Econometric Theory, **26**, 2010, 1663–1682.
doi:10.1017/S0266466609990752

M-ESTIMATION FOR A SPATIAL UNILATERAL AUTOREGRESSIVE MODEL WITH INFINITE VARIANCE INNOVATIONS

S.M. ROKNOSSADATI
Central Bank of Iran
and
University of Ottawa

M. ZAREPOUR
University of Ottawa

We study the limiting behavior of the M-estimators of parameters for a spatial unilateral autoregressive model with independent and identically distributed 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 also shown that self-normalized M-estimators are asymptotically normal. A numerical example and a simulation study are also given.

1. INTRODUCTION

Consider the doubly geometric spatial autoregressive (AR) model, introduced by Martin (1979),

$$Z_{ij} = \phi_1 Z_{i-1,j} + \phi_2 Z_{i,j-1} - \phi_1\phi_2 Z_{i-1,j-1} + \epsilon_{ij}, \qquad (1)$$

where $\{\epsilon_{ij} = \epsilon_{ij}(\phi_1, \phi_2) : i, j = 1, \ldots, n\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with $Z_{ij} = 0$ when $i \wedge j \le 0$. In this paper, we study the weak limit behavior of the M-estimators for the parameters in model (1) when $\{\epsilon_{ij}\}$ is in the domain of attraction of a symmetric stable law with index of stability α, $0 < \alpha \le 2$ (denoted by $DS(\alpha)$). Both stationary and unit root models are considered, and it is shown that asymptotically, the distributions of M-estimators are functionals of a stable sheet (Resnick, 1986). For innovations in $DS(\alpha)$, self-normalized M-estimates are asymptotically normal.

Spatial models (e.g., model (1)) appear in many applications such as geography, agriculture, geology, biology, and economics. See, e.g., the work of Whittle

This paper is supported by the Natural Sciences and Engineering Research Council of Canada. The first author was also supported by the Central Bank of Iran. The authors also are greatly indebted to two anonymous referees for several helpful comments. Address correspondence to Mahmoud Zarepour, Department of Mathematics and Statistics, University of Ottawa, 585 King Edward Street P.O. Box 450 STN A, Ottawa, Ontario, K1N 6N5, Canada; e-mail: zarepour@uottawa.ca.

(1954), Kempton and Howes (1981), Martin (1990), Cullis and Gleeson (1991), and Basu and Reinsel (1993, 1994) in the study of agricultural field trials, Jain (1981), Geman and Geman (1984), Chellappa (1985), and Dass and Nair (2003) in image processing, Tjøstheim (1978, 1981) in system theory, and Bronars and Jansen (1987) in economics. Asymptotic properties of the spatial unilateral AR models, which are our interest in this paper, have been investigated by several authors. Martin (1979) introduces the symmetrically reflective spatial AR model given in (1) and considers the maximum likelihood (ML) estimators when the model is stationary. Martin (1990) considers model (1) thoroughly and argues that these models have a broad range of applications in practical modeling. Tjøstheim (1978, 1981, 1983) establishes the limiting behavior and consistency of the corresponding Yule–Walker estimators. Basu and Reinsel (1992a, 1992b) show that in the stationary setting ML estimators have less bias than the Yule–Walker estimators proposed by Tjøstheim. They (1993, 1994) 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 a stationary model are considered, and their performances are compared. Bhattacharyya, Khalil, and Richardson (1996) investigate the asymptotic properties of the sequence of Gauss–Newton estimators for model (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 $(\hat{\phi}_1, \hat{\phi}_2)$ of (ϕ_1, ϕ_2) in (1) is the minimizer of the objective function

$$g(\phi_1, \phi_2) = \sum_{i=2}^{n} \sum_{j=2}^{n} \rho(Z_{ij} - \phi_1 Z_{i-1,j} - \phi_2 Z_{i,j-1} + \phi_1 \phi_2 Z_{i-1,j-1}) \tag{2}$$

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 (1981) and Van de Geer (2000). 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 (1992). 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 (1996) also develops the limit theory for M-estimates in addition to Gauss–Newton estimates for ARMA processes with infinite variance showing the dominance of M-estimates, asymptotically.

Spatial and spatiotemporal models are of interest to economists. As a specific example in addition to the other existing evidence in the literature, Bronars and Jansen (1987) investigate the time series and spatial pattern of unemployment rate fluctuations across labor market areas in the United States over the period 1977–1983. They employ both one-quadrant simultaneous (unilateral) and four-quadrant conditional autoregressive models of spatial series over a regular lattice and conclude that the two approaches yield similar results. We shall emphasize that our approaches cover models with both infinite and finite variance innovations, and that more research is required to examine the other possible applications of the proposed models in this paper in economics. Any economic index linked to a geographical phenomenon can be analyzed by similar models.

The main results of the paper are stated in the forthcoming section and followed by a numerical example and a simulation study in Section 3. Finally, the proofs of the main theorems appear in the Appendix.

2. PRELIMINARIES AND MAIN RESULTS

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

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

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

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

for some $0 < p_0 \leq 1$. 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 $0 < t, s \leq 1$, where $[x]$ stands for the 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 (1971). Then

$$S_n(t,s) \to_d \mathcal{S}_\alpha(t,s), \quad \text{as } n \to \infty, \tag{3}$$

in D_2, where the limiting process is a stable sheet (Resnick, 1986) defined on a regular rectangular grid in two dimensions. In fact, when (3) holds with either $\alpha > 1$ and $\mathrm{E}(\epsilon_{11}) = 0$, or $\alpha < 1$, or $\{\epsilon_{ij}\}$ have a symmetric distribution about 0 we say that $\{\epsilon_{ij}\}$ is in $DS(\alpha)$. This assumption is a standard assumption similar to that made in Knight (1989) and Davis et al. (1992). In practical cases, $\alpha > 1$, and therefore symmetry is not required. Throughout the paper, $\to_d$ and $\to_p$ stand for convergence in distribution and probability, respectively, with respect to the relevant topologies. Integrals involving dS_n, dS_α, or dW are interpreted as Itô stochastic integrals.

It can be shown (Feller, 1971) that under (3), $a_n = n^{2/\alpha} L(n)$, where $0 < \alpha \le 2$ and $L(\cdot)$ is a slowly varying function at infinity. We assume $b_n = 0$. When $\alpha = 2$, $\mathcal{S}_\alpha(\cdot,\cdot)$ is a Brownian sheet. When $0 < \alpha < 2$, a symmetric α-stable random variable $\mathcal{S}_\alpha$ has the *series representation* $\mathcal{S}_\alpha = \sum_{i=1}^{\infty} \delta_i \Gamma_i^{-1/\alpha}$, where the δ_i's are i.i.d. random variables with $P(\delta_1 = 1) = P(\delta_1 = -1) = p_0 = \frac{1}{2}$. Here $\{\Gamma_i\}$ are the arrival times of a Poisson process with Lebesgue mean measure independent of $\{\delta_i\}$. In fact, we have $\Gamma_i = E_1 + \cdots + E_i$, $i \ge 1$, where the E_i's are i.i.d. exponential random variables with mean 1. For more details on the series representation see Le Page, Woodroofe, and Zinn (1981). More generally, the random field $X(\cdot,\cdot) = \sum_{i=1}^{\infty} \delta_i \Gamma_i^{-1/\alpha} Y_i(\cdot,\cdot)$ is an α-stable random field in D_2 when $\{Y_i\}$ is an i.i.d. sequence of D_2-valued random fields independent of $\{\delta_i\}$ and $\{\Gamma_i\}$ as defined before. If $\mathrm{E}(|Y(t_i,s_i)|^\alpha) < \infty$, $(t_i,s_i) \in [0,1]^2$, $i = 1,\ldots,k$, then the $\mathbb{R}^k$-valued infinite series $(X(t_1,s_1),\ldots,X(t_k,s_k))$ converges almost surely (a.s.) and represents a symmetric α-stable random vector. Refer to Davis and Mikosch (2008) for examples on the necessary and sufficient conditions for the almost sure convergence of the series in D_2 and the space of continuous functions on $[0,1]^2$ equipped with the uniform topology. We will assume that the infinite series converges a.s. in D_2. For other extensive expositions on stable random variables, processes, and related stochastic integrals that appear in the paper, consult Samorodnitsky and Taqqu (1994).

Denote $\psi(x) = d\rho(x)/dx$, $\psi'(x) = d^2\rho(x)/dx^2$ and impose the following assumptions on the function $\rho(\cdot)$ in (2), which are introduced for technical simplicity in the proofs of the theorems.

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

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

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

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

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

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

for some nonnegative constant k_2 and $0 < \lambda_2 < 1$.

In practice, the objective function $\psi(\cdot)$ might have a countable number of discontinuity points. For example, consider the Huber's loss function (Huber, 1981) given by

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

for a known constant c. The loss function $\rho_H(\cdot)$ satisfies conditions A1–A3 if we ignore the discontinuity of $\psi_H(x) = d\rho_H(x)/dx$ at $\pm c$. Another example is the loss function $\rho(x) = 2\log(3 + x^2)$. This sufficiently smooth loss function does not satisfy A2 and A3 and is just locally convex around its global minimum. However, this suffices to search out the unique solutions (the M-estimates). Actually, the performance of $\rho(\cdot)$ is as good as that of $\rho_H(\cdot)$, as shown in Section 3. Asymptotically there is no difference between different objective functions. In practice, the optimal choice of $\rho(\cdot)$ is another challenging problem. Following the simulation studies of Knight and Liu (see Davis et al., 1992, pp. 149–150), the optimal choice for the case $\rho(x) = |x|^\theta$ is obtained by taking $\theta = 1$ for $\alpha \in [1, 2)$ and $\theta = \alpha$ for $\alpha \in (0, 1)$. Therefore, as mentioned before, a good loss function is a convex function ρ with $\rho(x)/x^2 \to 0$ as $x \to \pm\infty$. Note that the asymptotic consistency does not depend on the form of the loss function given that it satisfies Assumptions A1–A3.

Now, consider a general spatial unilateral model

$$Z_{ij} = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} c_{kl}\,\epsilon_{i-k,j-l}, \tag{4}$$

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

$$\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} |c_{kl}|^\delta < \infty \quad \text{for some } \delta < \min(\alpha, 1). \tag{5}$$

It can be shown (see Resnick, 1987, Lem. 4.24) that under this condition, the series in (4) converges a.s. and

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

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

$$Z_{ij} = \sum_{k=0}^{p}\sum_{l=0}^{q} \beta_{kl} Z_{i-k,j-l} + \epsilon_{ij}, \qquad \beta_{00} = 0 \tag{6}$$

with $\{\epsilon_{ij}\}$ in $DS(\alpha)$ has a stationary solution if it can be rewritten in the form of (4) where $\{c_{kl}\}$ satisfies (5). As an example, note that model (1) is a special case of (6) with $p = q = 1$, $\beta_{10} = \phi_1$, $\beta_{01} = \phi_2$, and $\beta_{11} = -\phi_1\phi_2$ and that

$$\begin{aligned} Z_{ij} &= \phi_1 Z_{i-1,j} + \phi_2 Z_{i,j-1} - \phi_1\phi_2 Z_{i-1,j-1} + \epsilon_{ij} \\ &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{r=0}^{\infty} \frac{(k+l+r)!}{k!\,l!\,r!} \phi_1^k \phi_2^l (-\phi_1\phi_2)^r \epsilon_{i-k-r,j-l-r} \\ &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \phi_1^k \phi_2^l \epsilon_{i-k,j-l} \end{aligned}$$

$$= \sum_{k=1}^{i} \sum_{l=1}^{j} \phi_1^{i-k} \phi_2^{j-l} \epsilon_{k,l} \quad \text{when } Z_{ij} = 0 \text{ for } i \wedge j \le 0$$

is a unilateral model. Notice that the coefficient of $Z_{i-1,j-1}$ that is set to $-\phi_1\phi_2$ allows the model to admit a causal moving average (MA) representation, whereas AR random fields may not generally have such representation. Also see Section 3 on a numerical example and simulation. Therefore by (5) to ensure the existence of the stationary representation we shall have for $0 < \alpha < 2$,

$$\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \left| \phi_1^k \phi_2^l \right|^{\delta} < \infty \quad \text{for some } \delta < \min(\alpha, 1).$$

For $\alpha = 2$, see Proposition 1 of Basu and Reinsel (1993).

The following theorem deals with the M-estimators of the parameters of the general stationary model given in (6). Following Davis et al. (1992), recall that the M-estimate, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_{01}, \hat{\beta}_{02}, \ldots, \hat{\beta}_{0q}, \ldots, \hat{\beta}_{p0}, \hat{\beta}_{p1}, \ldots, \hat{\beta}_{pq})'$, of $\boldsymbol{\beta} = (\beta_{01}, \ldots, \beta_{pq})'$ minimizes the objective function

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

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

$$\begin{aligned} W_n(\boldsymbol{u}) &= \sum_{i=p+1}^{n} \sum_{j=q+1}^{n} \big[\rho(Z_{ij} - \varphi_{10} Z_{i-1,j} - \varphi_{01} Z_{i,j-1} \\ &\qquad - \varphi_{11} Z_{i-1,j-1} - \cdots - \varphi_{pq} Z_{i-p,j-q}) - \rho(\epsilon_{ij})\big] \\ &= \sum_{i=p+1}^{n} \sum_{j=q+1}^{n} \big[\rho(\epsilon_{ij} - u_{10} a_n^{-1} Z_{i-1,j} - u_{01} a_n^{-1} Z_{i,j-1} \\ &\qquad - u_{11} a_n^{-1} Z_{i-1,j-1} - \cdots - u_{pq} a_n^{-1} Z_{i-p,j-q}) - \rho(\epsilon_{ij})\big], \end{aligned} \tag{7}$$

where $u_{ij} = a_n(\varphi_{ij} - \beta_{ij}) \in \mathbb{R},\ i = 0, 1, \ldots, p,\ j = 0, 1, \ldots, q$.

The minimizer of the process $W_n(\cdot)$, i.e., $a_n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, has a weak limit that is the minimizer of the process $W(\cdot)$ given in the following theorem. Of course, we assume that $W_n(\cdot)$ converges and the minimizer of $W_n(\cdot)$ exists and it is unique with probability 1 (see also Davis et al., 1992, Rmks. 1 and 2).

THEOREM 1. *Consider the stationary spatial unilateral model (6) with* $\{\epsilon_{ij}\}$ *in* $DS(\alpha)$, $0 < \alpha < 2$ *and suppose that Assumptions A1 and A2 hold. Then on* $C\left(\mathbb{R}^{p+q+pq}\right)$, $W_n(\cdot) \to_d W(\cdot)$, *where*

$$\begin{aligned} W(\boldsymbol{u}) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \big[\rho\big(\epsilon_{kij} - (u_{10} c_{i-1,j} + u_{01} c_{i,j-1} + u_{11} c_{i-1,j-1} \\ + \cdots + u_{pq} c_{i-p,j-q}) \delta_k \Gamma_k^{-1/\alpha}\big) - \rho(\epsilon_{kij})\big] \end{aligned} \tag{8}$$

with $\{c_{kl}\}$ the same as in (4). 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 (8). Also, notice that if Assumption A3 holds with $\gamma > 0$ then $\rho(\cdot)$ is convex and the limiting process in Theorem 1 has an almost sure unique minimizer. One may replace A2 by A3 in Theorem 1, but A2 is a weaker assumption. A similar theorem can be stated for the special case when $\rho(x) = |x|$, corresponding to LAD estimators, provided we make some assumptions on ϵ_{11} similar to those made in Theorem 3.4 of Davis (1996).

In the next theorem, we investigate the asymptotic behavior of M-estimators of the parameters in the spatial unilateral model given in (1) when the model has a unit root. We consider two cases: (i) $\phi_1 = 1$ and $|\phi_2| < 1$ and (ii) $\phi_1 = \phi_2 = 1$. The limits are functionals of the α-stable sheet (Itô stochastic integrals) in (3). As in Knight (1989), to make sure that an almost sure unique limit exists we assume the convexity of the function $\rho(\cdot)$.

THEOREM 2. *Consider model (1) with $\{\epsilon_{ij}\}$ in $DS(\alpha)$, $0 < \alpha < 2$ and suppose that the loss function $\rho(\cdot)$ is strictly convex. Denote $(\hat{\phi}_1, \hat{\phi}_2)$ to be the M-estimator of (ϕ_1, ϕ_2) in that model. Then under Assumptions A1–A3,*

(i) if $\phi_1 = 1$ and $|\phi_2| < 1$,

$$\begin{pmatrix} a_n\sqrt{n}\,(\hat{\phi}_1 - 1) \\ a_n(\hat{\phi}_2 - \phi_2) \end{pmatrix} \to_d \xi := \begin{pmatrix} \dfrac{\sigma \int_0^1\int_0^1 \mathcal{S}_\alpha(t_1,t_2)\,dW(t_1,t_2)}{\gamma \int_0^1\int_0^1 \mathcal{S}_\alpha^2(t_1,t_2)\,dt_1dt_2} \\ \arg\min Z(u) \end{pmatrix},$$

where

$$Z(u) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\left[\rho\left(\epsilon_{ij} - u\phi_2^{j-1}\delta_i\Gamma_i^{-1/\alpha}\right) - \rho(\epsilon_{ij})\right],$$

(ii) and if $\phi_1 = \phi_2 = 1$,

$$a_n\sqrt{n}\begin{pmatrix} \hat{\phi}_1 - 1 \\ \hat{\phi}_2 - 1 \end{pmatrix} \to_d \eta := \frac{\sigma}{\gamma}\begin{pmatrix} \dfrac{\int_0^1\int_0^1 \mathcal{S}_\alpha(t_1,t_2)\,dW(t_1,t_2)}{\int_0^1\int_0^1 \mathcal{S}_\alpha^2(t_1,t_2)\,dt_1dt_2} \\ \dfrac{\int_0^1\int_0^1 \mathcal{S}_\alpha(t_1,t_2)\,dW(t_1,t_2)}{\int_0^1\int_0^1 \mathcal{S}_\alpha^2(t_1,t_2)\,dt_1dt_2} \end{pmatrix}, \tag{9}$$

where $\mathcal{S}_\alpha(\cdot,\cdot)$ is the α-stable sheet and $W(\cdot,\cdot)$ is a standard Brownian sheet.

Remark 1. When $\phi_1 = 1$ and $|\phi_2| < 1$, distribution of ξ is not necessarily symmetric, whereas for the second case we observe a symmetric distribution. Similar to the results given by Knight (1989) and Remark 3 of Zarepour and Roknossadati (2008) it is possible to give self-normalizing coefficients for the

M-estimators in Theorem 2. More precisely, using the Resnick and Greenwoods result (Resnick and Greenwood, 1979, Thms. 1 and 3) when ϵ_{11} is in $DS(\alpha)$, $0 < \alpha \leq 2$, we get the following, results.

(a) If $\phi_1 = 1$ and $|\phi_2| < 1$,

$$\left(\sum_{i=2}^{n}\sum_{j=1}^{n} X_{i-1,j}^2\right)^{1/2}\left(\hat{\phi}_1 - 1\right) \to_d N\left(0, \frac{\sigma^2}{\gamma^2}\right),$$

(b) and if $\phi_1 = \phi_2 = 1$,

$$\begin{pmatrix} \sqrt{\sum_{i=2}^{n}\sum_{j=1}^{n} X_{i-1,j}^2} & 0 \\ 0 & \sqrt{\sum_{i=1}^{n}\sum_{j=2}^{n} Y_{i,j-1}^2} \end{pmatrix} \begin{pmatrix} \hat{\phi}_1 - 1 \\ \hat{\phi}_2 - 1 \end{pmatrix} \to_d N(\mathbf{0}, \mathbf{\Gamma}), \tag{10}$$

where $\mathbf{\Gamma} = \operatorname{diag}\left(\sigma^2/\gamma^2, \sigma^2/\gamma^2\right)$, $X_{ij} := Z_{ij} - \phi_2 Z_{i,j-1}$, and $Y_{ij} := Z_{ij} - \phi_1 Z_{i-1,j}$.

Because $(\hat{\phi}_1, \hat{\phi}_2)$ is a consistent estimator of (ϕ_1, ϕ_2), $X_{i-1,j}$ and $Y_{i,j-1}$ can be replaced by $\hat{X}_{i-1,j}$ and $\hat{Y}_{i,j-1}$ where $\hat{X}_{ij} := Z_{ij} - \hat{\phi}_2 Z_{i,j-1}$, and $\hat{Y}_{ij} := Z_{ij} - \hat{\phi}_1 Z_{i-1,j}$.

Again a similar result can be derived for the special case when $\rho(x) = |x|$ using techniques similar to those employed in Knight (1989).

Remark 2. The limiting distributions established in Theorems 1 and 2 depend on the index of stability, α. However, it is generally not known in application. There are many different techniques to estimate the tail index. The most prevalent techniques are quantile-quantile estimation and plotting, Hill estimation and plotting, and Pickands estimation. Although the Hill and Pickands estimators are consistent, there are some difficulties when using these estimators in practice. For instance, the Hill estimator usually works well when the underlying distribution is close to Pareto. For discussion on these difficulties and more on choosing the right estimator consult Resnick (2007) and references therein. Another related problem is the choice of the normalizing constants a_n. One way to avoid estimating a_n is to use bootstrap. Another standard solution is the self-normalization. This solution is not applicable for the stationary process given in Theorem 1. However, the self-normalized M-estimator version for a nonstationary AR(1,1) is discussed in Remark 1. The advantage of using the self-normalized coefficients is the fact that the limiting vectors in Remark 1 have normal distribution in both finite and infinite variance; therefore, it is not necessary to estimate the value of the tail index. Normality of the limiting distribution simplifies construction of confidence intervals. See the normal quantile plots in Figure 1 and the density and contour plots in Figure 2. In the infinite variance stationary case, in practice, we can assume errors are Pareto type or in the normal domain of attraction of a stable law $(L(n) = 1)$.

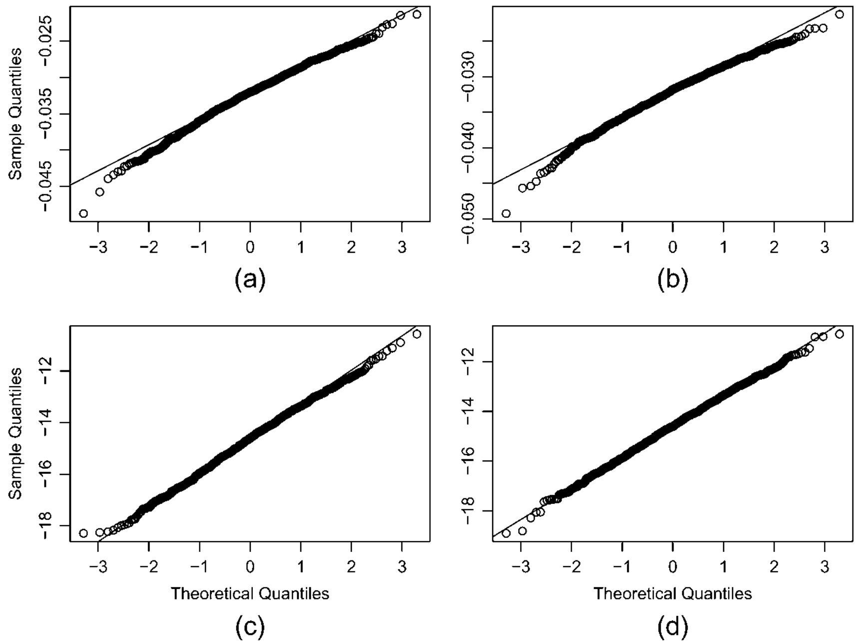

Figure 1. Normal Q-Q plots for the marginals of the limiting vector with fixed (parts (a) and (b)) and random normalization (parts (c) and (d)) based on LAD estimates for the parameters in model (1) with normal noise and $n = 60$. The replication size is 1,000.

Remark 3. The convergence rate in (9) when innovations have finite variance is the same as that given in Bhattacharyya et al. (1996) and Bhattacharyya, Richardson, and Franklin (1997), i.e., $n^{3/2}$. However, for an AR(1,1) model with infinite variance innovations the LAD and M-estimates have a higher rate of consistency compared to the OLS and Gauss–Newton estimates.

3. A NUMERICAL EXAMPLE AND SIMULATION

To investigate the performance of the M-estimation approach proposed in the preceding section, a numerical example for Theorem 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 (1981). This data set has also been analyzed by Basu and Reinsel (1993). We consider two loss functions, Huber and LAD. Because the data have a normal distribution, the $\alpha = 2$ is taken. Results tabulated in Tables 1 and 2 show that the doubly geometric spatial AR model given in (1) with a constant added to the model is more accurate than a general model with $-\phi_1\phi_2$ replaced by ϕ_3. Such a model is a special case of the unilateral first-order ARMA model of the quadrant

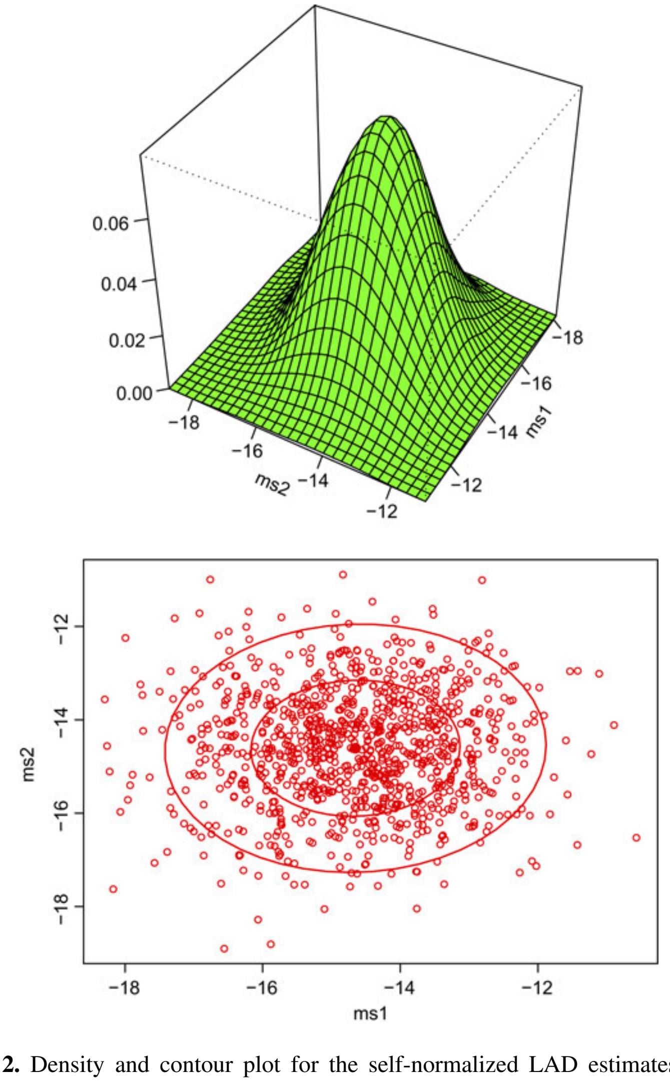

FIGURE 2. Density and contour plot for the self-normalized LAD estimates given in Figure 1.

type introduced by Basu and Reinsel (1993). Notice that the limit theory for this general model is the same as that for model (6) of Section 2 with different $\{c_{kl}\}$. From Table 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 (1993).

A simulation study is undertaken to investigate the result of Theorem 2 of Section 2. We present the results for five parameter settings for the index of

TABLE 1. M-estimates of the parameters for model (1) based on the yield of barley data (in kilograms) obtained from Kempton and Howes (1981)

	Const.	$\hat{\phi}_1$	$\hat{\phi}_2$	$-\hat{\phi}_1\hat{\phi}_2$	$\hat{\sigma}^2$
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 2. M-estimates of the parameters for model (1) with $-\phi_1\phi_2$ replaced by ϕ_3 based on the yield of barley data (in kilograms) obtained from Kempton and Howes (1981)

	Const.	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3(\neq -\hat{\phi}_1\hat{\phi}_2)$	$\hat{\sigma}^2$
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

stability $\alpha = 2, 1.8, 1.5, 1, 0.5$. The unit root model (1) with $\phi_1 = \phi_2 = 1$ and symmetric α-stable errors on an $n \times n$ regular grid with different values of n are generated. The replication size of 1,000 is chosen, and both mean and standard deviation of M-estimates of $\phi_1 - 1$ are calculated. The simulation results are summarized in Table 3 with Huber's loss function and in Table 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 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 gets smaller.

To explore the bivariate normality of the limits with the random normalization factors given in (10), we consider a simulated unit root model (1) with normal innovations using a replication size of 1,000 and $\rho(x) = |x|$ as the loss function in the M-estimation. To get the normality we need to take $n = 60$. Figure 1 portrays the normal quantile-quantile plots for the marginals of the limiting vectors in (9) (Figures 1a and 1b) and (10) (Figures 1c and 1d) 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 2. Notice that the contour plot confirms that the marginal normal variables of the limiting vector in (10) are independent.

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

TABLE 3. Mean and mean standard deviations (in parentheses) for M-estimates of the parameter $(\phi_1 - 1)$ for the unit root model (1) with $\phi_1 = \phi_2 = 1$ using Huber's loss function with different α-stable noises. The replication size is 1,000.

	Index of stability α				
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 4. Mean and mean standard deviations (in parentheses) for M-estimates of the parameter $(\phi_1 - 1)$ for the unit root model (1) with $\phi_1 = \phi_2 = 1$ using the loss function $\rho(x) = 2\log(x^2 + 3)$ with different α-stable noises. The replication size is 1,000.

	Index of stability α				
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)

REFERENCES

Basu, S. & G.C. Reinsel (1992a) Regression Models with Spatially Correlated Errors. Technical report, Department of Statistics, University of Wisconsin-Madison.

Basu, S. & G.C. Reinsel (1992b) A note on properties of spatial Yule-Walker estimators. *Journal of Statistical Computation and Simulation* 41, 243–255.

Basu, S. & G.C. Reinsel (1993) Properties of the spatial unilateral first-order ARMA model. *Advances in Applied Probability* 25, 631–648.

Basu, S. & G.C. Reinsel (1994) Regression models with spatially correlated errors. *Journal of the American Statistical Association* 89, 88–99.

Bhattacharyya, B.B., T.M. Khalil, & G.D. Richardson (1996) Gauss-Newton estimation of parameters for a spatial autoregression model. *Statistics and Probability Letters.* 28, 173–179.

Bhattacharyya, B.B., G.D. Richardson, & L.A. Franklin (1997) Asymptotic inference for near unit roots in spatial autoregression. *Annals of Statistics* 25, 1709–1724.

Bickel, P.J. & M.J. Wichura (1971) Convergence criteria for multiparameter stochastic processes and some applications. *Annals of Mathematical Statistics* 42, 1656–1670.

Bronars, S.G. & D.W. Jansen (1987) The geographic distribution of unemployment in the U.S.: A spatial-time series analysis. *Journal of Econometrics* 36, 251–279.

Brown, B.M. (1971) Martingale central limit theorems. *Annals of Mathematical Statistics* 42, 59–66.

Chellappa, R. (1985) Two-dimensional discrete Gaussian Markov random field models for image processing. In L.N. Kanal & A. Rosenfeld (eds.), *Progress in Pattern Recognition 2*, pp. 79–112. North-Holland.

Cullis, B.R. & A.C. Gleeson (1991) Spatial analysis of field experiments—An extension to two dimensions. *Biometrics* 47, 1449–1460.

Dass, S.C. & V.N. Nair (2003) Edge detection, spatial smoothing, and image reconstruction with partially observed multivariate data. *Journal of the American Statistical Association* 98, 77–89.

Davis, R.A. (1996) Gauss-Newton and *M*-estimates for ARMA processes with infinite variance. *Stochastic Processes and Their Applications* 63, 75–95.

Davis, R.A., K. Knight, & J. Liu (1992) *M*-estimation for autoregressions with infinite variance. *Stochastic Processes and Their Applications* 40, 145–180.

Davis, R. & S.I. Resnick (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities. *Annals of Probability* 13, 179–195.

Davis, R.A. & T. Mikosch (2008) Extreme value theory for space-time processes with heavy-tailed distributions. *Stochastic Processes and Their Applications* 118, 560–584.

Feller, W. (1971) *An Introduction to Probability Theory and Its Application*, vol. 2, 2nd ed. Wiley.

Geman, S. & D. Geman (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721–741.

Huber, P.J. (1981) *Robust Statistics*, Wiley Series in Probability and Mathematical Statistics. Wiley.

Jain, A.K. (1981) Advances in mathematical models for image processing. *Proceedings of IEEE* 69, 502–528.

Kallenberg, O. (1983) *Random Measures*. Akademie-Verlag.

Kempton, R.A. & C.W. Howes (1981) The use of neighbouring plot values in the analysis of variety trials. *Applied Statistics* 30, 59–70.

Knight, K. (1989) Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. *Canadian Journal of Statistics* 17, 261–278.

Le Page, R., M. Woodroofe, & J. Zinn (1981) Convergence to a stable distribution via order statistics. *Annals of Probability* 9, 624–632.

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

Martin, R.J. (1990) The use of time-series models and methods in the analysis of agricultural field trials. *Communications in Statistical Theory and Methods* 19, 55–81.

McLeish, D.L. (1974) Dependent central limit theorems and invariance principles. *Annnals of Probability* 2, 620–628.

Resnick, S.I. & P. Greenwood (1979) A bivariate stable characterization and domain of attraction. *Journal of Multivariate Analysis* 9, 206–221.

Resnick, S.I. (1986) Point processes, regular variation and weak convergence. *Advances in Applied Probability* 18, 66–138.

Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag.

Resnick, S.I. (2007) *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer-Verlag.
Samorodnitsky, G. & M. Taqqu (1994) *Stable Non-Gaussian Random Processes*. Chapman and Hall.
Tjøstheim, D. (1978) Statistical spatial series modelling. *Advances in Applied Probability* 10, 130–154.
Tjøstheim, D. (1981) Autoregressive modeling and spectral analysis of array data in the plane. *IEEE Transactions on Geoscience and Remote Sensing* 19, 15–24.
Tjøstheim, D. (1983) Statistical spatial series modelling. II. Some further results on unilateral lattice processes. *Advances in Applied Probability* 15, 562–584.
Van de Geer, S. (2000) *Empirical Processes in M-Estimation*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
Walsh, J.B. (1986) Martingales with a multidimensional parameter and stochastic integrals in the plane. In *Lectures in Probability and Statistics*, Lecture Notes in Mathematics 1215, pp. 329–491. Springer-Verlag.
Whittle, P. (1954) On stationary processes in the plane. *Biometrika* 41, 434–449.
Zarepour, M. & S.M. Roknossadati (2008) Multivariate autoregression of order one with infinite variance innovations. *Econometric Theory* 24, 677–695.

APPENDIX

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 and separable metric space, Resnick, 2007), 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 $f \in C_K^+(\mathrm{E})$, 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

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

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 $f_i \in C_K^+(\mathrm{E})$, 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.

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. As an example, the weak limits in Proposition 1 and Corollary 1 later in this Appendix are point processes. For more details on random measures and point processes see Kallenberg (1983) and Resnick (2007).

To get ready to prove Theorem 1 we need to provide two propositions that are similar to Proposition A.1 and Proposition A.2 of Davis et al. (1992). The proof of Proposition A.2 in Davis et al. (1992) extends in a straightforward way to MA random fields and therefore is 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 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 $\mathbb{E} = \mathbb{R} \times (\overline{\mathbb{R}}\backslash\{0\})$ where $\overline{\mathbb{R}} = [-\infty, \infty]$ equipped with vague topology. Also, let $\mathcal{E}$ be the σ-algebra generated by the open subsets of $\mathbb{E}$. For each x in $\mathbb{E}$, define a set function $\varepsilon_x(\cdot)$ on $\mathcal{E}$ by

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

where $A \in \mathcal{E}$.

PROPOSITION 1. *Suppose that the sequence of random variables* $\{Z_{ij}\}$ *is given by (4) with* $\{\epsilon_{ij}\}$ *in* $DS(\alpha)$. *Then*

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

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

Proof. The proof of this proposition is a straightforward extension of Theorem 2.4 in Davis and Resnick (1985) or Proposition 4.27 in Resnick (1987). Davis and Mikosch (2008) also extend the results for an infinite-order MA space-time process given by $X_t(\boldsymbol{s}) = \sum_{i=0}^{\infty} \psi_i(\boldsymbol{s}) Z_{t-i}(\boldsymbol{s})$, $\boldsymbol{s} \in [0,1]^2$, where $\{Z_t\}_{t\in\mathbb{Z}}$ is an i.i.d. sequence of random fields on $[0,1]^2$ with values in D_2. They express the regular variation and point process convergence and other relevant results in terms of $\hat{\omega}$ convergence of boundedly finite measures on a space that is not locally compact. Because our state space $\mathbb{E}$ is locally compact, a boundedly finite measure is a Radon measure, and $\hat{\omega}$ convergence coincides with vague convergence (Kallenberg, 1983). Therefore, an adaption of the results in Davis and Mikosch (2008), i.e., Theorem 5.5, can be considered as the proof. ∎

As a consequence we obtain the following corollary.

COROLLARY 1. *Consider model (4) and a nonnegative continuous function* f *on* $\mathbb{R} \times (\overline{\mathbb{R}}\backslash\{0\})$ *with compact support; then*

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

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

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

$$\lim_{\delta\to 0}\lim_{M\to\infty}\limsup_{n} P\left(\Big| \sum_{i=1}^{n}\sum_{j=1}^{n} f(\epsilon_{ij}, a_n^{-1} Z_{ij})[1 - I(|\epsilon_{ij}| \le M) \times I(|Z_{ij}| > \delta a_n)] \Big| > \epsilon \right) = 0 \qquad \textbf{(A.1)}$$

and

$$\sum_{i=1}^{\infty}\sum_{l=1}^{\infty}\sum_{k=1}^{\infty} f(\epsilon_{ikl}, c_{kl}\delta_i\Gamma_i^{-1/\alpha}) I(|\epsilon_{ikl}| \le M) I(|c_{kl}\delta_i\Gamma_i^{-1/\alpha}| > \delta)$$
$$\to_p \sum_{i=1}^{\infty}\sum_{l=1}^{\infty}\sum_{k=1}^{\infty} f(\epsilon_{ikl}, c_{kl}\delta_i\Gamma_i^{-1/\alpha}) \tag{A.2}$$

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

PROPOSITION 2. *Suppose that* $\{Z_{ij}\}$ *is given by (4) with* $\{\epsilon_{ij}\}$ *in* $DS(\alpha)$. *Let* $\{V_{ij}\}$ *be an 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* $\eta > 0$,

(i) $\limsup_{n\to\infty} P\left[\sum_{i=1}^{n}\sum_{j=1}^{n} |V_{ij}||a_n^{-1}Z_{ij}|^{\nu} I(|Z_{ij}| \le \delta a_n) > \eta\right] \le \eta^{-1} c_1 \mathrm{E}|V_{11}|\delta^{\nu-\alpha}$
for all $\nu > \alpha$,

(ii) $\limsup_{n\to\infty} P\left[\sum_{i=1}^{n}\sum_{j=1}^{n} |V_{ij}||a_n^{-1}Z_{ij}|^{\nu} I(|Z_{ij}| > \delta a_n) > \eta\right] \le c_2\delta^{-\alpha} P(|V_{11}| > 0)$
for all $\nu > 0$ *and constants* c_1 *and* c_2. *If, in addition,* V_{11} *has 0 mean and finite variance and* $1 \le \alpha < 2$, *then*

(iii) $\mathrm{Var}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} V_{ij}(a_n^{-1}Z_{ij}) I(|Z_{ij}| \le \delta a_n)\right] \to 0$
as $n \to \infty$, *and then* $\delta \to 0$.

Proof of Theorem 1. First, we show that finite-dimensional distribution of $W_n(\cdot)$ converges weakly. Here, we mimic the proof of Theorem 1 of Davis et al. (1992), omitting some details. 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 1 it follows that

$$W_n(\boldsymbol{u};\delta,M) = \sum_{i=1}^{n}\sum_{j=1}^{n} [\rho(\epsilon_{ij} - Y_{ni,nj}(\boldsymbol{u})) - \rho(\epsilon_{ij})] I(|\epsilon_{ij}| \le M, |Z_{ij}| > a_n\delta) \to_d,$$

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

It suffices to show that (A.1) and (A.2) hold with the function $f(x,y) = \rho(x-y) - \rho(x)$. Using the Taylor series expansion around ϵ_{ij} for each term of $W_n(\cdot)$ and Proposition 2, we get the required result.

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

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

Observe that by Lipschitz continuity of ψ we get

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

which implies

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

Now using Proposition 2 along with a modification of Proposition 1 we get

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

and

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

This completes the proof. ■

The following lemma from Knight (1989, p. 276) is used to prove Theorem 2 (see also Davis et al., 1992, Lem. 2.2).

LEMMA 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* $(u_1,\ldots,u_k)$*,*

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

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

Proof of Theorem 2. We prove part (i) of the theorem. Because $\rho(\cdot)$ is convex, by Lemma 1 it is enough to show weak convergence for the finite-dimensional distribution (denoted by $\to_{fidi}$). Let $(\hat{\phi}_1,\hat{\phi}_2)$ be the M-estimator of (ϕ_1,ϕ_2) in model (1) and define $X_{ij} := Z_{ij} - \phi_2 Z_{i,j-1}$ and $Y_{ij} := Z_{ij} - \phi_1 Z_{i-1,j}$. Then $X_{ij} = \phi_1 X_{i-1,j} + \epsilon_{ij}$, and $Y_{ij} = \phi_2 Y_{i,j-1} + \epsilon_{ij}$, which yields the nonstationary part $X_{ij} = \sum_{k=1}^{i}\epsilon_{kj}$, assuming $\phi_1 = 1$, and the stationary part $Y_{ij} = \sum_{l=1}^{j}\phi_2^{j-l}\epsilon_{il}$. Also recall from before that $\epsilon_{ij} = \epsilon_{ij}(\phi_1,\phi_2)$ and observe that

$$\begin{aligned}\epsilon_{ij}\left(\phi_1 + a_n^{-1}n^{-1/2}u_1, \phi_2 + a_n^{-1}u_2\right) &= \epsilon_{ij}(\phi_1,\phi_2) - a_n^{-1}n^{-1/2}u_1 X_{i-1,j} - a_n^{-1}u_2 Y_{i,j-1}\\ &\quad + a_n^{-2}n^{-1/2}u_1u_2 Z_{i-1,j-1}.\end{aligned}$$

Thus, we shall minimize the function

$$K_n(u_1,u_2) = \sum_{i=2}^{n}\sum_{j=2}^{n}\left[\rho\left(\epsilon_{ij}\left(\phi_1 + a_n^{-1}n^{-1/2}u_1, \phi_2 + a_n^{-1}u_2\right)\right) - \rho\left(\epsilon_{ij}(\phi_1,\phi_2)\right)\right]$$

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

Notice that because ρ is a strictly convex function, the objective function $g(\phi_1,\phi_2)$ in (2) can be easily shown to be jointly strictly convex in ϕ_1 and ϕ_2. Therefore $K_n(u_1,u_2)$ has a unique minimum at

$$(u_1,u_2) = \big(\sqrt{n}\,a_n(\hat{\phi}_1 - 1), a_n(\hat{\phi}_2 - \phi_2)\big).$$

We break up the rest of the proof into two steps.

Step 1. First we demonstrate that the process K_n can be approximately decomposed into two marginal convex processes $K_n^{(1)}$ and $K_n^{(2)}$. The first process involves the marginal unit root process X, whereas the second process involves the marginal stationary process Y, both defined as before. This decomposition makes the limit theory almost clear.

More precisely, we show

$$K_n(u_1,u_2) = K_n^{(1)}(u_1) + K_n^{(2)}(u_2) + o_p(1), \tag{A.3}$$

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 the Lipschitz continuity of ψ, we get

$$\sum_{i=2}^{n}\sum_{j=2}^{n}\int_0^{a_n^{-1}u_2Y_{i,j-1} - a_n^{-2}n^{-1/2}u_1u_2Z_{i-1,j-1}}$$

$$\times \left|\psi(\epsilon_{ij} - a_n^{-1} n^{-1/2} u_1 X_{i-1,j} - s) - \psi(\epsilon_{ij} - s)\right| ds$$

$$\leq k_1 \sum_{i=2}^{n} \sum_{j=2}^{n} \left|a_n^{-1} u_2 Y_{i,j-1} - a_n^{-2} n^{-1/2} u_1 u_2 Z_{i-1,j-1}\right| \left|- a_n^{-1} n^{-1/2} u_1 X_{i-1,j}\right|^{\lambda_1}$$

$$\leq k_1 \sup_{2\leq i\leq n} \left|u_1^{\lambda_1} a_n^{-\lambda_1} \sum_{j=2}^{n} X_{i-1,j}^{\lambda_1}\right| o_p(1) \to_p 0$$

using Assumption A2 and noting that

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

Therefore,

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

Continuing in this manner we obtain

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

and finally because

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

we get (A.3).

Step 2. Now, by an argument similar to that in Theorem 2 of Knight (1989) it can be shown that $K_n^{(1)}(u_1) \to_{fidi} K^{(1)}(u_1)$, where

$$K^{(1)}(u_1) = -\sigma u_1 \int_0^1 \int_0^1 \mathcal{S}_\alpha(t_1, t_2)\, dW(t_1, t_2) + \frac{\gamma}{2} u_1^2 \int_0^1 \int_0^1 \mathcal{S}_\alpha^2(t_1, t_2)\, dt_1 dt_2. \tag{A.4}$$

Also, an application of Theorem 1 in Section 2 implies $K_n^{(2)}(u_2) \to_{fidi} K^{(2)}(u_2)$, where

$$K^{(2)}(u_2) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \rho_{ij} \left(u_2 \phi_2^{j-1} \delta_i \Gamma_i^{-1/\alpha}\right).$$

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

To prove the first statement, using the Taylor series expansion of $K_n^{(1)}(u_1)$ around $u_1 = 0$ we 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 $\left|\psi'(\epsilon_{ij})-\psi'(\epsilon^*_{ij})\right| \leq k_2 \left|u_1 a_n^{-1} n^{-1/2} X_{i-1,j}\right|$. Therefore, we can replace $\psi'(\epsilon^*_{ij})$ by $\psi'(\epsilon_{ij})$ because

$$
\begin{aligned}
&u_1^2 a_n^{-2} n^{-1} \sum_{i=2}^{n}\sum_{j=2}^{n} X_{i-1,j}^2 \left|\psi'(\epsilon^*_{ij})-\psi'(\epsilon_{ij})\right| \\
&\quad \leq k_2 u_1^2 a_n^{-2} n^{-1} \sum_{i=2}^{n}\sum_{j=2}^{n} X_{i-1,j}^2 \times \left|u_1 a_n^{-1} n^{-1/2} X_{i-1,j}\right| \\
&\quad \leq k_2 u_1^3 n^{-1/2} \left(n^{-1} \sum_{i=2}^{n} a_n^{-3} \sum_{j=2}^{n} |X_{i-1,j}|^3\right) \to_p 0
\end{aligned}
$$

uniformly over u_1 in compact sets, using the fact that

$$
n^{-1} a_n^{-3} \sum_{i=2}^{n}\sum_{j=2}^{n} |X_{i-1,j}|^3 = O_p(1).
$$

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

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

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

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

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 Bhattacharyya et al. (1997), it is easy to see that $\{U_t(n),\mathfrak{F}_t(n), t\in F\}$ is a strong martingale array (for the related details, see Walsh, 1986). Therefore, applying a weak law of large numbers for martingales (Brown, 1971; McLeish, 1974) we get

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

The other term, the approximation error, is an $o_p(1)$ term due to the truncation at M:

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

as $n\to\infty$ first and then $M\to\infty$, because if $a_n^{-2}X_{i-1,j}^2\leq M$ for each i,j; $1\leq i,j\leq n$, then the term inside $|\cdot|$ cannot be greater than δ; i.e., it equals zero. Thus $K_n^{(1)}(u_1)\to_{fidi} K^{(1)}(u_1)$, with $K^{(1)}(u_1)$ as given in (A.4). ■