t)\\
&- n^{-1}\sum_{t=3}^{n}\cos(2(t-1)\theta)\\
&\times\left(T_{1,n}^2\left(\frac{t-2}{n}\right)-T_{2,n}^2\left(\frac{t-2}{n}\right)\right)\psi'(\epsilon_t)\\
&- 2n^{-1}\sum_{t=3}^{n}\sin(2(t-1)\theta)\\
&\times T_{1,n}\left(\frac{t-2}{n}\right)T_{2,n}\left(\frac{t-2}{n}\right)\psi'(\epsilon_t).
\end{aligned}
$$

The proofs of the preceding equations are straightforward applications of trigonometric identities. Details are omitted. Furthermore, asymptotically each $\psi'(\epsilon_t)$ can be replaced by $\mathrm{E}\left(\psi'(\epsilon_t)\right)$ in (4.3.4). To appreciate why, again we consider the third term of A_n where

$$
\sum_{t=3}^{n}X_{t-1}^2\psi'(\epsilon_t)=\sum_{t=3}^{n}X_{t-1}^2\left[\psi'(\epsilon_t)-\mathrm{E}\left(\psi'(\epsilon_t)\right)+\mathrm{E}\left(\psi'(\epsilon_t)\right)\right].
$$

Therefore,

$$
\begin{aligned}
n^{-1}a_n^{-2}\sin^2\theta\sum_{t=3}^{n}X_{t-1}^2\psi'(\epsilon_t) &= n^{-1}a_n^{-2}\sin^2\theta\sum_{t=3}^{n}X_{t-1}^2\left[\psi'(\epsilon_t)-\mathrm{E}\left(\psi'(\epsilon_t)\right)\right]\\
&+ n^{-1}a_n^{-2}\mathrm{E}\left(\psi'(\epsilon_1)\right)\sin^2\theta\sum_{t=3}^{n}X_{t-1}^2. \qquad (4.3.5)
\end{aligned}
$$

Note that the first term of the right hand side in (4.3.5) approaches to zero as $n\to\infty$. To see that, by (4.3.3) we have

$$
2a_n^{-2}\sin^2\theta\sum_{t=3}^{n}X_{t-1}^2\left[\psi'(\epsilon_t)-\mathrm{E}\left(\psi'(\epsilon_t)\right)\right]=\sum_{t=3}^{n}T_{1,n}^2\left(\frac{t-1}{n}\right)\left[\psi'(\epsilon_t)-\mathrm{E}\left(\psi'(\epsilon_t)\right)\right]
$$

$$
\begin{aligned}
&+\sum_{t=3}^{n} T_{2,n}^{2}\left(\frac{t-1}{n}\right)\left[\psi'(\epsilon_t)-\mathrm{E}\left(\psi'(\epsilon_t)\right)\right] \\
&-\sum_{t=3}^{n} \cos(2t\theta) T_{1,n}^{2}\left(\frac{t-1}{n}\right)\left[\psi'(\epsilon_t)-\mathrm{E}\left(\psi'(\epsilon_t)\right)\right] \\
&+\sum_{t=3}^{n} \cos(2t\theta) T_{2,n}^{2}\left(\frac{t-1}{n}\right)\left[\psi'(\epsilon_t)-\mathrm{E}\left(\psi'(\epsilon_t)\right)\right] \\
&-2\sum_{t=3}^{n} \sin(2t\theta) T_{1,n}\left(\frac{t-1}{n}\right) T_{2,n}\left(\frac{t-1}{n}\right) \\
&\times\left[\psi'(\epsilon_t)-\mathrm{E}\left(\psi'(\epsilon_t)\right)\right]. \qquad (4.3.6)
\end{aligned}
$$

Notice that (4.2.4) implies that the first two terms in (4.3.6) are $O_p(n^{1/2})$. To seek the rate of convergence for the other terms, from Lemma 4.3.3 we can conclude that

$$
\sup_{0<j\le n}\left|\sum_{k=1}^{j} e^{ik\theta} T_{i,n}^{2}\left(\frac{k}{n}\right)\left[\psi'(\epsilon_t)-\mathrm{E}\left(\psi'(\epsilon_t)\right)\right]\right| = o_p(n),
$$

for $i = 1, 2$. Consequently, we have

$$
n^{-1}a_n^{-2}\sin^2\theta \sum_{t=3}^{n} X_{t-1}^{2}\left[\psi'(\epsilon_t)-\mathrm{E}\left(\psi'(\epsilon_t)\right)\right] \xrightarrow{p} 0.
$$

By the same justification, we can show that the preceding result holds for the last two terms of (4.3.4).

To find the limiting distribution of A_n, notice that (4.2.7), (4.3.2), (4.3.3), and Lemma 4.3.2 along with Proposition 4.3.1, guarantee that

$$
2a_n^{-1}\sin\theta \sum_{t=3}^{n} X_{t-1}\epsilon_t \xrightarrow{d} \int_0^1 T_1(t)dT_2(t) - \int_0^1 T_2(t)dT_1(t),
$$

$$
\begin{aligned}
n^{-1/2}a_n^{-1}\sin\theta \sum_{t=3}^{n} X_{t-1}\psi(\epsilon_t) &\xrightarrow{d} \mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right) \\
&\left[\cos\theta\left(\int_0^1 T_1(t)dR_1(t) - \int_0^1 T_2(t)dR_2(t)\right)\right. \\
&\left.+\sin\theta\left(\int_0^1 T_1(t)dR_2(t) + \int_0^1 T_2(t)dR_1(t)\right)\right],
\end{aligned}
$$

$$n^{-1/2}a_n^{-1}\sin\theta\sum_{t=3}^{n}X_{t-2}\psi(\epsilon_t)\xrightarrow{d}\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\left(\int_0^1 T_1(t)dR_1(t)-\int_0^1 T_2(t)dR_2(t)\right),$$

$$n^{-1}a_n^{-2}\sin^2\theta\sum_{t=3}^{n}X_{t-1}^2\psi'(\epsilon_t)\xrightarrow{d}\frac{\mathrm{E}\left(\psi'(\epsilon_1)\right)}{2}\left(\int_0^1 T_1^2(t)d(t)+\int_0^1 T_2^2(t)d(t)\right),$$

$$n^{-1}a_n^{-2}\sin^2\theta\sum_{t=3}^{n}X_{t-2}^2\psi'(\epsilon_t)\xrightarrow{d}\frac{\mathrm{E}\left(\psi'(\epsilon_1)\right)}{2}\left(\int_0^1 T_1^2(t)d(t)+\int_0^1 T_2^2(t)d(t)\right),$$

$$n^{-1}a_n^{-2}\sin^2\theta\sum_{t=3}^{n}X_{t-1}X_{t-2}\psi'(\epsilon_t)\xrightarrow{d}\frac{\mathrm{E}\left(\psi'(\epsilon_1)\right)\cos\theta}{2}\left(\int_0^1 T_1^2(t)d(t)+\int_0^1 T_2^2(t)d(t)\right),$$

where $\mathbf{T}_i$ and $\mathbf{R}_i$, $i=1,2$, are defined in (4.2.7). Thus, the finite-dimensional distributions of A_n converge weakly to those of A where

$$\begin{aligned}
A(u,v) = &-u\frac{\mathrm{E}^{1/2}(\psi^2(\epsilon_1))}{\sin\theta}\left[\cos\theta\left(\int_0^1 T_1(t)dR_1(t)-\int_0^1 T_2(t)dR_2(t)\right)\right.\\
&\left.+\sin\theta\left(\int_0^1 T_1(t)dR_2(t)+\int_0^1 T_2(t)dR_1(t)\right)\right]\\
&-v\frac{\mathrm{E}^{1/2}(\psi^2(\epsilon_1))}{\sin\theta}\left(\int_0^1 T_1(t)dR_1(t)-\int_0^1 T_2(t)dR_2(t)\right)\\
&+u^2\frac{\mathrm{E}(\psi'(\epsilon_1))}{4\sin^2\theta}\left(\int_0^1 T_1^2(t)d(t)+\int_0^1 T_2^2(t)d(t)\right)\\
&+v^2\frac{\mathrm{E}(\psi'(\epsilon_1))}{4\sin^2\theta}\left(\int_0^1 T_1^2(t)d(t)+\int_0^1 T_2^2(t)d(t)\right)\\
&+uv\frac{\mathrm{E}(\psi'(\epsilon_1))\cos\theta}{2\sin^2\theta}\left(\int_0^1 T_1^2(t)d(t)+\int_0^1 T_2^2(t)d(t)\right).
\end{aligned}$$

Due to convexity of the processes A_n and A, Lemma 3.2.1 in Chapter 3 shows that the minimizer of A_n converges in distribution to the minimizer of A. Therefore, by setting the derivative of $A(u,v)$ to 0 and solving for u and v, we obtain the limiting distribution for the case (i).

□

Proof of Theorem 4.3.1 Case (ii). Consider the model $X_t = 2X_{t-1} - X_{t-2} + \epsilon_t$ is equivalent to

$$X_t = \sum_{k=1}^{t} \sum_{i=1}^{k} \epsilon_i.$$

We also have the following property when we have root 1 with the multiplicity of 2:

$$X_t - X_{t-1} = \sum_{k=1}^{t} \epsilon_k.$$

Using an argument similar to the proof for case (i), consider the following process

$$\begin{aligned}
&\sum_{t=3}^{n} \left[\rho\left(\epsilon_t - (\hat{\phi}_1 - 2)X_{t-1} - (\hat{\phi}_2 + 1)X_{t-2}\right) - \rho(\epsilon_t)\right] \\
&= \sum_{t=3}^{n} \Big[\rho\Big(\epsilon_t - (\hat{\phi}_1 - 2)X_{t-1} - (\hat{\phi}_2 + 1)X_{t-2} \\
&\qquad + (\hat{\phi}_1 - 2)X_{t-2} - (\hat{\phi}_1 - 2)X_{t-2}\Big) - \rho(\epsilon_t)\Big].
\end{aligned}$$

Therefore, we define the process

$$A_n(u, v) = \sum_{t=3}^{n} \left[\rho\left(\epsilon_t - n^{-1/2}a_n^{-1}u\left(X_{t-1} - X_{t-2}\right) - n^{-3/2}a_n^{-1}vX_{t-2}\right) - \rho(\epsilon_t)\right],$$

where $(u, v) = \left(n^{1/2}a_n(\hat{\phi}_1 - 2), n^{3/2}a_n\left((\hat{\phi}_1 - 2) + (\hat{\phi}_2 + 1)\right)\right)^T$ is the minimizer of $A_n(u, v)$. Using the Taylor series expansion of each summand of A_n around $u = 0$ and $v = 0$, we get

$$\begin{aligned}
A_n(u, v) = &-un^{-1/2}a_n^{-1}\sum_{t=3}^{n}\left(X_{t-1} - X_{t-2}\right)\psi(\epsilon_t) \\
&- vn^{-3/2}a_n^{-1}\sum_{t=3}^{n} X_{t-2}\psi(\epsilon_t) \\
&+ \frac{1}{2}u^2 n^{-1}a_n^{-2}\sum_{t=3}^{n}\left(X_{t-1} - X_{t-2}\right)^2\psi'(c_t^n) \\
&+ \frac{1}{2}v^2 n^{-3}a_n^{-2}\sum_{t=3}^{n} X_{t-2}^2\psi'(c_t^n) \\
&+ uvn^{-2}a_n^{-2}\sum_{t=3}^{n} X_{t-2}\left(X_{t-1} - X_{t-2}\right)\psi'(c_t^n). \qquad (4.3.7)
\end{aligned}$$

Asymptotically, we can substitute $\psi'(c_t^n)$ by $\psi'(\epsilon_t)$ and each $\psi'(\epsilon_t)$ can be replaced by $E\left(\psi'(\epsilon_t)\right)$ in the sums in (4.3.7). We do not provide the proof here since it is similar to case (i). Thus, the finite-dimensional distributions of A_n converge weakly to those of A where

$$\begin{aligned}
A(u,v) = &-u\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\int_0^1 S(t)dW(t)\\
&- v\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\int_0^1\int_0^t S(s)dsdW(t)\\
&+ \frac{u^2}{2}\mathrm{E}\left(\psi'(\epsilon_1)\right)\int_0^1 S^2(t)dt\\
&+ \frac{v^2}{2}\mathrm{E}\left(\psi'(\epsilon_1)\right)\int_0^1\left(\int_0^t S(s)ds\right)^2 dt\\
&+ uv\mathrm{E}\left(\psi'(\epsilon_1)\right)\int_0^1 S(t)\int_0^t S(s)dsdt.
\end{aligned}$$

The result of Theorem 4.3.1 Case (ii) follows by setting the derivative of $A(u,v)$ to 0 and solving for u and v. □

Proof of Theorem 4.3.1 Case (iii). For the model $X_t = -2X_{t-1} - X_{t-2} + \epsilon_t$, we can write

$$X_t = \sum_{k=1}^{t}\sum_{i=1}^{k}(-1)^{t-i}\epsilon_i,$$

and

$$(-1)^t X_t = \sum_{k=1}^{t}\sum_{i=1}^{k}(-1)^{i}\epsilon_i. \qquad (4.3.8)$$

The formula (4.3.8) shows that the results in this case are similar to those of case (ii) except for replacing ϵ_i by $(-1)\epsilon_i$, for $i = 1, 2, \ldots, t$. Notice that (4.3.8) implies

$$\begin{aligned}
(-1)^t(X_t + X_{t-1}) &= \sum_{i=1}^{t}(-1)^i\epsilon_i,\\
\sum_{t=3}^{n} X_{t-1}^2 &= \sum_{t=3}^{n}(-1)^{t-1}X_{t-1}(-1)^{t-1}X_{t-1},\\
\sum_{t=3}^{n} X_{t-1}X_{t-2} &= -\sum_{t=3}^{n}(-1)^{t-1}X_{t-1}(-1)^{t-2}X_{t-2}.
\end{aligned}$$

Consider the following process

$$\begin{aligned}
&\sum_{t=3}^{n}\left[\rho\left(\epsilon_t-(\hat{\phi}_1+2)X_{t-1}-(\hat{\phi}_2+1)X_{t-2}\right)-\rho(\epsilon_t)\right] \\
&\quad=\sum_{t=3}^{n}\left[\rho\left(\epsilon_t-(\hat{\phi}_1+2)X_{t-1}-(\hat{\phi}_2+1)X_{t-2}\right.\right. \\
&\qquad\left.\left.+(\hat{\phi}_2+1)X_{t-1}-(\hat{\phi}_2+1)X_{t-1}\right)-\rho(\epsilon_t)\right].
\end{aligned}$$

Then, we define the process

$$A_n(u,v)=\sum_{t=3}^{n}\left[\rho\left(\epsilon_t-n^{-3/2}a_n^{-1}uX_{t-1}-n^{-1/2}a_n^{-1}v\left(X_{t-1}+X_{t-2}\right)\right)-\rho(\epsilon_t)\right],$$

where $(u,v)=\left(n^{3/2}a_n\left((\hat{\phi}_1+2)-(\hat{\phi}_2+1)\right),n^{1/2}a_n(\hat{\phi}_2+1)\right)^T$ is the minimizer of $A_n(u,v)$. Using the Taylor series expansion of each summand of A_n around $u=0$ and $v=0$, we get

$$\begin{aligned}
A_n(u,v)&=-un^{-3/2}a_n^{-1}\sum_{t=3}^{n}X_{t-1}\psi(\epsilon_t) \\
&\quad-vn^{-1/2}a_n^{-1}\sum_{t=3}^{n}\left(X_{t-1}+X_{t-2}\right)\psi(\epsilon_t) \\
&\quad+\frac{1}{2}u^2n^{-3}a_n^{-2}\sum_{t=3}^{n}X_{t-1}^2\psi'(c_t^n) \\
&\quad+\frac{1}{2}v^2n^{-1}a_n^{-2}\sum_{t=3}^{n}\left(X_{t-1}+X_{t-2}\right)^2\psi'(c_t^n) \\
&\quad+uvn^{-2}a_n^{-2}\sum_{t=3}^{n}X_{t-1}\left(X_{t-1}+X_{t-2}\right)\psi'(c_t^n).
\end{aligned}\tag{4.3.9}$$

Similar to case (i) and (ii), we can show that the finite-dimensional distributions of A_n in (4.3.9) converge weakly to those of A where

$$\begin{aligned}
A(u,v)&=-u\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\int_0^1\int_0^t S^{(1)}(t)dsdW(t) \\
&\quad-v\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\int_0^1 S^{(1)}(s)dW(t) \\
&\quad+\frac{u^2}{2}\mathrm{E}\left(\psi'(\epsilon_1)\right)\int_0^1\left[\int_0^t S^{(1)}(s)ds\right]^2dt \\
&\quad+\frac{v^2}{2}\mathrm{E}\left(\psi'(\epsilon_1)\right)\int_0^1\left(S^{(1)}(t)\right)^2dt
\end{aligned}$$

$$+ uv\mathrm{E}\left(\psi'(\epsilon_1)\right) \int_0^1 S^{(1)}(t) \int_0^t S^{(1)}(s) ds dt.$$

Therefore, the limiting distribution in case (iii) follows by setting the derivative of $A(u, v)$ to 0 and solving for u and v.

□

Proof of Theorem 4.3.1 Case (iv). Consider the model

$$X_t = X_{t-2} + \epsilon_t.$$

We can conclude the following property when we have roots 1 and -1:

$$X_t + X_{t-1} = \sum_{k=1}^{t} \epsilon_k,$$

$$X_t - X_{t-1} = \sum_{k=1}^{t} (-1)^{t-i} \epsilon_k.$$

Using an argument similar to the proof for case (i), we have

$$\begin{aligned}
&\sum_{t=3}^{n} \left[\rho\left(\epsilon_t - \hat{\phi}_1 X_{t-1} - (\hat{\phi}_2 - 1) X_{t-2}\right) - \rho(\epsilon_t)\right] \\
&= \sum_{t=3}^{n} \left[\rho\left(\epsilon_t - \frac{1}{2}(\hat{\phi}_1 + (\hat{\phi} - 1))(X_{t-1} + X_{t-2})\right.\right. \\
&\qquad \left.\left. - \frac{1}{2}(\hat{\phi}_1 - (\hat{\phi}_2 - 1))(X_{t-1} - X_{t-2})\right) - \rho(\epsilon_t)\right].
\end{aligned}$$

thus, we consider the following process

$$A_n(u, v) = \sum_{t=3}^{n} \left[\rho\left(\epsilon_t - \frac{1}{2} n^{-1/2} a_n^{-1} u (X_{t-1} + X_{t-2}) - \frac{1}{2} n^{-1/2} a_n^{-1} v (X_{t-1} - X_{t-2})\right) - \rho(\epsilon_t)\right],$$

where $(u, v) = \left(n^{1/2} a_n (\hat{\phi}_1 + (\hat{\phi} - 1)), n^{1/2} a_n \left(\hat{\phi}_1 - (\hat{\phi}_2 - 1)\right)\right)$ is the minimizer of $A_n(u, v)$. Using the Taylor series expansion of each summand of A_n around $u = 0$ and $v = 0$, we get

$$\begin{aligned}
A_n(u, v) = &-\frac{1}{2} u n^{-1/2} a_n^{-1} \sum_{t=3}^{n} \left(X_{t-1} + X_{t-2}\right) \psi(\epsilon_t) \\
&- \frac{1}{2} v n^{-1/2} a_n^{-1} \sum_{t=3}^{n} (X_{t-1} - X_{t-2}) \psi(\epsilon_t)
\end{aligned}$$

$$+\frac{1}{8}u^2n^{-1}a_n^{-2}\sum_{t=3}^{n}\left(X_{t-1}+X_{t-2}\right)^2\psi'(c_t^n)$$
$$+\frac{1}{8}v^2n^{-1}a_n^{-2}\sum_{t=3}^{n}(X_{t-1}-X_{t-2})^2\psi'(c_t^n)$$
$$+\frac{1}{4}uvn^{-2}a_n^{-2}\sum_{t=3}^{n}\left(X_{t-1}+X_{t-2}\right)\left(X_{t-1}-X_{t-2}\right)\psi'(c_t^n). \tag{4.3.10}$$

Asymptotically, we can substitute $\psi'(c_t^n)$ by $\psi'(\epsilon_t)$ and each $\psi'(\epsilon_t)$ can be replaced by $E\left(\psi'(\epsilon_t)\right)$ in the sums in (4.3.10). We do not provide the proof here since it is similar to case (i). Moreover, the limit of very last term in (4.3.10) tends to 0 as $n \to \infty$ since

$$n^{-2}a_n^{-2}\sum_{t=3}^{n}\left(X_{t-1}+X_{t-2}\right)\left(X_{t-1}-X_{t-2}\right)\psi'(\epsilon_t^n)$$
$$= n^{-2}a_n^{-2}\sum_{t=3}^{n}\left(X_{t-1}^2-X_{t-2}^2\right)\psi'(\epsilon_t^n) \to 0.$$

Thus, the finite-dimensional distributions of A_n converge weakly to those of A where

$$A(u,v) = -\frac{1}{2}u\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\int_0^1 S(t)dW(t)$$
$$+\frac{1}{2}v\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\int_0^1 S^{(1)}(t)dW(t)$$
$$+\frac{1}{8}u^2\mathrm{E}\left(\psi'(\epsilon_1)\right)\int_0^1 S^2(t)dt$$
$$+\frac{1}{8}v^2\mathrm{E}\left(\psi'(\epsilon_1)\right)\int_0^1 \left(S^{(1)}(t)\right)^2 dt.$$

The result of Theorem 4.3.1 Case (iv) follows by setting the derivative of $A(u,v)$ to 0 and solving for u and v. □

4.4 Simulation

To investigate the results given in Theorem 4.3.1, we carry out a simulation for model (4.2.1), when $\phi(B) = 1 - \phi_1 B - \phi_2 B^2$ has complex conjugate unit roots. In other word, we illustrate the asymptotic result of case (i) in Theorem 4.3.1 by simulating the following

AR(2) process

$$X_t = 2\cos\theta X_{t-1} - X_{t-2} + \epsilon_t, \tag{4.4.1}$$

where the innovations are i.i.d. symmetric α-stable random variables. For the sake of brevity, we only present the M-estimates for case (i) in Theorem 4.3.1 when $\theta = \pi/4$. The simulations for the other unstable cases are similar.

In our simulation study, we consider the time series $\{X_t\}_{t=0}^{n}$ in model (4.4.1), for $n = 10$, 20, 30, 40, and 50 with $\alpha = 0.5$, 1, 1.3, 1.7 and 2. The time series are generated 10,000 times for each choice of n and α. The M-estimates of ϕ_1 and ϕ_2 in the AR(2) model are calculated for each replicate. Here, we use the Huber loss function presented in (3.4.1), and $\rho^*(x) = \ln(b + x^2)$, for $b = 3$. Note that the value of c for the Huber loss function is chosen based on the results of Table 3.1. According to values of α, the applied truncation values for c are 0.5, 1, 1.5, 2, and 3. The sample median and the sample 90% inter-percentile range (IPR) (95th percentile - 5th percentile) for $|\hat{\phi}_1 - 2\cos\theta|$ are tabulated in Tables 4.1 and 4.2 for the Huber loss function and the loss function ρ^*, respectively. To compare the efficiency of the M-estimates to the LS estimates, the simulated results of the sample median and 90% IPR for $|\hat{\phi}_1 - 2\cos\theta|$ with the LS method are shown in Table 4.3. Furthermore, the sample median and 90% IPR for $|\hat{\phi}_2 + 1|$ are presented in Tables 4.4, 4.5, and 4.6 for the Huber loss function, ρ^*, and the LS method, respectively.

As seen in all tables, M-estimates in both choices for the Huber loss function and ρ^* are significantly closer to the actual values as n gets larger compared to the LS estimates, specially for small values of α. Moreover, the 90% IPR is significantly smaller than the LS estimates. The 90% IPR values of M-estimates show fairly small deviations, which confirms that M-estimates are more precise than LS estimates.

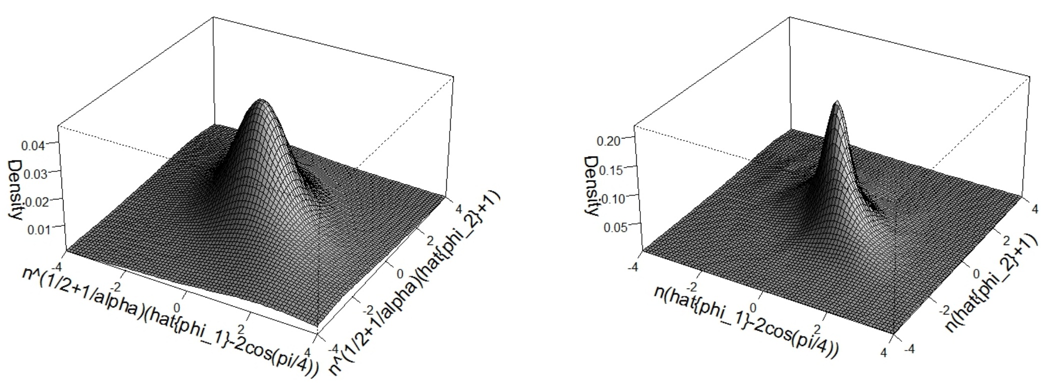
Table 4.1: Median and 90% IPR (in parentheses) for $|\hat{\phi}_1 - 2\cos\theta|$ in model (4.4.1) by the M-estimate method using the Huber loss function

	α				
n	0.5	1.0	1.3	1.7	2.0
10	0.0266	0.1114	0.1519	0.1833	0.1992
	(0.3876)	(0.3028)	(0.2623)	(0.2309)	(0.2150)
20	0.0030	0.0344	0.0548	0.0760	0.0836
	(0.0793)	(0.2219)	(0.2553)	(0.2729)	(0.2776)
30	0.0009	0.0180	0.0308	0.0454	0.0526
	(0.0238)	(0.1182)	(0.1515)	(0.1708)	(0.1879)
40	0.0004	0.0115	0.0217	0.0328	0.0391
	(0.0110)	(0.0729)	(0.1054)	(0.1277)	(0.1390)
50	0.0002	0.0080	0.0163	0.0263	0.0312
	(0.0059)	(0.0520)	(0.0769)	(0.0990)	(0.1138)

Furthermore, to illustrate the efficiency of M-estimates, we generate samples from model (4.4.1) for $\alpha = 1.3$, and 1.7 with sample size of $1,000$ and $\theta = \pi/4$. We estimate the parameters by the LS and M-estimate methods with $10,000$ simulations. Then we estimate joint densities of $(n^{1/2}a_n(\hat{\phi}_1 - 2\cos\theta), n^{1/2}a_n(\hat{\phi}_2 + 1))$ for both methods. Figures 4.1, 4.2, and 4.3 reflect the estimated limiting distribution for the M-estimate method using the Huber loss function, and the LS method when $\alpha = 1.3$, 1, and 1.7, respectively. These show that the density of the LS estimates gets thinner with bigger range as α decreases whereas the density of the M-estimates has fatter tails. The figures confirm our theoretical and simulated results.

Table 4.2: Median and 90% IPR (in parentheses) for $|\hat{\phi}_1 - 2\cos\theta|$ in model (4.4.1) by the M-estimate method using the loss function ρ^*

	α				
n	0.5	1.0	1.3	1.7	2.0
10	0.0194	0.0853	0.1303	0.1824	0.2104
	(1.0928)	(0.3289)	(0.2839)	(0.2318)	(0.2039)
20	0.0016	0.0224	0.0426	0.0726	0.0894
	(0.0788)	(0.1483)	(0.2102)	(0.2639)	(0.2852)
30	0.0004	0.0111	0.0232	0.0438	0.0585
	(0.0242)	(0.0735)	(0.1138)	(0.1573)	(0.1982)
40	0.0002	0.0068	0.0163	0.0312	0.0439
	(0.0239)	(0.0425)	(0.0760)	(0.1158)	(0.1431)
50	0.0001	0.0047	0.0121	0.0248	0.0337
	(0.0220)	(0.0296)	(0.0530)	(0.0904)	(0.1199)



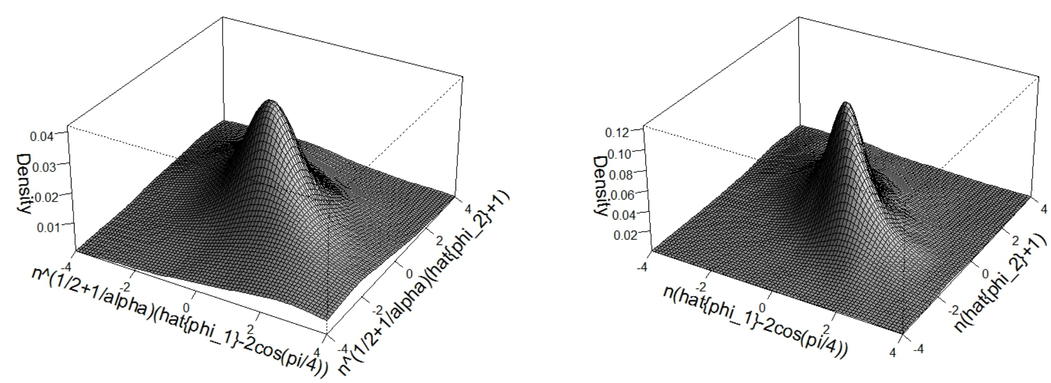
(a) M-estimate method

(b) LS method

Figure 4.1: Joint density estimation for the estimated parameters in model (4.4.1) when $\alpha = 1$

Table 4.3: Median and 90% IPR (in parentheses) for $|\hat{\phi}_1 - 2\cos\theta|$ in model (4.4.1) by the LS estimate method

	α				
n	0.5	1.0	1.3	1.7	2.0
10	0.0806	0.1351	0.1615	0.1858	0.1992
	(1.6076)	(0.7979)	(0.6529)	(0.6077)	(0.5757)
20	0.0313	0.0552	0.0667	0.0789	0.0836
	(0.3564)	(0.2919)	(0.2823)	(0.2808)	(0.2776)
30	0.0185	0.0349	0.0405	0.0482	0.0526
	(0.1748)	(0.1742)	(0.1747)	(0.1788)	(0.1879)
40	0.0129	0.0253	0.0311	0.0357	0.0391
	(0.1205)	(0.1288)	(0.1344)	(0.1359)	(0.1390)
50	0.0106	0.0202	0.0246	0.0290	0.0312
	(0.0964)	(0.1056)	(0.1044)	(0.1078)	(0.1138)



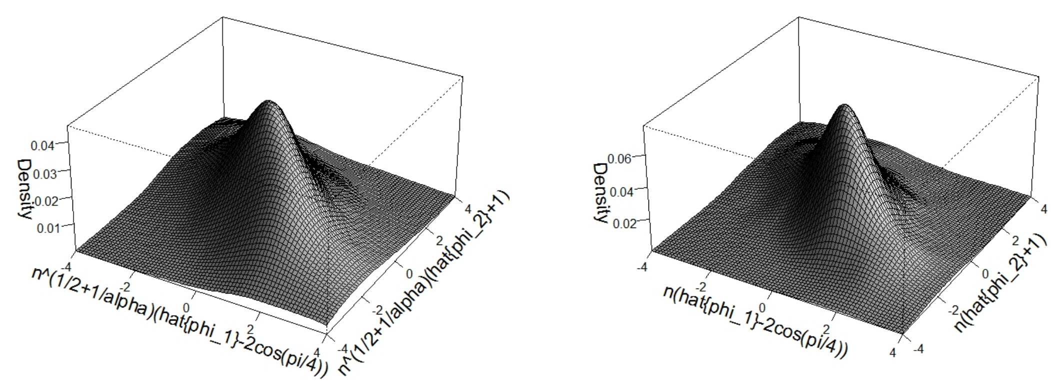
(a) M-estimate method

(b) LS method

Figure 4.2: Joint density estimation for the estimated parameters in model (4.4.1) when $\alpha = 1.3$

Table 4.4: Median and 90% IPR (in parentheses) for $|\hat{\phi}_2 + 1|$ in model (4.4.1) by the M-estimate method using the Huber loss function

	α				
n	0.5	1.0	1.3	1.7	2.0
10	0.0381	0.1341	0.1714	0.2031	0.2119
	(1.3185)	(0.9194)	(0.7634)	(0.6246)	(0.5596)
20	0.0036	0.0386	0.0580	0.0803	0.0869
	(0.1035)	(0.2515)	(0.2794)	(0.2946)	(0.2965)
30	0.0011	0.0193	0.0323	0.0468	0.0554
	(0.0280)	(0.1353)	(0.1774)	(0.1963)	(0.2101)
40	0.0004	0.0121	0.0226	0.0346	0.0401
	(0.0119)	(0.0774)	(0.1205)	(0.1483)	(0.1655)
50	0.0002	0.0083	0.0172	0.0266	0.0323
	(0.0065)	(0.0592)	(0.0843)	(0.1201)	(0.1365)



(a) M-estimate method

(b) LS method

Figure 4.3: Joint density estimation for the estimated parameters in model (4.4.1) when $\alpha = 1.7$

Table 4.5: Median and 90% IPR (in parentheses) for $|\hat{\phi}_2+1|$ in model (4.4.1) by the M-estimate method using the loss function ρ^*

	α				
n	0.5	1.0	1.3	1.7	2.0
10	0.0273	0.1061	0.1474	0.1997	0.2257
	(1.5441)	(0.8069)	(0.6843)	(0.6228)	(0.5739)
20	0.0019	0.0246	0.0451	0.0777	0.0959
	(0.1053)	(0.1700)	(0.2297)	(0.2833)	(0.3114)
30	0.0005	0.0116	0.0246	0.0446	0.0603
	(0.0241)	(0.0778)	(0.1301)	(0.1807)	(0.2201)
40	0.0002	0.0070	0.0172	0.0330	0.0438
	(0.0214)	(0.0457)	(0.0802)	(0.1346)	(0.1760)
50	0.0001	0.0049	0.0124	0.0257	0.0352
	(0.0166)	(0.0315)	(0.0571)	(0.1048)	(0.1405)

Table 4.6: Median and 90% IPR (in parentheses) for $|\hat{\phi}_2+1|$ in model (4.4.1) by the LS estimate method

	α				
n	0.5	1.0	1.3	1.7	2.0
10	0.0978	0.1591	0.1852	0.2098	0.2172
	(0.3779)	(0.3435)	(0.3327)	(0.3193)	(0.3011)
20	0.0332	0.0597	0.0689	0.0835	0.0871
	(0.1014)	(0.1222)	(0.1224)	(0.1274)	(0.1267)
30	0.0197	0.0363	0.0426	0.0500	0.0553
	(0.0556)	(0.0713)	(0.0759)	(0.0783)	(0.0789)
40	0.0132	0.0262	0.0320	0.0374	0.0401
	(0.0374)	(0.0516)	(0.0559)	(0.0583)	(0.0611)
50	0.0104	0.0206	0.0251	0.0293	0.0323
	(0.0298)	(0.0402)	(0.0436)	(0.0475)	(0.0497)

4.5 The bootstrap procedure and a numerical example

The asymptotic distributions for the proposed estimator of $\Phi = (\phi_1, \ldots, \phi_p)^T$ by Chan and Zhang [12] and this chapter are not computationally tractable. Due to the complexity of the limiting distributions, to make inferences based on $\hat{\phi}_1, \ldots, \hat{\phi}_p$, one may consider a resampling scheme. In this section, we briefly suggest using the result of Theorem 1 of Moreno and Romo [37], Theorem 3.1 of Zarepour and Knight [56], and Feigin and Resnick [18] in model (4.2.1), when $\{\epsilon_t\}$ is a sequence of i.i.d. random variables and belongs to the domain of attraction of a stable law with index α.

When $\alpha = 2$ and $\{\epsilon_1^*, \epsilon_2^*, \cdots, \epsilon_n^*\}$ is an i.i.d. sample from

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(\epsilon_i - \bar{\epsilon} \leq x),$$

then

$$\frac{1}{\sqrt{\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})^2}}\sum_{i=1}^{[nt]} \epsilon_i^* \xrightarrow{p} W(t),$$

in probability, where W is a Brownian motion. When $0 < \alpha < 2$ and $\{\epsilon_1^*, \epsilon_2^*, \cdots, \epsilon_n^*\}$ is an i.i.d. sample from

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(\epsilon_i \leq x),$$

then

$$a_n^{-1}\sum_{i=1}^{[nt]} \epsilon_i^* \xrightarrow{d} \sum_{i=1}^{\infty} M_i^*(t)\delta_i\Gamma_i^{-1/\alpha}, \tag{4.5.1}$$

in distribution. Here, $\{M_i^*(t)\}$ is a sequence of independent Poisson processes with Lebesgue mean measure and $\{\delta_k, \Gamma_k\}$ are the same as before. Note that in this case, the regular bootstrap asymptotically fails since the limit in (4.5.1) contains the Poisson point process $M_i^*(t)$, which is different from what we expect as in (4.1.6). Several studies have shown that subsampling of size m such that $m/n \to 0$ as $n \to \infty$ resolves the asymptotic failure of the regular bootstrap. For example, see [5], [18], and [57]. In summary, if $\{\epsilon_1^*, \epsilon_2^*, \cdots, \epsilon_m^*\}$ is an i.i.d. sample from $F_n(x) = \frac{1}{n}\sum_{i=1}^n I(\epsilon_i \leq x)$ with the sample size $m_n = m = o(n) \to \infty$, then

$$a_m^{-1}\sum_{i=1}^{[mt]} \epsilon_i^* \xrightarrow{d} S(\cdot),$$

in probability, where $S(\cdot)$ is the stable process defined in (4.1.6). To prove that the results of the limiting distribution are asymptotically valid when we use the bootstrap scheme with $m_n = m = o(n)$ resampling sample size, for example when the model has complex conjugate unit roots, it is necessary to show that

$$m^{1/2}a_m \begin{pmatrix} \hat{\phi}_1^* - \hat{\phi}_1 \\ \\ \hat{\phi}_2^* - \hat{\phi}_2 \end{pmatrix} \xrightarrow{d} \Gamma_1^{-1} \begin{pmatrix} \frac{2\mathrm{E}^{1/2}\left(\psi^2(\varepsilon_1)\right)\digamma_1 \sin\theta}{\mathrm{E}(\psi'(\varepsilon_1))\left(\int_0^1 T_1^2(t)d(t)+\int_0^1 T_2^2(t)d(t)\right)} \\ \\ \frac{2\mathrm{E}^{1/2}\left(\psi^2(\varepsilon_1)\right)\digamma_2 \sin\theta}{\mathrm{E}(\psi'(\varepsilon_1))\left(\int_0^1 T_1^2(t)d(t)+\int_0^1 T_2^2(t)d(t)\right)} \end{pmatrix}$$

in probability, where Γ_1, $\digamma_1$ and $\digamma_2$ are defined in Theorem 4.3.1. This convergence can be derived with a complex and tedious application of the continuous mapping theorem from the result of Zarepour and Knight [57]; for more details see also [37] and [56].

An illustration of the validity of the bootstrap scheme is given by a simulation study in the following example.

Example (*Bootstrap simulation study*) We consider bootstrapping for the parameters ϕ_1 and ϕ_2 in model (4.4.1) when $\theta = \pi/4$. We generate the time series $\{X_t\}_{t=0}^n$ in model (4.4.1), for the actual sample sizes n = 50, 100, and 200 with α = 1.3, and 1.7 (the cases with $\alpha > 1$ are of practical interest). The behaviour of the other unstable cases are more or less similar. Then we implement the following algorithm for each choice of n and α.

(i) We find M-estimates of the parameters ϕ_1 and ϕ_2 in model (4.4.1) using the Huber loss function. Then take

$$\hat{\epsilon}_t = X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2}.$$

(ii) We draw a sample of size m from centered residuals denoted by $\hat{\epsilon}_1^*, \ldots, \hat{\epsilon}_m^*$, and find $\{X_t^*\}_{t=0}^m$ from (4.4.1). Then we estimate the parameters ϕ_1 and ϕ_2 using the bootstrap observations by the same minimization of the objective function used in step (i).

(iii) We repeat step (ii) 3,000 times to get $(\hat{\phi}_1^{1*}, \hat{\phi}_2^{1*})^T, \ldots, (\hat{\phi}_1^{3,000*}, \hat{\phi}_2^{3,000*})^T$. To find a naive 95% confidence interval for ϕ_1 and ϕ_2, we obtain the 2.5th and 97.5th percentiles of 3,000 bootstrap estimates as the lower and upper bound of our confidence interval.

In order to compute the coverage rate of the bootstrap confidence intervals, the original time series are generated 10,000 times for each choice of n and α. Then by applying (i)-(iii), the naive 95% bootstrap confidence intervals for ϕ_1 and ϕ_2 are calculated for each replicate. Moreover, to study how the selection of the resampling size would affect our estimation, we perform the second step with three different resampling sizes $m = n/\ln(\ln(n))$, $n^{0.9}$, and $n^{0.95}$. Finally, the resulting coverage percentages of the naive 95% bootstrap confidence intervals for the parameters ϕ_1 and ϕ_2 for different values of m, n, and α are presented in Tables 4.7 and 4.8. Those tables show that the coverage percentages of the naive 95% bootstrap confidence intervals for ϕ_1 and ϕ_2 are very close to 95%. This illustrates that the bootstrap scheme with $m_n = m = o(n)$ resampling sample size is approximately valid when we have a non-stationary time series with innovations in the domain of attraction of a stable law.

Table 4.7: Coverage for the naive 95% bootstrap confidence interval for ϕ_1 in model (4.4.1)

α	1.3			1.7		
n	50	100	200	50	100	200
$m = n/\ln(\ln(n))$	96.1%	96.5%	97.4%	95.1%	96.5%	96.8%
$m = n^{(0.9)}$	97.2%	97.8%	97.7%	96.6%	96.9%	97.4%
$m = n^{(0.95)}$	95.7%	96.6%	96.3%	94.0%	94.8%	94.8%

Table 4.8: Coverage for the naive 95% bootstrap confidence interval for ϕ_2 in model (4.4.1)

α	1.3			1.7		
n	50	100	200	50	100	200
$m = n/\ln(\ln(n))$	94.4%	95.0%	96.4%	94.0%	94.1%	94.3%
$m = n^{(0.9)}$	95.3%	95.7%	96.5%	93.8%	93.1%	94.6%
$m = n^{(0.95)}$	93.8%	94.8%	94.0%	93.0%	93.6%	94.0%

Chapter 5

Asymptotic Theory for M-Estimates in Unstable AR(p) Processes with Infinite Variance Innovations

In this chapter, we derive the asymptotic distribution of M-estimators for the parameters in an unstable AR(p) process, where the innovations are in the domain of attraction of a stable law with index $0 < \alpha \leq 2$. Throughout this chapter, we assume that all definitions and conditions A1-A4, which were imposed in Chapters 3 and 4, hold.

5.1 The limiting distribution for the parameters of unstable AR(p) processes

The limiting distribution for the M-estimates of the parameters for an infinite-variance AR(2) processes is obtained in Chapter 4. To generalize, we extend those results to the AR(p) process when the characteristic root may have different multiplicities and lie on the unit circle. To derive the asymptotic behaviour of the M-estimates, consider the AR(p) model in (4.1.1) when the errors satisfy assumption A4. Define the process

$$A_n(q_1, \dots, q_p) = \sum_{t=p+1}^{n} \left[\rho(\epsilon_t - q_1 X_{t-1} - \cdots - q_p X_{t-p}) - \rho(\epsilon_t)\right]. \tag{5.1.1}$$

The M-estimate, $\hat{\phi} = (\hat{\phi}_1, \dots, \hat{\phi}_p)^T$, of $\Phi = (\phi_1, \dots, \phi_p)^T$ minimizes the objective function (5.1.1) with respect to $(q_1, \dots, q_p)^T = \mathcal{N}_n(\hat{\phi}_1 - \phi_1, \dots, \hat{\phi}_p - \phi_p)^T$. Here, $\mathcal{N}_n$ is a $p \times p$ diagonal matrix whose diagonal entries are appropriate normalizing constants. Thus, it is reasonable to expect that the minimizer of the process Z_n can be written as

$$\mathcal{N}_n \begin{pmatrix} \hat{\phi}_1 - \phi_1 \\ \hat{\phi}_2 - \phi_2 \\ \vdots \\ \hat{\phi}_p - \phi_p \end{pmatrix} = \mathcal{D}_n^{-1} \mathcal{B}_n. \tag{5.1.2}$$

Here,

$$\mathcal{B}_n = \mathcal{N}_n^{-1} \begin{pmatrix} \sum_{t=p+1}^n X_{t-1}\psi(\epsilon_t) \\ \sum_{t=p+1}^n X_{t-2}\psi(\epsilon_t) \\ \vdots \\ \sum_{t=p+1}^n X_{t-p}\psi(\epsilon_t) \end{pmatrix}$$

and $\mathcal{D}_n = (d_{i,j}^{(n)})$ is a $p \times p$ matrix such that

$$\mathcal{D}_n = \mathcal{N}_n \begin{pmatrix} \sum_{t=p+1}^n X_{t-1}^2 \psi'(c_t^{(1,1)}) & \sum_{t=p+1}^n X_{t-1}X_{t-2}\psi'(c_t^{(n)}) & \cdots & \sum_{t=p+1}^n X_{t-1}X_{t-p}\psi'(c_t^{(n)}) \\ \sum_{t=p+1}^n X_{t-1}X_{t-2}\psi'(c_t^{(n)}) & \sum_{t=p+1}^n X_{t-2}^2 \psi'(c_t^{(n)}) & \cdots & \sum_{t=p+1}^n X_{t-2}X_{t-p}\psi'(c_t^{(n)}) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=p+1}^n X_{t-1}X_{t-p}\psi'(c_t^{(n)}) & \sum_{t=p+1}^n X_{t-2}X_{t-p}\psi'(c_t^{(n)}) & \cdots & \sum_{t=p+1}^n X_{t-p}^2 \psi'(c_t^{(n)}) \end{pmatrix} \mathcal{N}_n^T. \tag{5.1.3}$$

Similar to AR(2) processes in Chapter 4, with the appropriate norming constants, $\psi'(c_t^{(n)})$ asymptotically can be replaced by $\psi'(\epsilon_t)$ in (5.1.3). Furthermore, in Theorem 5.1.1 we will prove that asymptotically each $\psi'(\epsilon_t)$ can be replaced by $E\left(\psi'(\epsilon_t)\right)$. Therefore, as $n \to \infty$, the matrix $\mathcal{D}_n$ will be nonsingular with probability 1; for more details see [11]. However, similar results were already proven for AR(2) processes in Chapter 4.

The limiting distribution of the M-estimates of the parameters of an AR(p) process cannot be obtained directly from (5.1.2). To do so, we consider the following steps. First, since time series with different characteristic roots are expected to behave differently, we can decompose the characteristic polynomial defined in (4.1.2) into the following polynomial:

$$\phi(z) = (1-z)^r(1+z)^s \prod_{k=1}^{l}(1 - 2\cos(\theta_k)z + z^2)^{d_k}\varphi(z),$$

where polynomial $\varphi(\cdot)$ corresponds to the q roots which are outside the unit circle and $q + r + s + 2\sum_{k=1}^{l} d_k = p$. Similar to Chan and Wei [11], we transform $\{X_t\}$ into various components based on the location of their roots. Then we find the limiting behaviour of each component individually. Davis et al. [14] consider the asymptotic behavior of the M-estimates for the parameters in a stationary AR(p) process. Here, we take $\varphi(z) = 1$. Define

$$\begin{aligned} u_t &= \phi(B)(1-B)^{-r}X_t, \\ v_t &= \phi(B)(1+B)^{-s}X_t, \\ w_t(k) &= \phi(B)(1 - 2\cos(\theta_k)B + B^2)^{-d_k}X_t, \end{aligned} \tag{5.1.4}$$

for $k = 1, 2, \ldots, l$. Equivalently,

$$\epsilon_t = (1-B)^r u_t = (1+B)^s v_t = (1 - 2\cos(\theta_k)B + B^2)^{d_k} w_t(k).$$

From Chan and Wei [11], there exists a nonsingular $p \times p$ matrix Q such that

$$Q\mathbf{X_t} = \left(\mathbf{u}_t^T, \mathbf{v}_t^T, \mathbf{w}_t^T(1), \ldots, \mathbf{w}_t^T(l)\right)^T,$$

where

$$\begin{aligned} \mathbf{X_t} &= (X_t, \ldots, X_{t-p+1})^T, \\ \mathbf{u_t} &= (u_t, \ldots, u_{t-r+1})^T, \\ \mathbf{v_t} &= (v_t, \ldots, v_{t-s+1})^T, \\ \mathbf{w_t}(k) &= (w_t(k), \ldots, w_{t-2d_k+1}(k))^T, \end{aligned}$$

for $k = 1, 2, \ldots, l$. Moreover, let

$$G_n = \operatorname{diag}\left(J_n, K_n, L_n(1), \ldots, L_n(l)\right)$$

be a normalization matrix, where J_n and K_n are as specified in (5.1.7) and (5.1.14) of Theorem 5.1.1. Also, $L_n(i)$ for $1 \le i \le l \le p$ are defined similar to L_n in (5.1.22) in Theorem 5.1.1. Then, we have

$$G_n Q \mathbf{X_t} \overset{p}{\sim} \operatorname{diag}\left(J_n \mathbf{u_t}, K_n \mathbf{v_t}, L_n(1)\mathbf{w_t}(1), \ldots, L_n(l)\mathbf{w_t}(l)\right)$$

and

$$\left(Q^T G_n^T\right)^{-1}(\hat{\Phi} - \Phi) \overset{p}{\sim} \begin{pmatrix} (J_n^T)^{-1}\left(\sum_{t=r+1}^n \mathbf{u_{t-1}}\mathbf{u_{t-1}}^T\psi'(\epsilon_t)\right)^{-1}\sum_{t=r+1}^n \mathbf{u_{t-1}}\psi(\epsilon_t) \\ (K_n^T)^{-1}\left(\sum_{t=s+1}^n \mathbf{v_{t-1}}\mathbf{v_{t-1}}^T\psi'(\epsilon_t)\right)^{-1}\sum_{t=s+1}^n \mathbf{v_{t-1}}\psi(\epsilon_t) \\ (L_n(1)^T)^{-1}\left(\sum_{t=2d_1+1}^n \mathbf{w_{t-1}}(1)\mathbf{w_{t-1}}^T(1)\psi'(\epsilon_t)\right)^{-1}\sum_{t=2d_1+1}^n \mathbf{w_{t-1}}(1)\psi(\epsilon_t) \\ \vdots \\ (L_n(l)^T)^{-1}\left(\sum_{t=2d_l+1}^n \mathbf{w_{t-1}}(l)\mathbf{w_{t-1}}^T(l)\psi'(\epsilon_t)\right)^{-1}\sum_{t=2d_l+1}^n \mathbf{w_{t-1}}(l)\psi(\epsilon_t) \end{pmatrix}, \tag{5.1.5}$$

where $a_n \overset{p}{\sim} b_n$ means $a_n = b_n + o_p(1)$. From (5.1.5), we can find the weak limit behaviour of the M-estimates for the AR(p) processes defined in (4.2.1). The result is presented in Theorem 5.1.1.

Theorem 5.1.1 *Suppose $\{X_t\}$ satisfies (4.2.1) and conditions A1-A4 in Chapters 3 and 4 hold. Then*

$$\left(Q^T G_n^T)\right)^{-1}(\hat{\Phi} - \Phi) \xrightarrow{d} \left((\Gamma^{-1}\mathcal{F})^T, (\Upsilon^{-1}\mathcal{H})^T, (\Lambda_1^{-1}\mathcal{G}_1)^T, \ldots, (\Lambda_l^{-1}\mathcal{G}_l)^T\right)^T,$$

where $(\Gamma^{-1}\mathcal{F})$ and $(\Upsilon^{-1}\mathcal{H})$, respectively, are defined in (5.1.11) and (5.1.17). Also, $(\Lambda_i^{-1}\mathcal{G}_i)$ for $i = 1, \ldots, l$ are given similar to $\Lambda^{-1}\mathcal{G}$ in (5.1.25). Note that these limiting distributions are functionals of multiple stochastic integrals of stable processes.

Proof of Theorem 5.1.1 with roots equal to 1. First, we consider the unit root model

$$(1 - B)^r u_t = \epsilon_t.$$

To avoid singular limiting distributions, similar to Chan and Wei [11], we define

$$u_t(j) = (1 - B)^{r-j} u_t \quad \text{for } j = 1, 2, \ldots, r,$$

or equivalently

$$u_t(1) = \sum_{j=1}^{t} \epsilon_j \qquad \text{and} \qquad u_t(j) = \sum_{k=1}^{t} u_k(j-1)$$

for $j = 2, \ldots, r$. From Definition 4.2.3, we have

$$a_n^{-1} u_{[nt]}(1) = S_n(t) =: \mathcal{S}_{1,n}(t) \stackrel{d}{\rightarrow} S(t) =: \mathcal{S}_1(t).$$

Since $u_t(j) = \sum_{k=1}^{t} u_k(j-1)$ for $j = 2, \ldots, r$, the continuous mapping theorem implies that

$$a_n^{-1} n^{-(j-1)} u_{[nt]}(j) = \int_0^t \mathcal{S}_{j-1,n}(s) ds =: \mathcal{S}_{j,n}(t) \stackrel{d}{\rightarrow} \mathcal{S}_j(t) \tag{5.1.6}$$

for $j = 2, \ldots, r$. We also define

$$J_n = N_n^{-1} C, \tag{5.1.7}$$

where

$$C = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (-1)\binom{r-1}{1} & (-1)^2\binom{r-1}{2} & \cdots & (-1)^{r-1} \end{pmatrix}$$

and the appropriate normalizing constant is

$$N_n = \operatorname{diag}\left(n^{r-1/2} a_n, n^{(r-1)-1/2} a_n, \ldots, n^{1/2} a_n\right).$$

First, notice that

$$C \begin{pmatrix} u_t \\ u_{t-1} \\ \vdots \\ u_{t-r+1} \end{pmatrix} = \begin{pmatrix} u_t(r) \\ u_t(r-1) \\ \vdots \\ u_t(1) \end{pmatrix}.$$

Thus, we have

$$C \left(\sum_{t=r+1}^{n} \mathbf{u_{t-1}} \mathbf{u_{t-1}}^T \psi'(\epsilon_t) \right) C^T = \Pi,$$

where

$$\Pi = \begin{pmatrix} \sum_{t=r+1}^{n} u_{t-1}^{2}(r)\psi'(\epsilon_t) & \sum_{t=r+1}^{n} u_{t-1}(r)u_{t-1}(r-1)\psi'(\epsilon_t) & \cdots & \sum_{t=r+1}^{n} u_{t-1}(r)u_{t-1}(1)\psi'(\epsilon_t) \\ \sum_{t=r+1}^{n} u_{t-1}(r-1)u_{t-1}(r)\psi'(\epsilon_t) & \sum_{t=r+1}^{n} u_{t-1}^{2}(r-1)\psi'(\epsilon_t) & \cdots & \sum_{t=r+1}^{n} u_{t-1}(r-1)u_{t-1}(1)\psi'(\epsilon_t) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=r+1}^{n} u_{t-1}(1)u_{t-1}(r)\psi'(\epsilon_t) & \sum_{t=r+1}^{n} u_{t-1}(1)u_{t-1}(r-1)\psi'(\epsilon_t) & \cdots & \sum_{t=r+1}^{n} u_{t-1}^{2}(1)\psi'(\epsilon_t) \end{pmatrix}.$$

The joint behavior of Π can be studied through some tedious calculations. Therefore, for simplicity, we only calculate the limiting behavior of each term of Π individually by using the following steps. As discussed before, each $\psi'(\epsilon_t)$ can be replaced by $\mathrm{E}\left(\psi'(\epsilon_t)\right)$ when $n \to \infty$. To see this, note that

$$\begin{aligned} \sum_{t=r+1}^{n} u_{t-1}(i)u_{t-1}(j)\psi'(\epsilon_t) &= \sum_{t=r+1}^{n} u_{t-1}(i)u_{t-1}(j)\left[\psi'(\epsilon_t) - \mathrm{E}\left(\psi'(\epsilon_t)\right) + \mathrm{E}\left(\psi'(\epsilon_t)\right)\right] \\ &= \sum_{t=r+1}^{n} u_{t-1}(i)u_{t-1}(j)\left[\psi'(\epsilon_t) - \mathrm{E}\left(\psi'(\epsilon_t)\right)\right] \\ &\quad + \mathrm{E}\left(\psi'(\epsilon_1)\right)\sum_{t=r+1}^{n} u_{t-1}(i)u_{t-1}(j) \end{aligned} \tag{5.1.8}$$

for $i, j = 1, 2, \ldots, r$. Notice that by (4.2.4) and (5.1.6), the first term on the right side of (5.1.8) is $O_p(a_n^2 n^{(i+j-3/2)})$. Then we have

$$a_n^{-2} n^{-(i+j-1)} \sum_{t=r+1}^{n} u_{t-1}(i)u_{t-1}(j)[\psi'(\epsilon_t) - \mathrm{E}\left(\psi'(\epsilon_t)\right)] \ \xrightarrow{p}\ 0.$$

From (5.1.6), along with applying the continuous mapping theorem and Lemma 3.2.1, the limit of each term of Π is as follows:

$$a_n^{-2} n^{-(i+j-1)} \sum_{t=r+1}^{n} u_{t-1}(i)u_{t-1}(j)\psi'(\epsilon_t) \ \xrightarrow{d}\ \mathrm{E}\left(\psi'(\epsilon_1)\right)\int_0^1 \mathcal{S}_i(t)\mathcal{S}_j(t)dt$$

for $i, j = 1, 2, \ldots, r$. Therefore,

$$J_n\left(\sum_{t=r+1}^{n} \mathbf{u_{t-1}}\mathbf{u_{t-1}}^T\psi'(\epsilon_t)\right)J_n^T \ \xrightarrow{d}\ \Gamma,$$

where $\Gamma = (\gamma_{i,j})$ is a $r \times r$ random matrix such that

$$\Gamma = E\left(\psi'(\epsilon_1)\right)\begin{pmatrix} \int_0^1 \mathcal{S}_{r-1}^2(t)dt & \int_0^1 \mathcal{S}_{r-1}(t)\mathcal{S}_{r-2}(t)dt & \cdots & \int_0^1 \mathcal{S}_{r-1}(t)\mathcal{S}_1(t)dt \\ \int_0^1 \mathcal{S}_{r-2}(t)\mathcal{S}_{r-1}(t)dt & \int_0^1 \mathcal{S}_{r-2}^2(t)dt & \cdots & \int_0^1 \mathcal{S}_{r-2}(t)\mathcal{S}_1(t)dt \\ \vdots & \vdots & \ddots & \vdots \\ \int_0^1 \mathcal{S}_1(t)\mathcal{S}_{r-1}(t)dt & \int_0^1 \mathcal{S}_1(t)\mathcal{S}_{r-2}(t)dt & \cdots & \int_0^1 \mathcal{S}_1^2(t)dt \end{pmatrix}. \tag{5.1.9}$$

Moreover, note that

$$J_n \sum_{t=r+1}^{n} \mathbf{u_{t-1}} \psi(\epsilon_t) \overset{d}{\to} \mathcal{F},$$

where

$$\mathcal{F} = \mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right) \begin{pmatrix} \int_0^1 \mathcal{S}_r(t)dW(t) \\ \int_0^1 \mathcal{S}_{r-1}(t)dW(t) \\ \vdots \\ \int_0^1 \mathcal{S}_1(t)dW(t) \end{pmatrix}. \tag{5.1.10}$$

Now, summarize the results to obtain the following limiting distribution

$$\left(J_n^T\right)^{-1} \left(\sum_{t=r+1}^{n} \mathbf{u_{t-1}} \mathbf{u_{t-1}}^T \psi'(\epsilon_t)\right)^{-1} \sum_{t=r+1}^{n} \mathbf{u_{t-1}} \psi(\epsilon_t) \overset{d}{\to} \Gamma^{-1}\mathcal{F}, \tag{5.1.11}$$

where $\Gamma = (\gamma_{i,j})_{r\times r}$ and $\mathcal{F}$, respectively, are defined in (5.1.9) and (5.1.10).

□

Proof of Theorem 5.1.1 with roots equal to -1. Consider the following model

$$(1+B)^s v_t = \epsilon_t.$$

The limiting distribution in this case is similar to the case in which the time series has unit root 1, except that ϵ_i is replaced by $(-1)^i\epsilon_i$ for $i = 1, 2, \ldots, t$. Define

$$v_t(j) = (1+B)^{s-j} v_t \quad \text{for } j = 1, 2, \ldots, s. \tag{5.1.12}$$

Notice that (5.1.12) implies that $(-1)^t v_t(j+1) = \sum_{i=1}^{t} (-1)^i v_t(j)$. By (4.2.4) and (4.2.7), we have

$$a_n^{-1}(-1)^t v_{[nt]}(1) = S_n^{(1)}(t) =: \mathcal{S}_{1,n}^{(1)}(t) \overset{d}{\to} S^{(1)}(t) =: \mathcal{S}_1^{(1)}(t),$$

which implies that

$$a_n^{-1} n^{-(j-1)} (-1)^t v_t(j) = \int_0^t \mathcal{S}_{j-1,n}^{(1)}(s)ds =: \mathcal{S}_{j,n}^{(1)}(t) \overset{d}{\to} \mathcal{S}_j^{(1)}(t) \tag{5.1.13}$$

for $j = 2, \ldots, s$. Then, let

$$K_n = \aleph_n^{-1} Ç, \tag{5.1.14}$$

where

$$Ç = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \binom{s-1}{1} & \binom{s-1}{2} & \cdots & 1 \end{pmatrix}$$

and the appropriate normalizing constant is

$$\aleph_n = \operatorname{diag}\left(n^{s-1/2}a_n, n^{(s-1)-1/2}a_n, \ldots, n^{1/2}a_n\right).$$

Therefore, similar to the case with root 1, we have

$$K_n \left(\sum_{t=s+1}^{n} \mathbf{v_{t-1}} \mathbf{v_{t-1}}^T \psi'(\epsilon_t)\right) K_n^T \xrightarrow{d} \Upsilon,$$

where $\Upsilon = (\upsilon_{i,j})$ is a $s \times s$ random matrix such that:

$$\Upsilon = \mathrm{E}\left(\psi'(\epsilon_1)\right) \begin{pmatrix} \int_0^1 (\mathcal{S}_{s-1}^{(1)}(t))^2 dt & \int_0^1 \mathcal{S}_{s-1}^{(1)}(t)\mathcal{S}_{s-2}^{(1)}(t)dt & \cdots & \int_0^1 \mathcal{S}_{s-1}^{(1)}(t)\mathcal{S}_1^{(1)}(t)dt \\ \int_0^1 \mathcal{S}_{s-2}^{(1)}(t)\mathcal{S}_{s-1}^{(1)}(t)dt & \int_0^1 (\mathcal{S}_{s-2}^{(1)})^2(t)dt & \cdots & \int_0^1 \mathcal{S}_{s-2}^{(1)}(t)\mathcal{S}_1^{(1)}(t)dt \\ \vdots & \vdots & \ddots & \vdots \\ \int_0^1 \mathcal{S}_1^{(1)}(t)\mathcal{S}_{s-1}^{(1)}(t)dt & \int_0^1 \mathcal{S}_1^{(1)}(t)\mathcal{S}_{s-2}^{(1)}(t)dt & \cdots & \int_0^1 (\mathcal{S}_1^{(1)})^2(t)dt \end{pmatrix}. \tag{5.1.15}$$

We also have

$$K_n \sum_{t=s+1}^{n} \mathbf{v_{t-1}} \psi(\epsilon_t) \xrightarrow{d} \mathcal{H},$$

where

$$\mathcal{H} = -\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right) \begin{pmatrix} \int_0^1 \mathcal{S}_s^{(1)}(t)dW(t) \\ \int_0^1 \mathcal{S}_{s-1}^{(1)}(t)dW(t) \\ \vdots \\ \int_0^1 \mathcal{S}_1^{(1)}(t)dW(t) \end{pmatrix}. \tag{5.1.16}$$

Finally, we get

$$\left(K_n^T\right)^{-1} \left(\sum_{t=s+1}^{n} \mathbf{v_{t-1}} \mathbf{v_{t-1}}^T \psi'(\epsilon_t)\right)^{-1} \sum_{t=s+1}^{n} \mathbf{v_{t-1}} \psi(\epsilon_t) \xrightarrow{d} \Upsilon^{-1}\mathcal{H}, \tag{5.1.17}$$

where $\Upsilon = (\upsilon_{i,j})_{s\times s}$ and $\mathcal{H}$ are defined in (5.1.15) and (5.1.16), respectively.

□

Proof of Theorem 5.1.1 with complex conjugate unit roots. Consider the model

$$(1 - 2\cos\theta B + B^2)^d w_t = \epsilon_t.$$

Let

$$y_t(j) = (1 - 2\cos\theta B + B^2)^{d-j} w_t, \quad \text{for} \quad j = 1, 2, \ldots, d \tag{5.1.18}$$

and

$$Y_t = (y_t(1), y_{t-1}(1), \ldots, y_{t-1}(d), y_{t-1}(d))^T.$$

Therefore, we have

$$(1 - 2\cos\theta B + B^2) y_t(j+1) = y_t(j)$$

and

$$y_t(j+1) = \frac{1}{\sin\theta} \sum_{k=1}^{t} \sin\left(\theta(t-k+1)\right) y_k(j) \quad \text{for} \quad j = 0, 1, \ldots, d-1.$$

We can find a $2d \times 2d$ matrix D such that $D\mathbf{w}_t = Y_t$, where $\mathbf{w}_t = (w_t, \ldots, w_{t-2d+1})$. For more details see Chan and Wei [11]. By applying trigonometric identities, we have

$$\sin(\theta) y_t(j) = a_n \sin\left((t+1)\theta\right) T_{1,t}(j-1) - a_n \cos\left((t+1)\theta\right) T_{2,t}(j-1), \tag{5.1.19}$$

where

$$T_{1,t}(j) = a_n^{-1} \sum_{k=1}^{t} \cos(k\theta) y_k(j)$$

and

$$T_{2,t}(j) = a_n^{-1} \sum_{k=1}^{t} \sin(k\theta) y_k(j).$$

Note that by Lemma 4.3.2 in Chapter 4, we have

$$\begin{pmatrix} T_{1,t}(0) \\ T_{2,t}(0) \end{pmatrix} \xrightarrow{d} \mathbf{T}(\cdot) = \begin{pmatrix} T_1(t) \\ T_2(t) \end{pmatrix}, \tag{5.1.20}$$

where $\mathbf{T}(\cdot)$ is defined in Lemma 4.3.2.

Now, consider that $\Xi = D(\sum_{t=2d+1}^{n} \mathbf{w_{t-1}}\mathbf{w_{t-1}}^T \psi'(\epsilon_t))D^T$ has the following representation:

$$\Xi = \begin{pmatrix} \sum_{t=2d+1}^{n} y_t^2(1)\psi'(\epsilon_t) & \sum_{t=2d+1}^{n} y_t(1)y_{t-1}(1)\psi'(\epsilon_t) & \cdots & \sum_{t=2d+1}^{n} y_t(1)y_{t-1}(d)\psi'(\epsilon_t) \\ \sum_{t=2d+1}^{n} y_{t-1}(1)y_t(1)\psi'(\epsilon_t) & \sum_{t=2d+1}^{n} y_{t-1}^2(1)\psi'(\epsilon_t) & \cdots & \sum_{t=2d+1}^{n} y_{t-1}(1)y_{t-1}(d)\psi'(\epsilon_t) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=2d+1}^{n} y_{t-1}(d)y_t(1)\psi'(\epsilon_t) & \sum_{t=2d+1}^{n} y_{t-1}(d)y_{t-1}(1)\psi'(\epsilon_t) & \cdots & \sum_{t=2d+1}^{n} y_{t-1}^2(d)\psi'(\epsilon_t) \end{pmatrix}.$$

To find the limiting distribution of Ξ, by using Lemma 4.3.3 in Chapter 4, we have

$$\sup_{0<j\leq n} \left| \sum_{k=1}^{j} e^{ik\theta} T_{i,n}\left(\frac{k}{n}\right) \right| = o_p(n) \tag{5.1.21}$$

for $i = 1, \ldots, d$. Moreover, by letting

$$L_n = M_n^{-1} D, \tag{5.1.22}$$

where

$$M_n = \operatorname{diag}\left(n^{1/2} a_n I, \ldots, n^{(d-1)/2} a_n I\right) \quad \text{and} \quad I = \operatorname{diag}(1,1),$$

we have

$$L_n \left(\sum_{t=2d+1}^{n} \mathbf{w_{t-1}}\mathbf{w_{t-1}}^T \psi'(\epsilon_t) \right) L_n^T \xrightarrow{d} \Lambda.$$

Here $\Lambda = (\lambda_{i,j})$ is a $2d \times 2d$ random matrix, where $\lambda_{2i-1,2j-1} = \lambda_{2i,2j}$ and $\lambda_{2i-1,2j} = \lambda_{2i,2j-1}$ for $i, j = 1, 2, \ldots, 2d$. By using (5.1.20) and (5.1.21) along with the continuous mapping theorem, we can show that $\lambda_{i,j}$ is presented as follows:

$$\begin{aligned} \lambda_{2i-1,2j-1} = \lambda_{2i,2j} = {} & \frac{E\left(\psi'(\epsilon_1)\right)}{2} \\ & \times \left(\int_0^1 T_{1,t}(i-1)T_{1,t}(j-1)d(t) + \int_0^1 T_{2,t}(i-1)T_{2,t}(j-1)d(t) \right), \end{aligned} \tag{5.1.23}$$

$$\begin{aligned} \lambda_{2i-1,2j} = \lambda_{2i,2j-1} = {} & \frac{E\left(\psi'(\epsilon_1)\right)}{2} \\ & \times \Bigg[\cos\theta \left(\int_0^1 T_{1,t}(i-1)T_{1,t}(j-1)d(t) + \int_0^1 T_{2,t}(i-1)T_{2,t}(j-1)d(t) \right) \end{aligned}$$

$$-\sin\theta\left(\int_0^1 T_{1,t}(i-1)T_{2,t}(j-1)d(t)-\int_0^1 T_{1,t}(j-1)T_{2,t}(i-1)d(t)\right)\Bigg].$$

Also, we have

$$L_n\sum_{t=2d+1}^{n}\mathbf{w_{t-1}}\psi(\epsilon_t)\ \xrightarrow{d}\ \mathcal{G},$$

where $\mathcal{G}=(\mathcal{G}_1,\mathcal{G}_2,\ldots,\mathcal{G}_{2d})^T$. Note that $\mathcal{G}_i$ for $i=1,\ldots,2d$ can be expressed as follows:

$$\begin{aligned}\mathcal{G}_{2i-1}=\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\Bigg[\cos\theta&\left(\int_0^1 T_{1,t}(j-1)dR_1(t)-\int_0^1 T_{2,t}(j-1)dR_2(t)\right)\\&+\sin\theta\left(\int_0^1 T_{1,t}(j-1)dR_2(t)+\int_0^1 T_{2,t}(j-1)dR_1(t)\right)\Bigg],\end{aligned}\tag{5.1.24}$$

$$\mathcal{G}_{2i}=\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\times\left(\int_0^1 T_{1,t}(j-1)dR_1(t)-\int_0^1 T_{2,t}(j-1)dR_2(t)\right),$$

where $\mathbf{R}=(R_1(\cdot),R_2(\cdot))^T$ is defined in (4.2.6). Thus, we have the following result:

$$\left(L_n^T\right)^{-1}\left(\sum_{t=2d+1}^{n}\mathbf{w_{t-1}}\mathbf{w_{t-1}}^T\psi'(\epsilon_t)\right)^{-1}\sum_{t=2d+1}^{n}\mathbf{w_{t-1}}\psi(\epsilon_t)\ \xrightarrow{d}\ \Lambda^{-1}\mathcal{G},\tag{5.1.25}$$

where $\Lambda=(\lambda_{i,j})$ and $\mathcal{G}=(\mathcal{G}_1,\mathcal{G}_2,\ldots,\mathcal{G}_{2d})^T$ are defined in (5.1.23) and (5.1.24), respectively.

□

5.2 The limiting distribution of the cross product terms

In the following theorem, we show that the limiting distributions of the cases involving the cross product terms converge to zero in probability.

Theorem 5.2.1 *Take $X_0=\cdots=X_{-p+1}=0$. Let J_n and K_n be matrices defined in* (5.1.7) *and* (5.1.14)*. Also, $L_n(i)$ for $1\le i\le l\le p$ are defined similar to L_n in* (5.1.22)*. Then, we have*

(i)

$$J_n \sum_{t=1}^{n} \mathbf{u_{t-1}} \mathbf{v_{t-1}}^T K_n{}^T \xrightarrow{p} 0,$$

(ii)

$$J_n \sum_{t=1}^{n} \mathbf{u_{t-1}} \mathbf{w_{t-1}}^T(j) L_n(j)^T \xrightarrow{p} 0,$$

(iii)

$$K_n \sum_{t=1}^{n} \mathbf{v_{t-1}} \mathbf{w_{t-1}}^T(j) L_n(j)^T \xrightarrow{p} 0,$$

(iv)

$$L_n(i) \sum_{t=1}^{n} \mathbf{w_{t-1}}(i) \mathbf{w_{t-1}}^T(j) L_n(j)^T \xrightarrow{p} 0.$$

Proof: It can be shown that all elements in (i)-(iv) can be expressed as following:

$$a_n^{-2} n^{-(i+j-1)} \sum_{t=1}^{n} u_t(i) v_t(j) = a_n^{-2} n^{-(i+j-1)} \sum_{t=1}^{n} \cos(t\pi) u_t(i) \left((-1)^t v_t(j)\right),$$

$$\begin{aligned} a_n^{-2} n^{-(i+j-1)} \sum_{t=1}^{n} uF_t(i) y_t(j,m) &= a_n^{-2} n^{-(i+j-1)} (\sin\theta_j)^{-1} \\ &\times \sum_{t=1}^{n} [u_t(i) \sin((t+1)\theta_j) T_{1,t}(j,m-1) \\ &\qquad - u_t(i) \cos((t+1)\theta_j) T_{2,t}(j,m-1)], \end{aligned}$$

$$\begin{aligned} a_n^{-2} n^{-(i+j-1)} \sum_{t=1}^{n} v_t(i) y_t(j,m) &= a_n^{-2} n^{-(i+j-1)} (\sin\theta_j)^{-1} \\ &\times \sum_{t=1}^{n} \left[\left((-1)^t v_t(i)\right) \cos(t\pi) \sin((t+1)\theta_j) T_{1,t}(j,m-1) \right. \\ &\qquad \left. - \left((-1)^t v_t(i)\right) \cos(t\pi) \cos((t+1)\theta_j) T_{1,t}(j,m-1)\right], \end{aligned}$$

$$a_n^{-2} n^{-(i+j-1)} \sum_{t=1}^{n} y_t(i,h) y_t(j,m) = a_n^{-2} n^{-(i+j-1)} (\sin\theta_i \sin\theta_j)^{-1}$$

$$\times \sum_{t=1}^{n} \left[(T_{1,t}(i,h-1)\sin((t+1)\theta_i) - T_{2,t}(i,h-1)\cos((t+1)\theta_i)) \right.$$
$$\left. \times \, (T_{1,t}(j,m-1)\sin((t+1)\theta_j) - T_{2,t}(i,m-1)\cos((t+1)\theta_j)) \right],$$

where $y_t(j,m)$ is defined in (5.1.18) by replacing w_t by $w_t(j)$ and $T_{1,t}$ and $T_{2,t}$ are defined in (5.1.19). Using the trigonometric identities, all the preceding forms can be expressed as the sum of $a_n^{-2}n^{-k+1}\sum_{t=1}^{n}\sin(t\theta)Z_tP_t$ and $a_n^{-2}n^{-k+1}\sum_{t=1}^{n}\cos(t\theta)Z_tP_t$ for some positive number k, random variables Z_t, P_t and some θ such that $e^{i\theta} \neq 1$ ([11]). Moreover, by (4.2.7), (5.1.6), (5.1.13), and (5.1.19) there exist a process $S(t)$ such that

$$a_n^{-2}n^{-k+2}Z_tP_t \xrightarrow{d} S(t).$$

By Lemma 4.3.3, it can be shown that each element in (i)-(iv) converges to zero in probability.

□

Chapter 6

Autoregressive Models with a Location Parameter

It is common that authors assume the location parameter is zero in time series models. This assumption is not usually true. Ignoring the location parameter does not usually contradict the loss of generality in time series with finite variance. This is due to the fact that simply centering the data shifts the observations to the zero location parameter. In infinite variance models, the assumption of nonzero location parameter in time series is not necessarily trivial.

Suppose we have the following AR(p) process

$$X_t = \mu + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t, \quad t = 1, 2, \ldots, n, \tag{6.0.1}$$

where $\{\epsilon_t\}$ satisfy in condition A4 and (4.2.3) and conditions A1-A3 in Chapter 3 hold. In Chapters 3 and 5, we estimated the location parameter (mean when $1 < \alpha \leq 2$) and $\Phi = (\phi_1, \ldots, \phi_p)^T$ by the M-estimate method. In this chapter, we show that we can estimate the location parameter μ simultaneously with Φ. In Section 6.1, we derive the limiting distribution of M-estimates for the parameters in model (6.0.1). A simulation study is carried out in Section 6.2 to analyze the behaviour of the M-estimates.

6.1 The limiting distribution for M-estimates

The limiting distribution for the infinite-variance random-walk processes with the location parameter is derived by Knight (1989). To find the asymptotic behaviour of the M-estimates of the parameters for AR(p) processes, for simplicity, we first consider $p = 2$. As an example, the limiting distribution for an AR(2) process with roots 1 is given as follows. Generalizing to larger p is straightforward, and it will be discussed after Example 6.1.1.

Example 6.1.1 *Consider the model*

$$X_t = \mu + 2X_{t-1} - X_{t-2} + \epsilon_t. \tag{6.1.1}$$

To estimate μ, ϕ_1, and ϕ_2 simultaneously, we consider the following sequence of stochastic processes

$$A_n(u,v,h) = \sum_{t=3}^{n} \left[\rho\left(\epsilon_t - n^{-1/2} a_n^{-1} u \left(X_{t-1} - X_{t-2} \right) - n^{-3/2} a_n^{-1} v X_{t-2} - n^{-1/2} h \right) - \rho(\epsilon_t) \right].$$

Using the Taylor series expansion of each summand of A_n around $u = 0$, $v = 0$, and $h = 0$, we get

$$\begin{aligned} A_n(u,v,h) = & -un^{-1/2}a_n^{-1} \sum_{t=3}^{n} \left(X_{t-1} - X_{t-2} \right) \psi(\epsilon_t) \\ & - vn^{-3/2} a_n^{-1} \sum_{t=3}^{n} X_{t-2} \psi(\epsilon_t) \\ & - hn^{-1/2} \sum_{t=3}^{n} \psi(\epsilon_t) \\ & + \frac{1}{2} u^2 n^{-1} a_n^{-2} \sum_{t=3}^{n} \left(X_{t-1} - X_{t-2} \right)^2 \psi'(c_t^n) \\ & + \frac{1}{2} v^2 n^{-3} a_n^{-2} \sum_{t=3}^{n} X_{t-2}^2 \psi'(c_t^n) \\ & + \frac{1}{2} h^2 n^{-1} \sum_{t=3}^{n} \psi'(c_t^n) \\ & + uvn^{-2} a_n^{-2} \sum_{t=3}^{n} X_{t-2} \left(X_{t-1} - X_{t-2} \right) \psi'(c_t^n) \end{aligned} \tag{6.1.2}$$

$$+\, uhn^{-1}a_n^{-1}\sum_{t=3}^{n}\left(X_{t-1}-X_{t-2}\right)\psi'(c_t^n)$$
$$+\, vhn^{-2}a_n^{-1}\sum_{t=3}^{n}X_{t-2}\psi'(c_t^n).$$

Similar to the preceding proofs in Chapters 4 and 5, we can substitute $\psi'(c_t^n)$ *by* $\psi'(\epsilon_t)$ *and each* $\psi'(\epsilon_t)$ *can be replaced by* $E\left(\psi'(\epsilon_t)\right)$ *in the sums in* (6.1.2). *Thus, the finite-dimensional distributions of* A_n *converge weakly to those of* A *where*

$$\begin{aligned}
A(u,v,h) = &-u\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\int_0^1 S(t)dW(t)\\
&- v\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\int_0^1\int_0^t S(s)dsdW(t)\\
&- hW(1)\\
&+ \frac{u^2}{2}\mathrm{E}\left(\psi'(\epsilon_1)\right)\int_0^1 S^2(t)dt\\
&+ \frac{v^2}{2}\mathrm{E}\left(\psi'(\epsilon_1)\right)\int_0^1\left(\int_0^t S(s)ds\right)^2 dt\\
&+ \frac{h^2}{2}\mathrm{E}\left(\psi'(\epsilon_1)\right)\\
&+ uv\mathrm{E}\left(\psi'(\epsilon_1)\right)\int_0^1 S(t)\int_0^t S(s)dsdt.\\
&+ uh\mathrm{E}\left(\psi'(\epsilon_1)\right)\int_0^1 S(t)dt\\
&+ vh\mathrm{E}\left(\psi'(\epsilon_1)\right)\int_0^1\int_0^t S(s)dsdt.
\end{aligned}$$

By setting the derivative of $A(u,v,h)$ *to 0 and solving for* u, v, *and* h, *we get*

$$\begin{pmatrix} n^{1/2}a_n(\hat{\phi}_1-2)\\ n^{3/2}a_n(\hat{\phi}_1-2)+n^{3/2}a_n(\hat{\phi}_2+1)\\ n^{1/2}\left(\hat{\mu}_M-\mu\right)\end{pmatrix} \xrightarrow{d} \tilde{\Gamma}^{-1}\begin{pmatrix}\mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\int_0^1 S(t)\, dW(t)\\ \mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\int_0^1\int_0^t S(s)\, ds\, dW(t)\\ \mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)W(1)\end{pmatrix},$$

where

$$\tilde{\Gamma} = \mathrm{E}\left(\psi'(\epsilon_1)\right)\begin{pmatrix}\int_0^1 S^2(t)\, dt & \int_0^1 S(t)\int_0^t S(s)\, ds\, dt & \int_0^1 S(t)dt\\ \int_0^1 S(t)\int_0^t S(s)\, ds\, dt & \int_0^1\left(\int_0^t S(s)\, ds\right)^2 dt & \int_0^1\int_0^t S(s)dsdt\\ \int_0^1 S(t)dt & \int_0^1\int_0^t S(s)dsdt & 1\end{pmatrix}.$$

We can extend the results of Example 6.1.1 to an AR(p) process when the characteristic root may have different multiplicities and lie on the unit circle. To derive the asymptotic behaviour of the M-estimates, consider the AR(p) model in (6.0.1), when the errors are an i.i.d. sequence of random variables in the domain of attraction of a symmetric stable law with index $0 < \alpha \leq 2$. Let u_t, v_t, and $w_t(k)$ for $k = 1, \ldots, l$ be the same as in (5.1.4). Thus, model (6.0.1) is equivalent to

$$\epsilon_t + \mu = (1-B)^r u_t = (1+B)^s v_t = (1 - 2\cos(\theta_k)B + B^2)^{d_k} w_t(k).$$

Here, for simplicity, we assume that the AR(p) model has only root 1 with the multiplicity of p, $\epsilon_t + \mu = (1-B)^p u_t$. Similar to Chapter 5, we define

$$\tilde{J}_n = \tilde{N}_n^{-1} \tilde{C},$$

where

$$\tilde{C} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & (-1)\binom{p-1}{1} & (-1)^2\binom{p-1}{2} & \cdots & (-1)^{p-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

and the appropriate normalizing constant is

$$\tilde{N}_n = \operatorname{diag}\left(n^{p-1/2} a_n, n^{(p-1)-1/2} a_n, \ldots, n^{1/2} a_n, n^{1/2}\right).$$

To find the limiting distribution for the M-estimates of the location and Φ simultaneously, we define

$$\tilde{\mathbf{u}}_{\mathbf{t}} = (u_t, \ldots, u_{t-p+1}, 1)^T.$$

It is easy to see that

$$\tilde{C} \begin{pmatrix} u_t \\ u_{t-1} \\ \vdots \\ u_{t-p+1} \\ 1 \end{pmatrix} = \begin{pmatrix} u_t(p) \\ u_t(p-1) \\ \vdots \\ u_t(1) \\ 1 \end{pmatrix}.$$

Thus, we have

$$\tilde{C}\left(\sum_{t=p+1}^{n} \mathbf{\tilde{u}_{t-1}}\mathbf{\tilde{u}_{t-1}^{T}}\psi'(\epsilon_t)\right)\tilde{C}^T = \tilde{\Pi},$$

where

$$\tilde{\Pi} = \begin{pmatrix} \sum_{t=p+1}^{n} u_{t-1}^2(p)\psi'(\epsilon_t) & \sum_{t=p+1}^{n} u_{t-1}(p)u_{t-1}(p-1)\psi'(\epsilon_t) & \cdots & \sum_{t=p+1}^{n} u_{t-1}(p)\psi'(\epsilon_t) \\ \sum_{t=p+1}^{n} u_{t-1}(p-1)u_{t-1}(p)\psi'(\epsilon_t) & \sum_{t=p+1}^{n} u_{t-1}^2(p-1)\psi'(\epsilon_t) & \cdots & \sum_{t=p+1}^{n} u_{t-1}(p-1)\psi'(\epsilon_t) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=p+1}^{n} u_{t-1}(1)u_{t-1}(p)\psi'(\epsilon_t) & \sum_{t=p+1}^{n} u_{t-1}(1)u_{t-1}(p-1)\psi'(\epsilon_t) & \cdots & \sum_{t=p+1}^{n} u_{t-1}(1)\psi'(\epsilon_t) \\ \sum_{t=p+1}^{n} u_{t-1}(p)\psi'(\epsilon_t) & \sum_{t=p+1}^{n} u_{t-1}(p-1)\psi'(\epsilon_t) & \cdots & \sum_{t=p+1}^{n} \psi'(\epsilon_t) \end{pmatrix}.$$

Therefore, we can prove that

$$\begin{aligned} &\left(\tilde{J}_n^T\right)^{-1}\left(\hat{\phi}_1-\phi_1,\ldots,\hat{\phi}_p-\phi_p,\hat{\mu}_M-\mu\right)^T \\ &\quad \overset{p}{\sim} \left(\tilde{J}_n^T\right)^{-1}\left(\sum_{t=p+1}^{n}\mathbf{\tilde{u}_{t-1}}\mathbf{\tilde{u}_{t-1}^{T}}\psi'(\epsilon_t)\right)^{-1}\sum_{t=p+1}^{n}\mathbf{\tilde{u}_{t-1}}\psi(\epsilon_t) \\ &\quad \overset{d}{\to} \tilde{\Gamma}^{-1}\tilde{\mathcal{F}}, \end{aligned}$$

where $\tilde{\Gamma} = (\tilde{\gamma}_{i,j})$ is a $(p+1)\times(p+1)$ random matrix such that

$$\tilde{\Gamma} = \mathrm{E}\left(\psi'(\epsilon_1)\right)\begin{pmatrix} \int_0^1 \mathcal{S}_{p-1}^2(t)dt & \int_0^1 \mathcal{S}_{p-1}(t)\mathcal{S}_{p-2}(t)dt & \cdots & \int_0^1 \mathcal{S}_{p-1}(t)dt \\ \int_0^1 \mathcal{S}_{p-2}(t)\mathcal{S}_{p-1}(t)dt & \int_0^1 \mathcal{S}_{p-2}^2(t)dt & \cdots & \int_0^1 \mathcal{S}_{p-2}(t)dt \\ \vdots & \vdots & \ddots & \vdots \\ \int_0^1 \mathcal{S}_1(t)\mathcal{S}_{p-1}(t)dt & \int_0^1 \mathcal{S}_1(t)\mathcal{S}_{p-2}(t)dt & \cdots & \int_0^1 \mathcal{S}_1(t)dt \\ \int_0^1 \mathcal{S}_{p-1}(t)dt & \int_0^1 \mathcal{S}_{p-2}(t)dt & \cdots & 1 \end{pmatrix}$$

and

$$\tilde{\mathcal{F}} = \mathrm{E}^{1/2}\left(\psi^2(\epsilon_1)\right)\begin{pmatrix} \int_0^1 \mathcal{S}_p(t)dW(t) \\ \int_0^1 \mathcal{S}_{p-1}(t)dW(t) \\ \vdots \\ \int_0^1 \mathcal{S}_1(t)dW(t) \\ W(1) \end{pmatrix}.$$

The limiting distributions of the other roots with the location parameter are more or less similar to this case and, due to similarity, we do not provide their proofs here.

6.2 Simulation

In this section, a simulation study is carried out to get some appreciation of the nature of the sampling distributions derived in Section 6.1 when the location exists in the model. Consider model (6.1.1) with $\mu = 4$. We present the results for the three parameters ϕ_1, ϕ_2, and μ for indices of stability $\alpha = 0.5, 1, 1.5, 1.8$, and 2.

We generate the time series $\{X_t\}_{t=0}^{n}$ in model (6.1.1), for n = 20, 50, and 100. Then, the M-estimates of ϕ_1, ϕ_2, and μ in AR(2) with roots 1 are calculated with the Huber loss function presented in (3.4.1). Note that the value of c for the Huber loss function is chosen based on the results of Table 3.1 in Chapter 3. According to values of α, the applied truncation values for c are 0.5, 1, 1.5, 2, and 3. We repeat the simulation 10,000 times for each choice of n and α. The average of the estimations of the three parameters are summarized in Tables 6.1, 6.2, and 6.3.

The simulation results show that, as expected, the M-estimates of μ, ϕ_1, and ϕ_2 are consistent and near the actual values as n gets large. However, due to the rate of convergence of $\hat{\mu}_M$ which is $\sqrt{n}$, the estimates of μ are not as precise as other parameters when n is small. Thus, to have an accurate estimation for the location, a large sample size is needed. Moreover, for all sample sizes, the estimates of the parameters are more accurate when $\alpha = 0.5$.

Table 6.1: M-estimates of the parameters in model (6.1.1) with sample size $n = 20$ and replication size $10,000$.

α	ϕ_1	ϕ_2	μ
0.5	1.9811	-0.9796	4.2711
1.0	1.9588	-0.9552	4.5645
1.5	1.9635	-0.9596	4.5279
1.8	1.9716	-0.9686	4.4169
2.0	1.9759	-0.9733	4.3496

Table 6.2: M-estimates of the parameters in model (6.1.1) with sample size $n = 50$ and replication size $10,000$.

α	ϕ_1	ϕ_2	μ
0.5	1.9985	-0.9984	4.0673
1.0	1.9931	-0.9929	4.2863
1.5	1.9930	-0.9927	4.2833
1.8	1.9946	-0.9944	4.2075
2.0	1.9958	-0.9957	4.1661

Table 6.3: M-estimates of the parameters in model (6.1.1) with sample size $n = 100$ and replication size $10,000$.

α	ϕ_1	ϕ_2	μ
0.5	1.9999	-0.9999	4.0105
1.0	1.9983	-0.9982	4.1414
1.5	1.9982	-0.9981	4.1504
1.8	1.9986	-0.9986	4.1122
2.0	1.9989	-0.9989	4.0934

Chapter 7

Future work

1. Liu, Parelius, and Singh [30] consider a nonparametric multivariate method based on the concept of data depth to analyze multivariate distributional characteristics such as location, scale, bias, and skewness. However, this approach is not applicable if the moments do not exist. Similar results can be derived and extended to the observations in the domain of attraction of a multivariate stable law.

2. A competitive method for constructing the confidence interval of the mean is the empirical likelihood method, introduced by Owen [38, 39]. Peng [41] derive an empirical-likelihood-based confidence interval for the mean when the underlying distribution has heavy tails. The performance of the model presented in Chapter 2 can be compared with empirical likelihood method.

3. Hall and LePage [22] propose a bootstrap method for estimating the distribution of the mean vector when the sampling distribution is not in the domain of attraction of any limit law. This method can be employed to make an inference for the mean vector of observations which are not in the domain of attraction of any limit law.

4. In this thesis, we assumed that the left and right tail indices are the same for all cases. The results of this thesis can be generalized to the case where $\alpha_l \neq \alpha_r$.

5. Generalizing the results in Chapter 4 and 5 to the case when the innovations $\{\epsilon_i\}$ are

weakly dependent such as innovations generated from MA(∞).

6. Stable innovations are one example of many Lévy jump distributions. There have been several activities in financial modeling to model innovations using variance gamma random variables; for example see Madan and Seneta [31]. Therefore, the unit root models whose innovations are variance gamma random variables would be of practical interest.

Appendix A

A multivariate stable characterization and the domain of attraction

Multivariate stable distributions are defined as those having a domain of attraction, where vectors $\mathbf{a}_n$ and $\mathbf{b}_n \in \mathbb{R}^p$ are used for normalizing and centering, respectively (Definition 3.1.1). A multivariate convergence in (3.1.2) shows that $\mathbf{Y}$ must be infinitely divisible. The class of all possible limits of an infinitely divisible bivariate distribution in (3.1.2) is discussed in Resnick and Greenwood [44].

The limiting distribution of $\mathbf{Y}$ can be a combination of stable laws with different indices of stability. When $\alpha_j = 2$, $j = 1, \ldots, p$, then $\mathbf{Y}$ has a multivariate normal distribution. For the other possibilities, the stable limits with indices equal to 2, will be independent from the stable limits with indices less than 2. For $0 < \alpha_j < 2$, $j = 1, \ldots, p$, the random vector $\mathbf{Y}$ is an infinitely divisible with Radon measure ν on $\mathbb{R}^p$ defined by $\tilde{\nu} = \nu \circ \tau$, where

$$\tau \mathbf{x} = \Big((\operatorname{sign} x_1)|x_1|^{1/\alpha_1}, \ldots, (\operatorname{sign} x_p)|x_p|^{1/\alpha_p} \Big)',$$

with $\tilde{\nu}$ given by

$$\tilde{\nu}\{\mathbf{x} : |x| > r, \theta(x) \in H\} = r^{-1} G(H) \tag{A.0.1}$$

and G is a finite measure on $[0, 2\pi]$. By Resnick and Greenwood [44] and Resnick [48], we have the following theorem.

Theorem A.0.1 *There exist sequences* $\mathbf{a}_n = (a_n^{(1)}, \ldots, a_n^{(p)})', \mathbf{b}_n \in \mathbb{R}^p$ *with* $a_n^{(j)} > 0$ *such that the following are equivalent:*

(i)

$$\left(S_n^{(1)}/a_n^{(1)}, \ldots, S_n^{(p)}/a_n^{(p)}\right)' - \mathbf{b}_n \xrightarrow{d} \mathbf{Y},$$

where $\mathbf{Y}$ *is a random vector on* $\mathbb{R}^p$ *with stable distribution.*

(ii) For all $A \in \mathcal{B}(\mathbb{R}^p - \{0\})$ *such that* $\nu(\partial A) = 0$ *and* $\nu(A) < \infty$*, we have*

$$\lim_{n\to\infty} nP\left(\mathbf{X}_{n,1} \in A\right) = \nu(A),$$

where $\mathbf{X}_{n,1} = ((a_n^{(1)})^{-1}X_{11}, \ldots, (a_n^{(p)})^{-1}X_{1p})'$ *and* ν *is a Radon measure defined by* $\tilde{\nu} = \nu \circ \tau$ *where* $\tau\mathbf{x} = \left((\operatorname{sign} x_1)|x_1|^{1/\alpha_1}, \ldots, (\operatorname{sign} x_p)|x_p|^{1/\alpha_p}\right)'$ *and* $\tilde{\nu}$ *is given in* (A.0.1).

If we can take $a_n^{(1)} = \cdots = a_n^{(p)}$, Theorem A.0.1 (ii) is equivalent to

$$\lim_{x\to\infty} \frac{P\left(|\mathbf{X}_1 > xr|\right)}{P\left(|\mathbf{X}_1 > x|\right)} = r^{-\alpha}, \qquad 0 < \alpha < 2 \tag{A.0.2}$$

and

$$\lim_{x\to\infty} P\left(\theta(\mathbf{X}_1) \in \cdot|\ |\mathbf{X}_1| > x\right) = \frac{G(\cdot)}{G([0, 2\pi])}, \tag{A.0.3}$$

where for any set $\{x : |x| > r, \theta(x) \in H\}$ (where $r > 0$, $H \in \mathcal{B}([0, 2\pi])$), $(|x|, \theta(x))$ are polar coordinate of x.

If it is not possible to take $a_n^{(i)} = a_n^{(j)}$ for $i \neq j = 1, \ldots, p$, we can use a function of random variables $\mathbf{X}_i$ to get the same results as (A.0.2) and (A.0.3); for more details see [44] and the following discussion.

If X_{1j} is in the domain of attraction of a stable law in $\mathbb{R}$, then

$$\lim_{x\to\infty} \frac{P(|X_{1j}| > xr)}{P(|X_{1j}| > x)} = r^{-\alpha_j}$$

and

$$\lim_{x\to\infty} \frac{P(X_{1j} > x)}{P(|X_{1j}| > x)} = p_j,$$

$$\lim_{x\to\infty} \frac{P(X_{1j} \leq -x)}{P(|X_{1j}| > x)} = q_j,$$

where $p_j \in (0,1)$ and $q_j = 1 - p_j$, $j = 1, \ldots, p$. Define

$$\begin{aligned} U_j(x) &= p_j U_{j+}(x) \qquad \text{if } x > 0, \\ &= q_j U_{j-}(-x) \quad \text{if } x < 0, \end{aligned}$$

where, for $x \geq 0$

$$U_{j+}(x) = \frac{1}{P(X_{1j} > x)}$$

and

$$U_{j-}(x) = \frac{1}{P(X_{1j} \leq -x)}.$$

Thus, it can be shown that Theorem A.0.1 (ii) is equivalent to

$$\lim_{n\to\infty} P\left((U_1(X_{11})/n, \ldots, U_p(X_{1p})/n) \in A\right) = \tilde{\nu}(A), \tag{A.0.4}$$

where $\tilde{\nu}$ is given in (A.0.1). Note that (A.0.4) is equivalent to (A.0.2) and (A.0.3) when $\mathbf{X}_1$ is replaced by $((U_1(X_{11}), \ldots, U_p(X_{1p}))$ and $\alpha = 1$.

Appendix B

$D[0,1]$, the space of càdlàg functions

In the uniform topology, two functions $x(\cdot)$ and $y(\cdot)$ are near one other if their graphs are uniformly close. Now consider $D[0,1]$, the space of right-continuous functions on $[0,1)$ that have finite left limits on $(0,1]$. In the Skorohod topology on $D[0,1]$, we allow a uniformly small deformation of the time scale; i.e., x and y are close if after deforming the time scale of one of them, for example, y, the resulting graphs are close. Most of the following definitions are collected from [7], [46], and [49].

Lemma B.0.1 *([7]) For each $x \in D[0,1]$ and each $\epsilon > 0$, there exist $k \geq 1$ and times $t_0, \ldots, t_k$, $0 = t_0 < \cdots < t_k = 1$, such that*

$$\sup\{|x(s) - x(t)| : t_{i-1} \leq s, t < t_i\} < \epsilon$$

for $i = 1, \ldots, k$.

From Lemma B.0.1, we can conclude that for any function in $D[0,1]$, the number of discontinuities of x is at most finite. Furthermore, functions in $D[0,1]$ are locally bounded ([7]).

Denote the uniform distance between x and y as

$$\|x - y\| := \sup_{0 \leq t \leq 1} |x(t) - y(t)|.$$

Definition B.0.1 *Let $e(t) \in \Lambda$ be the identity transformation and the time deformation is defined by*

$$\Lambda = \{\lambda : [0,1] \mapsto [0,1] : \lambda(0) = 0, \lambda(1) = 1, \lambda(\cdot) \textit{ is continuous and strictly increasing}\}.$$

The Skorohod metric $d(x,y)$ between two functions $x, y \in D[0,1]$ is

$$\begin{aligned} d(x,y) &= \inf\{\epsilon > 0 : \exists \lambda \in \Lambda \textit{ such that} \|\lambda - e\| \vee \|x - y \circ \lambda\| \leq \epsilon\} \\ &= \inf_{\lambda \in \Lambda} \{\|\lambda - e\| \vee \|x - y \circ \lambda\|\}, \end{aligned}$$

where $x \vee y = \max\{x, y\}$.

It is easy to show that d is a metric generating the Skorohod topology on $D[0,1]$ and with respect to this metric topology, D is separable but it is not complete ([7]). To resolve this problem, Skorohod introduced an equivalent metric (i.e. topologies are the same) which makes this space complete and separable. Since the topologies are the same, if $x_n \xrightarrow{d} x$ with respect to one metric then this convergence will hold with respect to the other metric as well. Thus for the Skorohod topology we can use either of these metrics. For detailed account of this subject see Bilingsley [7].

An immediate consequence of Definition B.0.1 is as follows. Given $x_n \in D[0,1]$, $n \geq 0$, we have $d(x_n, x_0) \to 0$ iff there exist $\lambda_n \in \Lambda$ such that

$$\|\lambda_n - e\| \to 0, \qquad \|x_n \circ \lambda_n - x_0\| \to 0.$$

From the definitions, we also get

$$d(x,y) \leq \|x - y\|.$$

So that uniform convergence always implies Skorohod convergence (take $\lambda(t) \equiv t$). While, the opposite is false. For more details see [7] and [46].

Bibliography

[1] Adler, R., Feldman, A., and Taqqu. M. (1998) *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Springer-Verlag.

[2] Anderson, T.W. (1959) On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics*, 30, 676-687.

[3] Arcones, M. A. and Giné, E. (1989) The bootstrap of the mean with arbitrary sample size. *Annales de l'Institut Henri Poincaré*, 25, 457-481.

[4] Athreya, K.B. (1987) Bootstrap of the mean in the infinite variance case. *Annals of Statistics*, 15, 724-731.

[5] Athreya, K.B., Lahiri, S.N., and Wu, W. (1998) Inference for heavy tailed distributions. *Journal of Statistical Planning and Inference*, 66, 61-75.

[6] Bickel, P.J. and Sakov, A. (2008) On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica*, 18, 967-985.

[7] Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley, New York.

[8] Banjevic, D., Ishwaran, H., and Zarepour, M. (2002) A recursive method for functionals of Poisson processes. *Bernoulli*, 8, 295-311.

[9] Caner, M. (1998) Tests for cointegration with infinite variance errors. *Journal of Econometrics*, 86, 155-175.

[10] Chan, N.H. and Tran, L.T. (1989) On the first order autoregressive process with infinite variance. *Econometric Theory*, 5, 354-862.

[11] Chan, N.H. and Wei, C.Z. (1988) Limiting distribution of least square estimates of unstable autoregressive processes. *Annals of Statistics*, 16, 367-401.

[12] Chan, N.H. and Zhang, R.M. (2012) Non-stationary autoregressive processes with infinite variance. *Journal of Time Series Analysis*, 33, 916-934.

[13] Chan, N. H. and Terrin, N. (1995) Inference for unstable long-memory processes with applications to fractional unit root autoregressions. *Annals of Statistics*, 23, 1662-1683.

[14] Davis, R.A., Knight, K., and Liu, J. (1992) M-estimation for autoregressions with infinite variance. *Stochastic Processes and Their Applications*, 40, 145-180.

[15] Dickey, D.A. and Fuller, W. (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, 427-431.

[16] Diciccio, T.T. and Efron, B. (1996) Bootstrap confidence intervals. *Statistical Science*, 11, 180-228.

[17] Fama, E. (1965) The behavior of stock market prices. *The Journal of Business*, 38 34-105

[18] Feigin, P.D. and Resnick, S.I. (1997) Linear programming estimators and bootstrapping for heavy tailed phenomena. *Advances in Applied Probability*, 29, 759-805.

[19] Feller, W. (1971) *An Introduction to Probability Theory and its Applications*. Volume 2, second edition, Wiley, New York.

[20] Hall, P. (1990) Asymptotic properties of the bootstrap for heavy-tailed distributions. *Annals of Probability*, 18, 1342-1360.

[21] Haan, L. de and Resnick, S. I. (1977) Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheor. Verw. Gebiete*, 40 , 317-337.

[22] Hall, P. and LePage, R. (1996) On bootstrap estimation of the distribution of the studentized mean. *Annals of the Institute of Statistical Mathematics*, 48, 403-421.

[23] Huber, P.J. (1981) *Robust statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

[24] Ito, K. (1969) *Stochastic Processes*. Lecture Notes Series No. 16, Mathematics Institute, Aarhus Universitet.

[25] Jeganathan, P. (1991) On the asymptotic behavior of least-squares estimators in AR time series with unit roots near the unit circle. *Econometric Theory*, 7, 269-306.

[26] Knight, K. (1989a). On the bootstrap of the sample mean in the infinite variance case. *Annals of Statistics*. 17, 1168-1175.

[27] Knight, K. (1989b) Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. *Canadian Journal of Statistics*, 17, 261-278.

[28] Kurtz, T. G. and Protter, P. (1991) Weak limit theorem for stochastic integrals and stochastic differential equations. *Annals of Probability*, 19, 1035-1070.

[29] LePage, R., Woodroofe, M., and Zinn, J. (1981) Convergence to a stable distribution via order statistics. *Annals of Probability*, 9, 624-632.

[30] Liu, R.Y., Parelius, J.M., and Singh, K (1999) Multivariate analysis by data depth: descriptive statistics, graphics and inference. *The Annals of Statistics*, 27, 783-858.

[31] Madan, D. and Seneta, E. (1990) The variance gamma model for share market returns. *The Journal of Business*, 63, 511-524.

[32] Mandelbrot, B.B. (1963) The variation of certain speculative prices. *Journal of Business*, 36, 394.

[33] Mandelbrot, B.B. (1967) How long is the coast of Britain? Statistical self-similarity and fractional dimension. *American Association for the Advancement of Science*, 156, 636-638.

[34] Mann, H.B. and Wald, A. (1943) On the statistical treatment of linear stochastic difference equations. *Econometrica*, 11, 173-220.

[35] Mikosch, T. (1999) *Regular Variation, Subexponentiality and Their Applications in Probability Theory*. Eindhoven University of Technology.

[36] Mittnik, S., Paulauskas, V., and Rachev, S.T. (2001) Statistical inference in regression with heavy-tailed integrated variables. *Mathematical and Computer Modelling*, 34, 1145-1158.

[37] Moreno, M. and Romo, J. (2012) Unit root bootstrap tests under infinite variance. *Journal of Time Series Analysis*, 33, 32-47.

[38] Owen, A. B. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75, 237-249.

[39] Owen, A. B. (1990) Empirical likelihood ratio confidence regions. *Annals of Statistics*, 18, 90-120.

[40] Paulauskas, V. and Rachev, S.T. (1998) Cointegrated processes with infinite variance innovations. *Annals of Applied Probability*, 8, 775-792.

[41] Peng, L. (2004) Empirical-likelihood-based confidence interval for the mean with a heavy-tailed distribution. *The Annals of Statistics*, 32, 1192-1214.

[42] Phillips, P. C. B. (1990) Time series regression with a unit root and infinite-variance errors. *Econometric Theory*, 6, 44-62.

[43] Phillips, P. C. B. and Perron, P. (1988) Testing for a unit root in time series regression. *Biometrika*, 75, 335-346.

[44] Resnick, S. and Greenwood, P. (1979) A bivariate stable characterization and domains of attraction. *Journal of Multivariate Analysis*, 9, 206-221.

[45] Resnick, S.I. (1986) Regular variation and weak convergence. *Applied Probability Trust*, 18, 66-138.

[46] Resnick, S.I. (1987) *Extreme Values, Regular Variation and Point Processes.* Springer-Verlag, New York.

[47] Resnick, S. (1997) Heavy tail modeling and teletraffic data. *Annals of Statistics.* 25, 1805-1869.

[48] Resnick, S. (2004) On the foundations of multivariate heavy-tail analysis. *Journal of Applied Probability*, 41, 191-212.

[49] Resnick, S. (2007) *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.* Springer, New York.

[50] Samorodnitsky, G. and Taqqu, M. (1994) *Stable Non-Gaussian Random Processes.* Chapman and Hall, New York.

[51] Samarakoon, D.M. and Knight, K. (2009) A note on unit root tests with infinite variance noise. *Econometric Reviews*, 28, 314-334.

[52] Singh, K. (1981) On the asyptotic accuracy of efron bootstrap. *The Annals of Statistics*, 9, 1187-1195.

[53] Tanaka, K. (2008) Analysis of models with complex roots on the unit circle. *Journal of the Japan Statistical Society*, 38, 145-155.

[54] Willinger, W., Taqqu, M.S., Leland, W.E., and Wilson, D.V. (1995) Self-similarity in high-speed packet traffic: analysis and modeling of ethernet traffic measurements. *Statistical Science*, 10, 67-85.

[55] White, J.S. (1958) The limiting distribution of the serial correlation coefficient in the explosive case. *The Annals of Mathematical Statistics*, 29, 1188-1197.

[56] Zarepour, M. and Knight, K. (1999a) Bootstrapping unstable first order autoregressive process with errors in the domain of attraction of stable law. *Stochastic Models*, 15, 11-27.

[57] Zarepour, M. and Knight, K. (1999b) Bootstrapping point processes with some applications. *Stochastic processes and their applications*, 84, 81-90.

[58] Zarepour, M. and Roknossadati, S.M. (2008) Multivariate autoregression of order one with infinite variance innovations. *Econometric Theory*, 24, 677-695.