Econometric Theory, **24**, 2008, 677–695. Printed in the United States of America.
doi: 10.1017/S0266466608080286

MULTIVARIATE AUTOREGRESSION OF ORDER ONE WITH INFINITE VARIANCE INNOVATIONS

M. ZAREPOUR AND S.M. ROKNOSSADATI
University of Ottawa

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.

1. INTRODUCTION

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

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

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 paper, 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, 1997).

For $d = 1$, in the case that the errors (innovations) have infinite variance, several studies have been developed. Chan and Tran (1989) and Chan (1990) 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 ordinary least squares (OLS) estimates whereas Knight (1989) suggested the *M*-estimates under the same model. Phillips (1990) studied the OLS estimates under weakly dependent innovations with finite variance, and Knight (1991) took into account the *M*-estimates in the case that the innova-

The authors are greatly indebted to two anonymous referees and the co-editor for several helpful comments. This paper is supported by the Natural Sciences and Engineering Research Council of Canada. 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.

© 2008 Cambridge University Press 0266-4666/08 $15.00

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

tions 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 (1992). 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 nonnormal innovations having finite variance, Cox and Llatas (1991) and Lucas (1995) developed *M*-estimation under i.i.d. and weakly dependent errors, respectively.

Caner (1997) 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 (2001).

In this paper, 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 (1997). 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. (1992) for AR processes with the usual stationarity constraints, which differs from our case. It seems likely that the findings of this paper can be extended to weakly dependent errors such as MA(∞), but we do not address this question here.

2. PRELIMINARIES

Consider the model (1). The OLS estimator for $\mathbf{\Phi}$ in this model is given by

$$\hat{\mathbf{\Phi}} = \left(\sum_{t=1}^{n} \mathbf{U}_t \mathbf{U}'_{t-1} \right) \left(\sum_{t=1}^{n} \mathbf{U}_{t-1} \mathbf{U}'_{t-1} \right)^{-1}. \tag{2}$$

Throughout this paper we denote the transpose of the matrix A by A'.

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

Define the stochastic processes

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

$$\mathbf{V}_n'(\cdot) = (V_n^{(1)}(\cdot), V_n^{(2)}(\cdot), \ldots, V_n^{(d)}(\cdot)), \qquad V_n^{(j)}(\cdot) = (a_n^{(j)})^{-2} \sum_{i=1}^{[n\cdot]} (\epsilon_i^{(j)})^2, \tag{3}$$

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 paper, $\rightarrow_d$ stands for weak convergence, and all 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

$$\mathbf{S}_n'(\cdot) \rightarrow_d \mathbf{S}'(\cdot) := (S_1(\cdot), \ldots, S_d(\cdot)) \tag{4}$$

in D^d, where $\mathbf{S}_n'(\cdot)$ is defined as in (3). Following Resnick and Greenwood (1979), 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

$$\mathbf{V}_n'(\cdot) \rightarrow_d \mathbf{V}'(\cdot) := (V_1(\cdot), \ldots, V_d(\cdot)) \tag{5}$$

in D^d, where $\mathbf{V}_n'(\cdot)$ is defined as in (3), and $\mathbf{V}'(\cdot)$ is a stable process in D^d. As Paulauskas and Rachev (1998) demonstrate, we can generalize the convergence in (4) and (5) to weak convergence in D_d. Note that for $1 < \alpha_j \leq 2$, $j = 1, \ldots, d$, (4) does not hold unless the coordinates of $\{\boldsymbol{\epsilon}_t\}$ are in the domain of attraction of a symmetric stable law (see, e.g., Knight, 1989).

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 $\mathbb{R}^d \backslash \{\mathbf{0}\}$ defined by $\nu \circ \tau = \tilde{\nu}$, where

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

with $\tilde{\nu}$ given by

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

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

and S is a finite measure on $[0, 2\pi]$. The limiting stable processes in (4) and (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 $\mathbf{j}_i = (j_i^{(1)}, \ldots, j_i^{(d)})'$ and $\mathbf{j}_i^2 = ((j_i^{(1)})^2, \ldots, (j_i^{(d)})^2)'$, where $\mathbf{j}_i \in \mathbb{R}^d \backslash \{\mathbf{0}\}$, $i \geq 1$, 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 and Greenwood (1979) and Resnick (1986).

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, 1971). 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, 1979). 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 (1980) for more details.

The classical M-estimator $\hat{\mathbf{\Phi}}_M$ of $\mathbf{\Phi}$ in (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 (1), from Knight (1989) we know that

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

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, 1989)

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

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

Therefore, the fact that $\sqrt{n}/a_n \to 0$ 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

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

and

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

where L is a slowly varying function at infinity for some random variables X_1 and X_2. When the innovations have stable or Pareto-like distributions, we have $n^{1/\alpha}L(n)/a_n \to 0$ (for more details, see Davis et al., 1992; Brockwell and Davis, 1991).

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, 1981) given by

$$\rho_H(x) = \begin{cases} \dfrac{1}{2}x^2 & \text{if } |x| \le c, \\ c|x| - \dfrac{1}{2}c^2 & \text{if } |x| > c, \end{cases} \tag{6}$$

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 in the next section.

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

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

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

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

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

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

A3–A6 of Knight (1989). Convexity of $\rho(\cdot)$ follows from the positive definiteness of the diagonal Hessian matrix

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

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, 1989, p. 266). For illustration, we consider the bivariate Huber's function

$$\rho^*(x, y) = \rho_H(x) + \rho_H(y) \tag{7}$$

and

$$\rho^{**}(x, y) = \rho_T(x) + \rho_T(y), \tag{8}$$

where $\rho_H(\cdot)$ is Huber's loss function defined in (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., 1992, Remarks 1 and 2). Actually, the performance of $\rho^{**}(\cdot)$ is as good as that of $\rho^*(\cdot)$, as shown in Section 4. Notice that we require just one such function $\rho(\cdot)$ satisfying the conditions A2–A4. Figure 1 illustrates the behavior of both loss functions.

The following proposition, which is in fact Proposition 2 of Paulauskas and Rachev (1998), 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, 1991, Thm. 2.7).

PROPOSITION 1. *Suppose that (4) holds. Then for* $0 \le t \le 1$, *as* $n \to \infty$,

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

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

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

Proof. See page 782 of Paulauskas and Rachev (1998) and the related discussions therein and Proposition 2 of Mittnik et al. (2001). ■

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

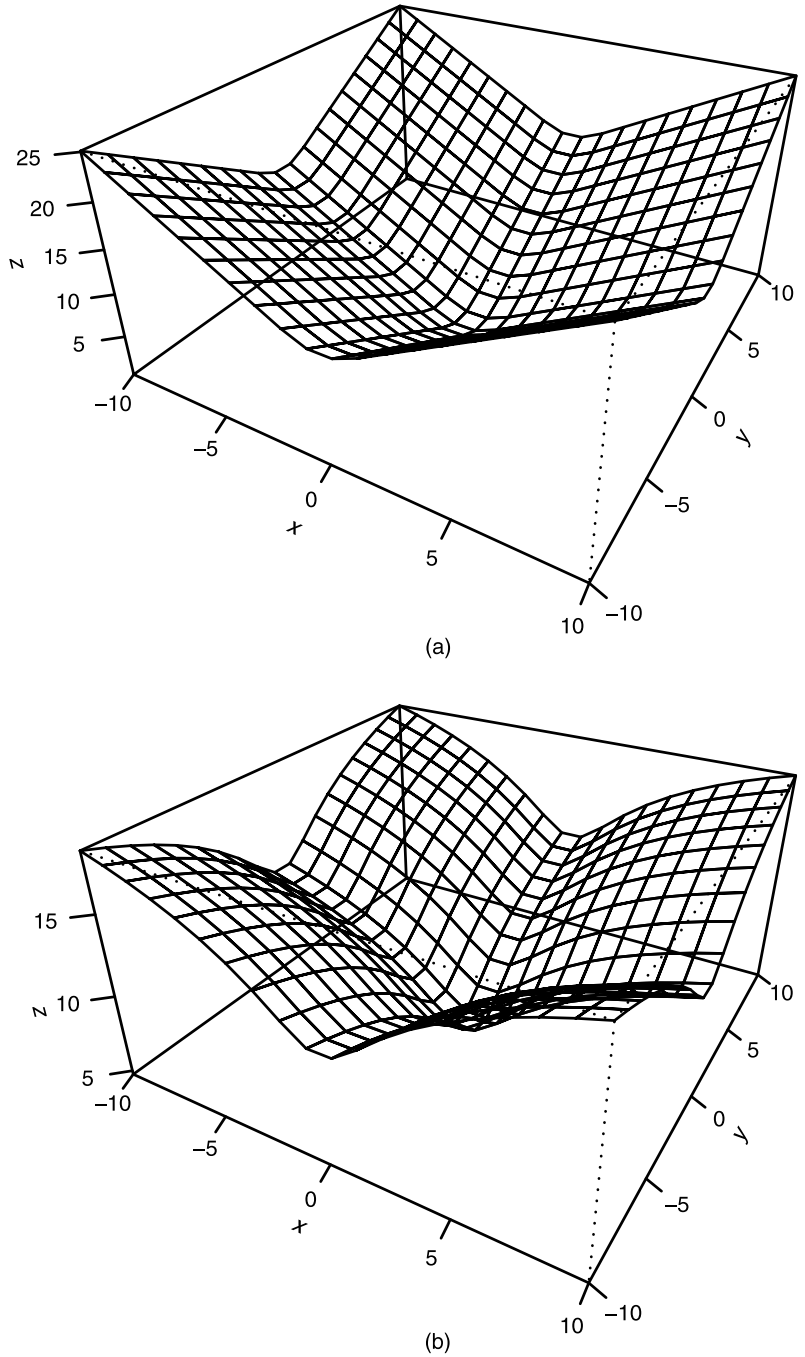


FIGURE 1. (a) Bivariate Huber's loss function; (b) bivariate loss function given in (8).

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

Finally, we present the following lemma from Knight (1989, p. 276), which will be used to prove Theorem 2 in Section 3 (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 $T(\cdot)$ *has a unique minimum* $\hat{u}$. *If* $\hat{u}_n$ *minimizes* $T_n(\cdot)$, *then* $\hat{u}_n \to_d \hat{u}$.

3. OLS AND *M*-ESTIMATES FOR MULTIVARIATE RANDOM WALKS

In the first theorem in this section, for the model (1) we find the weak limit behavior of $\mathbf{A}_n \times (\hat{\mathbf{\Phi}} - \mathbf{I})$ 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 proved in the bivariate case; $d = 2$ (see also Remark 2 in this section).

THEOREM 1. *Consider the model (1) with* $\mathbf{\Phi} = \mathbf{I}$ *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*

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

where $\hat{\mathbf{\Phi}}$ *is the OLS estimator given in (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 (4).*

Remark 1. Theorem 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 (1997). On the other hand, if $a_n^{(1)} \neq a_n^{(2)}$, then one of the OLS estimates might be inconsistent. More precisely, let $\hat{\mathbf{\Phi}} = \begin{pmatrix} \hat{\alpha} & \hat{\beta} \\ \hat{\gamma} & \hat{\theta} \end{pmatrix}$ be the OLS estimator of $\mathbf{\Phi} = \mathbf{I}_2$ in model (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 $\alpha_1 \geq \frac{2}{3}$. Note that if $\alpha_1, \alpha_2 \geq 1$, the OLS estimates remain consistent.

Proof. Let $\hat{\mathbf{\Phi}} = \begin{pmatrix} \hat{\alpha} & \hat{\beta} \\ \hat{\gamma} & \hat{\theta} \end{pmatrix}$ be the OLS estimator of $\mathbf{\Phi} = \mathbf{I}_2$ in model (1). From (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:

$$\hat{\alpha} - 1 = \frac{\sum_{t=2}^{n} Y_{t-1}^2 \sum_{t=2}^{n} X_{t-1}\epsilon_t^{(1)} - \sum_{t=2}^{n} X_{t-1}Y_{t-1} \sum_{t=2}^{n} Y_{t-1}\epsilon_t^{(1)}}{\sum_{t=2}^{n} X_{t-1}^2 \sum_{t=2}^{n} Y_{t-1}^2 - \left(\sum_{t=2}^{n} X_{t-1}Y_{t-1}\right)^2}, \tag{9}$$

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

$$\hat{\beta} = \frac{\sum_{t=2}^{n} X_{t-1}^2 \sum_{t=2}^{n} Y_{t-1}\epsilon_t^{(1)} - \sum_{t=2}^{n} X_{t-1}Y_{t-1} \sum_{t=2}^{n} X_{t-1}\epsilon_t^{(1)}}{\sum_{t=2}^{n} X_{t-1}^2 \sum_{t=2}^{n} Y_{t-1}^2 - \left(\sum_{t=2}^{n} X_{t-1}Y_{t-1}\right)^2}, \tag{10}$$

$$\hat{\gamma} = \frac{\sum_{t=2}^{n} Y_{t-1}^2 \sum_{t=2}^{n} X_{t-1}\epsilon_t^{(2)} - \sum_{t=2}^{n} X_{t-1}Y_{t-1} \sum_{t=2}^{n} Y_{t-1}\epsilon_t^{(2)}}{\sum_{t=2}^{n} X_{t-1}^2 \sum_{t=2}^{n} Y_{t-1}^2 - \left(\sum_{t=2}^{n} X_{t-1}Y_{t-1}\right)^2}, \tag{11}$$

$$\hat{\theta} - 1 = \frac{\sum_{t=2}^{n} X_{t-1}^2 \sum_{t=2}^{n} Y_{t-1}\epsilon_t^{(2)} - \sum_{t=2}^{n} X_{t-1}Y_{t-1} \sum_{t=2}^{n} X_{t-1}\epsilon_t^{(2)}}{\sum_{t=2}^{n} X_{t-1}^2 \sum_{t=2}^{n} Y_{t-1}^2 - \left(\sum_{t=2}^{n} X_{t-1}Y_{t-1}\right)^2}. \tag{12}$$

Next, we obtain the normalization coefficients for the sums in (9)–(12). In fact, we have

$$(a_n^{(1)})^{-2} \sum_{t=2}^{n} X_{t-1}\epsilon_t^{(1)} = \frac{1}{2}[(S_n^{(1)}(1))^2 - V_n^{(1)}(1)] = \int_0^1 S_n^{(1)}(s)\, dS_n^{(1)}(s), \tag{13}$$

$$(a_n^{(1)})^{-1}(a_n^{(2)})^{-1} \sum_{t=2}^{n} Y_{t-1}\epsilon_t^{(1)} = \int_0^1 S_n^{(2)}(s)\, dS_n^{(1)}(s), \tag{14}$$

$$n^{-1}(a_n^{(2)})^{-2} \sum_{t=2}^{n} Y_{t-1}^2 = \int_0^1 (S_n^{(2)})^2(s)\, ds, \tag{15}$$

$$n^{-1}(a_n^{(1)})^{-1}(a_n^{(2)})^{-1} \sum_{t=2}^{n} X_{t-1}Y_{t-1} = \int_0^1 S_n^{(1)}(s) S_n^{(2)}(s)\, ds, \tag{16}$$

$$n^{-1}(a_n^{(1)})^{-2} \sum_{t=2}^{n} X_{t-1}^2 = \int_0^1 (S_n^{(1)})^2(s)\, ds, \tag{17}$$

$$(a_n^{(1)})^{-1}(a_n^{(2)})^{-1} \sum_{t=2}^{n} X_{t-1}\epsilon_t^{(2)} = \int_0^1 S_n^{(1)}(s)\, dS_n^{(2)}(s), \tag{18}$$

and

$$(a_n^{(2)})^{-2} \sum_{t=2}^{n} Y_{t-1}\epsilon_t^{(2)} = \int_0^1 S_n^{(2)}(s)\, dS_n^{(2)}(s). \tag{19}$$

Therefore, using (9)–(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.

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

Thus, we can write all the elements of $\mathbf{A}_n \times (\hat{\mathbf{\Phi}} - \mathbf{I}_2)$ as a function of the bivariate stable process $\mathbf{S}_n$. The rest of the proof is based on Proposition 1. As in Mittnik et al. (2001), we define a map $f: D_6 \to \mathbb{R}^4$, $f = (f_1, f_2, f_3, f_4)$ 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 $g_i: D_6 \to \mathbb{R}$, $i = 1, \ldots, 5$, 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 1 (with $d = 2$) to complete the proof. ■

Remark 2. Theorem 1 also extends the result of Caner (1997) in two directions. First, Theorem 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 1 is modified as follows:

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

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 paper, the innovation vector $\boldsymbol{\epsilon}_t$ has dependent components, which arises in more practical situations and differs from Assumption 1 of Caner (1997).

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

The next theorem is the natural extension of the M-estimates for a multivariate random walk model. Consider the model (1) with

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

and

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

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

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

Furthermore, put

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

where

$$\mathbf{u}_n^{(1)} = \sqrt{n}\,(a_n^{(1)}(\hat{\alpha}_M^{(11)} - 1), a_n^{(2)}\hat{\alpha}_M^{(12)}, \ldots, a_n^{(d)}\hat{\alpha}_M^{(1d)}),$$
$$\mathbf{u}_n^{(2)} = \sqrt{n}\,(a_n^{(1)}\hat{\alpha}_M^{(21)}, a_n^{(2)}(\hat{\alpha}_M^{(22)} - 1), \ldots, a_n^{(d)}\hat{\alpha}_M^{(2d)}), \ldots,$$
$$\mathbf{u}_n^{(d)} = \sqrt{n}\,(a_n^{(1)}\hat{\alpha}_M^{(d1)}, a_n^{(2)}\hat{\alpha}_M^{(d2)}, \ldots, a_n^{(d)}(\hat{\alpha}_M^{(dd)} - 1)).$$

Then we obtain the following result.

THEOREM 2. *Under Assumptions A1–A4,*

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

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

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

$\mathbf{A}_{ij} = E(\rho_{ij}(\boldsymbol{\epsilon}_1))\int_0^1 \mathbf{S}(s)\mathbf{S}'(s)\,ds, i, j = 1, \ldots, d,$ *and* $\mathbf{b} = (\mathbf{b}_1', \ldots, \mathbf{b}_d')'$ *with* $\mathbf{b}_i' = \sqrt{E(\rho_i^2(\boldsymbol{\epsilon}_1))}\int_0^1 \mathbf{S}'(s)\,dW_i(s), i = 1, \ldots, d.$ *Here, the stable random vector* $\mathbf{S}(\cdot)$ *is defined as in (4), and* $W_i(\cdot)$ *is the standard Brownian motion.*

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

Proof. Here we mimic the proof of Theorem 2 of Knight (1989). First, observe that the vector $-\mathbf{u}_n$ minimizes

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

where

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

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

$$\begin{aligned} R_n(u_{11},\ldots,u_{dd}) &= n^{-1/2} \sum_{t=2}^{n} \sum_{j=1}^{d} \sum_{i=1}^{d} u_{ij} (a_n^{(j)})^{-1} \rho_i(\boldsymbol{\epsilon}_t) X_{t-1}^{(j)} \\ &\quad + (2n)^{-1} \sum_{t=2}^{n} \sum_{j=1}^{d} \sum_{i=1}^{d} u_{ij}^2 (a_n^{(j)})^{-2} \rho_{ii}(\mathbf{c}_t) (X_{t-1}^{(j)})^2 \\ &\quad + n^{-1} \sum_{t=2}^{n} \sum_{k=1}^{d} \sum_{j=2}^{d} \sum_{i=1}^{j-1} u_{ik} u_{jk} (a_n^{(k)})^{-2} \rho_{ij}(\mathbf{c}_t) (X_{t-1}^{(k)})^2 \\ &\quad + n^{-1} \sum_{t=2}^{n} \sum_{k=1}^{d} \sum_{j=1}^{d} \sum_{i=k+1}^{d} u_{jk} u_{ji} (a_n^{(i)})^{-1} (a_n^{(k)})^{-1} \rho_{jj}(\mathbf{c}_t) X_{t-1}^{(i)} X_{t-1}^{(k)} \\ &\quad + n^{-1} \sum_{t=2}^{n} \sum_{k=1}^{d} \sum_{j=2}^{d} \sum_{\substack{l=1 \\ l<j}}^{d-1} \sum_{\substack{i=1 \\ i\neq k}}^{d} u_{lk} u_{ji} (a_n^{(k)})^{-1} (a_n^{(i)})^{-1} \rho_{lj}(\mathbf{c}_t) X_{t-1}^{(k)} X_{t-1}^{(i)}, \end{aligned}$$

where

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

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

$$\begin{pmatrix} \mathbf{S}_n(\cdot) \\ \mathbf{W}_n(\cdot) \end{pmatrix} \to_d \begin{pmatrix} \mathbf{S}(\cdot) \\ \mathbf{W}(\cdot) \end{pmatrix}, \tag{20}$$

where $\mathbf{S}_n(\cdot)$ and $\mathbf{S}(\cdot)$ are defined as in (3) and (4) and

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

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

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 $(E(\rho_i(\boldsymbol{\epsilon}_1)\rho_j(\boldsymbol{\epsilon}_1)))_{i,j=1,\ldots,d}$. Because

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

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

$$\begin{aligned} &n^{-1}\sum_{t=2}^{n}\sum_{j=1}^{d}\sum_{i=1}^{d} u_{ij}^2 (a_n^{(j)})^{-2}(X_{t-1}^{(j)})^2 |\rho_{ii}(\mathbf{c}_t) - \rho_{ii}(\boldsymbol{\epsilon}_t)| \\ &\quad \le \lambda \sum_{j=1}^{d}\sum_{i=1}^{d}\sum_{r=1}^{d}\sum_{s=1}^{d} c_r u_{ij}^2 |u_{rs}| n^{-1/2} \left[n^{-1} \sum_{t=2}^{n} (a_n^{(j)})^{-2}(a_n^{(s)})^{-1} |(X_{t-1}^{(j)})^2 X_{t-1}^{(s)}| \right] \\ &\quad \to_p 0 \end{aligned}$$

uniformly over u_{ij} in compact sets. Here, $\to_p$ stands for convergence in probability. Furthermore, in the limit each $\rho_{ij}(\boldsymbol{\epsilon}_t)$ can be substituted by $E(\rho_{ij}(\boldsymbol{\epsilon}_t))$. To see this for the second term of R_n, by applying a weak law of large numbers for martingales we derive

$$\begin{aligned} &\sum_{j=1}^{d} u_{ij}^2 n^{-1} \sum_{t=2}^{n} (a_n^{(j)})^{-2}(X_{t-1}^{(j)})^2 I((a_n^{(j)})^{-2}(X_{t-1}^{(j)})^2 \le M) \\ &\quad \times [\rho_{ii}(\boldsymbol{\epsilon}_t) - E(\rho_{ii}(\boldsymbol{\epsilon}_t))] \to_p 0, \end{aligned}$$

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

$$\begin{aligned} &P\left(\left| \sum_{j=1}^{d} u_{ij}^2 n^{-1} \sum_{t=2}^{n} (a_n^{(j)})^{-2}(X_{t-1}^{(j)})^2 I((a_n^{(j)})^{-2}(X_{t-1}^{(j)})^2 > M)[\rho_{ii}(\boldsymbol{\epsilon}_t) - E(\rho_{ii}(\boldsymbol{\epsilon}_t))] \right| > \delta \right) \\ &\quad \le P\left(\max_{1\le j\le d} \max_{2\le t\le n} (a_n^{(j)})^{-2}(X_{t-1}^{(j)})^2 > M \right) \to 0, \end{aligned}$$

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

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

Merging the preceding arguments along with applying the continuous mapping theorem and using Proposition 1, we can show that the finite-dimensional distribution of $R_n(\cdot)$ converges weakly to those of $R(\cdot)$ where

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

$$
\begin{aligned}
R(u_{11},\ldots,u_{dd}) = {} & \sum_{j=1}^{d}\sum_{i=1}^{d} u_{ij}\sqrt{E(\rho_i^2(\boldsymbol{\epsilon}_1))}\int_0^1 S_j(s)\,dW_i(s) \\
& + \frac{1}{2}\sum_{j=1}^{d}\sum_{i=1}^{d} u_{ij}^2 E(\rho_{ii}(\boldsymbol{\epsilon}_1))\int_0^1 S_j^2(s)\,ds \\
& + \sum_{k=1}^{d}\sum_{j=2}^{d}\sum_{i=1}^{j-1} u_{ik}u_{jk}E(\rho_{ij}(\boldsymbol{\epsilon}_1))\int_0^1 S_k^2(s)\,ds \\
& + \sum_{k=1}^{d}\sum_{j=1}^{d}\sum_{i=k+1}^{d} u_{jk}u_{ji}E(\rho_{jj}(\boldsymbol{\epsilon}_1))\int_0^1 S_i(s)S_k(s)\,ds \\
& + \sum_{k=1}^{d}\sum_{j=2}^{d}\sum_{\substack{l=1\\ l<j}}^{d-1}\sum_{\substack{i=1\\ i\neq k}}^{d} u_{lk}u_{ji}E(\rho_{lj}(\boldsymbol{\epsilon}_1))\int_0^1 S_i(s)S_k(s)\,ds.
\end{aligned}
$$

Thus, employing Lemma 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 3. Because $\mathbf{S}$ and $\mathbf{W}$ given in (20) are independent, using the Cramér–Wold device (see Billingsley, 1968), we achieve the following result: Define

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

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

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

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 $\{\mathbf{S}(s); 0 \le s \le 1\}$, $\mathbf{T}_i$, $i = 1,\ldots,d$, are d-variate random variables with normally distributed marginals, and the $\mathbf{T}_i$'s are not independent. Therefore, we get

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

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

where $r_{ij} = E(\rho_{ij}(\boldsymbol{\epsilon}_1))$, $r_i = \sqrt{E(\rho_i^2(\boldsymbol{\epsilon}_1))}$, $i, j = 1, \ldots, d$, and

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

is the random normalizing matrix.

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

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

The loss functions given in (7) and (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 (21) have normal distributions. See the normal quantile plots in Figure 3 of the forthcoming section.

4. SIMULATION

A simulation study is undertaken to investigate the results given in Theorems 1 and 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 (1) with the aforementioned indices of stability, we consider the set of points $\{\mathbf{S}_i = (\cos\theta_i, \sin\theta_i)' : \theta_i = i\pi/4, i = 0, 1, \ldots, 8\}$ 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 $M = 10{,}000$ times to get the random samples $\mathbf{S}_i = (S_{1i}, S_{2i})'$, $i = 1, \ldots, 10{,}000$. 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 $\boldsymbol{\epsilon}_t, t = 1, \ldots, n$, we perform this procedure again n times independently. The observations $\mathbf{U}_t =$

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

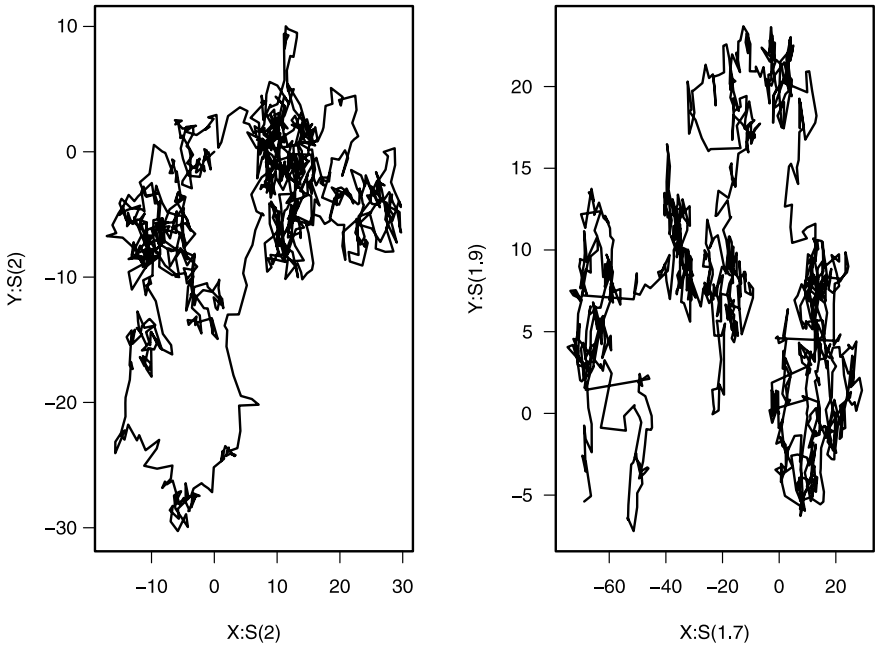


FIGURE 2. Sample path of bivariate random walks with innovations belonging to $DS(2,2)$ (left) and $DS(1.7,1.9)$.

$(X_t, Y_t)'$ 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 1, are calculated using (2). Figure 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 (2002), and Samorodnitsky and Taqqu (1994).

For the bivariate convex function ρ, we employ the loss functions $\rho^*(x,y)$ and $\rho^{**}(x,y)$ given in (7) and (8). For both the OLS and the M-estimates provided in Tables 1 and 2 we choose $n = 50, 100$, and a replication size of 1,000 is chosen. 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 1, as expected, the OLS estimates are consistent and tend to the actual values as n gets large whereas where

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

TABLE 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 1,000.

	$n = 50$				$n = 100$			
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$
$\boldsymbol{\alpha}_1 = (1.0, 1.0)$	0.9154	0.2162	−0.0708	0.9569	0.9301	−0.0419	0.0565	1.0041
$\boldsymbol{\alpha}_2 = (1.0, 2.0)$	0.9338	−0.0031	−0.0025	0.9251	0.9685	0.0071	−0.0010	0.9621
$\boldsymbol{\alpha}_3 = (0.8, 1.1)$	0.9515	−0.0887	−0.0007	0.9711	0.9857	−0.0117	0.0009	0.9806
$\boldsymbol{\alpha}_4 = (0.6, 2.0)$	0.9452	11.5710*	0.0001	0.9218	0.9628	1100.633*	0.0000	0.9591
$\boldsymbol{\alpha}_5 = (0.2, 0.9)$	0.9642	6.8606e+14**	0.0000	0.9166	0.9508	1.3737e+17**	0.0000	0.9634
$\boldsymbol{\alpha}_6 = (0.2, 1.9)$	0.9217	−1.1088e+18**	0.0000	0.9213	0.9337	−2.6234e+24**	−0.0000	0.9626

Note: * and ** indicate near inconsistency (low rate of convergence) and inconsistency cases, respectively.

TABLE 2. The *M*-estimates using Huber's function ρ^* (up) and ρ^{**} (down). The replication size is 1,000.

	$n = 50$				$n = 100$			
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$
$\boldsymbol{\alpha}_1 = (1.0, 1.0)$	0.9827	−0.0012	0.0006	0.9841	0.9947	0.0004	0.0003	0.9839
$\boldsymbol{\alpha}_2 = (1.0, 2.0)$	0.9875	0.0033	0.0003	0.9356	0.9954	−0.0015	−0.0002	0.9696
$\boldsymbol{\alpha}_3 = (0.8, 1.1)$	0.9922	0.0045	−0.0002	0.9793	0.9981	−0.0003	−0.0001	0.9935
$\boldsymbol{\alpha}_4 = (0.6, 2.0)$	0.9986	0.0008	−0.0001	0.9668	0.9997	0.0003	−0.0000	0.9892
$\boldsymbol{\alpha}_1 = (1.0, 1.0)$	0.9892	−0.0021	−0.0094	0.9907	0.9975	0.0006	0.0009	0.9945
$\boldsymbol{\alpha}_2 = (1.0, 2.0)$	0.9919	0.0010	0.0017	0.9375	0.9981	0.0002	0.0001	0.9706
$\boldsymbol{\alpha}_3 = (0.8, 1.1)$	0.9975	−0.0011	0.0004	0.9843	0.9995	0.0009	−0.0003	0.9959
$\boldsymbol{\alpha}_4 = (0.6, 2.0)$	0.9940	0.0006	0.0004	0.9698	0.9982	0.0005	−0.0003	0.9907

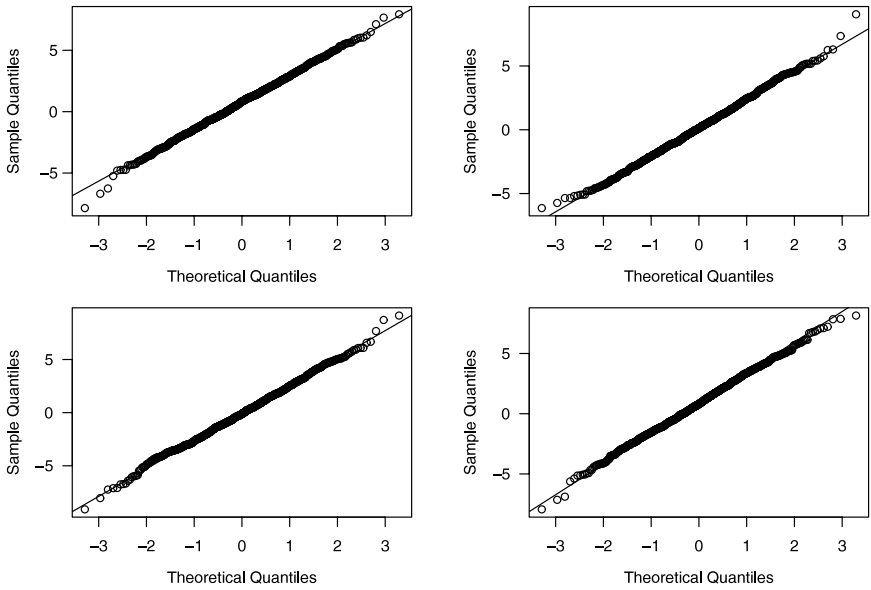


FIGURE 3. Normal Q-Q plots for the marginals of the limiting vector in (21) using model (1) with innovations in $DS(1.7, 2)$. The sample and replication size are 100 and 1,000, respectively.

$\alpha_1 < \alpha_2(1 + \alpha_2)^{-1}$, $\hat{\beta}$ is inconsistent, and the other estimates remain consistent. The results obtained from Table 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 1.

To explore the marginal normality of the limits with the random normalization factors given in (21), we consider a simulated bivariate random walk model (1) with innovations belonging to $DS(1.7, 2)$ using a sample size of $n = 100$, replication size of 1,000, and $\rho^*(\cdot)$ as the loss function in the M-estimation. Figure 3 portrays the normal quantile-quantile plots for the marginals of the limiting vector in (21), illustrating the normality of marginals.

REFERENCES

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

Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.

Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods*, 2nd ed. Springer Series in Statistics. Springer-Verlag.

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press

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

Chan, N.H. (1990) Inference for near-integrated time series with infinite variance. *Journal of the American Statistical Association* 85, 1069–1074.

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

Cox, D.D. & I. Llatas (1991) Maximum likelihood type estimation for nearly nonstationary autoregressive time series. *Annals of Statistics* 19, 1109–1128.

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

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

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

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

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

Kurtz, T.G. & P. Protter (1991) Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* 19, 1035–1070.

LePage, R. (1980) Multidimensional infinitely divisible variables and processes, I. Stable case. *Lecture Notes in Mathematics* 1391, 153–163.

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

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

Paulauskas, V. & S.T. Rachev (1998) Cointegrated processes with infinite variance innovations. *Annals of Applied Probability* 8, 775–792.

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

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

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

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

https://doi.org/10.1017/S0266466608080286 Published online by Cambridge University Press