uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Mohammad Roknossadati
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Mathematics)
GRADE / DEGREE

Department of Mathematics and Statistics
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Robust Estimation of Multivariate and Spatial Autoregression

TITRE DE LA THÈSE / TITLE OF THESIS

Mahmoud Zarepour
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Mayer Alvo | Chul Gyu Park

Rafal Kulik | Andrei Volodin

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Robust Estimation of Multivariate and Spatial Autoregression

by

Mohammad Roknossadati, B.Sc., M.Sc.

A thesis submitted to
the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

Department of Mathematics and Statistics
Ottawa-Carleton Institute for Mathematics and Statistics

University of Ottawa
Ottawa, Ontario, Canada
May, 2008

Your file Votre référence
ISBN: 978-0-494-46520-2
Our file Notre référence
ISBN: 978-0-494-46520-2

Abstract

This dissertation consists of five chapters. In Chapter 1, we collect some fundamental concepts and definitions employed in the forthcoming chapters.

In Chapter 2, we consider the limiting behavior of a vector autoregressive model of order one (VAR(1)) with independent and identically distributed (i.i.d.) innovations vector with dependent components in the domain of attraction of a multivariate stable law with possibly different indices of stability. It is shown that in some cases the ordinary least squares (OLS) estimates are inconsistent. This inconsistency basically originates from the fact that each coordinate of the partial sum processes of dependent i.i.d. vectors of innovations in the domain of attraction of stable laws needs a different normalizer to converge to a limiting process. It is also revealed that certain M-estimates, with some regularity conditions, as an appropriate alternative, not only resolve inconsistency of the OLS estimates but also give higher consistency rates in all cases.

In Chapter 3, we study the limiting behavior of the M-estimators of parameters for a spatial unilateral autoregressive model with i.i.d. innovations in the domain of attraction of a stable law with index $\alpha \in (0, 2]$. Both stationary and unit root models and some extensions are considered. It is shown that self-normalized M-estimators are asymptotically normal.

In Chapter 4, we investigate the limit theory of the M-estimators of parameters for a near unit root spatial autoregressive model considered in Chapter 3.

Finally, some suggestions for future research are presented in Chapter 5.

Acknowledgements

I am grateful to Professor Mahmoud Zarepour, my thesis supervisor, for his constant supports and his tolerance. His assistance has been invaluable to me during these years. I have benefited greatly from his talent, wisdom, and experience for which I greatly appreciate his generosity.

My sincere gratitude goes to my thesis examining committee. I would like to thank Professors A. Volodin, C.G. Park, M. Alvo and R. Kulik for accepting to be my examiners. Their very helpful guidance and comments have improved the thesis and my knowledge. My thanks also go to the Faculty members of the Ottawa-Carleton institute who provided the excellent academic environment and taught me new ideas.

How can I thank my family for their love and concerns? How can I thank my teachers who have affected my life, those who taught me how to love and how to be loved? I can never compensate their favors but I wish to put their memories in my heart and soul.

Dedication

This thesis is dedicated to the memory of my mother Sahebjan Dashti.

Contents

List of Tables

List of Figures

Chapter 1

Definitions and Preliminaries

Stable processes appear in a wide range of applications which are indeed heavy tailed phenomena. Some of examples include applications in economics, finance, telecommunications, hydrology, biology and physics [50, 51]. In this dissertation, we consider several autoregressive and other related models with innovations (error terms) having common distribution with infinite variance. Our approaches also cover cases with finite variance such as models with normal innovations. We study the limit theory for the M-estimators of parameters of the corresponding models.

In this chapter, we introduce some preliminary concepts which would be used in the forthcoming chapters. We assume familiarity with basic concepts such as metric spaces, for instance $C[0,1]$ and $D[0,1]$, and convergence concepts as weak convergence in these metric spaces [10, 34]. Most of the definitions and related theorems are collected from [28, 50, 51].

In what follows and throughout this dissertation, $\to_d$ and $\to_p$ stand for convergence in distribution and probability, respectively.

1.1 Regular Variation

The theory of regularly varying functions is an analytical tool for dealing with heavy tails and domains of attraction. To start with, we introduce such functions.

Definition 1.1. A measurable function $U : \mathbb{R}_+ \longmapsto \mathbb{R}_+$ is a regularly varying function at ∞ with index $\rho \in \mathbb{R}$, written $U \in RV_\rho$, if for $x > 0$,

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\rho.$$

If $\rho = 0$, U is called slowly varying (at ∞) and is denoted by $L(x)$. It is always possible to represent a ρ-varying function as $x^\rho L(x)$. A very useful property which makes most of the computations simple is that $\lim_{x\to\infty} x^\epsilon L(x) = \infty$ and $\lim_{x\to\infty} x^{-\epsilon} L(x) = 0$, for any $\epsilon > 0$. Some distributions have regularly varying tails; for example, the extreme-value distribution

$$\Phi_\alpha(x) = \exp[-x^{-\alpha}], \quad x \geq 0,$$

with the property $1 - \Phi_\alpha(x) \sim x^{-\alpha}$, $x \to \infty$ (as shorthand for $\lim_{x\to\infty} \frac{1-\Phi_\alpha(x)}{x^{-\alpha}} = 1$) and a stable law with index α, $0 < \alpha < 2$, (see Section 1.4 for definition) with the property

$$1 - G(x) \sim cx^{-\alpha}, \quad x \to \infty, \quad c > 0.$$

The sequential form of regular variation which is very useful is defined as follows:

Proposition 1.1 ([28]). A monotone function $U : \mathbb{R}_+ \longmapsto \mathbb{R}_+$ is regularly varying if there are two sequences $\{\lambda_n\}$ and $\{b_n\}$ of positive numbers satisfying

$$b_n \to \infty, \quad \lambda_n \sim \lambda_{n+1}, \quad n \to \infty,$$

and for all $x > 0$,

$$\lim_{n\to\infty} \lambda_n U(b_n x) =: \chi(x) \text{ exists and is positive and finite.}$$

In this case $\chi(x)/\chi(1) = x^\rho$ and $U \in RV_\rho$ for some $\rho \in \mathbb{R}$.

Typically, U is distribution tail, $\lambda_n = n$, and b_n is a distribution quantile. Another deep result is the so-called Karamata's theorem which says that a regularly varying function integrates the way you expect a power function to integrate.

Theorem 1.1 (Karamata's theorem [50]). Suppose $\rho \geq -1$ and $U \in RV_\rho$. Then $\int_0^x U(x)\, dt \in RV_{\rho+1}$ and

$$\lim_{x\to\infty} \frac{xU(x)}{\int_0^x U(t)\, dt} = \rho + 1.$$

If $\rho < -1$ (or if $\rho = -1$ and $\int_x^\infty U(s)\, ds < \infty$), then $U \in RV_\rho$ implies that $\int_x^\infty U(t)\, dt$ is finite, $\int_x^\infty U(t)\, dt \in RV_{\rho+1}$, and

$$\lim_{x\to\infty} \frac{xU(x)}{\int_x^\infty U(t)\, dt} = -\rho - 1.$$

A very applicable variant of Karamata's theorem is given as follows.

Theorem 1.2 (Variant of Karamata's theorem [50]). Suppose F is a distribution on $\mathbb{R}_+$ and

$$1 - F(x) \sim x^{-\alpha} L(x), \quad x \to \infty.$$

(i) For $\gamma > \alpha$,

$$\lim_{x\to\infty} \frac{\int_0^x u^\gamma F(du)}{x^\gamma(1-F(x))} = \frac{\alpha}{\gamma-\alpha},$$

(ii) and for $\gamma > 0$,

$$\lim_{x\to\infty} \frac{\int_x^\infty u^{-\gamma} F(du)}{x^{-\gamma}(1-F(x))} = \frac{\alpha}{\gamma+\alpha}.$$

As a corollary to the Karamata's theorem we have the following well-known representation of a regularly varying function.

Corollary 1.1 (the Karamata representation [28, 50]). The function L is slowly varying iff L can be represented as

$$L(x) = c(x) \exp\Big(\int_1^x t^{-1}\epsilon(t)\, dt \Big), \quad x > 0,$$

where $c : \mathbb{R}_+ \longmapsto \mathbb{R}_+$, $\epsilon : \mathbb{R}_+ \longmapsto \mathbb{R}_+$, and

$$\lim_{x\to\infty} c(x) = c \in (0,\infty), \quad \lim_{t\to\infty} \epsilon(t) = 0.$$

Finally we shall define a multivariate regularly varying function to be used in the next section. A subset $C \subset \mathbb{R}^d$ is called a *cone* if whenever $\mathbf{x} \in C$, then for any $t > 0$, $t\mathbf{x} \in C$. Now suppose $h \geq 0$ is a measurable function defined on C, and $\mathbf{1} = (1,...,1) \in C$. A function h is called *multivariate regularly varying at* ∞ *with limit function* τ if for all $\mathbf{x} \in C$,

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

where $\tau(\mathbf{x}) > 0$ for $\mathbf{x} \in C$.

1.2 Point Processes

Suppose $(\Omega, \mathcal{A}, P)$ is a probability space. Let $M_+(\mathbb{E})$ (and $M_p(\mathbb{E})$) be the space of Radon (and point) measures on a nice space $\mathbb{E}$, that is a locally compact topological space with countable base. A measure μ is called Radon if $\mu(K) < \infty$ for any compact subset K in $\mathbb{E}$. The space $M_+(\mathbb{E})$ is a complete separable metric space equipped with the vague topology (metrizable as a complete, separable metric space [50]), and $M_p(\mathbb{E})$ is a closed subset of $M_+(\mathbb{E})$. For $\mu, \mu_n \in M_+(\mathbb{E})$, let $\mu_n \to_v \mu$ denote the vague convergence of μ_n to μ; i.e. for all $f \in C_K^+(E)$, we have

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

as $n \to \infty$, where

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

In $M_+(\mathbb{E})$, the measures $\{\mu_n(\cdot) : n = 1, 2, ...\}$ converge weakly ($\mu_n \to_d \mu$) if and only if for any family $\{f_i\}$ with $f_i \in C_K^+(E)$, we have

$$(\mu_n(f_i), i \geq 1) \to_d (\mu(f_i), i \geq 1)$$

in $\mathbb{R}^\infty$, i.e. weak convergence occurs with respect to vague topology. In practice, one assume a sequence $\{f_i\}$ and proves $\mathbb{R}^\infty$ convergence. This reduces to proving finite dimensional convergence.

Regular variation of distribution tails can be reformulated in terms of vague convergence:

Theorem 1.3 ([50]). Let X be a nonnegative random variable with distribution function $F(x)$, and set $\bar{F} = 1 - F$. The following are equivalent:

(i) $\bar{F} \in RV_{-\alpha}, \alpha > 0$.

(ii) There exists a sequence $\{b_n\}$ with $b_n \to \infty$ such that

$$\lim_{n\to\infty} n\bar{F}(b_n x) = x^{-\alpha}, \quad x > 0.$$

(iii) There exists a sequence $\{b_n\}$ with $b_n \to \infty$ such that

$$\mu_n(\cdot) := nP\Big(\frac{X}{b_n} \in \cdot\Big) \to_v \nu_\alpha(\cdot)$$

in $M_+(0, \infty]$, where $\nu_\alpha(x, \infty] = x^{-\alpha}$.

Definition 1.2. A point process N is a map $N : (\Omega, \mathcal{A}) \mapsto (M_p(\mathbb{E}), \mathcal{M}_p(\mathbb{E}))$ with state space $\mathbb{E}$, where $\mathcal{M}_p(\mathbb{E})$ is the Borel σ-algebra of subsets of $M_p(\mathbb{E})$ generated by open sets in vague topology. The Borel subsets of $\mathbb{E}$ is denoted by $\mathcal{E}$.

Definition 1.3. A point process N is a Poisson point process with mean measure μ, or a Poisson random measure (PRM(μ)), if

(i) for $A \in \mathcal{E}$,

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

(ii) $N(A_1), ..., N(A_k)$ are independent random variables, where $A_1, ..., A_k$ are disjoint subsets of $\mathbb{E} \in \mathcal{E}$.

For example, it can be shown that $N = \sum_{n=1}^{\infty} \epsilon_{\Gamma_n}$, where ϵ as Dirac measure and $\Gamma_n = \sum_{i=1}^{n} E_i$ with iid E_i's having exp(1) distribution, is a Poisson point process on $[0, \infty)$ with the Lebesgue mean measure.

The following main theorem deals with multivariate regularly varying tail probabilities and other equivalent statements. Let $\mathbf{Z} \geq \mathbf{0}$ be a d-dimensional random vector that takes values in the nonnegative quadrant $[\mathbf{0},\infty) = [0,\infty)^d$ having distribution F with regularly varying tail (see the third statement in the following theorem).

Theorem 1.4 ([50], Theorems 6.1 and 6.2). Set $\mathbb{E} = [\mathbf{0},\infty)\setminus\{\mathbf{0}\}$. The following statements are equivalent.

1. Let $[\mathbf{0},\mathbf{x}] = [0,x_1]\times\cdots\times[0,x_d]$. There exists a Radon measure ν on $\mathbb{E}$ such that

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

for all points $\mathbf{x}\in\mathbb{E}$ which are continuity points of the function $\nu([\mathbf{0},\cdot]^c)$.

2. There is a function $b(t)\to\infty$ and a Radon measure ν on $\mathbb{E}$, the limit measure, such that in $M_+(\mathbb{E})$,

$$tP\Big[\frac{\mathbf{Z}}{b(t)}\in\cdot\Big]\to_v \nu(\cdot),\quad t\to\infty.$$

3. There exists a sequence $b_n\to\infty$ and a Radon measure ν on $\mathbb{E}$ such that in $M_+(\mathbb{E})$,

$$nP\Big[\frac{\mathbf{Z}}{b_n}\in\cdot\Big]\to_v \nu(\cdot),\quad n\to\infty.$$

4. Let $\mathbf{Z}_1,\mathbf{Z}_2,\ldots$ be iid copies of $\mathbf{Z}$. There exists $b_n\to\infty$ such that

$$\sum_{i=1}^n \epsilon_{\mathbf{Z}_i/b_n}\to_d N$$

in $M_p(\mathbb{E})$, such that N is PRM(ν).

5. For any sequence $k=k(n)\to\infty$ such that $n/k\to\infty$, we have

$$\frac{1}{k}\sum_{i=1}^n \epsilon_{\mathbf{Z}_i/b(\frac{n}{k})}\to_d \nu$$

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

Equivalent formulations in terms of polar coordinates in Theorem 1.4. are not given [50]. Multivariate regular variation also has an exact probabilistic equivalence in terms of convergence of empirical measures to a limiting PRM, which is used to prove weak convergence of partial sum processes or maximal processes in the space $D[0, \infty)$.

Theorem 1.5 ([49, 50]). Let $\{\mathbf{Z}, \mathbf{Z}_1, \mathbf{Z}_2, ...\}$ be iid random elements of $[\mathbf{0}, \infty)$. Then multivariate regular variation of the distribution of $\mathbf{Z}$ in $\mathbb{E} = [\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$,

$$nP\Big[\frac{\mathbf{Z}}{b_n} \in \cdot\Big] \to_v \nu_\alpha(\cdot),$$

is also equivalent to

$$\sum_i \epsilon_{(\frac{i}{n}, \mathbf{Z}_i/b_n)} \to_d \mathrm{PRM}(\lambda \times \nu_\alpha)$$

in $M_+([0, \infty) \times \mathbb{E})$, where λ denotes the Lebesgue measure.

It can be shown that the measure ν in Theorem 1.4 spreads mass onto each axe according to the one-dimensional measure ν_α given in Theorem 1.3 but assigns no mass off the axes. To see this and a general representation for the limiting measure ν based on the polar coordinate transformation, consult [50]. As a general example of multivariate regularly varying densities, the density of the form

$$F'(\mathbf{x}) = c(1 + \|\mathbf{x}\|^\gamma)^{-\beta}, \quad c, \gamma, \beta > 0, \ \mathbf{x} \in \mathbb{E},$$

where $\|\cdot\|$ denotes a norm on $\mathbb{R}^d$ can be illustrated. Two-dimensional Cauchy density and the bivariate t-density are of this special form.

Recall Theorem 1.4 with $\mathbb{E} = [-\infty, \infty) \setminus \{\mathbf{0}\}$. If $d = 1$, then the part (3) of the theorem implies for $x > 0$ that

$$nP\Big[\frac{Z_1}{b_n} > x\Big] \to \nu(x, \infty] \quad n \to \infty, \tag{1.1}$$

which is sequential regular variation, so $\nu(x,\infty] = px^{-\alpha},\ p \geq 0$. Similarly, for $x > 0$,

$$nP\Big[\frac{Z_1}{b_n} < -x\Big] \to \nu[-\infty, -x) \quad n \to \infty, \tag{1.2}$$

and we get $\nu[-\infty,-x) = qx^{-\alpha},\ q \geq 0$.

The parameter α for the right tail must be the same as the left tail because the same b_n successfully scales both tails and b_n relates to α by $b_n = b(n)$, where $b(\cdot) \in RV_{1/\alpha}$. Therefore,

$$\nu(dx) = \big(p\alpha x^{-\alpha-1}I(x>0) + q\alpha|x|^{-\alpha-1}I(x<0)\big)\,dx. \tag{1.3}$$

Sometimes (1.1) and (1.2) are written as

$$P[|Z_1| > x] \in RV_{-\alpha},$$

together with

$$\lim_{x\to\infty}\frac{P[Z_1 > x]}{P[|Z_1| > x]} = p, \quad \lim_{x\to\infty}\frac{P[Z_1 \leq -x]}{P[|Z_1| > x]} = 1 - p =: q, \quad 0 \leq p \leq 1.$$

The measure ν in (1.3) is a Lévy measure (see Definition 1.4).

1.3 Lévy Processes

In this section we set $\mathbb{E} = \mathbb{R}^d \setminus \{\mathbf{0}\}$ and denote the Euclidean metric by $\|\mathbf{x}\|$, $\mathbf{x} \in \mathbb{E}$.

Definition 1.4 (Lévy measure). A measure ν satisfying the following is called Lévy measure on $\mathbb{E}$.

(i) For every $x > 0$, $\nu\{\mathbf{u} \in \mathbb{E} : \|\mathbf{u}\| > x\} < \infty$.

(ii) $\int_{0<\|\mathbf{x}\|\leq 1} \|\mathbf{x}\|^2 \nu(d\mathbf{x}) < \infty$.

Lévy pure jump process

Let $\mathbf{j}_k \in \mathbb{E}$, $k \geq 1$, be the points of a Poisson point process with Lévy mean measure ν. Fix t and let $I \subset \mathbb{E}$ be a set bounded away from $\mathbf{0}$. Define

$$\mathbf{S}_I(t) = \sum_{\substack{t_k \leq t \\ \mathbf{j}_k \in I}} \mathbf{j}_k = \iint_{[0,t]\times I} \mathbf{u}\, N(ds, d\mathbf{u}),$$

where N is a PRM$(\lambda \times \nu)$, i.e. $N = \sum_k \epsilon_{(t_k, \mathbf{j}_k)}$, and λ denotes the Lebesgue measure. Suppose we have a sequence $\varepsilon_n \downarrow 0$ such that $1 = \varepsilon_0 > \varepsilon_1 > \dots$. Set

$$I_{j+1} := \{\mathbf{x} \in \mathbb{E} : \varepsilon_{j+1} < \|\mathbf{x}\| \leq \varepsilon_j\}, \quad j = 0, 1, 2, \dots,$$

and define the process

$$\begin{aligned} \mathbf{X}_{j+1}(t) &:= \mathbf{S}_{I_{j+1}}(t) - E(\mathbf{S}_{I_{j+1}}(t)) \\ &= \iint_{\substack{0 \leq s \leq t \\ \mathbf{u} \in I_{j+1}}} \mathbf{u}\, N(ds, d\mathbf{u}) - t \int_{\mathbf{u} \in I_{j+1}} \mathbf{u}\, \nu(d\mathbf{u}). \end{aligned}$$

Also, define the finite sum

$$\mathbf{X}_0(t) := \sum_{t_k \leq t} \mathbf{j}_k I(\|\mathbf{j}_k\| > 1) = \iint_{[0,t]\times\{\mathbf{x}:\|\mathbf{x}\|>1\}} \mathbf{u}\, N(ds, d\mathbf{u}).$$

The following convergent series is called *Lévy Pure Jump Process* with Lévy measure ν:

$$\mathbf{X}(t) = \mathbf{X}_0(t) + \sum_{j=0}^{\infty} \mathbf{X}_{j+1}(t).$$

The equivalent representation which is called *Itô representation* is given by

$$\begin{aligned} \mathbf{X}(t) = &\iint_{\substack{s \leq t \\ \|\mathbf{u}\|>1}} \mathbf{u}\, N(ds, d\mathbf{u}) \\ &+ \lim_{\varepsilon \downarrow 0} \left[\iint_{\substack{s \leq t \\ \|\mathbf{u}\| \in (\varepsilon, 1]}} \mathbf{u}\, N(ds, d\mathbf{u}) - \iint_{\substack{s \leq t \\ \|\mathbf{u}\| \in (\varepsilon, 1]}} \mathbf{u}\, ds\, \nu(d\mathbf{u}) \right]. \end{aligned}$$

Basic properties

It can be verified that the Lévy process has stationary independent increments. The process is also stochastic continuous; that is, if $t_n \to t$, then $\mathbf{X}(t_n) \to_p \mathbf{X}(t)$.

Recall the Lévy measure in (1.3) assuming $0 < \alpha < 2, 0 \leq p \leq 1$, and $q = 1 - p$. A Lévy process with this Lévy measure is called *stable Lévy motion.* A special case with $p = q = 1/2$ and a symmetric Lévy measure v_α is called *symmetric α-stable motion.* We study stable random variables in more detail in the next section. A very fundamental result about weak convergence of partial sum processes to Lévy pure jump processes is given here for d-dimensional iid random vectors. See [50] for its proof.

Theorem 1.6. Let $\{\mathbf{X}_{n,j} = (X_{n,j}^{(1)}, ..., X_{n,j}^{(d)}),\ j \geq 1\}$ for each $n \geq 1$ be iid random vectors such that

$$nP\Big[\mathbf{X}_{n,1} \in \cdot\Big] \to_v \nu(\cdot), \quad n \to \infty$$

in $M_+(\mathbb{E})$, where ν is a Lévy measure. Moreover, suppose that for each $j = 1, ..., d$,

$$\lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} nE\left((X_{n,1}^{(j)})^2 I(|X_{n,1}^{(j)}| \leq \varepsilon)\right) = 0.$$

Define the partial sum process based on the nth row of the array by

$$\mathbf{X}_n(t) := \sum_{k=1}^{[nt]} \left(\mathbf{X}_{n,k} - E(\mathbf{X}_{n,k} I(\|\mathbf{X}_{n,k}\| \leq 1))\right), \quad t \geq 0.$$

Then $\mathbf{X}_n \to_d \mathbf{X}$ in $D([0,\infty), \mathbb{R}^d)$, where the limiting process $\mathbf{X}(\cdot)$ is a Lévy pure jump process with Lévy measure ν.

1.4 Stable Processes

The high variability of the stable distributions make them an effective way to model diverse phenomena with high variability. We start with the definition of a stable distribution.

Definition 1.5. A random variable X is said to have a stable distribution if for any $n \geq 2$ and $X_1, X_2, ..., X_n$ independent copies of X, there is a positive number a_n and a real number b_n such that

$$X_1 + X_2 + \cdots + X_n \stackrel{d}{=} a_n X + b_n, \tag{1.4}$$

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

Such a random variable is called *strictly* stable if (1.4) holds with $b_n = 0$ and is called *symmetric* if its distribution is symmetric, i.e. $X \stackrel{d}{=} -X$, and is usually denoted by $S\alpha S$. It can also be shown [28] that $a_n = n^{1/\alpha}$ for some $0 < \alpha \leq 2$. The following equivalent definition states that the stable random variables are the only random variables having *domain of attraction.*

Definition 1.6 (Equivalent definition). A random variable X is said to have a stable distribution if there is a sequence of iid random variables $X_1, X_2, ...$ and sequences of positive numbers $\{a_n\}$ and real numbers $\{b_n\}$ such that

$$a_n^{-1} \sum_{i=1}^{n} (X_i - b_n) \rightarrow_d X. \tag{1.5}$$

Notice that the above definition is the statement of central limit theory when X is Gaussian and X_i's are iid with finite variance. Moreover, following [28], it is known that $a_n = n^{1/\alpha}L(n)$ for a slowly varying function $L(\cdot)$. When $a_n = n^{1/\alpha}$, the X_i's are said to belong to the *normal* domain of attraction of X. Also a_n and b_n in (1.5) can be defined by

$$a_n = \inf\{x : P(|X| > x) \leq 1/n\}, \quad b_n = E\left(XI(|X| \leq a_n)\right).$$

A stable random variable X has the following characteristic function:

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

where $0 < \alpha \leq 2$, $\sigma \geq 0$, $-1 \leq \beta \leq 1$, and μ is a real number. The parameter α is called the index of stability. The other unique parameters σ, β and μ are scale, skewness and shift parameters, respectively. The shift parameter equals the mean when $1 < \alpha \leq 2$. Denoting the corresponding stable distribution by $S_\alpha(\sigma, \beta, \mu)$, it can be revealed that the probability density of X exists and is continuous. The fact that the density of a α-stable random variable cannot be found in closed form, with a few exceptions, makes the need for the use of characteristic function clear. 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.

It is easy to verify that $X \sim S_\alpha(\sigma, \beta, \mu)$ is symmetric iff $\beta = \mu = 0$. It can also be seen that $X \sim S_\alpha(\sigma, \beta, \mu)$ with $\alpha \neq 1$ is strictly stable iff $\mu = 0$, and $X \sim S_1(\sigma, \beta, \mu)$ is strictly stable iff $\beta = 0$. Since μ reflects only location, it can be taken equal to 0. It is also worth noting that in some applications the α-stable random variable can be chosen to be totally skewed to the right, i.e. $\beta = 1$, because of the following property:

Let $X \sim S_\alpha(\sigma, \beta, 0)$ with $0 < \alpha < 2$. Then there exist two i.i.d. random variables

$Y_1, Y_2 \sim S_\alpha(\sigma, 1, 0)$ such that

$$X \stackrel{d}{=} \left(\frac{1+\beta}{2}\right)^{1/\alpha} Y_1 - \left(\frac{1-\beta}{2}\right)^{1/\alpha} Y_2 + \text{const.}$$

A very important property of stable random variables which makes many of the suitable techniques applicable for the Gaussian case invalid is the fact that if $X \sim S_\alpha(\sigma, \beta, \mu)$, then $E|X|^\delta = \infty$ for any $\delta \geq \alpha$, $0 < \alpha < 2$.

It sometimes, for instance in the dissertation, suffices to work with SαS random variables. A random variable X is SαS iff $X \sim S_\alpha(\sigma, 0, 0)$, with characteristic function $Ee^{i\theta X} = \exp\{-\sigma^\alpha|\theta|^\alpha\}$. Here is a necessary and sufficient condition for a sequence of random variables to be in the domain of attraction of a SαS random variable.

Corollary 1.2 ([49, 50]). Consider the special case where $\{X_n, n \geq 1\}$ are iid random variables on $\mathbb{R}$ and set $X_{n,i} = X_i/a_n$, where a_n is defined as in (1.5). Define ν to be the Lévy measure in (1.3) with $0 < \alpha < 2$. Then

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

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

$$nP\Big[a_n^{-1}X_1 \in \cdot\Big] \to_v \nu(\cdot),$$

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

For other extensive expositions on stable random variables, processes, multivariate extensions and related stochastic integrals which have appeared in the following chapters, consult [51].

In what follows, in Chapter 2 we consider a unit root vector autoregressive model with innovations in the domain of attraction of a d-variate $(\alpha_1, ..., \alpha_d)$-stable distribution with possibly different indexes of stability, and in Chapters 3 and 4 we investigate several stationary and unit root spatial autoregressive models with innovations in the domain of attraction of a SαS distribution. We focus on limiting behavior of the M-estimates of the parameters. We shall emphasize that our approaches cover models with both infinite and finite variance innovations. Although the least absolute deviation (LAD) estimates may be considered as a special case of M-estimates, to prove the related theorems it is necessary to apply different techniques. However, the proofs are not given in this thesis because we can modify, for example, the proof of Theorem 3 of [36] to get the desired result. For the same reason we do not consider any intercept term or any scale parameter in the underlying models.

Chapter 2

Multivariate Autoregression of Order One with Infinite Variance Innovations

2.1 Introduction

Consider the vector autoregressive model of order one (VAR(1))

$$\mathbf{U}_t = \mathbf{\Phi}\,\mathbf{U}_{t-1} + \boldsymbol{\epsilon}_t, \quad t = 1, 2, \ldots, n, \tag{2.1}$$

where $\{\mathbf{U}_t\}$ and $\{\boldsymbol{\epsilon}_t\}$ are sequences of multivariate random vectors in $\mathbb{R}^d$ and $\mathbf{U}_0 = \mathbf{0}$. In this chapter, we assume that $\mathbf{U}_t = \mathbf{\Phi}\,\mathbf{U}_{t-1} + \boldsymbol{\epsilon}_t$, where actually $\mathbf{\Phi} = \mathbf{I}_d$ (homogeneous unit root model, in the sense of Caner [14]).

For $d = 1$, in the case that the errors (innovations) have infinite variance, several studies have been developed. Chan and Tran [18] and Chan [16] considered the unit root tests for the AR(1) models with independent and identically distributed (i.i.d.) innovations. They obtained the asymptotic distribution of the OLS estimates whereas Knight [36] suggested the M-estimates under the same model. Phillips [47] studied the OLS estimates under weakly dependent innovations with finite variance, and Knight [37] took into account the M-estimates in the case that the innovations come from moving average sequences of order infinity with infinite variance terms. For a general autoregressive (AR) process with infinite variance innovations when a stationary solution exists, the M-estimates have been considered by Davis, Knight, and Liu [24]. They pointed out that the M-estimators, similar to their counterpart the least absolute deviation (LAD) estimators, give less weight to the outliers when the distribution of innovations belongs to the class of heavy-tailed distributions. These estimators essentially provide a faster rate of convergence in comparison with the other usual estimators. They also investigated consistency properties of the M-estimators corresponding to a general class of loss functions. For AR processes with one unit root and non-normal innovations having finite variance, Cox and Llatas [20] and Lucas [40] developed M-estimation under i.i.d. and weakly dependent errors, respectively.

Caner [14] extends the limit theory for the OLS estimation of a VAR(1) for a random walk model with i.i.d. errors with independent components in the domain of attraction of a stable law with the same index of stability α, $0 < \alpha < 2$ and concludes that the OLS estimators are consistent, and asymptotic distributions are derived. The asymptotic theory for a bivariate time series regression (cointegrated model) with the innovation vector of different heavy-tailed behavior using the OLS estimators has been considered by Mittnik, Paulauskas, and Rachev [44].

In this chapter, we develop the limiting theory for a multivariate random walk model with innovations in the domain of attraction of a multivariate stable law having *different* indices of stability. In our model, we suppose that the innovation vector has *dependent* components which is somehow more practical in comparison with Caner [14]. We also look into asymptotic properties of the OLS and the M-estimates. It is observed that in some cases the OLS estimates are inconsistent, whereas the M-estimates are always consistent with a higher rate of convergence in comparison with consistent OLS estimates. This confirms the superiority of the M-estimators, as has been noted by Davis et al. [24] for AR processes with the usual stationarity constraints, which differs from our case. It seems likely that the findings of this chapter can be extended to weakly dependent errors such as MA(∞), but we do not address this question here.

2.2 Preliminaries

Consider Model (2.1). The OLS estimator for $\mathbf{\Phi}$ in this model is given by

$$\widehat{\mathbf{\Phi}} = \left(\sum_{t=1}^{n} \mathbf{U}_t \mathbf{U}_{t-1}^{'}\right) \left(\sum_{t=1}^{n} \mathbf{U}_{t-1} \mathbf{U}_{t-1}^{'}\right)^{-1}. \tag{2.2}$$

Throughout the chapter we denote the transpose of the matrix A by $A^{'}$.

Define the stochastic processes

$$\mathbf{S}_n^{'}(\cdot) = \left(S_n^{(1)}(\cdot), S_n^{(2)}(\cdot), \ldots, S_n^{(d)}(\cdot)\right), \quad S_n^{(j)}(\cdot) = \left(a_n^{(j)}\right)^{-1} \sum_{i=1}^{[n\cdot]} \epsilon_i^{(j)},$$

$$\mathbf{V}_n'(\cdot) = \left(V_n^{(1)}(\cdot), V_n^{(2)}(\cdot), \ldots, V_n^{(d)}(\cdot)\right), \quad V_n^{(j)}(\cdot) = \left(a_n^{(j)}\right)^{-2} \sum_{i=1}^{[n\cdot]} \left(\epsilon_i^{(j)}\right)^2, \quad (2.3)$$

for $j = 1, \ldots, d$, where $[x]$ stands for integer part of x. Let $D = D[0,1]$ be the Skorohod space of càdlàg functions defined on $[0,1]$ and equipped with the Skorohod metric. Furthermore, let $D_d = D([0,1], \mathbb{R}^d)$ indicate the corresponding Skorohod space of $\mathbb{R}^d$-valued càdlàg functions and $D^d = D[0,1] \times \cdots \times D[0,1]$ the product of d topological spaces with the product topology. Throughout the chapter, integrals involving $dS_n^{(j)}$, dS_j or dW_j, $j = 1, \ldots, d$, are interpreted as Itô stochastic integrals. Moreover, we make the following assumption.

A1: $\{\boldsymbol{\epsilon}_t\}$ is a sequence of i.i.d. random vectors such that

$$\mathbf{S}_n'(\cdot) \to_d \mathbf{S}'(\cdot) := (S_1(\cdot), \ldots, S_d(\cdot)), \quad (2.4)$$

in D^d, where $\mathbf{S}_n'(\cdot)$ is defined as in (2.3). Following Resnick and Greenwood [48], we call $\mathbf{S}'(\cdot)$ a stable process.

When A1 holds, we say that the distribution of $\boldsymbol{\epsilon}_t$ belongs to the domain of attraction of a multivariate stable law with index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)$, $(DS(\alpha_1, \ldots, \alpha_d))$, where $\alpha_j \in (0, 2]$, for $j = 1, \ldots, d$. Assumption A1 also implies that

$$\mathbf{V}_n'(\cdot) \to_d \mathbf{V}'(\cdot) := (V_1(\cdot), \ldots, V_d(\cdot)) \quad (2.5)$$

in D^d, where $\mathbf{V}_n'(\cdot)$ is defined as in (2.3), and $\mathbf{V}'(\cdot)$ is a stable process in D^d. As Paulauskas and Rachev [45] demonstrate, we can generalize the convergence in (2.4) and (2.5) to weak convergence in D_d. Note that for $1 < \alpha_j \leq 2$, $j = 1, \ldots, d$, (2.4) does not hold unless the coordinates of $\{\boldsymbol{\epsilon}_t\}$ are in the domain of attraction of a symmetric stable law.

If $0 < \alpha_j < 2$, $j = 1, \ldots, d$, then $\mathbf{S}(\cdot)$, the multivariate stable process, is an infinitely divisible process with the Lévy measure ν on $\mathbb{R}^d \setminus \{\mathbf{0}\}$ defined by $\nu \circ \tau = \tilde{\nu}$, where

$$\tau \mathbf{x} = \left((\operatorname{sign} x_1) \, |x_1|^{1/\alpha_1}, \ldots, (\operatorname{sign} x_d) \, |x_d|^{1/\alpha_d} \right),$$

with $\tilde{\nu}$ given by

$$\tilde{\nu}\{\mathbf{x} : |x| > r, \, \theta(x) \in H\} = r^{-1} S(H),$$

and S is a finite measure on $[0, 2\pi]$. The limiting stable processes in (2.4) and (2.5) can be represented by the following series:

$$\mathbf{S}(\cdot) = \sum_{i=1}^{\infty} \mathbf{j}_i I(U_i \leq \cdot)$$

and

$$\mathbf{V}(\cdot) = \sum_{i=1}^{\infty} \mathbf{j}_i^2 I(U_i \leq \cdot),$$

with $\mathbf{j}_i = \left(j_i^{(1)}, \ldots, j_i^{(d)} \right)'$ and $\mathbf{j}_i^2 = \left(\left(j_i^{(1)} \right)^2, \ldots, \left(j_i^{(d)} \right)^2 \right)'$, where $\mathbf{j}_i \in \mathbb{R}^d \setminus \{\mathbf{0}\}$, $i \geq 1$, are the points of the corresponding Poisson point process with mean measure $\nu(\cdot)$, and are independent of the i.i.d. sequence $\{U_i\}$ of uniformly distributed random variables on $[0, 1]$. For more details see Resnick [49, 50].

It is well known that $a_n^{(j)} = n^{1/\alpha_j} L_j(n)$, $j = 1, \ldots, d$, where $L_j(\cdot)$'s are slowly varying functions at infinity (see Feller [28]). When $\alpha_j = 2$, $j = 1, \ldots, d$, $S_j(\cdot)$ is Brownian motion and $V_j(t) = t$. For example, if $d = 2$ and $\alpha_1 = 2$ and $\alpha_2 < 2$, then $S_1(\cdot)$ (Brownian motion) and $S_2(\cdot)$ (stable process) are independent (see Resnick and Greenwood [48]). If $0 < \alpha_1 = \cdots = \alpha_d = \alpha < 2$, then $\mathbf{S}(t) = \sum_{i=1}^{\infty} \mathbf{C}_i \Gamma_i^{-1/\alpha} I(U_i \leq t)$, where $\mathbf{C}_i$'s are i.i.d. random vectors on the boundary of the unit sphere. Here $\{\Gamma_i\}$ are the arrival times of Poisson processes with Lebesgue mean measure independent of $\{\mathbf{C}_i\}$. Also $\{U_i\}$ is an i.i.d. sequence of random variables with uniform distribution in $[0,1]$ and $\{U_i, \Gamma_i, \mathbf{C}_i\}$ are independent. See LePage [39] for more details.

The classical M-estimator $\widehat{\mathbf{\Phi}}_M$ of $\mathbf{\Phi}$ in (2.1) is the minimizer of $\sum_{t=2}^{n} \rho(\mathbf{U}_t - \mathbf{\Phi}\,\mathbf{U}_{t-1})$ for some convex and almost everywhere differentiable function $\rho(\cdot)$, which guarantees the uniqueness of the solutions. The M-estimates are usually utilized as robust estimates and have a faster rate of convergence than the OLS estimates in the presence of infinite variance innovations. To explain this, for the univariate case $(d=1)$ of Model (2.1), from Knight [36] we know that

$$a_n\sqrt{n}(\hat{\phi}_M - 1) \rightarrow_d \frac{\sqrt{E\psi^2(\epsilon_1)}\int_0^1 S(s)\,dW(s)}{E\psi'(\epsilon_1)\int_0^1 S^2(s)\,ds},$$

where ψ is the derivative of ρ and $W(\cdot)$ is the standard Brownian motion, whereas for the OLS estimator $\hat{\phi}$ of ϕ we have (Chan and Tran [18])

$$n(\hat{\phi} - 1) \rightarrow_d \frac{S^2(1) - V(1)}{2\int_0^1 S^2(s)\,ds}.$$

Therefore, the fact that $\sqrt{n}/a_n \rightarrow 0$, shows clearly that the M-estimators have a better rate of convergence. It is interesting to notice that in some cases the M-estimators also have a better rate of convergence than the OLS estimators in the stationary model with $\phi < 1$. In fact, we have

$$a_n(\hat{\phi}_M - \phi) \rightarrow_d X_1,$$

and

$$n^{1/\alpha}L(n)(\hat{\phi} - \phi) \rightarrow_d X_2,$$

where L is a slowly varying function at infinity for some r.v.'s X_1 and X_2. When the innovations have stable or Pareto-like distributions, we have $n^{1/\alpha}L(n)/a_n \rightarrow 0$ (for more details, see Davis et al. [24], and Brockwell and Davis [11]).

Usually, $\rho(x)$ grows at a slower rate than x^2 as $|x|$ gets large. An example for $\rho(\cdot)$ is the Huber's loss function (Huber [31]) given by

$$\rho_H(x) = \begin{cases} \frac{1}{2}x^2 & \text{if} \quad |x| \le c, \\ c|x| - \frac{1}{2}c^2 & \text{if} \quad |x| > c, \end{cases} \tag{2.6}$$

for a known constant c.

We also impose the following assumptions on the function $\rho(\cdot)$, which are introduced for technical simplicity in the proof of Theorem 2.2 in the next section.

A2: $\rho(\cdot)$ is a convex and twice differentiable function from $\mathbb{R}^d$ to $\mathbb{R}$,

A3: $E\left(\rho_i^2(\epsilon_1^{(1)}, \ldots, \epsilon_1^{(d)})\right) < \infty$ and $E\left(\rho_i(\epsilon_1^{(1)}, \ldots, \epsilon_1^{(d)})\right) = 0$, where $\rho_i(x_1, \ldots, x_d) = \partial\rho(x_1, \ldots, x_d)/\partial x_i,\ i = 1, \ldots, d,$

A4: $\left|E\left(\rho_{ij}(\epsilon_1^{(1)}, \ldots, \epsilon_1^{(d)})\right)\right| < \infty$, where

$$\rho_{ij}(x_1, \ldots, x_d) = \partial^2\rho(x_1, \ldots, x_d)/\partial x_i \partial x_j \quad i, j = 1, \ldots, d,$$

and $\rho_{ij}(\cdot)$ satisfies the condition

$$|\rho_{ij}(x_1, \ldots, x_d) - \rho_{ij}(y_1, \ldots, y_d)| \le \sum_{k=1}^{d} c_k |x_k - y_k|,$$

for some nonnegative constants c_k and all x_k and y_k, $k = 1, \ldots, d$.

Example. Assumptions A2-A4 hold if we use $\rho(x_1, \ldots, x_d) = a_1 g_1(x_1) + \cdots + a_d g_d(x_d)$, where $a_i > 0$, $i = 1, \ldots, d$, are known constants and the g_i's fulfill conditions A3-A6 of Knight [36]. Convexity of $\rho(\cdot)$ follows from the positive definiteness of the diagonal Hessian matrix

$$H = \text{diag}\left(a_1 \partial^2 g_1(x_1)/\partial x_1^2, \ldots, a_d \partial^2 g_d(x_d)/\partial x_d^2\right).$$

Moreover, $\rho(\cdot)$ inherits the other assumptions from conditions A3-A6 of Knight. The only restrictive Assumption, A4, can be weakened if we let the function $\rho_{ij}(\cdot)$ be discontinuous at a countable number of points (Knight [36], page 266). For illustration, we consider the bivariate Huber's function

$$\rho^*(x, y) = \rho_H(x) + \rho_H(y) \tag{2.7}$$

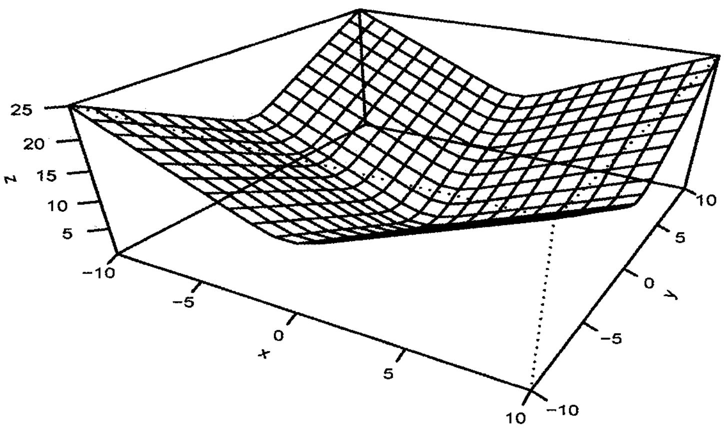
and

$$\rho^{**}(x, y) = \rho_T(x) + \rho_T(y), \tag{2.8}$$

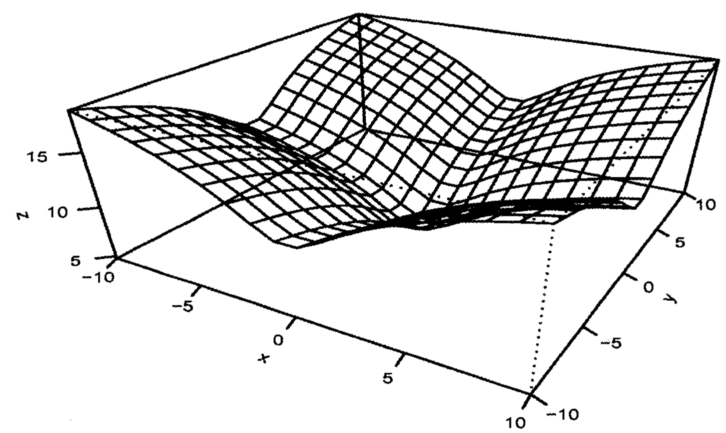
where $\rho_H(\cdot)$ is Huber's loss function defined in (2.6) and $\rho_T(x) = 2\,ln(3 + x^2)$. The function $(d/dx)\rho_T(x) = 4x/(3 + x^2)$ is the score function of a t-distribution with 3 degrees of freedom. The loss function $\rho^*(\cdot)$ satisfies conditions A2-A4 if we ignore the discontinuity of Huber's loss function at $\pm c$. The sufficiently smooth loss function $\rho^{**}(\cdot)$ does not satisfy A4 and is just locally convex around its global minimum. However, this suffices to search out the unique solutions (the M-estimates), and this illustrates that Assumption A2 might be too strong (see Davis et al. [24], Remark 1 and 2). Actually, the performance of $\rho^{**}(\cdot)$ is as good as that of $\rho^*(\cdot)$, as shown in Section 2.4. Notice that we require just one such function $\rho(\cdot)$ satisfying the conditions A2-A4. Figure 2.1 illustrates the behavior of both loss functions.

The following proposition, which is in fact Proposition 2 of Paulauskas and Rachev [45], is used in the next section. The proof is based on the results on the convergence of stochastic integrals for semimartingales (see also Kurtz and Protter [38], Theorem 2.7).

Proposition 2.1. Suppose that (2.4) holds. Then for $0 \le t \le 1$, as $n \to \infty$,



(a)



(b)

Figure 2.1: (a) Bivariate Huber's loss function; (b) Bivariate loss function given in (2.8).

$$\left(\mathbf{S}_n'(t), \int_0^t S_n^{(i)}(s)\, dS_n^{(j)}(s),\ i,j = 1,\dots,d\right) \to_d \left(\mathbf{S}'(t), \int_0^t S_i(s)\, dS_j(s),\ i,j = 1,\dots,d\right)$$

in D_s, where $s = d(d+1)$.

Proof. See page 782 of Paulauskas and Rachev [45] and the related discussions therein and Proposition 2 of Mittnik et al. [44].

Finally, we present the following Lemma from Knight [36], page 276, which will be used to prove Theorem 2.2 in Section 2.3 (see also Davis et al. [24], Lemma 2.2).

Lemma 2.1. Suppose that $\{T_n(\cdot)\}$ is a sequence of convex stochastic processes from $\mathbb{R}^d$ to $\mathbb{R}$ and that for any k-tuple of vectors $(u_1, \dots, u_k)$,

$$(T_n(u_1), \dots, T_n(u_k)) \to_d (T(u_1), \dots, T(u_k)),$$

where the stochastic process $T(\cdot)$ has a unique minimum at $\hat{u}$. If $\hat{u}_n$ minimizes $T_n(\cdot)$, then $\hat{u}_n \to_d \hat{u}$.

2.3 OLS and M-estimates for multivariate random walks

In the first theorem in this section, for the Model (2.1) we find the weak limit behavior of $\mathbf{A}_n \times (\widehat{\mathbf{\Phi}} - \mathbf{I})$ for some sequence of matrices $\{\mathbf{A}_n\}$, where $\times$ stands for the coordinate-wise product. The following Theorem deals with inconsistency of the OLS estimators and is proven in the bivariate case; $d = 2$ (see also Remark 2.2 in this section).

Theorem 2.1. Consider the Model (2.1) with $\mathbf{\Phi} = \mathbf{I}_2$ and suppose that A1 holds. If $\mathbf{A}_n = \begin{pmatrix} n & na_n^{(2)}/a_n^{(1)} \\ na_n^{(1)}/a_n^{(2)} & n \end{pmatrix}$, then

$$\mathbf{A}_n \times (\widehat{\mathbf{\Phi}} - \mathbf{I}) \to_d \mathbf{L}'\mathbf{K}^{-1},$$

where $\widehat{\mathbf{\Phi}}$ is the OLS estimator given in (2.2), $\mathbf{L} = \int_0^1 \mathbf{S}\, d\mathbf{S}'$, $\mathbf{K} = \int_0^1 \mathbf{S}(s)\mathbf{S}'(s)\, ds$, and $\mathbf{S}(\cdot)$ is the stable random vector defined in (2.4).

Remark 2.1. Theorem 2.1 establishes that if $a_n^{(1)} = a_n^{(2)}$, then the OLS estimates are consistent and the rate of convergence is n. This coincides with the results presented in Caner [14]. On the other hand, if $a_n^{(1)} \neq a_n^{(2)}$, then one of the OLS estimates might be inconsistent. More precisely, let $\widehat{\mathbf{\Phi}} = \begin{pmatrix} \hat{\alpha} & \hat{\beta} \\ \hat{\gamma} & \hat{\theta} \end{pmatrix}$ be the OLS estimator of $\mathbf{\Phi} = \mathbf{I}_2$ in Model (2.1). If $\alpha_1 < \alpha_2(1+\alpha_2)^{-1}$, $\hat{\beta}$ is not consistent. A similar conclusion can be made for $\hat{\gamma}$. For example if $\alpha_2 = 2$, for consistency of estimates we need to have $\alpha_1 \geq 2/3$. Note that if $\alpha_1, \alpha_2 \geq 1$, the OLS estimates remain consistent.

Proof. Let $\widehat{\mathbf{\Phi}} = \begin{pmatrix} \hat{\alpha} & \hat{\beta} \\ \hat{\gamma} & \hat{\theta} \end{pmatrix}$ be the OLS estimator of $\mathbf{\Phi} = \mathbf{I}_2$ in Model (2.1). From (2.2), assuming that $\mathbf{U}_t = (X_t, Y_t)'$ and $\boldsymbol{\epsilon}_t = (\epsilon_t^{(1)}, \epsilon_t^{(2)})'$, we derive the following representations:

$$\hat{\alpha} - 1 = \frac{\sum_{t=2}^n Y_{t-1}^2 \sum_{t=2}^n X_{t-1}\epsilon_t^{(1)} - \sum_{t=2}^n X_{t-1}Y_{t-1} \sum_{t=2}^n Y_{t-1}\epsilon_t^{(1)}}{\sum_{t=2}^n X_{t-1}^2 \sum_{t=2}^n Y_{t-1}^2 - \left(\sum_{t=2}^n X_{t-1}Y_{t-1}\right)^2}, \tag{2.9}$$

$$\hat{\beta} = \frac{\sum_{t=2}^n X_{t-1}^2 \sum_{t=2}^n Y_{t-1}\epsilon_t^{(1)} - \sum_{t=2}^n X_{t-1}Y_{t-1} \sum_{t=2}^n X_{t-1}\epsilon_t^{(1)}}{\sum_{t=2}^n X_{t-1}^2 \sum_{t=2}^n Y_{t-1}^2 - \left(\sum_{t=2}^n X_{t-1}Y_{t-1}\right)^2}, \tag{2.10}$$

$$\hat{\gamma} = \frac{\sum_{t=2}^n Y_{t-1}^2 \sum_{t=2}^n X_{t-1}\epsilon_t^{(2)} - \sum_{t=2}^n X_{t-1}Y_{t-1} \sum_{t=2}^n Y_{t-1}\epsilon_t^{(2)}}{\sum_{t=2}^n X_{t-1}^2 \sum_{t=2}^n Y_{t-1}^2 - \left(\sum_{t=2}^n X_{t-1}Y_{t-1}\right)^2}, \tag{2.11}$$

$$\hat{\theta} - 1 = \frac{\sum_{t=2}^n X_{t-1}^2 \sum_{t=2}^n Y_{t-1}\epsilon_t^{(2)} - \sum_{t=2}^n X_{t-1}Y_{t-1} \sum_{t=2}^n X_{t-1}\epsilon_t^{(2)}}{\sum_{t=2}^n X_{t-1}^2 \sum_{t=2}^n Y_{t-1}^2 - \left(\sum_{t=2}^n X_{t-1}Y_{t-1}\right)^2}. \tag{2.12}$$

Next, we obtain the normalization coefficients for the sums in (2.9)-(2.12). In fact, we have

$$\left(a_n^{(1)}\right)^{-2}\sum_{t=2}^{n} X_{t-1}\epsilon_t^{(1)} = 1/2\left[\left(S_n^{(1)}(1)\right)^2 - V_n^{(1)}(1)\right] = \int_0^1 S_n^{(1)}(s)\,dS_n^{(1)}(s), \tag{2.13}$$

$$\left(a_n^{(1)}\right)^{-1}\left(a_n^{(2)}\right)^{-1}\sum_{t=2}^{n} Y_{t-1}\epsilon_t^{(1)} = \int_0^1 S_n^{(2)}(s)\,dS_n^{(1)}(s), \tag{2.14}$$

$$n^{-1}\left(a_n^{(2)}\right)^{-2}\sum_{t=2}^{n} Y_{t-1}^2 = \int_0^1 \left(S_n^{(2)}\right)^2(s)\,ds, \tag{2.15}$$

$$n^{-1}\left(a_n^{(1)}\right)^{-1}\left(a_n^{(2)}\right)^{-1}\sum_{t=2}^{n} X_{t-1}Y_{t-1} = \int_0^1 S_n^{(1)}(s)S_n^{(2)}(s)\,ds, \tag{2.16}$$

$$n^{-1}\left(a_n^{(1)}\right)^{-2}\sum_{t=2}^{n} X_{t-1}^2 = \int_0^1 \left(S_n^{(1)}\right)^2(s)\,ds, \tag{2.17}$$

$$\left(a_n^{(1)}\right)^{-1}\left(a_n^{(2)}\right)^{-1}\sum_{t=2}^{n} X_{t-1}\epsilon_t^{(2)} = \int_0^1 S_n^{(1)}(s)\,dS_n^{(2)}(s), \tag{2.18}$$

and

$$\left(a_n^{(2)}\right)^{-2}\sum_{t=2}^{n} Y_{t-1}\epsilon_t^{(2)} = \int_0^1 S_n^{(2)}(s)\,dS_n^{(2)}(s). \tag{2.19}$$

Therefore, using (2.9)-(2.19), we observe that the normalization factors for $\hat{\alpha}-1$, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\theta}-1$ are n, $na_n^{(2)}/a_n^{(1)}$, $na_n^{(1)}/a_n^{(2)}$, and n, respectively.

Thus, we can write all the elements of $\mathbf{A}_n \times (\widehat{\boldsymbol{\Phi}} - \mathbf{I}_2)$ as a function of the bivariate stable process $\mathbf{S}_n$. The rest of the proof is based on Proposition 2.1. As in Mittnik et al. [44], we define a map $f : D_6 \longrightarrow \mathbb{R}^4$, $f = (f_1, f_2, f_3, f_4)$ with coordinates $f_i(\mathbf{x}) = g_i(\mathbf{x})\,(g_5(\mathbf{x}))^{-1}$, $i = 1, \ldots, 4$. Here, $\mathbf{x} = (x_1, \ldots, x_6) \in D_6$, and the functions $g_i : D_6 \longrightarrow \mathbb{R}$, $i = 1, \ldots, 5$, are defined as follows:

$$g_1(\mathbf{x}) := x_3(1)\left(\int_0^1 x_2^2(s)\,ds\right) - x_4(1)\left(\int_0^1 x_1(s)x_2(s)\,ds\right),$$

$$g_2(\mathbf{x}) := x_4(1)\left(\int_0^1 x_1^2(s)\,ds\right) - x_3(1)\left(\int_0^1 x_1(s)x_2(s)\,ds\right),$$

$$g_3(\mathbf{x}) := x_5(1)\left(\int_0^1 x_2^2(s)\,ds\right) - x_6(1)\left(\int_0^1 x_1(s)x_2(s)\,ds\right),$$

$$g_4(\mathbf{x}) := x_6(1)\left(\int_0^1 x_1^2(s)\,ds\right) - x_5(1)\left(\int_0^1 x_1(s)x_2(s)\,ds\right),$$

and

$$g_5(\mathbf{x}) := \left(\int_0^1 x_1^2(s)\,ds\right)\left(\int_0^1 x_2^2(s)\,ds\right) - \left(\int_0^1 x_1(s)x_2(s)\,ds\right)^2.$$

If μ denotes the distribution of the limiting vector $\mathbf{S}(\cdot)$ and D_f the set of discontinuity points of f, it is easy to see that $\mu(D_f) = 0$. So, we can apply the continuous mapping theorem for f using Proposition 2.1 (with $d = 2$) to complete the proof.

Remark 2.2. Theorem 2.1 also extends the result of Caner [14] in two directions. First, Theorem 2.1 can be applied to a d-variate AR model of order one when the innovations are i.i.d. vectors in the domain of attraction of a stable law with possibly different indices between 0 and 2. In fact, the normalizing matrix in Theorem 2.1 is modified as follows:

$$\mathbf{A}_n = n\begin{pmatrix} 1 & a_n^{(2)}/a_n^{(1)} & \dots & a_n^{(d)}/a_n^{(1)} \\ a_n^{(1)}/a_n^{(2)} & 1 & \dots & a_n^{(d)}/a_n^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ a_n^{(1)}/a_n^{(d)} & a_n^{(2)}/a_n^{(d)} & \dots & 1 \end{pmatrix}.$$

Hence, if $0 < a_n^{(1)} \neq a_n^{(2)} \neq \cdots \neq a_n^{(d)} \leq 2$, then at most $d(d-1)/2$ of the OLS estimates might be inconsistent (near inconsistency occurs frequently). Moreover, throughout the chapter, the innovation vector $\boldsymbol{\epsilon}_t$ has *dependent* components, which arises in more practical situations and differs from Assumption 1 of Caner [14].

The next theorem is the natural extension of the M-estimates for a multivariate random walk model. Consider the Model (2.1) with

$$\mathbf{U}_t = \left(X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(d)}\right)',$$

and

$$\boldsymbol{\epsilon}_t = \left(\epsilon_t^{(1)}, \epsilon_t^{(2)}, \ldots, \epsilon_t^{(d)}\right)'.$$

Denote the M-estimator of $\mathbf{\Phi}$ in (2.1) as the minimizer of $\sum_{t=2}^{n} \rho(\mathbf{U}_t - \mathbf{\Phi}\,\mathbf{U}_{t-1})$ by

$$\widehat{\mathbf{\Phi}}_M := \begin{pmatrix} \hat{\alpha}_M^{(11)} & \hat{\alpha}_M^{(12)} & \ldots & \hat{\alpha}_M^{(1d)} \\ \hat{\alpha}_M^{(21)} & \hat{\alpha}_M^{(22)} & \ldots & \hat{\alpha}_M^{(2d)} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\alpha}_M^{(d1)} & \hat{\alpha}_M^{(d2)} & \ldots & \hat{\alpha}_M^{(dd)} \end{pmatrix}.$$

Furthermore, put

$$\mathbf{u}_n := \left(\mathbf{u}_n^{(1)}, \mathbf{u}_n^{(2)}, \ldots, \mathbf{u}_n^{(d)}\right)',$$

where

$$\mathbf{u}_n^{(1)} = \sqrt{n}\left(a_n^{(1)}(\hat{\alpha}_M^{(11)} - 1), a_n^{(2)}\hat{\alpha}_M^{(12)}, \ldots, a_n^{(d)}\hat{\alpha}_M^{(1d)}\right),$$

$$\mathbf{u}_n^{(2)} = \sqrt{n}\left(a_n^{(1)}\hat{\alpha}_M^{(21)}, a_n^{(2)}(\hat{\alpha}_M^{(22)} - 1), \ldots, a_n^{(d)}\hat{\alpha}_M^{(2d)}\right), \ldots,$$

$$\mathbf{u}_n^{(d)} = \sqrt{n}\left(a_n^{(1)}\hat{\alpha}_M^{(d1)}, a_n^{(2)}\hat{\alpha}_M^{(d2)}, \ldots, a_n^{(d)}(\hat{\alpha}_M^{(dd)} - 1)\right).$$

Then we obtain the following result.

Theorem 2.2. Under Assumptions A1-A4,

$$\mathbf{u}_n \to_d \mathbf{u},$$

where $\mathbf{u} = \mathbf{A}^{-1}\,\mathbf{b}$ with

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \dots & \mathbf{A}_{1d} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \dots & \mathbf{A}_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}_{d1} & \mathbf{A}_{d2} & \dots & \mathbf{A}_{dd} \end{pmatrix},$$

$\mathbf{A}_{ij} = E\left(\rho_{ij}(\boldsymbol{\epsilon}_1)\right) \int_0^1 \mathbf{S}(s)\mathbf{S}'(s)\,ds$, $i, j = 1, \dots, d$, and $\mathbf{b} = \left(\mathbf{b}_1', \dots, \mathbf{b}_d'\right)'$ with $\mathbf{b}_i' = \sqrt{E\left(\rho_i^2(\boldsymbol{\epsilon}_1)\right)} \int_0^1 \mathbf{S}'(s)\,dW_i(s)$, $i = 1, \dots, d$. Here, the stable random vector $\mathbf{S}(\cdot)$ is defined as in (2.4), and $W_i(\cdot)$ is the standard Brownian motion.

Proof. Here we mimic the proof of Theorem 2 of Knight [36]. First, observe that the vector $-\mathbf{u}_n$ minimizes

$$R_n(u_{11}, \dots, u_{dd}) = \sum_{t=2}^{n} \rho\Big(\epsilon_t^{(1)} + \mathbf{u}_{1,n}'\mathbf{U}_{t-1},\ \epsilon_t^{(2)} + \mathbf{u}_{2,n}'\mathbf{U}_{t-1}$$
$$\dots,\ \epsilon_t^{(d)} + \mathbf{u}_{d,n}'\mathbf{U}_{t-1}\Big) - \sum_{t=2}^{n} \rho\left(\epsilon_t^{(1)}, \epsilon_t^{(2)}, \dots, \epsilon_t^{(d)}\right),$$

where

$$\mathbf{u}_{j,n} = \left(\left(a_n^{(1)}\right)^{-1} n^{-1/2} u_{j1}, \dots, \left(a_n^{(d)}\right)^{-1} n^{-1/2} u_{jd}\right)', \quad j = 1, 2, \dots, d.$$

Using the Taylor series expansion of each summand of R_n around $(u_{ij})_{i,j=1,\dots,d} = \mathbf{0}$, we obtain

$$R_n(u_{11}, \dots, u_{dd}) = n^{-1/2} \sum_{t=2}^{n} \sum_{j=1}^{d} \sum_{i=1}^{d} u_{ij} \left(a_n^{(j)}\right)^{-1} \rho_i(\boldsymbol{\epsilon}_t) X_{t-1}^{(j)}$$
$$+ (2n)^{-1} \sum_{t=2}^{n} \sum_{j=1}^{d} \sum_{i=1}^{d} u_{ij}^2 \left(a_n^{(j)}\right)^{-2} \rho_{ii}(\mathbf{c}_t) \left(X_{t-1}^{(j)}\right)^2$$
$$+ n^{-1} \sum_{t=2}^{n} \sum_{k=1}^{d} \sum_{j=2}^{d} \sum_{i=1}^{j-1} u_{ik}u_{jk} \left(a_n^{(k)}\right)^{-2} \rho_{ij}(\mathbf{c}_t) \left(X_{t-1}^{(k)}\right)^2$$

$$+ n^{-1} \sum_{t=2}^{n} \sum_{k=1}^{d} \sum_{j=1}^{d} \sum_{i=k+1}^{d} u_{jk} u_{ji} \left(a_n^{(i)}\right)^{-1} \left(a_n^{(k)}\right)^{-1} \rho_{jj}(\mathbf{c}_t) X_{t-1}^{(i)} X_{t-1}^{(k)}$$

$$+ n^{-1} \sum_{t=2}^{n} \sum_{k=1}^{d} \sum_{j=2}^{d} \sum_{\substack{l=1 \\ l<j}}^{d-1} \sum_{\substack{i=1 \\ i \neq k}}^{d} u_{lk} u_{ji} \left(a_n^{(k)}\right)^{-1} \left(a_n^{(i)}\right)^{-1} \rho_{lj}(\mathbf{c}_t) X_{t-1}^{(k)} X_{t-1}^{(i)},$$

where

$$\mathbf{c}_t = \boldsymbol{\epsilon}_t + \lambda \left(\epsilon_t^{(1)} + \mathbf{u}_{1,n}^{'} \mathbf{U}_{t-1},\ \epsilon_t^{(2)} + \mathbf{u}_{2,n}^{'} \mathbf{U}_{t-1}, \ldots,\ \epsilon_t^{(d)} + \mathbf{u}_{d,n}^{'} \mathbf{U}_{t-1} \right)^{'},$$

for some $\lambda \in [0, 1]$. For the first term of R_n, we have

$$\begin{pmatrix} \mathbf{S}_n(\cdot) \\ \mathbf{W}_n(\cdot) \end{pmatrix} \rightarrow_d \begin{pmatrix} \mathbf{S}(\cdot) \\ \mathbf{W}(\cdot) \end{pmatrix}, \tag{2.20}$$

where $\mathbf{S}_n(\cdot)$ and $\mathbf{S}(\cdot)$ are defined as in (2.3) and (2.4), and

$$\mathbf{W}_n(\cdot) = \left(\left(nE[\rho_1^2(\boldsymbol{\epsilon}_1)]\right)^{-1/2} \sum_{i=1}^{[n\cdot]} \rho_1(\boldsymbol{\epsilon}_i), \ldots, \left(nE[\rho_d^2(\boldsymbol{\epsilon}_1)]\right)^{-1/2} \sum_{i=1}^{[n\cdot]} \rho_d(\boldsymbol{\epsilon}_i) \right)^{'}$$

converges in distribution to the d-variate Brownian motion $\mathbf{W}(\cdot) = (W_1(\cdot), \ldots, W_d(\cdot))'$. Notice that $\mathbf{S}(\cdot)$, the d-variate stable process, is independent from $\mathbf{W}(\cdot)$ with covariance matrix $\Big(E\big(\rho_i(\boldsymbol{\epsilon}_1)\rho_j(\boldsymbol{\epsilon}_1)\big)\Big)_{i,j=1,\ldots,d}$.

Because

$$|\rho_{ij}(\boldsymbol{\epsilon}_t) - \rho_{ij}(\mathbf{c}_t)| \leq \lambda \sum_{r=1}^{d} \sum_{s=1}^{d} c_r \left| u_{rs} \left(a_n^{(s)}\right)^{-1} n^{-1/2} X_{t-1}^{(s)} \right|$$

for $i, j = 1, 2, \ldots, d$, it is easy to see that in the limit we can replace $\rho_{ij}(\mathbf{c}_t)$ by $\rho_{ij}(\boldsymbol{\epsilon}_t)$, for all $i, j = 1, 2, \ldots, d$. We only develop a proof for the second term of R_n but similar arguments for the other terms will complete the proof. We have

$$n^{-1} \sum_{t=2}^{n} \sum_{j=1}^{d} \sum_{i=1}^{d} u_{ij}^2 \left(a_n^{(j)}\right)^{-2} \left(X_{t-1}^{(j)}\right)^2 |\rho_{ii}(\mathbf{c}_t) - \rho_{ii}(\boldsymbol{\epsilon}_t)| \leq \lambda \sum_{j=1}^{d} \sum_{i=1}^{d} \sum_{r=1}^{d}$$

$$\sum_{s=1}^{d} c_r u_{ij}^2 |u_{rs}| n^{-1/2} \left[n^{-1} \sum_{t=2}^{n} \left(a_n^{(j)}\right)^{-2} \left(a_n^{(s)}\right)^{-1} \left| \left(X_{t-1}^{(j)}\right)^2 X_{t-1}^{(s)} \right| \right] \rightarrow_p 0$$

uniformly over u_{ij} in compact sets.

Furthermore, in the limit each $\rho_{ij}(\epsilon_t)$ can be substituted by $E\left(\rho_{ij}(\epsilon_t)\right)$. To see this for the second term of R_n, by applying a weak law of large numbers for martingales we derive

$$\sum_{j=1}^{d} u_{ij}^2\, n^{-1} \sum_{t=2}^{n} \left(a_n^{(j)}\right)^{-2} \left(X_{t-1}^{(j)}\right)^2 I\left(\left(a_n^{(j)}\right)^{-2} \left(X_{t-1}^{(j)}\right)^2 \le M\right)$$
$$\left[\rho_{ii}(\epsilon_t) - E\left(\rho_{ii}(\epsilon_t)\right)\right] \to_p 0,$$

for a fixed i, $i = 1, \ldots, d$. Also

$$P\Big(\Big|\sum_{j=1}^{d} u_{ij}^2\, n^{-1} \sum_{t=2}^{n} \left(a_n^{(j)}\right)^{-2} \left(X_{t-1}^{(j)}\right)^2$$
$$\times I\left(\left(a_n^{(j)}\right)^{-2} \left(X_{t-1}^{(j)}\right)^2 > M\right) \left[\rho_{ii}(\epsilon_t) - E\left(\rho_{ii}(\epsilon_t)\right)\right]\Big| > \delta\Big) \le$$
$$P\Big(\max_{1\le j\le d}\max_{2\le t\le n} \left(a_n^{(j)}\right)^{-2} \left(X_{t-1}^{(j)}\right)^2 > M\Big) \to 0,$$

by letting $n \to \infty$ and $M \to \infty$, respectively. Therefore,

$$n^{-1} \sum_{j=1}^{d}\sum_{i=1}^{d}\sum_{t=2}^{n} u_{ij}^2 \left(a_n^{(j)}\right)^{-2} \left(X_{t-1}^{(j)}\right)^2 \left[\rho_{ii}(\epsilon_t) - E\left(\rho_{ii}(\epsilon_t)\right)\right] \to_p 0.$$

Merging the preceding arguments along with applying the continuous mapping theorem ([10]) and using Proposition 2.1, we can show that the finite-dimensional distribution of $R_n(\cdot)$ converges weakly to those of $R(\cdot)$ where

$$R(u_{11}, \ldots, u_{dd}) = \sum_{j=1}^{d}\sum_{i=1}^{d} u_{ij} \sqrt{E\left(\rho_i^2(\epsilon_1)\right)} \int_0^1 S_j(s)\, dW_i(s)$$
$$+ \frac{1}{2} \sum_{j=1}^{d}\sum_{i=1}^{d} u_{ij}^2 E\left(\rho_{ii}(\epsilon_1)\right) \int_0^1 S_j^2(s)\, ds$$

$$+\sum_{k=1}^{d}\sum_{j=2}^{d}\sum_{i=1}^{j-1} u_{ik}u_{jk}E\big(\rho_{ij}(\boldsymbol{\epsilon}_1)\big)\int_0^1 S_k^2(s)\,ds$$

$$+\sum_{k=1}^{d}\sum_{j=1}^{d}\sum_{i=k+1}^{d} u_{jk}u_{ji}E\big(\rho_{jj}(\boldsymbol{\epsilon}_1)\big)\int_0^1 S_i(s)S_k(s)\,ds$$

$$+\sum_{k=1}^{d}\sum_{j=2}^{d}\sum_{\substack{l=1\\ l<j}}^{d-1}\sum_{\substack{i=1\\ i\neq k}}^{d} u_{lk}u_{ji}E\big(\rho_{lj}(\boldsymbol{\epsilon}_1)\big)\int_0^1 S_i(s)S_k(s)\,ds.$$

Thus, employing Lemma 2.1, we get $\mathbf{u}_n \to_d \mathbf{u}$, where $\mathbf{u}$ is the minimizer of $R(\cdot)$. By finding the unique minimizer of $R(\cdot)$ we can derive the limiting distribution of $\mathbf{u}_n$.

Remark 2.3. Because $\mathbf{S}$ and $\mathbf{W}$ given in (2.20) are independent, using the Cramer-Wold device (see Billingsley [10]), we achieve the following result:

Define

$$\mathbf{v}_1 := \left((\hat{\alpha}_M^{(11)} - 1), \hat{\alpha}_M^{(12)}, \ldots, \hat{\alpha}_M^{(1d)}\right)',$$

$$\mathbf{v}_2 := \left(\hat{\alpha}_M^{(21)}, (\hat{\alpha}_M^{(22)} - 1), \ldots, \hat{\alpha}_M^{(2d)}\right)', \ldots,$$

$$\mathbf{v}_d := \left(\hat{\alpha}_M^{(d1)}, \hat{\alpha}_M^{(d2)}, \ldots, (\hat{\alpha}_M^{(dd)} - 1)\right)',$$

and $\mathbf{T}_i := \int_0^1 \mathbf{f}(s)\,dW_i(s),\ i = 1, \ldots, d$, where

$$\mathbf{f}(s) = \left(\frac{S_1(s)}{\sqrt{\int_0^1 S_1^2(s)\,ds}}, \frac{S_2(s)}{\sqrt{\int_0^1 S_2^2(s)\,ds}}, \ldots, \frac{S_d(s)}{\sqrt{\int_0^1 S_d^2(s)\,ds}}\right)',$$

and $W_i(\cdot),\ i = 1, \ldots, d$, are standard Brownian motions. Notice that conditional on $\{\mathbf{S}(s);\ 0 \leq s \leq 1\}$, $\mathbf{T}_i,\ i = 1, \ldots, d$, are d-variate random variables with normally distributed marginals, and the $\mathbf{T}_i$'s are not independent. Therefore, we get

$$\begin{pmatrix} r_{11}\mathbf{B}_n & r_{12}\mathbf{B}_n & \dots & r_{1d}\mathbf{B}_n \\ r_{21}\mathbf{B}_n & r_{22}\mathbf{B}_n & \dots & r_{2d}\mathbf{B}_n \\ \vdots & \vdots & \vdots & \vdots \\ r_{d1}\mathbf{B}_n & r_{d2}\mathbf{B}_n & \dots & r_{dd}\mathbf{B}_n \end{pmatrix} \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_d \end{pmatrix} \to_d \begin{pmatrix} r_1\mathbf{T}_1 \\ r_2\mathbf{T}_2 \\ \vdots \\ r_d\mathbf{T}_d \end{pmatrix}, \tag{2.21}$$

where $r_{ij} = E\left(\rho_{ij}(\epsilon_1)\right)$, $r_i = \sqrt{E\big(\rho_i^2(\epsilon_1)\big)}$, $i, j = 1, \dots, d$, and

$$\mathbf{B}_n = \left(\frac{\sum_{t=2}^n X_{t-1}^{(i)} X_{t-1}^{(j)}}{\sqrt{\sum_{t=2}^n \left(X_{t-1}^{(i)}\right)^2}} \right)_{ij}, \quad i, j = 1, \dots, d,$$

is the random normalizing matrix.

If $r_{ij} = 0$ for all $i \neq j$, $i, j = 1, \dots, d$, then we derive

$$\begin{pmatrix} \mathbf{B}_n & 0 & \dots & 0 \\ 0 & \mathbf{B}_n & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \mathbf{B}_n \end{pmatrix} \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_d \end{pmatrix} \to_d \begin{pmatrix} \frac{r_1}{r_{11}}\mathbf{T}_1 \\ \frac{r_2}{r_{22}}\mathbf{T}_2 \\ \vdots \\ \frac{r_d}{r_{dd}}\mathbf{T}_d \end{pmatrix}.$$

The loss functions given in (2.7) and (2.8) are of this special type. The advantage of using the self-normalized matrix $\mathbf{B}_n$ is the fact that marginals of the limiting vector in (2.21) have normal distributions. See the normal quantile plots in Figure 2.3 of the forthcoming section.

2.4 Simulation

A simulation study is undertaken to investigate the results given in Theorems 2.1 and 2.2. Taking $d = 2$, we present simulation results for six parameter settings for $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$: $\boldsymbol{\alpha}_1 = (1,1)$, $\boldsymbol{\alpha}_2 = (1,2)$, $\boldsymbol{\alpha}_3 = (0.8, 1.1)$, $\boldsymbol{\alpha}_4 = (0.6, 2)$, $\boldsymbol{\alpha}_5 = (0.2, 0.9)$, $\boldsymbol{\alpha}_6 = (0.2, 1.9)$. All the corresponding distributions of innovations come from symmetric bivariate stable laws. To generate the random numbers $\boldsymbol{\epsilon}_t$ in (2.1) with the aforementioned indices of stability, we consider the set of points $\{\mathbf{S}_i = (cos\,\theta_i, sin\,\theta_i)' : \theta_i = i\pi/4,\ i = 0, 1, \cdots, 7\}$ on the boundary of the unit circle and take a sample with uniform probability, say, $\mathbf{S} = (S_1, S_2)'$. We repeat this procedure $M = 10000$ times to get the random samples $\mathbf{S}_i = (S_{1i}, S_{2i})'$, $i = 1, ..., 10000$. The random number $\boldsymbol{\epsilon}_1$ is obtained from

$$\boldsymbol{\epsilon}_1 = (\epsilon_1^{(1)}, \epsilon_1^{(2)})' = \left(\sum_{i=1}^{M} S_{1i} \Gamma_i^{-1/\alpha_1}, \sum_{i=1}^{M} S_{2i} \Gamma_i^{-1/\alpha_2} \right)',$$

where $\Gamma_i = E_1 + \cdots + E_i$ for i.i.d. exp(1) random variables E_i. In the exact formula we need $M \to \infty$. To generate n random numbers $\boldsymbol{\epsilon}_t$, $t = 1, ..., n$, we perform this procedure again n times independently. The observations $\mathbf{U}_t = (X_t, Y_t)'$ are simulated from $\mathbf{U}_t = \mathbf{U}_{t-1} + \boldsymbol{\epsilon}_t$ using $\mathbf{U}_0 = \mathbf{0}$, and the OLS estimates, presented in Table 2.1, are calculated using (2.2). Figure 2.2 depicts the sample path of two bivariate random walks $\mathbf{U}_t$ simulated by this method. The characteristic function for $\boldsymbol{\epsilon}_t$ when $M \to \infty$ is

$$\psi(\theta_1, \theta_2) = \exp\left[\iint_{\mathcal{C}} (1 - e^{i\theta_1 x^{-1/\alpha_1} u + i\theta_2 x^{-1/\alpha_2} v})\, dH(u, v)\, dx \right],$$

where $H(\cdot, \cdot)$ is the distribution function of $\mathbf{S}_i$ defined on the boundary of the unit circle $\mathcal{C} = \{x \in \mathbb{R}^2 : ||x|| = 1\}$. For more details see Banjevic, Ishwaran, and Zarepour [1], and Samorodnitsky and Taqqu [51].

For the bivariate convex function ρ, we employ the loss functions $\rho^*(x, y)$ and $\rho^{**}(x, y)$ given in (2.7) and (2.8). For both the OLS and the M-estimates provided in Tables 2.1 and 2.2 we choose $n = 50, 100$ and a replication size of 1000 is chosen.

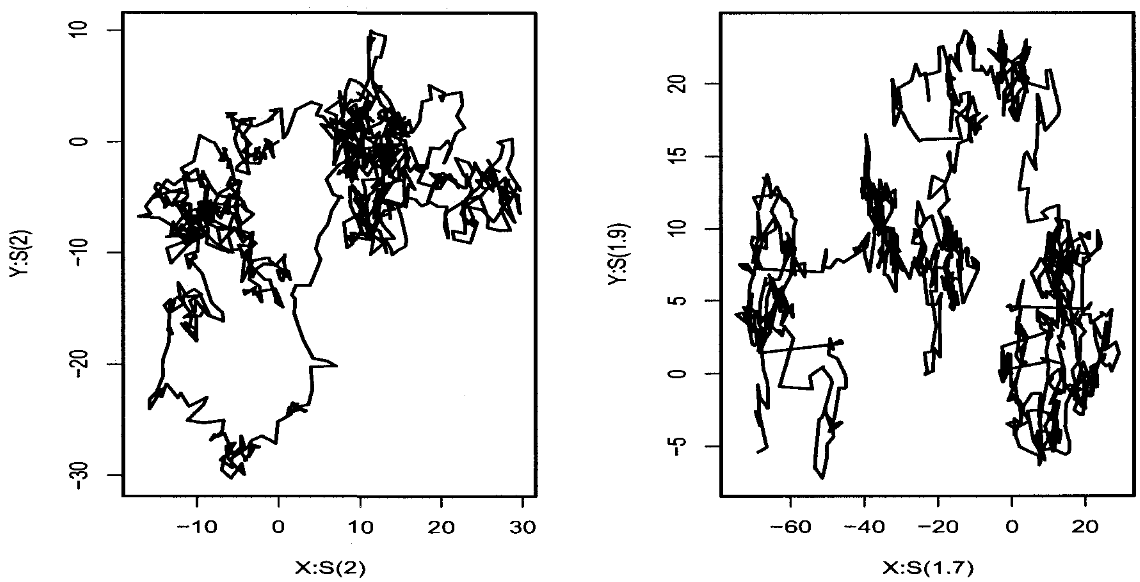


Figure 2.2: Sample path of bivariate random walks with innovations belonging to $DS(2,2)$ (left) and $DS(1.7,1.9)$.

Table 2.1: The OLS estimates of parameters when $\hat{\beta}$ is consistent $(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)$ and when it is not. The replication size is 1000. Note: (*) and (**) indicate near inconsistency (low rate of convergence) and inconsistency cases, respectively.

	$n = 50$				$n = 100$			
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$
$\boldsymbol{\alpha}_1$	0.9154	0.2162	-0.0708	0.9569	0.9301	-0.0419	0.0565	1.0041
$\boldsymbol{\alpha}_2$	0.9338	-0.0031	-0.0025	0.9251	0.9685	0.0071	-0.0010	0.9621
$\boldsymbol{\alpha}_3$	0.9515	-0.0887	-0.0007	0.9711	0.9857	-0.0117	0.0009	0.9806
$\boldsymbol{\alpha}_4$	0.9452	11.5710*	0.0001	0.9218	0.9628	1100.633*	0.0000	0.9591
$\boldsymbol{\alpha}_5$	0.9642	6.8e+14**	0.0000	0.9166	0.9508	1.3e+17**	0.0000	0.9634
$\boldsymbol{\alpha}_6$	0.9217	-1.1e+18**	0.0000	0.9213	0.9337	-2.6e+24**	-0.0000	0.9626

The numbers tabulated in both tables are the averages of the replications.

The simulation results give rise to the following observations. In the cases for which $\alpha_1 \geq \alpha_2(1+\alpha_2)^{-1}$ presented in Table 2.1, as expected, the OLS estimates are consistent and tend to the actual values as n gets large whereas when $\alpha_1 < \alpha_2(1+\alpha_2)^{-1}$, $\hat{\beta}$ is inconsistent, and the other estimates remain consistent. The results obtained from Table 2.2 are quite satisfactory. As n gets large, the M-estimates are in proximity of the actual values in both choices for $\rho^*(\cdot)$ and $\rho^{**}(\cdot)$, confirming the consistency of the M-estimates with better rates of convergence in almost all entries in comparison with the OLS estimates given in Table 2.1.

To explore the marginal normality of the limits with the random normalization factors given in (2.21), we consider a simulated bivariate random walk Model (2.1)

Table 2.2: The M-estimates using Huber's function ρ^* (up) and ρ^{**} (down). The replication size is 1000.

	$n = 50$				$n = 100$			
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$
$\boldsymbol{\alpha}_1$	0.9827	-0.0012	0.0006	0.9841	0.9947	0.0004	0.0003	0.9839
$\boldsymbol{\alpha}_2$	0.9875	0.0033	0.0003	0.9356	0.9954	-0.0015	-0.0002	0.9696
$\boldsymbol{\alpha}_3$	0.9922	0.0045	-0.0002	0.9793	0.9981	-0.0003	-0.0001	0.9935
$\boldsymbol{\alpha}_4$	0.9986	0.0008	-0.0001	0.9668	0.9997	0.0003	-0.0000	0.9892
$\boldsymbol{\alpha}_1$	0.9892	-0.0021	-0.0094	0.9907	0.9975	0.0006	0.0009	0.9945
$\boldsymbol{\alpha}_2$	0.9919	0.0010	0.0017	0.9375	0.9981	0.0002	0.0001	0.9706
$\boldsymbol{\alpha}_3$	0.9975	-0.0011	0.0004	0.9843	0.9995	0.0009	-0.0003	0.9959
$\boldsymbol{\alpha}_4$	0.9940	0.0006	0.0004	0.9698	0.9982	0.0005	-0.0003	0.9907

with innovations belonging to $DS(1.7, 2)$ using a sample size of $n = 100$, replication size of 1000, and $\rho^*(\cdot)$ as the loss function in the M-estimation. Figure 2.3 portrays the normal quantile-quantile plots for the marginals of the limiting vector in (2.21), illustrating the normality of marginals.

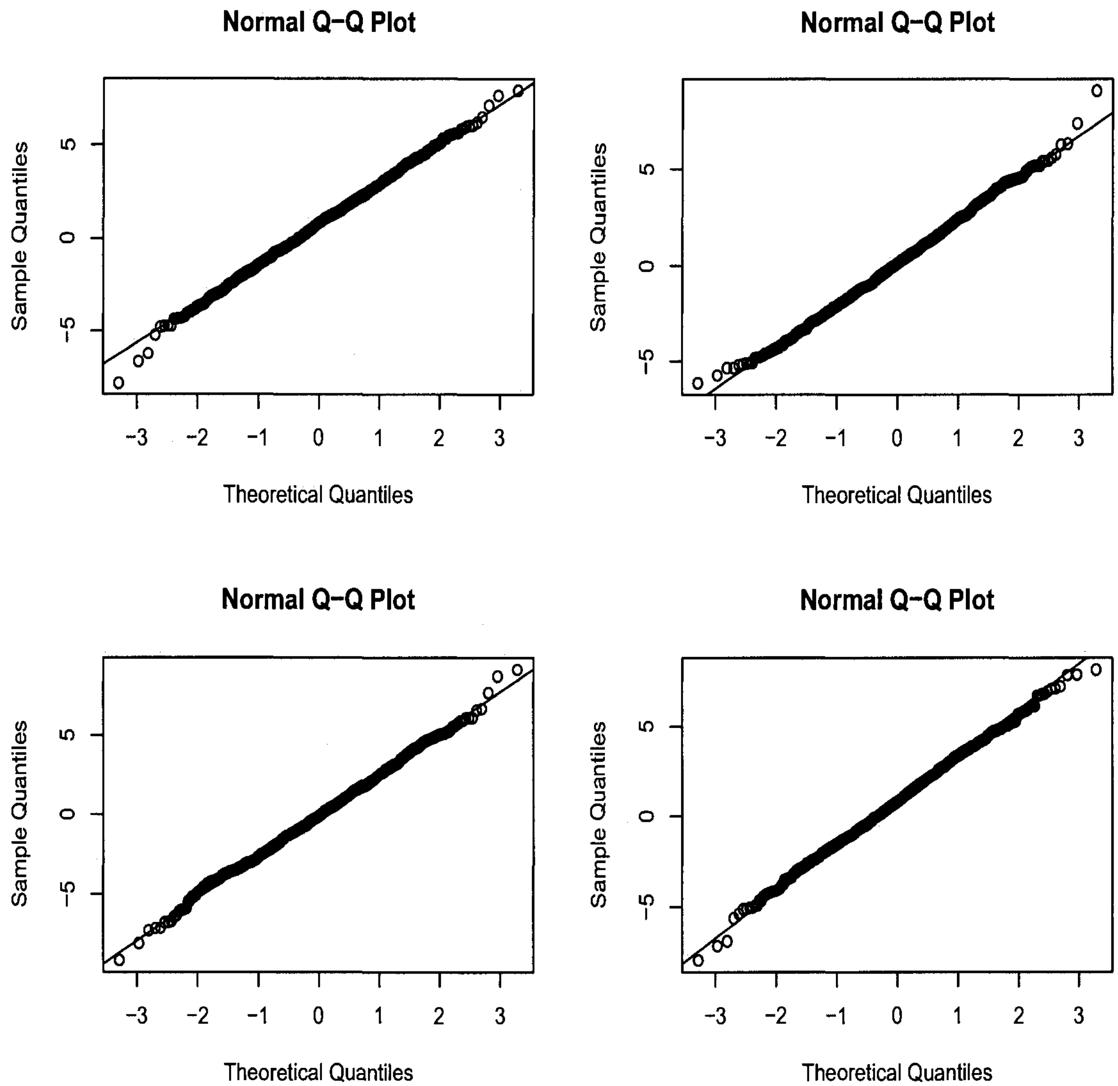


Figure 2.3: Normal Q-Q plots for the marginals of the limiting vector in (2.21) using Model (2.1) with innovations in $DS(1.7, 2)$. The sample and replication size are 100 and 1000, respectively.

Chapter 3

M-estimation for a Spatial Unilateral Autoregressive Model with Infinite Variance Innovations

3.1 Introduction

Consider the doubly geometric spatial autoregressive (AR) model, introduced by Martin [41],

$$Z_{ij} = \alpha Z_{i-1,j} + \beta Z_{i,j-1} - \alpha\beta Z_{i-1,j-1} + \epsilon_{ij}, \tag{3.1}$$

where $\{\epsilon_{ij} = \epsilon_{ij}(\alpha, \beta) : i, j = 1, \ldots, n\}$ is a sequence of i.i.d. random variables with $Z_{ij} = 0$ when $i \wedge j \leq 0$. In this chapter, we study the weak limit behavior of the M-estimators for the parameters in Model (3.1) when $\{\epsilon_{ij}\}$ is in the domain of attraction of a symmetric stable law with index of stability α_0, $0 < \alpha_0 \leq 2$ (denoted by $DS(\alpha_0)$). Both stationary and unit root models are considered, and it is shown that asymptotically, distribution of M-estimators are functionals of stable sheet [49]. For innovations in $DS(\alpha_0)$, self-normalized M-estimates are asymptotically normal.

Spatial models (for example, Model 3.1) appear in many applications such as geography, agriculture, geology, biology and economics. See for example the work of Whittle [58], Kempton and Howes [35], Martin [42], Cullis and Gleeson [21], and Basu and Reinsel [4, 5] in the study of agricultural field trials, Jain [32], Geman and Geman [29], Chellappa [19], and Dass and Nair [22] in image processing, Tjøstheim [52, 53] in system theory and Bronars and Jansen [12] in economics. Asymptotic properties of the spatial unilateral AR models, which are of our interest in this chapter, have been investigated by several authors. Martin [41] introduces the symmetrically reflective spatial AR model given in (3.1) and considers the maximum likelihood (ML) estimators when model is stationary. Tjøstheim [52, 53, 54] establishes the limiting behavior and consistency of the corresponding Yule-Walker estimators. Basu and Reinsel [2, 3] show that in the stationary setting ML estimators have less bias than the Yule-Walker estimators proposed by Tjøstheim. They [4, 5] also define and study the spatial unilateral autoregressive moving average (ARMA) model of first order. Several estimators such as ML, restricted ML (REML), generalized least squares (GLS) and ordinary least squares (OLS) for stationary model are considered and their performances are compared. Bhattacharyya et al. [6] investigate the asymptotic properties of the sequence of Gauss-Newton estimators for the Model (3.1) with unit roots and with finite variance innovations showing the bivariate normality of the weak limits. They also show that two estimators are asymptotically independent.

The M-estimator $(\hat{\alpha}, \hat{\beta})$ of (α, β) in (3.1) is the minimizer of the objective function

$$g(\alpha, \beta) = \sum_{i=2}^{n} \sum_{j=2}^{n} \rho(Z_{ij} - \alpha Z_{i-1,j} - \beta Z_{i,j-1} + \alpha\beta Z_{i-1,j-1}) \tag{3.2}$$

for some function $\rho(\cdot)$. Usually $\rho(x)$ grows at a slower rate than x^2 as $|x|$ gets large. Many of the developments in the M-estimation method can be found in the books by Huber [31] and Van de Geer [55]. For a univariate AR model with infinite variance innovations, several studies have been carried out. For a general AR process with infinite variance innovations when a stationary solution exists, the M-estimates have been considered by Davis, Knight, and Liu [24]. They point out that the M-estimators, similar to their counterpart, the least absolute deviation (LAD) estimators, give less weight to the outliers when the distribution of innovations belongs to the class of heavy tailed distributions. These estimators essentially provide a faster rate of convergence in comparison with the other usual estimators. Davis [23] also develops the limit theory for M-estimates as well as Gauss-Newton estimates for ARMA processes with infinite variance showing the dominance of M-estimates, asymptotically.

The main results of the chapter are stated in the forthcoming section and followed by a numerical example and a simulation study in Section 3.3. Finally, the proofs of the main theorems appear in the Appendix.

3.2 Preliminaries and main results

Let $\{\epsilon_{ij} : i, j = 1, \ldots, n\}$ be a sequence of i.i.d. random variables in $DS(\alpha_0)$. Such random variables have distributions with regularly varying tails, i.e.

$$P(|\epsilon_{11}| > x) = x^{-\alpha_0} L(x),$$

where $L(\cdot)$ is a slowly varying function at infinity, and

$$\lim_{x \to \infty} \frac{P(\epsilon_{11} > x)}{P(|\epsilon_{11}| > x)} = p_0,$$

where $0 < p_0 \le 1$. This implies that there exist constants a_n and b_n such that the stochastic process

$$S_n(t, s) = a_n^{-1} \sum_{i=1}^{[nt]} \sum_{j=1}^{[ns]} \epsilon_{ij} - [n^2 ts] b_n,$$

converges in distribution to a process for $0 < t, s \le 1$, where $[x]$ stands for integer part of x. More precisely, let D_2 be the space of càdlàg functions defined on the unit square $[0, 1] \times [0, 1]$ and equipped with the metric introduced by Bickel and Wichura ([9], see also Appendix of the next chapter). Then

$$S_n(t, s) \to_d S_{\alpha_0}(t, s), \quad \text{as } n \to \infty, \tag{3.3}$$

in D_2, where the limiting process is a stable sheet (Resnick [49]) defined on a regular rectangular grid in two dimensions. In fact, when (3.3) holds with either $\alpha_0 > 1$ and $E(\epsilon_{11}) = 0$, or $\alpha_0 < 1$, or $\{\epsilon_{ij}\}$ have a symmetric distribution about 0 we say that $\{\epsilon_{ij}\}$ is in $DS(\alpha_0)$. Throughout the chapter, all integrals involving dS_n, dS_{α_0} or dW are interpreted as Itô stochastic integrals.

It can be shown (Feller [28]) that under (3.3), $a_n = n^{2/\alpha_0} L(n)$, where $0 < \alpha_0 \le 2$ and $L(\cdot)$ is a slowly varying function at infinity. We assume $b_n = 0$. When $\alpha_0 =$

2, $S_{\alpha_0}(\cdot,\cdot)$ is a Brownian sheet. When $0 < \alpha_0 < 2$, a symmetric α_0-stable random variable S_{α_0} has the series representation $S_{\alpha_0} = \sum_{i=1}^{\infty} \delta_i \Gamma_i^{-1/\alpha_0}$, where δ_i's are i.i.d. random variables with $P(\delta_1 = 1) = P(\delta_1 = -1) = p_0 = 1/2$. Here $\{\Gamma_i\}$ are the arrival times of a Poisson process with Lebesgue mean measure independent of $\{\delta_i\}$ (Samorodnitsky and Taqqu [51]).

Denote $\psi(x) = \partial\rho(x)/\partial x$, $\psi'(x) = \partial^2\rho(x)/\partial x^2$ and impose the following assumptions on the function $\rho(\cdot)$ in (3.2):

A1: $\sigma^2 := E\left(\psi^2(\epsilon_{11})\right) < \infty$ and $E\left(\psi(\epsilon_{11})\right) = 0$ if $\alpha_0 \geq 1$; and $E\left(|\psi(\epsilon_{11})|\right) < \infty$ if $\alpha_0 < 1$.

A2: $\psi(\cdot)$ satisfies the Lipschitz continuity condition

$$|\psi(x) - \psi(y)| \leq k_1 |x - y|^{\lambda_1},$$

for some nonnegative constant k_1 and $\lambda_1 > \max(\alpha_0 - 1, 0)$.

A3: $\gamma := E\left(\psi'(\epsilon_{11})\right) < \infty$, and

$$|\psi'(x) - \psi'(y)| \leq k_2 |x - y|^{\lambda_2},$$

for some nonnegative constant k_2 and $\lambda_2 > 0$.

The loss function $\rho(\cdot)$ might have a countable number of discontinuity points (e.g. Huber's loss function) or might be locally convex with a global minimum.

Now, consider a general spatial unilateral model

$$Z_{ij} = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} c_{kl}\, \epsilon_{i-k,j-l}, \tag{3.4}$$

where $\{\epsilon_{ij}\}$ is in $DS(\alpha_0)$ and $\{c_{kl}\}$ is a sequence of constants satisfying

$$\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}|c_{kl}|^{\delta} < \infty \text{ for some } \delta < \min(\alpha_0, 1). \tag{3.5}$$

It can be shown (see Davis et al [24]) that under this condition, the series in (3.4) converges almost surely and

$$\lim_{x\to\infty}\frac{P(|Z_{11}| > x)}{P(|\epsilon_{11}| > x)} = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}|c_{kl}|^{\alpha_0}.$$

Therefore, the spatial unilateral ARMA(p,q) model

$$Z_{ij} = \sum_{k=0}^{p}\sum_{l=0}^{q}\beta_{kl}Z_{i-k,j-l} + \epsilon_{ij}, \quad \beta_{00} = 0, \tag{3.6}$$

with $\{\epsilon_{ij}\}$ in $DS(\alpha_0)$ has a stationary solution if it can be rewritten in the form of (3.4) where $\{c_{kl}\}$ satisfies (3.5). As an example, note that Model (3.1) is a special case of (3.6) with $p = q = 1$, $\beta_{10} = \alpha$, $\beta_{01} = \beta$ and $\beta_{11} = -\alpha\beta$, and that

$$\begin{aligned}
Z_{ij} &= \alpha Z_{i-1,j} + \beta Z_{i,j-1} - \alpha\beta Z_{i-1,j-1} + \epsilon_{ij} \\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{r=0}^{\infty}\frac{(k+l+r)!}{k!l!r!}\alpha^k\beta^l(-\alpha\beta)^r\epsilon_{i-k-r,j-l-r} \\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\alpha^k\beta^l\epsilon_{i-k,j-l} \\
&= \sum_{k=1}^{i}\sum_{l=1}^{j}\alpha^{i-k}\beta^{j-l}\epsilon_{k,l} \quad \text{when } Z_{ij} = 0 \text{ for } i \wedge j \le 0,
\end{aligned}$$

is a unilateral model. Therefore by (3.5) to ensure the existence of the stationary representation we shall have for $0 < \alpha_0 < 2$,

$$\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\left|\alpha^k\beta^l\right|^{\delta} < \infty \text{ for some } \delta < \min(\alpha_0, 1).$$

For $\alpha_0 = 2$, see Proposition 1 of Basu and Reinsel [4].

The following theorem deals with the M-estimators of the parameters of the general stationary model given in (3.6). Following [24], recall that the M-estimate, $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_{01}, \hat{\beta}_{02}, \ldots, \hat{\beta}_{0q}, \ldots, \hat{\beta}_{p0}, \hat{\beta}_{p1}, \ldots, \hat{\beta}_{pq}\right)'$, of $\boldsymbol{\beta} = (\beta_{01}, \ldots, \beta_{pq})'$ minimizes the objective function

$$\sum_{i=p+1}^{n} \sum_{j=q+1}^{n} \rho(Z_{ij} - \beta_{10}Z_{i-1,j} - \beta_{01}Z_{i,j-1} - \beta_{11}Z_{i-1,j-1} - \cdots - \beta_{pq}Z_{i-p,j-q}),$$

with respect to $\{\beta_{ij}\}$. This is equivalent to minimizing the sequence of stochastic processes

$$W_n(\boldsymbol{u}) = \sum_{i=p+1}^{n} \sum_{j=q+1}^{n} \Big[\rho(\epsilon_{ij} - u_{10}a_n^{-1}Z_{i-1,j} - u_{01}a_n^{-1}Z_{i,j-1} - u_{11}a_n^{-1}Z_{i-1,j-1} - \cdots - u_{pq}a_n^{-1}Z_{i-p,j-q}) - \rho(\epsilon_{ij})\Big]. \quad (3.7)$$

The minimizer of the process $W_n(\cdot)$, that is $a_n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, has a weak limit which is the minimizer of the process $W(\cdot)$ given in the following theorem.

Theorem 3.1. Consider the stationary spatial unilateral Model (3.6) with $\{\epsilon_{ij}\}$ in $DS(\alpha_0)$, $0 < \alpha_0 < 2$ and suppose that A1 and A2 hold. Then on $C\left(\mathbb{R}^{p+q+pq}\right)$, $W_n(\cdot) \rightarrow_d W(\cdot)$, where

$$W(\boldsymbol{u}) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \Big[\rho\big(\epsilon_{kij} - (u_{10}c_{i-1,j} + u_{01}c_{i,j-1} + u_{11}c_{i-1,j-1} + \cdots + u_{pq}c_{i-p,j-q})\delta_k\Gamma_k^{-1/\alpha}\big) - \rho(\epsilon_{kij})\Big]. \quad (3.8)$$

Here $\epsilon_{kij} \stackrel{d}{=} \epsilon_{11}$, and δ_k and Γ_k are defined as before, and $C\left(\mathbb{R}^{p+q+pq}\right)$ is the metric space of continuous functions on $\mathbb{R}^{p+q+pq}$ equipped with the uniform metric.

Note that we cannot find a closed form for the minimizer of $W(\cdot)$ in (3.8). A similar theorem can be stated for the special case when $\rho(x) = |x|$, corresponding to LAD estimators, provided on some assumptions on ϵ_{11} similar to those made in Theorem 3.4 of [23].

In the next theorem, we investigate asymptotic behavior of M-estimators of the parameters in the spatial unilateral model given in (3.1) when the model has a unit root. We consider two cases; (i) $\alpha = 1$ and $|\beta| < 1$, and (ii) $\alpha = \beta = 1$. The limits are functionals of α_0-stable sheet (Itô stochastic integrals) in (3.3). As in [36], to make sure that an a.s. unique limit exists we assume the convexity of the function $\rho(\cdot)$.

Theorem 3.2. Consider Model (3.1) with $\{\epsilon_{ij}\}$ in $DS(\alpha_0)$, $0 < \alpha_0 < 2$ and suppose that the loss function $\rho(\cdot)$ is strictly convex. Denote $(\hat{\alpha}, \hat{\beta})$ to be the M-estimator of (α, β) in that model. Then under assumptions A1-A3,
(i) if $\alpha = 1$ and $|\beta| < 1$,

$$\begin{pmatrix} a_n\sqrt{n}\,(\hat{\alpha}-1) \\ a_n(\hat{\beta}-\beta) \end{pmatrix} \to_d \xi := \begin{pmatrix} \frac{\sigma\int_0^1\int_0^1 S_{\alpha_0}(t_1,t_2)\,dW(t_1,t_2)}{\gamma\int_0^1\int_0^1 S^2_{\alpha_0}(t_1,t_2)\,dt_1dt_2} \\ \arg\min Z(u) \end{pmatrix},$$

where

$$Z(u) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\Big[\rho\big(\epsilon_{ij} - u\beta^{j-1}\delta_i\Gamma_i^{-1/\alpha_0}\big) - \rho(\epsilon_{ij})\Big],$$

(ii) and if $\alpha = \beta = 1$,

$$a_n\sqrt{n}\begin{pmatrix} \hat{\alpha}-1 \\ \hat{\beta}-1 \end{pmatrix} \to_d \eta := \frac{\sigma}{\gamma}\begin{pmatrix} \frac{\int_0^1\int_0^1 S_{\alpha_0}(t_1,t_2)\,dW(t_1,t_2)}{\int_0^1\int_0^1 S^2_{\alpha_0}(t_1,t_2)\,dt_1dt_2} \\ \frac{\int_0^1\int_0^1 S_{\alpha_0}(t_1,t_2)\,dW(t_1,t_2)}{\int_0^1\int_0^1 S^2_{\alpha_0}(t_1,t_2)\,dt_1dt_2} \end{pmatrix}, \tag{3.9}$$

where $S_{\alpha_0}(\cdot,\cdot)$ is the α_0-stable sheet and $W(\cdot,\cdot)$ is a standard Brownian sheet.

Remark 3.1. When $\alpha = 1$ and $|\beta| < 1$, distribution of ξ is not necessarily symmetric while for the second case we observe a symmetric distribution. Similar to the results given by Knight [36] and Remark 3 of Zarepour and Roknossadati [60] (see Remark 2.3) it is possible to give self-normalizing coefficients for M-estimators in Theorem 3.2. More precisely, using Resnick-Greenwood's result (Theorem 1 and Theorem 3 of [48]) when ϵ_{11} is in $DS(\alpha_0)$, $0 < \alpha_0 \leq 2$, we get:

(i) If $\alpha = 1$ and $|\beta| < 1$,

$$\left(\sum_{i=2}^{n}\sum_{j=1}^{n} X_{i-1,j}^2\right)^{\frac{1}{2}} (\hat{\alpha} - 1) \rightarrow_d N\left(0, \frac{\sigma^2}{\gamma^2}\right),$$

(ii) and if $\alpha = \beta = 1$,

$$\begin{pmatrix} \sqrt{\sum_{i=2}^{n}\sum_{j=1}^{n} X_{i-1,j}^2} & 0 \\ 0 & \sqrt{\sum_{i=1}^{n}\sum_{j=2}^{n} Y_{i,j-1}^2} \end{pmatrix} \begin{pmatrix} \hat{\alpha} - 1 \\ \hat{\beta} - 1 \end{pmatrix} \rightarrow_d N(\mathbf{0}, \mathbf{\Gamma}), \tag{3.10}$$

where $\mathbf{\Gamma} = \text{diag}\left(\frac{\sigma^2}{\gamma^2}, \frac{\sigma^2}{\gamma^2}\right)$, $X_{ij} := Z_{ij} - \beta Z_{i,j-1}$, and $Y_{ij} := Z_{ij} - \alpha Z_{i-1,j}$.
Since $(\hat{\alpha}, \hat{\beta})$ is a consistent estimator of (α, β), $X_{i-1,j}$ and $Y_{i,j-1}$ can be replaced by $\hat{X}_{i-1,j}$ and $\hat{Y}_{i,j-1}$ where $\hat{X}_{ij} := Z_{ij} - \hat{\beta} Z_{i,j-1}$, and $\hat{Y}_{ij} := Z_{ij} - \hat{\alpha} Z_{i-1,j}$.

Again a similar result can be derived for the special case when $\rho(x) = |x|$ using techniques similar to those employed in Knight [36].

3.3 A numerical example and simulation

To investigate the performance of the M-estimation approach proposed in the preceding section, a numerical example regarding Theorem 3.1 and some simulations regarding the unit root model are presented. We first consider the yield of barley (in kilograms) data set from a 7×28 regular grid given in Kempton and Howes [35]. This data set has also been analyzed by Basu and Reinsel [4]. We consider two loss functions; Huber and least absolute deviation (LAD). Since the data has a normal distribution, the $\alpha_0 = 2$ is taken. Results tabulated in Table 3.1 and Table 3.2 show that the doubly geometric spatial AR model given in (3.1) with a constant added to the model is more accurate than a general model with $-\alpha\beta$ replaced by γ. Such a model is a special case of the unilateral first-order ARMA model of the quadrant type introduced by Basu and Reinsel [4]. Notice that the limit theory for this general model is the same as that for Model (3.6) of Section 3.2 with different $\{c_{kl}\}$. From Table 3.1 it can also be seen that the performance of the fit by LAD and Huber's loss functions is as good as that of the ML fit reported by Basu and Reinsel [4].

A simulation study is undertaken to investigate the result of Theorem 3.2 of Section 3.2. We present the results for five parameter settings for index of stability $\alpha_0 = 2, 1.8, 1.5, 1, 0.5$. The unit root Model (3.1) with $\alpha = \beta = 1$ and symmetric α_0-stable errors on an $n \times n$ regular grid with different values of n are generated. The replication size of 1000 is chosen and both mean and standard deviation of M-estimates of $\alpha - 1$ are calculated. The simulation results are summarized in Table 3.3 with Huber's loss function and in Table 3.4 with the loss function $\rho(x) = 2\log(x^2+3)$. Although the latter is just convex around its global minimum, the results tabulated in Table 3.4 are quite satisfactory. As n gets large, the M-estimates are close to the actual values for both choices for the loss function with better rates of convergence as α_0 gets smaller.

Table 3.1: M-estimates of the parameters for Model (3.1) based on the yield of barley data (in kg) obtained from Kempton and Howes (1981).

	const.	$\hat{\alpha}$	$\hat{\beta}$	$-\hat{\alpha}\hat{\beta}$	$\hat{\sigma}^2$
Huber	2.732	0.246	0.828	-0.204	0.0322
LAD	2.780	0.247	0.836	-0.207	0.0322
ML	2.627	0.241	0.812	-0.196	0.032

Table 3.2: M-estimates of the parameters for Model (3.1) with $-\alpha\beta$ replaced by γ based on the yield of barley data (in kg) obtained from Kempton and Howes (1981).

	const.	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}(\neq -\hat{\alpha}\hat{\beta})$	$\hat{\sigma}^2$
Huber	2.731	0.241	0.794	-0.106	0.0332
LAD	2.791	0.230	0.833	-0.107	0.0329
ML	2.663	0.240	0.796	-0.108	0.032

To explore the bivariate normality of the limits with the random normalization factors given in (3.10), we consider a simulated unit root Model (3.1) with normal innovations using a replication size of 1000, and $\rho(x) = |x|$ as the loss function in the M-estimation. To get the normality we need to take $n = 60$. Figure 3.1 portrays the normal quantile-quantile plots for the marginals of the limiting vectors in (3.9)-up and (3.10)-down illustrating a significant deviation from normality and the normality of marginals after using fixed and random normalizing coefficients. The bivariate normality of the data after using a random normalizer can be seen in Figure 3.2. Notice that the contour plot confirms that the marginal normal variables of the limiting vector in (3.10) are independent.

Some simulations, which are not reported here, demonstrate that for $\alpha_0 < 2$ to have the normality we need to take n significantly large. For instance, for a model with Cauchy innovations ($\alpha_0 = 1$) a sample size of $n = 500$ suffices to get bivariate normality.

Table 3.3: Mean and mean standard deviations (in parentheses) for M-estimates of the parameter $(\alpha - 1)$ for the unit root Model (3.1) with $\alpha = \beta = 1$ using Huber's loss function with different α_0-stable noises. The replication size is 1000.

	index of stability α_0				
n	2	1.8	1.5	1	0.5
5	-0.0202	-0.0165	-0.0138	-0.0113	-0.0130
	(0.0224)	(0.0200)	(0.0200)	(0.0641)	(0.1136)
10	-0.0034	-0.0028	-0.0016	-0.0004	-0.0033
	(0.0038)	(0.0036)	(0.0026)	(0.0009)	(0.0566)
15	-0.0011	-0.0009	-0.0005	-0.0001	-0.0044
	(0.0015)	(0.0011)	(0.0009)	(0.0002)	(0.0608)
20	-0.0006	-0.0004	-0.0002	-0.0000	-0.0001
	(0.0007)	(0.0006)	(0.0004)	(0.0003)	(0.0019)
25	-0.0003	-0.0002	-0.0001	-0.0000	-0.0000
	(0.0004)	(0.0003)	(0.0002)	(0.0002)	(0.0010)
30	-0.0002	-0.0001	-0.0000	-0.0000	-0.0000
	(0.0003)	(0.0002)	(0.0001)	(0.0001)	(0.0158)

Table 3.4: Mean and mean standard deviations (in parentheses) for M-estimates of the parameter $(\alpha - 1)$ for the unit root Model (3.1) with $\alpha = \beta = 1$ using the loss function $\rho(x) = 2\log(x^2 + 3)$ with different α_0-stable noises. The replication size is 1000.

	index of stability α_0				
n	2	1.8	1.5	1	0.5
5	-0.0128	-0.0106	-0.0087	-0.0029	0.0042
	(0.0173)	(0.0141)	(0.0140)	(0.0069)	(0.0105)
10	-0.0017	-0.0013	-0.0007	-0.0002	0.0002
	(0.0018)	(0.0016)	(0.0014)	(0.0008)	(0.0004)
15	-0.0006	-0.0004	-0.0002	-0.0000	0.0000
	(0.0006)	(0.0005)	(0.0004)	(0.0004)	(0.0001)
20	-0.0002	-0.0002	-0.0000	-0.0000	0.0000
	(0.0003)	(0.0003)	(0.0001)	(0.0004)	(0.0000)
25	-0.0001	-0.0000	-0.0000	-0.0000	0.0000
	(0.0001)	(0.0001)	(0.0000)	(0.0002)	(0.0000)
30	-0.0000	-0.0000	-0.0000	-0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)

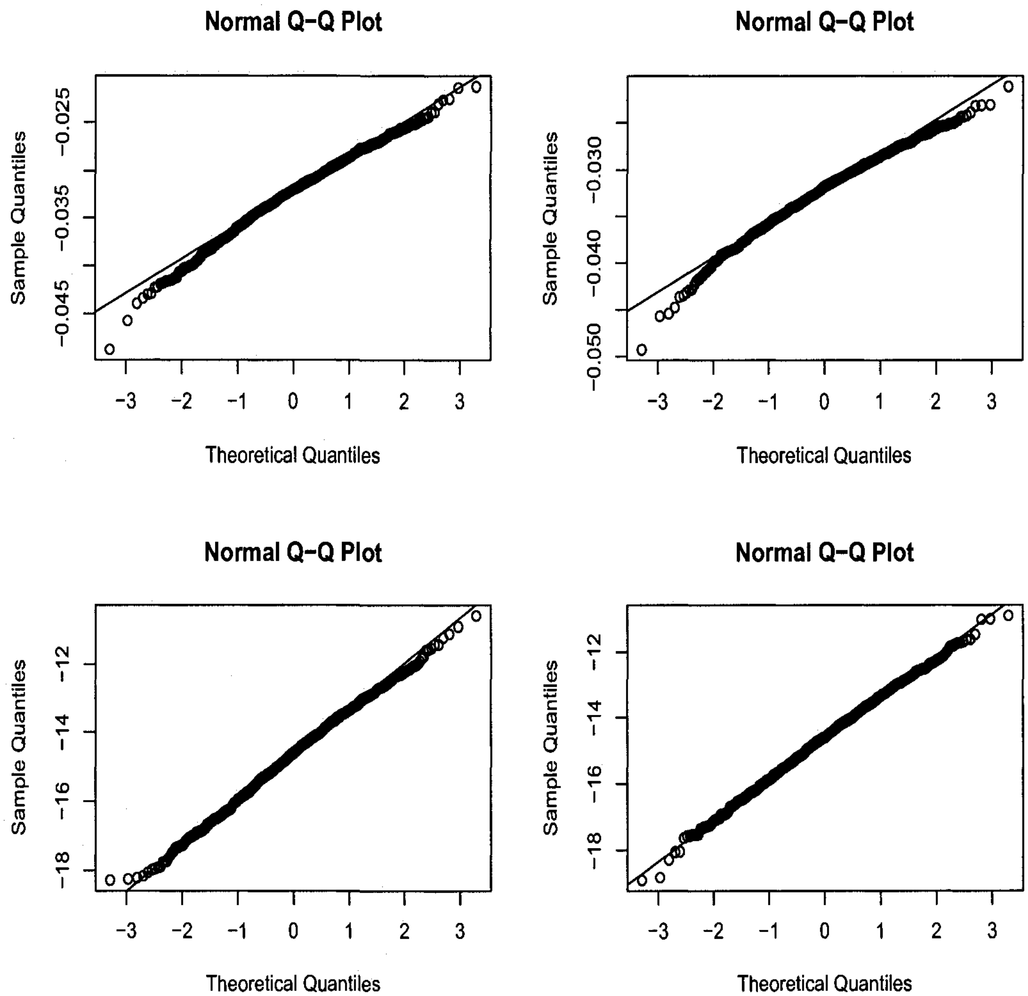


Figure 3.1: Normal Q-Q plots for the marginals of the limiting vector with fixed (up) and random normalization (down) based on LAD estimates for the parameters in Model (3.1) with normal noise and $n = 60$. The replication size is 1000.

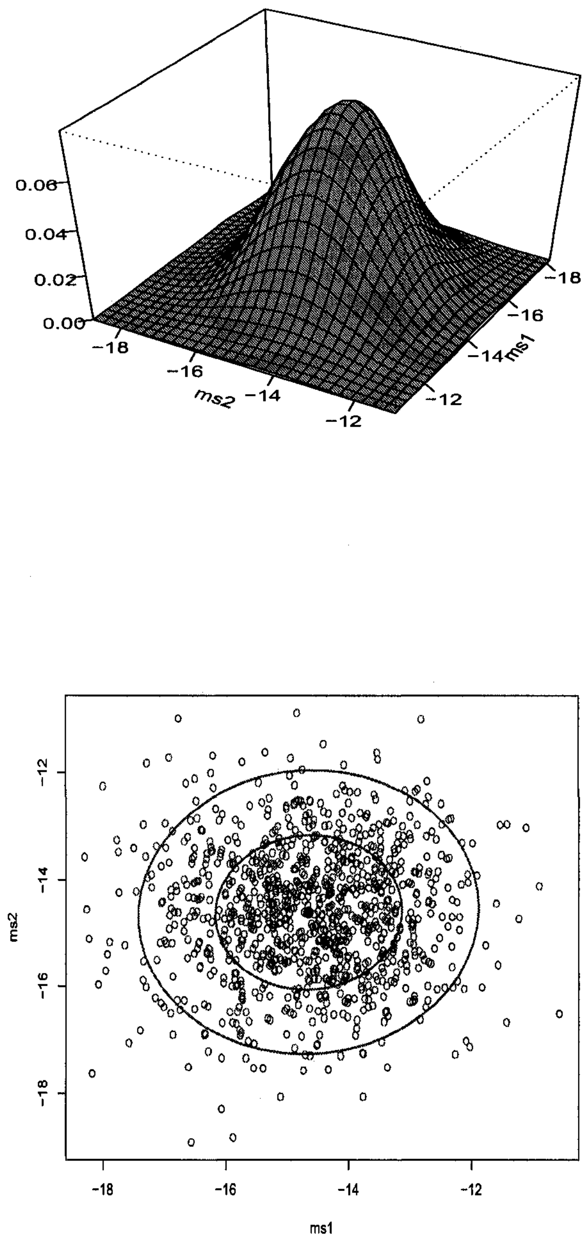


Figure 3.2: Density (up) and contour plot (down) for the self-normalized LAD estimates given in Figure 3.1.

3.4 Appendix

To get ready to prove the first theorem of Section 3.2 we need to provide two propositions which are similar to the Proposition 1 and Proposition 2 of Davis et al. [24] and therefore are given without proof. In what follows, the sequences of i.i.d. random variables $\{\epsilon_{ikl}\}$, with $\epsilon_{111} \stackrel{d}{=} \epsilon_{11}$, $\{\delta_i\}$ and $\{\Gamma_i\}$ are as specified in Section 3.2, and all the sequences are mutually independent. We denote by $M_p(\mathbb{R} \times (\overline{\mathbb{R}}\backslash\{0\}))$ the set of point processes defined on $\mathbb{R} \times (\overline{\mathbb{R}}\backslash\{0\})$ where $\overline{\mathbb{R}} = [-\infty, \infty]$ equipped with vague topology (Kallenberg [34]).

Proposition 3.1. Suppose that the sequence of random variables $\{Z_{ij}\}$ is given by (3.4) with $\{\epsilon_{ij}\}$ in $DS(\alpha_0)$. Then

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \varepsilon_{(\epsilon_{ij}, a_n^{-1} Z_{ij})} \rightarrow_d \sum_{i=1}^{\infty}\sum_{l=1}^{\infty}\sum_{k=1}^{\infty} \varepsilon_{(\epsilon_{ikl}, c_{kl}\delta_i \Gamma_i^{-1/\alpha_0})},$$

in $M_p(\mathbb{R} \times (\overline{\mathbb{R}}\backslash\{0\}))$ with respect to vague topology.

As a consequence we get the following corollary:

Corollary 3.1. Consider Model (3.4) and a nonnegative continuous function f on $\mathbb{R} \times (\overline{\mathbb{R}}\backslash\{0\})$ with compact support, then

$$\sum_{i=1}^{n}\sum_{j=1}^{n} f(\epsilon_{ij}, a_n^{-1} Z_{ij}) \rightarrow_d \sum_{i=1}^{\infty}\sum_{l=1}^{\infty}\sum_{k=1}^{\infty} f(\epsilon_{ikl}, c_{kl}\delta_i \Gamma_i^{-1/\alpha_0}),$$

where $c_{kl} = 0$ for $k, l \leq 0$.

For $f(x, y)I(|x| < M)I(|y| > \delta)$, where f is a continuous function, weak convergence follows if for all $\epsilon > 0$,

$$\lim_{\delta\to 0}\lim_{M\to\infty}\limsup_{n} P\Big(\Big|\sum_{i=1}^{n}\sum_{j=1}^{n} f(\epsilon_{ij}, a_n^{-1}Z_{ij})[1 - I(|\epsilon_{ij}| \leq M) \\ I(|Z_{ij}| > \delta a_n)]\Big| > \epsilon\Big) = 0, \quad (3.11)$$

and

$$\sum_{i=1}^{\infty}\sum_{l=1}^{\infty}\sum_{k=1}^{\infty} f(\epsilon_{ikl}, c_{kl}\delta_i\Gamma_i^{-1/\alpha_0}) I(|\epsilon_{ikl}| \leq M) I(|c_{kl}\delta_i\Gamma_i^{-1/\alpha_0}| > \delta) \to_p \\ \sum_{i=1}^{\infty}\sum_{l=1}^{\infty}\sum_{k=1}^{\infty} f(\epsilon_{ikl}, c_{kl}\delta_i\Gamma_i^{-1/\alpha_0}) \quad (3.12)$$

as $\delta \to 0$ and $M \to \infty$.

Proposition 3.2. Suppose that $\{Z_{ij}\}$ is given by (3.4) with $\{\epsilon_{ij}\}$ in $DS(\alpha_0)$. Let $\{V_{ij}\}$ be i.i.d. sequence of random variables with finite mean such that for every i and j, V_{ij} and Z_{ij} are independent. Then for all $\delta > 0$ and $\eta > 0$,

$$(a) \quad \limsup_{n\to\infty} P\Big[\sum_{i=1}^{n}\sum_{j=1}^{n} |V_{ij}||a_n^{-1}Z_{ij}|^{\nu} I(|Z_{ij}| \leq \delta a_n) > \eta\Big] \leq \eta^{-1}c_1 E|V_{11}|\delta^{\nu-\alpha_0}$$

for all $\nu > \alpha_0$,

$$(b) \quad \limsup_{n\to\infty} P\Big[\sum_{i=1}^{n}\sum_{j=1}^{n} |V_{ij}||a_n^{-1}Z_{ij}|^{\nu} I(|Z_{ij}| > \delta a_n) > \eta\Big] \leq c_2\delta^{-\alpha_0} P(|V_{11}| > 0)$$

for all $\nu > 0$ and constants c_1 and c_2. If in addition V_{11} has zero mean and finite variance and $1 \leq \alpha_0 < 2$, then

(c)

$$Var\Big[\sum_{i=1}^{n}\sum_{j=1}^{n} V_{ij}(a_n^{-1}Z_{ij}) I(|Z_{ij}| \leq \delta a_n)\Big] \to 0$$

as $n \to \infty$ and then $\delta \to 0$.

Proof of Theorem 3.1. In here, we mimic the proof of Theorem 1 of Davis et al. [24]. First, we show that finite dimensional distribution of $W_n(\cdot)$ converges weakly. Let

$$Y_{ni,nj}(\boldsymbol{u}) = u_{10}a_n^{-1}Z_{i-1,j} + u_{01}a_n^{-1}Z_{i,j-1} + u_{11}a_n^{-1}Z_{i-1,j-1} + \cdots + u_{pq}a_n^{-1}Z_{i-p,j-q}.$$

By Corollary 3.1 it follows that

$$\begin{aligned} W_n(\boldsymbol{u};\delta,M) = \sum_{i=1}^{n}\sum_{j=1}^{n}[\rho(\epsilon_{ij} - Y_{ni,nj}(\boldsymbol{u})) - \rho(\epsilon_{ij})]I(|\epsilon_{ij}| \le M,\ |Z_{ij}| > a_n\delta) \to_d \\ W(\boldsymbol{u};\delta,M) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\Big[\rho\big(\epsilon_{kij} - (u_{10}c_{i-1,j} + u_{01}c_{i,j-1} + u_{11}c_{i-1,j-1} \\ + \cdots + u_{pq}c_{i-p,j-q})\delta_k\Gamma_k^{-1/\alpha_0}\big) - \rho(\epsilon_{kij})\Big]I(|\epsilon_{kij}| \le M,\ |(u_{10}c_{i-1,j} \\ + u_{01}c_{i,j-1} + u_{11}c_{i-1,j-1} + \cdots + u_{pq}c_{i-p,j-q})\Gamma_k^{-1/\alpha_0}| > \delta). \end{aligned}$$

It suffices to show that (3.11) and (3.12) hold with the function $f(x,y) = \rho(x-y) - \rho(x)$. Using the Taylor series expansion around ϵ_{ij} for each term of $W_n(\cdot)$, we get

$$\begin{aligned} \sum_{i=1}^{n}\sum_{j=1}^{n}[\rho(\epsilon_{ij} - Y_{ni,nj}(\boldsymbol{u})) - \rho(\epsilon_{ij})] = -\sum_{i=1}^{n}\sum_{j=1}^{n}Y_{ni,nj}(\boldsymbol{u})\psi(\xi_{ij}^n) = \\ -\sum_{i=1}^{n}\sum_{j=1}^{n}Y_{ni,nj}(\boldsymbol{u})\psi(\epsilon_{ij}) + \sum_{i=1}^{n}\sum_{j=1}^{n}Y_{ni,nj}(\boldsymbol{u})[\psi(\epsilon_{ij}) - \psi(\xi_{ij}^n)], \end{aligned}$$

where $|\xi_{ij}^n - \epsilon_{ij}| \le |Y_{ni,nj}(\boldsymbol{u})|$.

From Proposition 3.2-(b), with $V_{ij} = \psi(\epsilon_{ij})I(|\epsilon_{ij}| > M)$, we get

$$\lim_{M\to\infty}\limsup_{n\to\infty} P\Big[\Big|\sum_{i=1}^{n}\sum_{j=1}^{n}Y_{ni,nj}(\boldsymbol{u})I(|Y_{ni,nj}(\boldsymbol{u})| > \delta)\psi(\epsilon_{ij})I(|\epsilon_{ij}| > M)\Big| > \eta\Big] = 0.$$

Furthermore, employing Proposition 3.2-(a) if $\alpha_0 < 1$ and (c) if $\alpha_0 \ge 1$ with $V_{ij} = \psi(\epsilon_{ij})$ we deduce

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\Big[\Big|\sum_{i=1}^{n}\sum_{j=1}^{n}Y_{ni,nj}(\boldsymbol{u})\psi(\epsilon_{ij})I(|Y_{ni,nj}(\boldsymbol{u})| \le \delta)\Big| > \eta\Big] = 0.$$

Also using the Assumption A2 we have

$$\Big|\sum_{i=1}^{n}\sum_{j=1}^{n} Y_{ni,nj}(\boldsymbol{u})\Big(\psi(\epsilon_{ij})-\psi(\xi_{ij}^{n})\Big)I(|Y_{ni,nj}(\boldsymbol{u})|\le\delta)\Big|$$
$$\le k_1\sum_{i=1}^{n}\sum_{j=1}^{n}|Y_{ni,nj}(\boldsymbol{u})|^{1+\lambda_1}I(|Y_{ni,nj}(\boldsymbol{u})|\le\delta)\to_p 0,$$

by Proposition 3.2-(a) as $n\to\infty$ and then $\delta\to 0$, and similarly by Proposition 3.2-(b)

$$\Big|\sum_{i=1}^{n}\sum_{j=1}^{n} Y_{ni,nj}(\boldsymbol{u})\Big(\psi(\epsilon_{ij})-\psi(\xi_{ij}^{n})\Big)I(|Y_{ni,nj}(\boldsymbol{u})|>\delta)I(|\epsilon_{ij}|>M)\Big|$$
$$\le k_1\sum_{i=1}^{n}\sum_{j=1}^{n}|Y_{ni,nj}(\boldsymbol{u})|^{1+\lambda_1}I(|Y_{ni,nj}(\boldsymbol{u})|>\delta)I(|\epsilon_{ij}|>M)\to_p 0,$$

as $n\to\infty$ and $M\to\infty$. Combining the above arguments proves (3.11), and (3.12) is proven in a similar way.

Finally, to see the tightness of $W_n(\cdot)$ we shall show

$$\lim_{\delta\to 0}\limsup_{n\to\infty}P\big[\sup_{\|\boldsymbol{u}-\boldsymbol{v}\|\le\delta}|W_n(\boldsymbol{u})-W_n(\boldsymbol{v})|>\eta\big]=0.$$

Observe that by Lipschitz continuity of ψ we get

$$\big|\psi(\epsilon_{ij})-\psi(\xi_{ij}^{n})\big|\le k_1|Y_{ni,nj}(\boldsymbol{u}-\boldsymbol{v})|^{\lambda_1},$$

which implies

$$|W_n(\boldsymbol{u})-W_n(\boldsymbol{v})|\le\Big|\sum_{i=1}^{n}\sum_{j=1}^{n}Y_{ni,nj}(\boldsymbol{u}-\boldsymbol{v})\psi(\epsilon_{ij})\Big|+k_1\sum_{i=1}^{n}\sum_{j=1}^{n}|Y_{ni,nj}(\boldsymbol{u}-\boldsymbol{v})|^{\lambda_1+1}.$$

Now using Proposition 3.2 and the fact that

$$a_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}Z_{i-r,j-v}\psi(\epsilon_{ij})=O_p(1),$$

and

$$a_n^{-(1+\lambda_1)} \sum_{i=1}^{n} \sum_{j=1}^{n} |Z_{i-r,j-v}|^{1+\lambda_1} = O_p(1)$$

we complete the proof.

Proof of Theorem 3.2. We prove the part (i) of the theorem. Since $\rho(\cdot)$ is convex, by Lemma 2.1 it is enough to show the weak convergence for the finite dimensional distribution (denoted by $\rightarrow_{fidi}$). Let $(\hat{\alpha}, \hat{\beta})$ be the M-estimator of (α, β) in Model (3.1), and define $X_{ij} := Z_{ij} - \beta Z_{i,j-1}$, and $Y_{ij} := Z_{ij} - \alpha Z_{i-1,j}$. Then $X_{ij} = \alpha X_{i-1,j} + \epsilon_{ij}$, and $Y_{ij} = \beta Y_{i,j-1} + \epsilon_{ij}$ which yields the non-stationary part $X_{ij} = \sum_{k=1}^{i} \epsilon_{kj}$, assuming $\alpha = 1$, and the stationary part $Y_{ij} = \sum_{l=1}^{j} \beta^{j-l} \epsilon_{il}$. Also recall from before that $\epsilon_{ij} = \epsilon_{ij}(\alpha, \beta)$ and observe that

$$\begin{aligned} \epsilon_{ij}\big(\alpha + a_n^{-1} n^{-1/2} u_1, \beta + a_n^{-1} u_2\big) =& \epsilon_{ij}(\alpha, \beta) - a_n^{-1} n^{-1/2} u_1 X_{i-1,j} - a_n^{-1} u_2 Y_{i,j-1} \\ &+ a_n^{-2} n^{-1/2} u_1 u_2 Z_{i-1,j-1}. \end{aligned}$$

Thus, we shall minimize the function

$$\begin{aligned} K_n(u_1, u_2) &= \sum_{i=2}^{n} \sum_{j=2}^{n} \Big[\rho\Big(\epsilon_{ij}\big(\alpha + a_n^{-1} n^{-1/2} u_1, \beta + a_n^{-1} u_2\big)\Big) - \rho\big(\epsilon_{ij}(\alpha, \beta)\big)\Big] \\ &= \sum_{i=2}^{n} \sum_{j=2}^{n} \Big[\rho\big(\epsilon_{ij} - a_n^{-1} n^{-1/2} u_1 X_{i-1,j} - a_n^{-1} u_2 Y_{i,j-1} \\ &\quad + a_n^{-2} n^{-1/2} u_1 u_2 Z_{i-1,j-1}\big) - \rho(\epsilon_{ij})\Big]. \end{aligned}$$

Notice that $K_n(u_1, u_2)$ has a unique minimum at

$$(u_1, u_2) = \big(\sqrt{n}\, a_n(\hat{\alpha} - 1), a_n(\hat{\beta} - \beta)\big).$$

First we show that

$$K_n(u_1, u_2) = K_n^{(1)}(u_1) + K_n^{(2)}(u_2) + o_p(1), \tag{3.13}$$

where

$$K_n^{(1)}(u_1) = \sum_{i=2}^{n}\sum_{j=2}^{n} \rho_{ij}\big(a_n^{-1}n^{-1/2}u_1X_{i-1,j}\big)$$

and

$$K_n^{(2)}(u_2) = \sum_{i=2}^{n}\sum_{j=2}^{n} \rho_{ij}\big(a_n^{-1}u_2Y_{i,j-1}\big),$$

with $\rho_{ij}(u) = \rho(\epsilon_{ij} - u) - \rho(\epsilon_{ij})$. We have

$$\sum_{i=2}^{n}\sum_{j=2}^{n}\Big[\rho_{ij}\big(a_n^{-1}n^{-1/2}u_1X_{i-1,j} + a_n^{-1}u_2Y_{i,j-1} - a_n^{-2}n^{-1/2}u_1u_2Z_{i-1,j-1}\big)$$
$$- \rho_{ij}\big(a_n^{-1}n^{-1/2}u_1X_{i-1,j}\big)\Big] =$$
$$-\sum_{i=2}^{n}\sum_{j=2}^{n}\int_0^{a_n^{-1}u_2Y_{i,j-1}-a_n^{-2}n^{-1/2}u_1u_2Z_{i-1,j-1}} \psi(\epsilon_{ij} - a_n^{-1}n^{-1/2}u_1X_{i-1,j} - s)\,ds,$$

and

$$\sum_{i=2}^{n}\sum_{j=2}^{n}\rho_{ij}\big(a_n^{-1}u_2Y_{i,j-1} - a_n^{-2}n^{-1/2}u_1u_2Z_{i-1,j-1}\big) =$$
$$-\sum_{i=2}^{n}\sum_{j=2}^{n}\int_0^{a_n^{-1}u_2Y_{i,j-1}-a_n^{-2}n^{-1/2}u_1u_2Z_{i-1,j-1}} \psi(\epsilon_{ij} - s)\,ds.$$

Using Lipschitz continuity of ψ, we get

$$\sum_{i=2}^{n}\sum_{j=2}^{n}\int_0^{a_n^{-1}u_2Y_{i,j-1}-a_n^{-2}n^{-1/2}u_1u_2Z_{i-1,j-1}} \big|\psi(\epsilon_{ij} - a_n^{-1}n^{-1/2}u_1X_{i-1,j} - s) - \psi(\epsilon_{ij} - s)\big|\,ds$$
$$\leq k_1\sum_{i=2}^{n}\sum_{j=2}^{n}\big|a_n^{-1}u_2Y_{i,j-1} - a_n^{-2}n^{-1/2}u_1u_2Z_{i-1,j-1}\big|\,\big| - a_n^{-1}n^{-1/2}u_1X_{i-1,j}\big|^{\lambda_1}$$
$$\leq k_1 \sup_{2\leq i\leq n}\Big|u_1^{\lambda_1}a_n^{-\lambda_1}\sum_{j=2}^{n}X_{i-1,j}^{\lambda_1}\Big|o_p(1) \rightarrow_p 0,$$

noting that

$$\sup_{2\leq i\leq n}\Big|u_1^{\lambda_1}a_n^{-\lambda_1}\sum_{j=2}^{n}X_{i-1,j}^{\lambda_1}\Big| = O_p(1).$$

Therefore,

$$K_n(u_1,u_2) = K_n^{(1)}(u_1) + \sum_{i=2}^{n}\sum_{j=2}^{n} \rho_{ij}\big(a_n^{-1}u_2 Y_{i,j-1} - a_n^{-2}n^{-1/2}u_1u_2 Z_{i-1,j-1}\big) + o_p(1).$$

Continuing in this manner we obtain

$$K_n(u_1,u_2) = K_n^{(1)}(u_1) + K_n^{(2)}(u_2) + \sum_{i=2}^{n}\sum_{j=2}^{n} \rho_{ij}\big(- a_n^{-2}n^{-1/2}u_1u_2 Z_{i-1,j-1}\big) + o_p(1),$$

and finally since

$$a_n^{-2}n^{-1/2}\sum_{i=2}^{n}\sum_{j=2}^{n} \rho_{ij}\big(- u_1u_2 Z_{i-1,j-1}\big) = o_p(1)$$

we get (3.13). Now, by an argument similar to that in Theorem 2 of Knight [36] it can be shown that $K_n^{(1)}(u_1) \to_{fidi} K^{(1)}(u_1)$, where

$$K^{(1)}(u_1) = -\sigma u_1 \int_0^1\int_0^1 S_{\alpha_0}(t_1,t_2)\,dW(t_1,t_2) + \frac{\gamma}{2}u_1^2 \int_0^1\int_0^1 S_{\alpha_0}^2(t_1,t_2)\,dt_1dt_2. \quad (3.14)$$

To see this, using the Taylor series expansion of $K_n^{(1)}(u_1)$ around $u_1 = 0$ observe that

$$K_n^{(1)}(u_1) = -u_1 a_n^{-1}n^{-1/2}\sum_{i=2}^{n}\sum_{j=2}^{n} X_{i-1,j}\psi(\epsilon_{ij}) + \frac{1}{2}u_1^2 a_n^{-2}n^{-1}\sum_{i=2}^{n}\sum_{j=2}^{n} X_{i-1,j}^2\psi'(\epsilon_{ij}^*),$$

where $\left|\psi'(\epsilon_{ij}) - \psi'(\epsilon_{ij}^*)\right| \leq k_2\left|u_1a_n^{-1}n^{-1/2}X_{i-1,j}\right|$. Therefore, we can replace $\psi'(\epsilon_{ij}^*)$ by $\psi'(\epsilon_{ij})$ since

$$u_1^2a_n^{-2}n^{-1}\sum_{i=2}^{n}\sum_{j=2}^{n} X_{i-1,j}^2\left|\psi'(\epsilon_{ij}^*) - \psi'(\epsilon_{ij})\right| \leq k_2u_1^2n^{-1/2}\times \left(n^{-1}\sum_{i=2}^{n}a_n^{-3}\sum_{j=2}^{n}|X_{i-1,j}|^3\right) \to_p 0$$

uniformly over u_1 in compact sets.

Moreover, in the limit each $\psi'(\epsilon_{ij})$ can be substituted by $E\left(\psi'(\epsilon_{ij})\right)$. That is, we shall show that

$$u_1^2a_n^{-2}n^{-1}\sum_{i=2}^{n}\sum_{j=2}^{n} X_{i-1,j}^2\left[\psi'(\epsilon_{ij}) - E(\psi'(\epsilon_{ij}))\right] \to_p 0$$

uniformly over u_1 in compact sets. Denote the set of all ordered pairs of positive integers by I, and define $F = \{(t_1, t_2) \in I : t_1, t_2 \leq n\}$ for a fixed and positive integer n. Denote

$$U_t = \sum_{i=1}^{t_1} \sum_{j=1}^{t_2} \big(c_1 X_{i-1,j} + c_2 Y_{i,j-1}\big) I(c_1 X_{i-1,j} + c_2 Y_{i,j-1} \leq M) \big[\psi'(\epsilon_{ij}) - E(\psi'(\epsilon_{ij}))\big],$$

for each $t = (t_1, t_2) \in F$, and given constants c_1 and c_2. Let $\mathfrak{F}_t$ be the smallest σ-field making each ϵ_{ij} measurable, for $1 \leq i \leq t_1$ and $1 \leq j \leq t_2$. Then similar to Lemma 2.4 of [8], it is easy to see that $\{U_t, \mathfrak{F}_t, t \in F\}$ is a strong martingale (see Lemma 4.2 of the next chapter). Therefore, applying a weak law of large numbers for martingales (Brown [13] and McLeish [43]) we get

$$n^{-1} \sum_{i=2}^{n} a_n^{-2} \sum_{j=2}^{n} X_{i-1,j}^2 I(a_n^{-2} X_{i-1,j}^2 \leq M) \left[\psi'(\epsilon_{ij}) - E(\psi'(\epsilon_{ij}))\right] \rightarrow_p 0.$$

The other term, the approximation error, is an $o_p(1)$ term:

$$P\left[\left| n^{-1} a_n^{-2} \sum_{i=2}^{n} \sum_{j=2}^{n} X_{i-1,j}^2 I(a_n^{-2} X_{i-1,j}^2 > M) \left[\psi'(\epsilon_{ij}) - E(\psi'(\epsilon_{ij}))\right] \right| > \delta \right]$$
$$\leq P\left[\max_{1 \leq i \leq n} a_n^{-2} \sum_{j=2}^{n} X_{i-1,j}^2 > M\right] \rightarrow 0,$$

as $n \rightarrow \infty$ first and then $M \rightarrow \infty$. Thus $K_n^{(1)}(u_1) \rightarrow_{fidi} K^{(1)}(u_1)$, with $K^{(1)}(u_1)$ as given in (3.14).

Also, an application of Theorem 3.1 of Section 3.2 implies $K_n^{(2)}(u_2) \rightarrow_{fidi} K^{(2)}(u_2)$, where

$$K^{(2)}(u_2) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \rho_{ij}\big(u_2 \beta^{j-1} \delta_i \Gamma_i^{-1/\alpha_0}\big).$$

Minimizing $K^{(1)}(u_1)$ and $K^{(2)}(u_2)$ with respect to u_1 and u_2, respectively, we obtain the desired result.

Chapter 4

M-estimation for Near Unit Roots in Spatial Autoregression with Infinite Variance

4.1 Introduction and main result

Unit root and near unit root models have been investigated by several researchers. Beginning with the well known work of Dickey and Fuller [27] on unit root time series models, the work of Phillips [46] and Chan and Tran [17] focus on the near unit root model $x_t(n) = \alpha_n x_{t-1}(n) + \epsilon_t$, $1 \le t \le n$, where α_n is an unknown sequence of constants approaching to 1 as the sample size n gets large. They derive asymptotic behavior of the least squares (LS) estimates of the sequence α_n in a model with finite variance. Knight [36] investigates the asymptotic behavior of the M-estimates when errors are in the domain of attraction of stable laws.

As it is mentioned in Chapter 3, spatial models appear in many applications such as geography, agriculture, geology, biology and economics. Basu and Reinsel [4, 5] define and study the spatial unilateral autoregressive moving average (ARMA) model of first order which encompasses the spatial AR model of first order as a special case. Examples given by Cullis and Gleeson [21], and Basu and Reinsel [5] show the applicability of the above spatial AR model having unit roots ($\alpha = \beta = 1$) and one near unit root ($\beta = .947$), respectively. Bhattacharyya, Richardson and Franklin [8] consider the spatial autoregressive model

$$Z_{ij}(n) = \alpha_n Z_{i-1,j}(n) + \beta_n Z_{i,j-1}(n) - \alpha_n \beta_n Z_{i-1,j-1}(n) + \epsilon_{ij}, \tag{4.1}$$

$1 \leq i, j \leq n$, when α_n and β_n are near unit roots. They investigate Gauss-Newton estimators of (α_n, β_n) under some assumptions on the existence of second and fourth moments, and on the rate of convergence of initial estimators. Later, Bhattacharyya, Li, Pensky and Richardson [7] introduce a statistic (the normalized periodogram ordinate) to test for unit roots in Model (4.1) under the same assumptions made as before.

The M-estimator $(\hat{\alpha}_n, \hat{\beta}_n)$ of (α_n, β_n) in (4.1) is the minimizer of the objective function

$$g(\alpha_n, \beta_n) = \sum_{i=2}^{n} \sum_{j=2}^{n} \rho(Z_{ij} - \alpha_n Z_{i-1,j} - \beta_n Z_{i,j-1} + \alpha_n \beta_n Z_{i-1,j-1}) \tag{4.2}$$

for some function $\rho(\cdot)$. In (4.2) and thereafter, expressions such as $Z_{ij}(n)$ are denoted simply by $Z_{ij}, 1 \leq i, j \leq n$. In this chapter, we study the weak limit behavior of the M-estimators for the parameters in Model (4.1) when $\{\epsilon_{ij}\}$ is in the domain of attraction of a symmetric stable law with index of stability α, $0 < \alpha \leq 2$, denoted by $DS(\alpha)$.

Similar to Chapter 3, define the stochastic process

$$S_n(t,s) = a_n^{-1} \sum_{i=1}^{[nt]} \sum_{j=1}^{[ns]} \epsilon_{ij}, \quad 0 < t, s \leq 1,$$

where $[x]$ stands for integer part of x and a_n is a normalizing constant. Let D_2 be the space of càdlàg functions defined on the unit square $K := [0,1] \times [0,1]$ and equipped with the metric introduced by Bickel and Wichura [9]. This metric on D_q induces Skorohod's topology when $q = 1$, and D_2 is separable and complete (see Appendix for more details). The main assumption on $\{\epsilon_{ij}\}$ is the following:

B1:

$$S_n(t,s) \rightarrow_d S_\alpha(t,s), \quad \text{as } n \rightarrow \infty, \tag{4.3}$$

in D_2, where the limiting process is a stable sheet ($\alpha \in (0,2]$) defined on a regular rectangular grid in two dimensions.

Further assume that

B2: $\alpha_n = e^{c/n}$ and $\beta_n = e^{d/n}$, where c and d are nonzero unknown constants.

B3: $Z_{ij}(n) = 0$ when either $i \leq 0$ or $j \leq 0$.

Recall that the M-estimator $(\hat{\alpha}_n, \hat{\beta}_n)$ of (α_n, β_n) in Model (4.1) is the minimizer of the objective function given in (4.2), where

B4: $\rho(\cdot)$ is a convex and twice differentiable function, and Assumptions A1 through A3 of Chapter 3 (with α_0 replaced by α) hold. The principal result of this chapter is now given.

Theorem 4.1. Consider Model (4.1) and suppose that Assumptions A1-A3 of Chapter 3 and B1-B4 are satisfied. Then

$$a_n\sqrt{n}\begin{pmatrix} \hat{\alpha}_n - \alpha_n \\ \hat{\beta}_n - \beta_n \end{pmatrix} \to_d \eta := \frac{\sigma}{\gamma}\begin{pmatrix} \frac{\int_0^1\int_0^1 J_1(t,s)\,dW(t,s)}{\int_0^1\int_0^1 J_1^2(t,s)\,dtds} \\ \frac{\int_0^1\int_0^1 J_2(t,s)\,dW(t,s)}{\int_0^1\int_0^1 J_2^2(t,s)\,dtds} \end{pmatrix}, \tag{4.4}$$

where

$$J_1(t,s) := S_\alpha(t,s) + c\int_0^t S_\alpha(x,s)e^{(t-x)c}\,dx, \quad 0 < t, s \le 1,$$

$$J_2(t,s) := S_\alpha(t,s) + d\int_0^s S_\alpha(t,y)e^{(s-y)d}\,dy, \quad 0 < t, s \le 1,$$

and $S_\alpha(\cdot,\cdot)$ is the α-stable sheet appeared in (4.3) and $W(\cdot,\cdot)$ is a standard Brownian sheet.

Remark 4.1. Similar to the results given in Remark 3.1 it is possible to give self-normalizing coefficients for M-estimators in Theorem 4.1. That is, when ϵ_{11} is in $DS(\alpha)$, $0 < \alpha \le 2$, we get:

$$\begin{pmatrix} \sqrt{\sum_{i=2}^n\sum_{j=1}^n X_{i-1,j}^2} & 0 \\ 0 & \sqrt{\sum_{i=1}^n\sum_{j=2}^n Y_{i,j-1}^2} \end{pmatrix}\begin{pmatrix} \hat{\alpha}_n - \alpha_n \\ \hat{\beta}_n - \beta_n \end{pmatrix} \to_d N(\mathbf{0}, \mathbf{\Gamma}), \tag{4.5}$$

where $\mathbf{\Gamma} = \operatorname{diag}\left(\frac{\sigma^2}{\gamma^2}, \frac{\sigma^2}{\gamma^2}\right)$, $X_{ij} := Z_{ij} - \beta_n Z_{i,j-1}$, and $Y_{ij} := Z_{ij} - \alpha_n Z_{i-1,j}$. Since $(\hat{\alpha}_n, \hat{\beta}_n)$ is a consistent estimator of (α_n, β_n), $X_{i-1,j}$ and $Y_{i,j-1}$ can be replaced by $\hat{X}_{i-1,j}$ and $\hat{Y}_{i,j-1}$ where $\hat{X}_{ij} := Z_{ij} - \hat{\beta}_n Z_{i,j-1}$, and $\hat{Y}_{ij} := Z_{ij} - \hat{\alpha}_n Z_{i-1,j}$.

Remark 4.2. A similar result can be derived for the special case when $\rho(x) = |x|$, corresponding to the least absolute deviation (LAD) estimator.

4.2 Proof of the main theorem

Let $K := [0,1]^2$ and $S_\alpha(t,s)$, $(t,s) \in K$ denote a stable sheet on K. Define $X_{ij} := Z_{ij} - \beta_n Z_{i,j-1}$ and $Y_{ij} := Z_{ij} - \alpha_n Z_{i-1,j}$, and observe using (4.1) that $X_{ij} = \alpha_n X_{i-1,j} + \epsilon_{ij}$ and $Y_{ij} = \beta_n Y_{i,j-1} + \epsilon_{ij}$. Now, it follows from B2 and B3 that

$$X_{ij} = \sum_{k=1}^{i} \alpha_n^{i-k} \epsilon_{kj} \quad \text{and} \quad Y_{ij} = \sum_{l=1}^{j} \beta_n^{j-l} \epsilon_{il}. \tag{4.6}$$

The following Lemma is used to verify Theorem 4.1. Define the random elements J_n and J by

$$J_n(t,s) := a_n^{-1} \sum_{i=1}^{[nt]} \sum_{j=1}^{[ns]} \alpha_n^{[nt]-i} \beta_n^{[ns]-j} \epsilon_{ij} \tag{4.7}$$

and

$$\begin{aligned} J(t,s) := S_\alpha(t,s) &+ c \int_0^t S_\alpha(x,s) e^{(t-x)c}\, dx \\ &+ d \int_0^s S_\alpha(t,y) e^{(s-y)d}\, dy + cd \int_0^t \int_0^s S_\alpha(x,y) e^{(t-x)c} e^{(s-y)d}\, dxdy, \end{aligned} \tag{4.8}$$

which is a stable process with index of stability α (Remark 4.3).

Lemma 4.1. Assume that B1 and B2 are satisfied. Then $J_n \to_d J$ in D_2, where J_n and J are defined as in (4.7) and (4.8).

Proof. For $\alpha = 2$, see [7]. Take $\alpha < 2$ and let $\eta_0 = 1$ and $\eta_k \downarrow 0$ as $k \to \infty$. Set for any $\eta > 0$,

$$S_n^{(\eta)}(t,s) := a_n^{-1} \sum_{i=1}^{[nt]} \sum_{j=1}^{[ns]} \epsilon_{ij} I\left(a_n^{-1}|\epsilon_{ij}| > \eta\right),$$

$$J_n^{(\eta)}(t,s) := a_n^{-1} \sum_{i=1}^{[nt]} \sum_{j=1}^{[ns]} \alpha_n^{[nt]-i} \beta_n^{[ns]-j} \epsilon_{ij} I\left(a_n^{-1}|\epsilon_{ij}| > \eta\right)$$

and define $h(x,y) = e^{(t-x)c}e^{(s-y)d}$, where (t,s), $(x,y) \in K$. Using the relation

$$a_n^{-1}e^{(t-i/n)c}e^{(s-j/n)d}\epsilon_{ij}I\left(a_n^{-1}|\epsilon_{ij}| > \eta\right) = \int_{(i-1)/n}^{i/n}\int_{(j-1)/n}^{j/n} h(x,y)\,dS_n^{(\eta)}(x,y),$$

Model (4.1) and B2 observe that

$$\begin{aligned} J_n^{(\eta)}(t,s) =& a_n^{-1}\sum_{i=1}^{[nt]}\sum_{j=1}^{[ns]}\left(e^{([nt]-i)c/n}e^{([ns]-j)d/n} - e^{(t-i/n)c}e^{(s-j/n)d}\right)\epsilon_{ij}I\left(a_n^{-1}|\epsilon_{ij}| > \eta\right) \\ &+ \int_0^{[nt]/n}\int_0^{[ns]/n} h(x,y)\,dS_n^{(\eta)}(x,y), \end{aligned}$$

and therefore

$$J_n^{(\eta)}(t,s) = \int_0^t\int_0^s h(x,y)\,dS_n^{(\eta)}(x,y) + o_p(1).$$

Now, noting that both functions $S_n^{(\eta)}$ and h are of bounded variation and making use of the integration by parts formula given in Appendix we get

$$\int_0^t\int_0^s h(x,y)\,dS_n^{(\eta)}(x,y) = T_0 + T_1 + T_2,$$

where

$$T_0 = S_n^{(\eta)}(t,s)h(t,s) - S_n^{(\eta)}(t,0)h(t,0) - S_n^{(\eta)}(0,s)h(0,s) + S_n^{(\eta)}(0,0)h(0,0) = S_n^{(\eta)}(t,s),$$

$$\begin{aligned} T_1 =& -\int_0^t S_n^{(\eta)}(x,s)\,dh(x,s) + \int_0^t S_n^{(\eta)}(x,0)\,dh(x,0) - \int_0^s S_n^{(\eta)}(t,y)\,dh(t,y) \\ &+ \int_0^s S_n^{(\eta)}(0,y)\,dh(0,y) \\ =& \, c\int_0^t S_n^{(\eta)}(x,s)h(x,s)\,dx + d\int_0^s S_n^{(\eta)}(t,y)h(t,y)\,dy \end{aligned}$$

and

$$T_2 = \int_0^t\int_0^s S_n^{(\eta)}(x,y)\,dh(x,y) = cd\int_0^t\int_0^s S_n^{(\eta)}(x,y)h(x,y)\,dxdy.$$

Thus,

$$J_n^{(\eta)}(t,s) = S_n^{(\eta)}(t,s) + c\int_0^t S_n^{(\eta)}(x,s)h(x,s)\,dx$$
$$+ d\int_0^s S_n^{(\eta)}(t,y)h(t,y)\,dy + cd\int_0^t\int_0^s S_n^{(\eta)}(x,y)h(x,y)\,dxdy + o_p(1).$$

Now using a similar approach to that described in Section 1.3 (see also [49], page 111) it can be shown that $S_n^{(\eta)}(\cdot,\cdot) \to_d S_\alpha^{(\eta)}(\cdot,\cdot)$ in D_2, where

$$S_\alpha^{(\eta)}(\cdot,\cdot) = \sum_{\mathbf{t}_i \le (\cdot,\cdot)} \delta_i \Gamma_i^{-1/\alpha} I\left(|\Gamma_i^{-1/\alpha}| > \eta\right),$$

with $\mathbf{t}_i \in \mathbb{R}_+^2$ and $\{\delta_i\}$ and $\{\Gamma_i\}$ are as specified in Section 3.2.

Therefore, from continuous mapping theorem it follows for each fixed $(t,s) \in K$ and $\eta > 0$ that

$$J_n^{(\eta)}(t,s) \to_d J_\alpha^{(\eta)}(t,s),$$

in D_2, where

$$J_\alpha^{(\eta)}(t,s) := S_\alpha^{(\eta)}(t,s) + c\int_0^t S_\alpha^{(\eta)}(x,s)e^{(t-x)c}\,dx$$
$$+ d\int_0^s S_\alpha^{(\eta)}(t,y)e^{(s-y)d}\,dy + cd\int_0^t\int_0^s S_\alpha^{(\eta)}(x,y)e^{(t-x)c}e^{(s-y)d}\,dxdy.$$

Also, using $S_\alpha^{(\eta_k)}(\cdot,\cdot) \to_d S_\alpha(\cdot,\cdot)$ in D_2 as $k \to \infty$, we have

$$J_\alpha^{(\eta_k)}(t,s) \to_d J(t,s).$$

Thus, it suffices (Billingsley [10], Theorem 4.2) to show for $c > 0$ that

$$\lim_{k\to\infty}\limsup_{n\to\infty} P\Big[\sup_{(t,s)\le(1,1)} |J_n^{(\eta_k)}(t,s) - J_n(t,s)| > 4c\Big] = 0$$

or equivalently

$$\lim_{k\to\infty}\limsup_{n\to\infty} P\Big[\sup_{(t,s)\le(1,1)} \Big|\sum_{i=1}^{[nt]}\sum_{j=1}^{[ns]} a_n^{-1}\alpha_n^{[nt]-i}\beta_n^{[ns]-j}\epsilon_{ij} I\left(a_n^{-1}|\epsilon_{ij}| \le \eta_k\right)\Big| > 4c\Big]$$
$$= 0. \quad (4.9)$$

By Wichura's [57] maximal inequality, the probability in (4.9) is for large k

$$O\Big(\mathrm{Var}\sum_{i=1}^{n}\sum_{j=1}^{n} a_n^{-1}\alpha_n^{n-i}\beta_n^{n-j}\epsilon_{ij} I\left(a_n^{-1}|\epsilon_{ij}| \leq \eta_k\right)\Big).$$

Now an application of Karamata's theorem (apply Theorem 1.2 with $\gamma = 2$) implies

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_n^{-1}\alpha_n^{n-i}\beta_n^{n-j}\epsilon_{ij} I\left(a_n^{-1}|\epsilon_{ij}| \leq \eta_k\right)\right) &\leq n^2 E\big\{a_n^{-2}\epsilon_{11}^2 I\left(a_n^{-1}|\epsilon_{11}| \leq \eta_k\right)\big\}\\ &\leq \alpha(2-\alpha)^{-1} n^2 P\big(|\epsilon_{11}| > a_n\eta_k\big)\eta_k^2\\ &\to \alpha(2-\alpha)^{-1}\eta_k^{2-\alpha} \quad \text{as } n \to \infty,\end{aligned}$$

and the last expression converges to 0 as $k \to \infty$. This finishes the proof.

As a consequence of Lemma 4.1, as $d \to 0$ we get

$$a_n^{-1}\sum_{j=1}^{[ns]} X_{[nt],j} \to_d J_1(t,s) := S_\alpha(t,s) + c\int_0^t S_\alpha(x,s)e^{(t-x)c}\,dx, \quad 0 < t, s \leq 1, \quad (4.10)$$

and as $c \to 0$,

$$a_n^{-1}\sum_{i=1}^{[nt]} Y_{i,[ns]} \to_d J_2(t,s) := S_\alpha(t,s) + d\int_0^s S_\alpha(t,y)e^{(s-y)d}\,dy, \quad 0 < t, s \leq 1, \quad (4.11)$$

where the processes X and Y are defined as in (4.6).

As in Chapter 3, denote the set of all ordered pairs of positive integers by I, and define $F = \{(t_1, t_2) \in I : t_1 \leq n,\ t_2 \leq n\}$ for a fixed and positive integer n. Set

$$U_t := \sum_{i=1}^{t_1}\sum_{j=1}^{t_2}\big(c_1 X_{i-1,j} + c_2 Y_{i,j-1}\big) I(c_1 X_{i-1,j} + c_2 Y_{i,j-1} \leq M)\big[\psi'(\epsilon_{ij}) - E(\psi'(\epsilon_{ij}))\big],$$

for each $t = (t_1, t_2) \in F$, and given constants c_1 and c_2. Let $\mathfrak{F}_t$ be the smallest σ-field making each ϵ_{ij} measurable, for $1 \leq i \leq t_1$ and $1 \leq j \leq t_2$. It is easy to see that

$\{U_t(n), \mathfrak{F}_t(n), t \in F\}$ is a strong martingale array (see Lemma 4.2 in Appendix). Now, we are ready to prove Theorem 4.1.

Proof of Theorem 4.1. Since $\rho(\cdot)$ is a convex function, it is enough to show the weak convergence for the finite dimensional distribution (denoted by $\rightarrow_{fidi}$). Recall the processes X and Y from (4.6) and observe that

$$\epsilon_{ij}\big(\alpha_n + a_n^{-1}n^{-1/2}u_1, \beta_n + a_n^{-1}n^{-1/2}u_2\big) = \epsilon_{ij}(\alpha_n, \beta_n) - a_n^{-1}n^{-1/2}u_1X_{i-1,j} - a_n^{-1}n^{-1/2}u_2Y_{i,j-1} + a_n^{-2}n^{-1}u_1u_2Z_{i-1,j-1},$$

where

$$\epsilon_{ij}(\alpha_n, \beta_n) = Z_{ij} - \alpha_n Z_{i-1,j} - \beta_n Z_{i,j-1} + \alpha_n\beta_n Z_{i-1,j-1}.$$

Thus, we shall minimize the function

$$\begin{aligned} K_n(u_1, u_2) &= \sum_{i=2}^{n}\sum_{j=2}^{n}\Big[\rho\Big(\epsilon_{ij}\big(\alpha_n + a_n^{-1}n^{-1/2}u_1, \beta_n + a_n^{-1}n^{-1/2}u_2\big)\Big) - \rho\big(\epsilon_{ij}(\alpha_n, \beta_n)\big)\Big] \\ &= \sum_{i=2}^{n}\sum_{j=2}^{n}\Big[\rho\big(\epsilon_{ij} - a_n^{-1}n^{-1/2}u_1X_{i-1,j} - a_n^{-1}n^{-1/2}u_2Y_{i,j-1} \\ &\quad + a_n^{-2}n^{-1}u_1u_2Z_{i-1,j-1}\big) - \rho(\epsilon_{ij})\Big]. \end{aligned}$$

Notice that $K_n(u_1, u_2)$ has a unique minimum at

$$(u_1, u_2) = \sqrt{n}\, a_n\big(\hat{\alpha}_n - \alpha_n, \hat{\beta}_n - \beta_n\big).$$

First similar to what we did to prove (3.13) we get

$$K_n(u_1, u_2) = K_n^{(1)}(u_1) + K_n^{(2)}(u_2) + o_p(1),$$

where

$$K_n^{(1)}(u_1) = \sum_{i=2}^{n}\sum_{j=2}^{n}\rho_{ij}\big(a_n^{-1}n^{-1/2}u_1X_{i-1,j}\big)$$

and

$$K_n^{(2)}(u_2) = \sum_{i=2}^{n}\sum_{j=2}^{n} \rho_{ij}\big(a_n^{-1}n^{-1/2}u_2Y_{i,j-1}\big),$$

with $\rho_{ij}(u) = \rho(\epsilon_{ij} - u) - \rho(\epsilon_{ij})$. Now, by a quite similar argument as in the proof of Theorem 3.2 of Chapter 3, using the fact that $\{U_t(n), \mathfrak{F}_t(n), t \in F\}$ is a strong martingale array, it can be shown that $K_n^{(1)}(u_1) \rightarrow_{fidi} K^{(1)}(u_1)$, where

$$K^{(1)}(u_1) = -\sigma u_1 \int_0^1\int_0^1 J_1(t_1,t_2)\,dW(t_1,t_2) + \frac{\gamma}{2}u_1^2 \int_0^1\int_0^1 J_1^2(t_1,t_2)\,dt_1dt_2$$

and $K_n^{(2)}(u_2) \rightarrow_{fidi} K^{(2)}(u_2)$, where

$$K^{(2)}(u_2) = -\sigma u_2 \int_0^1\int_0^1 J_2(t_1,t_2)\,dW(t_1,t_2) + \frac{\gamma}{2}u_2^2 \int_0^1\int_0^1 J_2^2(t_1,t_2)\,dt_1dt_2,$$

and J_1 and J_2 are defined as in (4.10) and (4.11), respectively.

Finally, minimizing $K^{(1)}(u_1)$ and $K^{(2)}(u_2)$ with respect to u_1 and u_2, respectively, we complete the proof.

4.3 Simulation

A simulation study is given to determine the speed of convergence of $a_n\sqrt{n}(\hat{\alpha}_n - \alpha_n)$ and $a_n\sqrt{n}(\hat{\beta}_n - \beta_n)$ to the limits as given in Theorem 4.1. Since $\hat{\alpha}_n$ and $\hat{\beta}_n$ have the same marginal asymptotic properties, only averages and mean standard deviations of $\hat{\alpha}_n - \alpha_n$, based on a replication size of 1000, are reported in the tables. The near unit root Model (4.1) with symmetric α-stable errors on an $n \times n$ regular grid with different values of n are generated. Here, we present the results for five parameter settings for $\alpha = 2, 1.8, 1.5, 1, 0.5$ and the four combinations $c = -1$, $d = 1$ and $c = d = -1$ with Huber's loss function (Table 4.1 and Table 4.3), and $c = -1$, $d = 1$

and $c = d = -1$ with $\rho(x) = 2\log(x^2 + 3)$ as a loss function in M-estimation (Table 4.2 and Table 4.4). Observe that though the loss function $\rho(x) = 2\log(x^2 + 3)$ is just convex around its global minimum, the results tabulated in Table 4.2 and Table 4.4 are quite satisfactory. As n gets large, the M-estimates are in proximity of the actual values for both choices for loss functions with better rates of convergence as α gets smaller.

A simulated near unit root model with normal innovations using a replication size of 1000, and $\rho(x) = |x|$ as the loss function in the M-estimation confirms the bivariate normality of the limit given in (4.5) after using the random normalizer. Simulations also show that to get bivariate normality we need to take $n = 60$. This is in accordance with the simulation results for normal innovations given in [8] as a special case compared to our results. Moreover, for $\alpha < 2$ we are still able to have bivariate normality. For instance, for a model with Cauchy innovations ($\alpha = 1$) a sample size of $n = 500$ suffices to get bivariate normality. Since results are similar to those given in Section 3.3, details and related plots for different combinations of c and d as well as different loss functions are not reported.

Table 4.1: Mean and mean standard deviations (in parentheses) for M-estimates of $(\hat{\alpha}_n - \alpha_n)$ for the near unit root Model (4.1) with $c = -1$, $d = 1$ using Huber's loss function with different α-stable noises. The replication size is 1000.

	index of stability α				
n	2	1.8	1.5	1	0.5
5	-0.0142	-0.0122	-0.0011	-0.0138	-0.0611
	(0.0302)	(0.0316)	(0.0519)	(0.1144)	(0.2406)
10	-0.0006	-0.0021	-0.0005	-0.0006	-0.0048
	(0.0068)	(0.0066)	(0.0068)	(0.0332)	(0.0548)
15	-0.0004	-0.0006	-0.0005	-0.0007	-0.0029
	(0.0032)	(0.0031)	(0.0030)	(0.0301)	(0.0520)
20	-0.0003	-0.0004	-0.0004	-0.0001	-0.0004
	(0.0017)	(0.0018)	(0.0017)	(0.0007)	(0.0141)
25	-0.0003	-0.0003	-0.0004	-0.0008	-0.0000
	(0.0012)	(0.0011)	(0.0013)	(0.0300)	(0.0001)
30	-0.0002	-0.0003	-0.0002	-0.0000	-0.0000
	(0.0009)	(0.0008)	(0.0007)	(0.0003)	(0.0221)

Table 4.2: Mean and mean standard deviations (in parentheses) for M-estimates of $(\hat{\alpha}_n - \alpha_n)$ for the near unit root Model (4.1) with $c = -1$, $d = 1$ using the loss function $\rho(x) = 2\log(x^2 + 3)$ with different α-stable noises. The replication size is 1000.

	index of stability α				
n	2	1.8	1.5	1	0.5
5	-0.0064	-0.0066	-0.0040	-0.0036	0.0033
	(0.0264)	(0.0282)	(0.0282)	(0.0264)	(0.0084)
10	0.0012	0.0016	0.0012	0.0002	0.0002
	(0.0066)	(0.0073)	(0.0062)	(0.0041)	(0.0003)
15	0.0011	0.0010	0.0008	0.0001	0.0000
	(0.0033)	(0.0032)	(0.0030)	(0.0014)	(0.0001)
20	0.0008	0.0007	0.0006	0.0001	0.0000
	(0.0020)	(0.0020)	(0.0018)	(0.0006)	(0.0000)
25	0.0006	0.0005	0.0004	0.0000	0.0000
	(0.0014)	(0.0013)	(0.0011)	(0.0003)	(0.0000)
30	0.0004	0.0004	0.0003	0.0000	0.0000
	(0.0010)	(0.0009)	(0.0008)	(0.0002)	(0.0000)

Table 4.3: Mean and mean standard deviations (in parentheses) for M-estimates of $(\hat{\alpha}_n - \alpha_n)$ for the near unit root Model (4.1) with $c = d = -1$ using Huber's loss function with different α-stable noises. The replication size is 1000.

	index of stability α				
n	2	1.8	1.5	1	0.5
5	-0.0497	-0.0474	-0.0412	-0.0377	-0.0215
	(0.0519)	(0.0489)	(0.0487)	(0.1104)	(0.1894)
10	-0.0095	-0.0085	-0.0063	-0.0031	-0.0057
	(0.0109)	(0.0109)	(0.0096)	(0.0173)	(0.0911)
15	-0.0036	-0.0032	-0.0020	-0.0008	-0.0019
	(0.0047)	(0.0043)	(0.0035)	(0.0019)	(0.0346)
20	-0.0018	-0.0016	-0.0011	-0.0003	-0.0036
	(0.0025)	(0.0022)	(0.0019)	(0.0010)	(0.0548)
25	-0.0011	-0.0010	-0.0006	-0.0002	-0.00003
	(0.0015)	(0.0014)	(0.0012)	(0.0006)	(0.0374)
30	-0.0007	-0.0006	-0.0004	-0.00005	-0.00009
	(0.0010)	(0.0009)	(0.0007)	(0.0003)	(0.0021)

Table 4.4: Mean and mean standard deviations (in parentheses) for M-estimates of $(\hat{\alpha}_n - \alpha_n)$ for the near unit root Model (4.1) with $c = d = -1$ using the loss function $\rho(x) = 2\log(x^2 + 3)$ with different α-stable noises. The replication size is 1000.

	index of stability α				
n	2	1.8	1.5	1	0.5
5	-0.0374	-0.0353	-0.0266	-0.0202	0.00791
	(0.0448)	(0.0436)	(0.0447)	(0.0489)	(0.0145)
10	-0.0070	-0.0059	-0.0045	-0.0017	0.00039
	(0.0092)	(0.0087)	(0.0077)	(0.0053)	(0.0007)
15	-0.0027	-0.0022	-0.0016	-0.0006	0.00006
	(0.0039)	(0.0035)	(0.0030)	(0.0017)	(0.0001)
20	-0.0013	-0.0011	-0.0008	-0.0001	0.00002
	(0.0020)	(0.0019)	(0.0015)	(0.0008)	(0.0000)
25	-0.0009	-0.0007	-0.0004	-0.0001	0.00001
	(0.0013)	(0.0012)	(0.0009)	(0.0004)	(0.0000)
30	-0.0005	-0.0004	-0.0003	-0.0007	0.00000
	(0.0008)	(0.0008)	(0.0006)	(0.0003)	(0.0000)

4.4 Appendix

In this section we first bring the integration by parts formula for bivariate functions of bounded variation from Hobson ([30], p. 666):

Let $f(x_1,x_2)$ and $g(x_1,x_2)$ both be functions of bounded variation. Let the functions be such that each of them has an RS-integral with respect to the other, taken over the cell $(a_1,a_2;b_1,b_2)$. Then:

$$\int_{(a_1,a_2)}^{(b_1,b_2)} g(x_1,x_2)\,df(x_1,x_2) = \Big[f(x_1,x_2)g(x_1,x_2)\Big]_{(a_1,a_2)}^{(b_1,b_2)} - \int_{a_1}^{b_1}\Big[f(x_1,x_2)\,dg(x_1,x_2)\Big]_{a_2}^{b_2}$$
$$- \int_{a_2}^{b_2}\Big[f(x_1,x_2)\,dg(x_1,x_2)\Big]_{a_1}^{b_1} + \int_{(a_1,a_2)}^{(b_1,b_2)} f(x_1,x_2)\,dg(x_1,x_2),$$

where

$$\Big[f(x_1,x_2)\,dg(x_1,x_2)\Big]_{a_1}^{b_1} = f(b_1,x_2)\,dg(b_1,x_2) - f(a_1,x_2)\,dg(a_1,x_2),$$

and

$$\Big[f(x_1,x_2)g(x_1,x_2)\Big]_{(a_1,a_2)}^{(b_1,b_2)} = f(b_1,b_2)g(b_1,b_2) - f(b_1,a_2)g(b_1,a_2) - f(a_1,b_2)g(a_1,b_2)$$
$$+ f(a_1,a_2)g(a_1,a_2).$$

Let $s=(s_1,s_2)$ and $t=(t_1,t_2)$ be members of I, the set of all ordered pairs of positive integers. Define $s<t$ to mean $s_1\le t_1$ and $s_2\le t_2$, and $s\ll t$ denote $s_1<t_1$ and $s_2<t_2$. Furthermore, given the underlying probability space $(\Omega,\mathfrak{F},P)$ and sub-σ-field $\mathfrak{F}_t$, let Z_t be an integrable, $\mathfrak{F}_t$-measurable random variable, where $t\in F\subseteq I$. Assume that $\mathfrak{F}_s\subseteq\mathfrak{F}_t\subseteq\mathfrak{F}$ when $s<t$. Then, following Walsh [56], $\{Z_t,\mathfrak{F}_t,t\in F\}$ is called a strong martingale if for each s and t in F,

(i) $$E(Z_t\mid\mathfrak{F}_s)=Z_s\quad\text{when } s<t$$

and

(ii) $$E(Z(s,t] \mid \mathfrak{F}^*_s) = 0 \quad \text{when } s \ll t,$$

where $Z(s,t] = Z_{t_1t_2} - Z_{s_1t_2} - Z_{t_1s_2} + Z_{s_1s_2}$ and $\mathfrak{F}^*_s$ denotes the smallest σ-field containing each $\mathfrak{F}_{ij}$ with either $i \le s_1$ or $j \le s_2$.

Now recall the set $F = \{(t_1,t_2) \in I : t_1 \le n,\, t_2 \le n\}$, the random variable U_t and $\mathfrak{F}_t$ the related σ-field defined in Section 4.2.

Lemma 4.2. $\{U_t, \mathfrak{F}_t, t \in F\}$ is a strong martingale.

Proof. We follow [8]. Define $\xi_{ij} = \big(c_1X_{i-1,j} + c_2Y_{i,j-1}\big)I(c_1X_{i-1,j} + c_2Y_{i,j-1} \le M)\big[\psi(\epsilon_{ij}) - E(\psi(\epsilon_{ij}))\big]$ and let $s = (s_1,s_2) < t = (t_1,t_2)$ be in F.

(i) From $U_t = \sum_{i=1}^{t_1}\sum_{j=1}^{t_2}\xi_{ij}$ observe that

$$U_t = U_s + \sum_{i=1}^{s_1}\sum_{j=s_2+1}^{t_2}\xi_{ij} + \sum_{i=s_1+1}^{t_1}\sum_{j=1}^{s_2}\xi_{ij} + \sum_{i=s_1+1}^{t_1}\sum_{j=s_2+1}^{t_2}\xi_{ij}.$$

Now since $X_{i-1,j} = \sum_{k=1}^{i-1}\alpha_n^{i-1-k}\epsilon_{kj}$ and $Y_{i,j-1} = \sum_{l=1}^{j-1}\beta_n^{j-1-l}\epsilon_{il}$ are independent of ϵ_{ij}, we get

$$E\Big(\sum_{i=1}^{s_1}\sum_{j=s_2+1}^{t_2}\xi_{ij} + \sum_{i=s_1+1}^{t_1}\sum_{j=1}^{s_2}\xi_{ij} + \sum_{i=s_1+1}^{t_1}\sum_{j=s_2+1}^{t_2}\xi_{ij}\Big|\mathfrak{F}_s\Big) = 0$$

and $E\big(U_t|\mathfrak{F}_s\big) = U_s$ follows.

(ii) For $s \ll t$ notice that $U(t,s] = \sum_{i=s_1+1}^{t_1}\sum_{j=s_2+1}^{t_2}\xi_{ij}$. Therefore, by definition of $\mathfrak{F}^*_s$,

$$E(U(t,s]|\mathfrak{F}^*_s) = \sum_{i=s_1+1}^{t_1}\sum_{j=s_2+1}^{t_2}E(\xi_{ij}) = 0.$$

Remark 4.3. We show that the stochastic process $J(t,s)$ in (4.8) is an α-stable process:

Considering the fact that $\{\epsilon_{ij}\}$ is in $DS(\alpha)$, from the invariance principle it suffices to show that $\lim_{n\to\infty} E(e^{-itJ_n}) = \exp(-\theta|t|^\alpha)$ for some positive constant θ. For $\alpha = 2$, it can be seen [7] that the J-process is a mean zero Gaussian process with $\mathrm{cov}\big(J(u,v), J(s,t)\big) = \big[\frac{e^{(u+s)c} - e^{|u-s|c}}{2c}\big].\big[\frac{e^{(v+t)d} - e^{|v-t|d}}{2d}\big]$; the same covariance structure as the product of two independent one-parameter Ornstein-Uhlenbeck process. To avoid the complexity in computation, assume that $a_n = n^{2/\alpha}$ when $0 < \alpha < 2$. We have

$$\begin{aligned} E(e^{-itJ_n}) &= E\Big(\exp\Big[-i\sum_{i=1}^{n}\sum_{j=1}^{n} t a_n^{-1}\alpha_n^{n-i}\beta_n^{n-j}\epsilon_{ij}\Big]\Big) \\ &= \prod_{i=1}^{n}\prod_{j=1}^{n}\exp\big(-|t a_n^{-1}\alpha_n^{n-i}\beta_n^{n-j}|^\alpha\big) \\ &= \exp\big(-\theta_n|t|^\alpha\big), \qquad \text{where} \quad \theta_n = a_n^{-\alpha}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_n^{\alpha(n-i)}\beta_n^{\alpha(n-j)}. \end{aligned}$$

Notice that as $n \to \infty$,

$$\begin{aligned} \theta_n &= n^{-2}\Big(\sum_{i=1}^{n}\alpha_n^{\alpha(n-i)}\sum_{j=1}^{n}\beta_n^{\alpha(n-j)}\Big) = \Big[\frac{1-e^{c\alpha}}{n(1-e^{c\alpha/n})}\Big]\Big[\frac{1-e^{d\alpha}}{n(1-e^{d\alpha/n})}\Big] \\ &\to \theta = \frac{(e^{c\alpha}-1)(e^{d\alpha}-1)}{cd\alpha^2}. \end{aligned}$$

Therefore, the characteristic function for J-process is in the form of $\exp(-\theta|t|^\alpha)$, where $\theta = \frac{(e^{c\alpha}-1)(e^{d\alpha}-1)}{cd\alpha^2}$. Observe that when $c, d \to 0$, we get $\theta \to 1$ which is in accordance with the results given in [6] for models with finite variance.

More on D_q the space of càdlàg functions ([9])

Let $K = [0,1]^q$ be the unit cube. A function $x : K \mapsto \mathbb{R}$ is said to be a step function if x is a linear combination of functions of the form

$$t \to I_{E_1 \times E_2 \times \cdots \times E_q}(t),$$

where each E_p is either left-closed, right-open subinterval of $[0, 1]$, or the singleton $\{1\}$. Let D_q be the uniform closure, in the space of all bounded functions from K to $\mathbb{R}$, of the vector subspace of simple functions. The functions in D_q are characterized by their continuity properties, as follows. If $t \in K$ and if, for $1 \leq p \leq q$, R_p is one of the relations $<$ and $\geq$, let $Q := Q_{R_1,\dots,R_q}(t)$ denote the quadrant

$$\{(s_1, \dots, s_q) \in K : s_p R_p t_p,\ 1 \leq p \leq q\}.$$

Then $x \in D_q$ iff for each $t \in K$, (i) $x_Q := \lim_{s \to t, s \in Q} x(s)$ exists for each of the 2_q quadrants Q, and (ii) $x(t) = x_{Q \geq \cdots \geq}$. In this sense, the functions of D_q are càdlàg; "continuous from above, with limits from below."

Bickel and Wichura [9] introduce a metric topology on D_q which for $q = 1$ coincides with Skorohod's J_1-topology (Billingsley [10]). Let Λ be the group of all transformations $\lambda : K \mapsto K$ of the form $\lambda(t_1, \dots, t_q) = (\lambda_1(t_1), \dots, \lambda_q(t_q))$, where each $\lambda_p : [0, 1] \mapsto [0, 1]$ is continuous, strictly increasing, and fixes zero and one. Define the *Skorohod* distance between x and y in D_q to be

$$d(x, y) = \inf\{\min\left(||x - y\lambda||, ||\lambda||\right) : \lambda \in \Lambda\},$$

where $||x - y\lambda|| = \sup\{|x(t) - y(\lambda(t))| : t \in K\}$ and $||\lambda|| = \sup\{|\lambda(t) - t| : t \in K\}$. With respect to this metric topology, D_q is separable and complete and the Borel σ-algebra D_q coincides with the σ-algebra generated by the coordinate mappings (Billingsley [10]).

Chapter 5

Future Works

1. As mentioned earlier in Chapter 2, the findings of that chapter can be extended to weakly dependent errors such as MA(∞). The univariate case is investigated by Knight [37]. To do so for multivariate autoregression with weakly dependent errors, we need to have a multivariate version of central limit theorem for dependent (ρ-mixing) sequences which has been studied, for example, by Jakubowski and Kobus [33].

2. What studied in Chapter 2 is a homogeneous unit root model, in the sense of Caner [14]. That is, in Model (2.1) we assumed $\mathbf{\Phi} = \mathbf{I}_d$. What happens if all the eigenvalues of $\mathbf{\Phi}$ equal 1 or some diagonal entries of $\mathbf{\Phi}$ are greater than one. Are the OLS estimates still inconsistent in some cases?

3. Caner [15] proposes some tests for detecting cointegrated time series models assuming innovation vectors to be in the domain of attraction of a multivariate α-stable distribution. One may study limiting behavior of the M-estimates of the parameters considering a more general assumption, that is, models with innovations in the domain of attraction of a multivariate stable distribution with possibly different indices of stability.

4. For the models investigated in Chapter 3 and 4, a test statistics for finding unit roots is suggested in [7] assuming models having innovations with finite forth moments. One may consider models with infinite variance innovations and apply M-estimation approach to propose a possibly similar statistics and investigate its asymptotic properties.

5. The performance of the models presented in Chapter 3 and 4 can be compared with the existing spatio-temporal methods in the literature such as those considered by [53, 22] in which the underlying models are Gaussian. Spatial smoothing techniques such as kriging and cokriging can be used to interpolate the missing data.

6. As noted earlier, estimation of the index of stability is a challenging problem. For instance, the Hill estimator of α is an efficient estimator only if the tail of the underlying distribution is Pareto-type. As a remedy to this problem, beside the normal limits given in the preceding Chapters, bootstrapping can be considered. Similar to the work of Zarepour and Knight [59] we can show that when $\alpha = \beta = 1$, bootstrap of the OLS estimate of the parameters of Model (3.1) with ϵ_{11} in $DS(\alpha)$ is asymptotically invalid. However, re-sampling sample size to $o(n)$ makes the bootstrap of the OLS and the M-estimates asymptotically valid.

7. Asymptotic properties of the sample covariance and correlation functions of a stationary sequence of infinite moving averages have been considered by Davis and Resnick [25, 26]. Similar results can be derived and extended to the spatial processes investigated in the preceding chapters.

Bibliography

[1] Banjevic, D., Ishwaran, H., Zarepour, M. (2002) A recursive method for functionals of Poisson processes, *Bernoulli* 8, 295-311.

[2] Basu, S., Reinsel, G.C. (1992a) regression models with spatially correlated errors, Technical Report, Department of Statistics, University of Wisconsin-Madison.

[3] Basu, S., Reinsel, G.C. (1992b) A note on properties of spatial Yule-Walker estimators, *J. Statist. Comput. Simulation* 41, 243-255.

[4] Basu, S., Reinsel, G.C. (1993) Properties of the spatial unilateral first-order ARMA model, *Adv. in Appl. Probab.* 25, 631-648.

[5] Basu, S., Reinsel, G.C. (1994) Regression models with spatially correlated errors, *J. Amer. Statist. Assoc.* 89, 88-99.

[6] Bhattacharyya, B.B., Khalil, T.M., Richardson, G.D. (1996) Gauss-Newton estimation of parameters for a spatial autoregression model, *Statist. Probab. Lett.* 28, 173-179.

[7] Bhattacharyya, B.B., Li, X., Pensky, M., Richardson, G.D. (2000) Testing for unit roots in a nearly nonstationary spatial autoregressive process, *Ann. Inst. Statist. Math.* 52, 71-83.

[8] Bhattacharyya, B.B., Richardson, G.D., Franklin L.A. (1997) Asymptotic inference for near unit roots in spatial autoregression, *Ann. Statist.* 25, 1709-1724.

[9] Bickel P.J., Wichura M.J. (1971) Convergence criteria for multiparameter stochastic processes and some applications, *Ann. Math. Statist.* 42, 1656-1670.

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

[11] Brockwell, P.J., Davis, R.A. (1991) *Time series: theory and methods*, Second edition, Springer Series in Statistics, Springer-Verlag, New York.

[12] Bronars S.G., Jansen D.W. (1987) The geographic distribution of unemployment in the U.S.: a spatial-time series analysis, *J. Econometrics* 36, 251-279.

[13] Brown, B.M. (1971) Martingale central limit theorems, *Ann. Math. Statist.* 42, 59-66.

[14] Caner, M. (1997) Weak convergence to a matrix stochastic integral with stable processes, *Econometric Theory* 13, 506-528.

[15] Caner, M. (1998) Tests for cointegration with infinite variance errors, *J. Econometrics* 86, no. 1, 155-175.

[16] Chan, N.H. (1990) Inference for near-integrated time series with infinite variance, *J. Amer. Statist. Assoc.* 85, 1069-1074.

[17] Chan, N.H., Tran, L.T. (1987) Asymptotic inference for nearly nonstationary AR(1) processes, *Ann. Statist.* 15, 1050-1063.

[18] Chan, N.H., Tran, L.T. (1989) On the first-order autoregressive process with infinite variance, *Econometric Theory* 5, 354-362.

[19] Chellappa, R. (1985) Two-dimensional discrete Gaussian Markov random feild models for image processing, *Progrss in Pattern Recognition 2*, ed. L. N. Kanal and A. Rosenfeld, pp. 79-112, North-Holland, Amsterdam.

[20] Cox, D.D., Llatas, I. (1991) Maximum likelihood type estimation for nearly nonstationary autoregressive time series, *Ann. Statist.* 19, 1109-1128.

[21] Cullis, B.R., Gleeson, A.C. (1991) Spatial analysis of feild expriments-an extension to two dimensions, *Biometrics* 47, 1449-1460.

[22] Dass, S.C., Nair, V.N. (2003) Edge detection, spatial smoothing, and image reconstruction with partially observed multivariate data, *J. Amer. Statist. Assoc.* 98, 77-89.

[23] Davis, R.A. (1996) Gauss-Newton and M-estimates for $ARMA$ processes with infinite variance, *Stochastic Process. Appl.* 63, 75-95.

[24] Davis, R.A., Knight, K., Liu, J. (1992) M-estimation for autoregressions with infinite variance, *Stochastic Process. Appl.* 40, 145-180.

[25] Davis, R., Resnick, S. (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities, *Ann. Probab.* 13, 179-195.

[26] Davis, R, Resnick, S. (1986) Limit theory for the sample covariance and correlation functions of moving averages, *Ann. Statist.* 14, 533-558.

[27] Dickey, D.A., Fuller, W.A. (1979) Distribution of the estimators for autoregressive time series with a unit root, *J. Amer. Statist. Assoc.* 74, 427-431.

[28] Feller, W. (1971) *An Introduction to Probability Theory and its Application, Vol. II (Second Edition)*, Wiley, New York.

[29] Geman, S., Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.* 6, 721-741.

[30] Hobson, E.W. (1957) *The Theory of Functions of a Real Variable and the Theory of Fourier Series*, Dover Publications, New York.

[31] Huber, Peter J. (1981) *Robust statistics*, Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York.

[32] Jain, A.K. (1981) Advances in mathematical models for image processing, *Proc. IEEE* 69, 502-528.

[33] Jakubowski, A., Kobus, M. (1989) α-stable limit theorems for sums of dependent random vectors, *J. Multivar. Anal.* 29, 219-251.

[34] Kallenberg, O. (1983) *Random Measures*, Berlin: Akademie-Verlag.

[35] Kempton, R.A., Howes, C.W. (1981) The use of neighbouring plot values in the analysis of variety trials, *Appl. Statist.* 30, 59-70.

[36] Knight, K. (1989) Limit theory for autoregressive-parameter estimates in an infinite-variance random walk, *Canadian J. Statist.* 17, 261-278.

[37] Knight, K. (1991) Limit theory for M-estimates in an integrated infinite variance process, *Econometric Theory* 7, 200-212.

[38] Kurtz, T.G., Protter, P. (1991) Weak limit theorems for stochastic integrals and stochastic differential equations, *Ann. Probab.* 19, 1035-1070.

[39] Le Page, R. (1980) Multidimensional Infinitely Divisible Variables and Processes. Part I: Stable case, *Lecture Notes in Math.* 1391, 153-163.

[40] Lucas, A. (1995) Unit root tests based on M-estimators, *Econometric Theory* 11, 331-346.

[41] Martin, R. J. (1979) A subclass of lattice processes applied to a problem in planar sampling, *Biometrika* 66, 209-217.

[42] Martin, R. J. (1990) The use of time-series models and methods in the analysis of agricultural field trials, *Comm. Statist. Theory Methods* 19, 55-81.

[43] McLeish, D.L. (1974) Dependent central limit theorems and invariance principles, *Ann. Probab.* 2, 620-628.

[44] Mittnik, S., Paulauskas, V., Rachev, S.T. (2001) Statistical inference in regression with heavy-tailed integrated variables, *Mathematical and Computer Modelling* 34, 1145-1158.

[45] Paulauskas, V., Rachev, S.T. (1998) Cointegrated processes with infinite variance innovations, *Ann. of Appl. Probab.* 8, 775-792.

[46] Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregression, *Biometrika* 74, 535-547.

[47] Phillips, P.C.B. (1990) Time series regression with a unit root and infinite-variance errors, *Econometric Theory* 6, 44-62.

[48] Resnick, S., Greenwood, P. (1979) A bivariate stable characterization and domain of attraction, *J. Multivar. Anal.* 9, 206-221.

[49] Resnick, S. (1986) Point processes, regular variation and weak convergence, *Adv. in Appl. Probab.* 18, 66-138.

[50] Resnick, S. (2007) *Heavy-Tail Phenomena: Probabilistic and statistical Modeling*, Springer, New York.

[51] Samorodnitsky, G., Taqqu, M. (1994) *Stable Non-Gaussian Random Processes*, Chapman & Hall, New York.

[52] Tjøstheim, D. (1978) Statistical spatial series modelling, *Adv. in Appl. Probab.* 10, 130-154.

[53] Tjøstheim, D. (1981) Autoregressive modeling and spectral analysis of array data in the plane, *IEEE Trans. Geosci. Rem. Sensing* 19, 15-24.

[54] Tjøstheim, D. (1983) Statistical spatial series modelling. II. Some further results on unilateral lattice processes, *Adv. in Appl. Probab.* 15, 562-584.

[55] Van de Geer, S. (2000) *Emprical Processes in M-Estimation*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

[56] Walsh, J.B. (1986) Martingales with a multidimensional parameter and stochastic integrals in the plane., *Lectures in Probability and Statistics. Lecture Notes in Math.* 1215, 329-491. Springer, New York.

[57] Wichura, M. (1969) Inequalities with applications to the weak convergence of random processes with multidimensional time parameters, *Ann. Math. Statist.* 40, 681-687.

[58] Whittle , P. (1954) On stationary processes in the plane, *Biometrika* 41, 434-449.

[59] Zarepour, M., Knight, K. (1999) Bootstrapping unstable first order autoregressive process with errors in the domain of attraction of stable law, *Comm. Statist. Stochastic Models* 15, 11-27.

[60] Zarepour, M., Roknossadati, S.M. (2008) Multivariate autoregression of order one with infinite variance innovations, *Econometric Theory* 24, 677-695.