



ELSEVIER

Stochastic Processes and their Applications 51 (1994) 135–165

stochastic
processes
and their
applications

Limit distributions for linear programming time series estimators[†]

Paul D. Feigin^{a,*}, Sidney I. Resnick^b

^a Faculty of Industrial Engineering and Management, Technion, Haifa 32000, Israel

^b Cornell University, School of Operations Research and Industrial Engineering, ETC Building, Ithaca, NY 14853 USA

Received 10 October 1992; revised 2 June 1993

Abstract

We consider stationary autoregressive processes of order p which have positive innovations. We propose consistent parameter estimators based on linear programming. Under conditions, including regular variation of either the left or right tail of the innovations distribution, we prove that the estimators have a limit distribution. The rate of convergence of our estimator is favorable compared with the Yule–Walker estimator under comparable circumstances.

Key words: Poisson processes; Linear programming; Autoregressive processes; Parameter estimation; Weak convergence; Consistency; Time series analysis

1. Introduction

Consider the stationary *autoregressive process of order p* , denoted by $AR(p)$, with positive innovations $\{Z_t\}$, and with autoregressive coefficients $\phi_1, \dots, \phi_p, \phi_p \neq 0$. These processes are defined by the following relation.

$$X_t = \sum_{k=1}^p \phi_k X_{t-k} + Z_t, \quad t = 0, \pm 1, \pm 2, \dots \quad (1.1)$$

where we assume that $\{Z_t\}$ is an independent and identically distributed (i.i.d.) sequence of random variables with left endpoint of their common distribution being 0.

*Corresponding author. E-mail: paulf@ie.technion.ac.il.

[†]This research was partially supported by the United States–Israel Binational Science Foundation (BSF) Grant No. 89-00323/1. Sidney Resnick was also partially supported by NSF Grant DMS-9100027 at Cornell University.

We assume the order p is known. Based on observation of $\{X_0, \dots, X_n\}$ we are interested in estimating the parameters, and in determining the asymptotic properties of these estimators as $n \rightarrow \infty$.

Assuming the Z 's have finite variance, the usual method of estimation would be to use the Yule–Walker estimators (see, for example, Brockwell and Davis (1991)), which would typically result in the estimators converging at the rate $n^{1/2}$. However, one can sometimes do better than the $n^{1/2}$ rate of convergence by exploiting the special nature of the innovations.

The case of $p = 1$ was discussed by Davis and McCormick (1989) who used a point process approach to obtain the asymptotic distribution of the natural estimator in the positive innovation context when the innovations distribution, F , varies regularly at 0 and satisfies a suitable moment condition. This estimator is

$$\hat{\phi}^{(n)} = \bigwedge_{j=1}^n \frac{X_j}{X_{j-1}}, \quad (1.2)$$

where \wedge denotes the minimum operator. For the case of $p > 1$, a straightforward generalization of (1.2) does not perform well, Andel (1989). Andel (1989) considered the case $p = 2$ and suggested two estimators of (ϕ_1, ϕ_2) . One is based on a maximum likelihood argument and is the one we consider in this paper for general p . As discussed in Feigin (1994) and Feigin and Resnick (1992), this estimator is obtained by solving equations which turn out to be examples of *generalized martingale estimating equations*. Andel (1989) found by simulation that this estimator converges at a faster rate than the Yule–Walker estimator. Andel's finding is explained in Feigin and Resnick (1992), who establish a *rate of consistency* for the estimators of ϕ_1 and ϕ_2 in the case $p = 2$. In this paper we consider general p and derive a limit distribution for the estimators under assumptions on the Z 's which involve moment conditions and regular variation conditions on either the left tail or the right tail. Precise statement of conditions and results is presented in Section 2.

Here is a rapid review of the derivation of our estimator: Suppose temporarily that the common distribution of the Z 's is unit exponential so that

$$P[Z_1 > x] = e^{-x}, \quad x > 0.$$

In this case, conditionally on $\{X_0 = x_0, X_{-1} = x_{-1}, \dots, X_{-p+1} = x_{-p+1}\}$, the likelihood is proportional to (using $I(\cdot)$ for the indicator function).

$$\begin{aligned} & I\left(\bigwedge_{i=1}^n \left(X_i - \sum_{i=1}^p \phi_i X_{i-i}\right) \geq 0\right) \exp\left\{\phi_1 \sum_1^n X_{i-1} + \phi_2 \sum_1^n X_{i-2} + \dots + \phi_n \sum_1^n X_{i-p}\right\} \\ & \approx I\left(\bigwedge_{i=1}^n \left(X_i - \sum_{i=1}^p \phi_i X_{i-i}\right) \geq 0\right) \exp\left\{\sum_{i=1}^p \phi_i \sum_{i=1}^n X_i\right\}. \end{aligned}$$

Assuming that $\sum_{t=1}^n X_t$ is ultimately positive, the corresponding maximum likelihood estimator will thus be approximately determined by solving the linear program (LP)

$$\max \left(\sum_{i=1}^p \phi_i \right)$$

subject to

$$X_t \geq \sum_{i=1}^p \phi_i X_{t-i}, \quad t = 1, \dots, n.$$

Note that the fact that $\sum_1^n X_{t-1} / \sum_1^n X_{t-i} \approx 1$ in the stationary case justifies the simplified (approximate) form of the objective function.

Even if the density of the Z 's is not exponential, the LP still gives an estimation procedure with good properties. Thus the form of the estimator considered in this paper is as follows: We suppose the true autoregressive parameter is $\phi^{(0)} = (\phi_1^{(0)}, \dots, \phi_p^{(0)})'$. Our estimator is obtained by solving the linear program to get

$$\hat{\phi}^{(n)} = \arg \max_{\delta \in D_n} \delta' \mathbf{1} \quad (1.3)$$

where $\mathbf{1}' = (1, \dots, 1)$ and where the feasible region D_n is defined as

$$D_n = \left\{ \delta \in \mathbb{R}^p: X_t - \sum_{i=1}^p \delta_i X_{t-i} \geq 0, \quad t = 1, \dots, n \right\}. \quad (1.4)$$

Here is an outline of how we will obtain a limit distribution for the estimator $\hat{\phi}^{(n)}$ given in (1.3).

(1) *Change variables:* For a suitable normalizing sequence $q_n \rightarrow \infty$ we seek a limit distribution for

$$q_n(\hat{\phi}^{(n)} - \phi^{(0)}).$$

By a change of variable we can express this as

$$q_n(\hat{\phi}^{(n)} - \phi^{(0)}) = \arg \max_{\delta \in A_n} \delta' \mathbf{1}$$

where

$$A_n = \left\{ \delta \in \mathbb{R}^p: \sum_{i=1}^p \delta_i \frac{X_{t-i}}{q_n Z_t} \leq 1, \quad t = 1, \dots, n; \delta' \mathbf{1} \geq -1 \right\}.$$

To see this, note that since $X_t = \sum_{i=1}^p \phi_i^{(0)} X_{t-i} + Z_t$, we have from the definition of D_n in (1.4) that

$$\begin{aligned} D_n &= \left\{ \delta \in \mathbb{R}^p: \sum_{i=1}^p (\phi_i^{(0)} - \delta_i) X_{t-i} + Z_t \geq 0, \quad t = 1, \dots, n \right\} \\ &= \left\{ \delta \in \mathbb{R}^p: \sum_{i=1}^p (\delta_i - \phi_i^{(0)}) \frac{X_{t-i}}{Z_t} \leq 1, \quad t = 1, \dots, n \right\}. \end{aligned}$$

Now $\hat{\phi}^{(n)}$ satisfies

$$(\hat{\phi}^{(n)})' \mathbf{1} \geq \delta' \mathbf{1}$$

for all δ such that

$$\sum_{i=1}^p (\delta_i - \phi_i^{(0)}) \frac{q_n X_{t-i}}{q_n Z_t} \leq 1, \quad t = 1, \dots, n.$$

Set $\eta = q_n(\delta - \phi^{(0)})$ so that $q_n^{-1}\eta + \phi^{(0)} = \delta$ and thus $\hat{\phi}^{(n)}$ satisfies

$$(\hat{\phi}^{(n)})' \mathbf{1} \geq (q_n^{-1}\eta + \phi^{(0)})' \mathbf{1},$$

or equivalently

$$q_n(\hat{\phi}^{(n)} - \phi^{(0)})' \mathbf{1} \geq \eta' \mathbf{1}$$

for all η such that

$$\sum_{i=1}^p \frac{\eta_i X_{t-i}}{q_n Z_t} \leq 1, \quad t = 1, \dots, n.$$

Finally, note that although the change of variables actually requires us to search for a maximum in the set

$$\left\{ \delta \in \mathbb{R}^p: \sum_{i=1}^p \delta_i \frac{X_{t-i}}{q_n Z_t} \leq 1, t = 1, \dots, n \right\},$$

since $\delta = \mathbf{0}$ is in this set, it does no harm to add the constraint $\delta' \mathbf{1} \geq -1$. We add this constraint since it preserves the polar nature of the feasible region.

(2) *Weak convergence of point processes:* Moment restrictions and regular variation assumptions on the distributions of the Z 's allow us to prove that a sequence of point processes $\{\mu_n\}$, where μ_n has points in \mathbb{R}^p

$$\left\{ \mathbf{u}_t^{(n)} = \left(\frac{X_{t-1}}{q_n Z_t}, \dots, \frac{X_{t-p}}{q_n Z_t} \right); 1 \leq t \leq n \right\},$$

converges weakly to a limit point process μ_∞ with points $\{\mathbf{u}_k\}$. The \mathbf{u}_k depend on the p -dimensional distribution of $\{X_t\}$ and on the regular variation assumptions.

(3) *Weak convergence of feasible regions:* Since the closed random set A_n given in step 1 is a function of the points of μ_n , we are hopeful that there is a closed random set A such that

$$A_n \Rightarrow A,$$

where the weak convergence is with respect to the Hausdorff metric on the compact subsets of \mathbb{R}^p . The limit set A is given by the polyhedral domain

$$A = \{ \delta \in \mathbb{R}^p: \delta' \mathbf{1} \geq -1, \delta' \mathbf{u}_k \leq 1, k \geq 1 \}.$$

(4) *Weak convergence of solutions:* We will show that the optimal solution of the linear program

$$\max_{\delta \in A_n} \delta' \mathbf{1}$$

converges weakly as $n \rightarrow \infty$ to the corresponding solution of the linear program determined by the points of the limit point process, namely

$$\max_{\delta \in A} \delta' \mathbf{1}.$$

Thus the limit distribution for $q_n(\hat{\phi}^{(n)} - \phi^{(0)})$ is the distribution of

$$\arg \max_{\delta \in A} \delta' \mathbf{1}.$$

Although this distribution cannot be calculated explicitly, the above formula gives a recipe that can be simulated. However, we emphasize that the limit distribution will turn out to depend on the distribution of (X_1, \dots, X_p) . We will see that q_n , being the inverse of a regularly varying function, will frequently go to ∞ at a rate which is faster than \sqrt{n} .

In Section 2, we state precisely our results and the conditions under which they hold. Section 3 discusses convergence of feasible regions and solutions and relates these ideas to convergence of point measures. We also consider carefully when our feasible regions are compact in \mathbb{R}^p . Section 4 gives some background results in time series analysis and weak convergence which we need for the proofs of our main results which come in Section 7. Sections 5 and 6 provide some important details regarding conditions for boundedness of A and for the uniqueness of solutions. We give some concluding remarks in Section 8.

2. Conditions and results

We need conditions which specify the model and guarantee stationarity. In order to obtain a limit distribution for our estimators, we impose regular variation and moment conditions on the distribution of the innovation sequence. We recall that a function $U: [0, \infty) \rightarrow (0, \infty)$ is regularly varying at ∞ with exponent $\rho \in \mathbb{R}$ if

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\rho, \quad x > 0.$$

We now state precisely conditions under which our results will hold.

(1) *Condition M (model specification):* The process $\{X_t: t = 0, \pm 1, \pm 2, \dots\}$ satisfies the equations

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + Z_t, \quad t = 0, \pm 1, \pm 2, \dots \quad (2.1)$$

where $\{Z_t\}$ is an independent and identically distributed sequence of random variables with essential infimum (left endpoint) equal to 0 and common distribution function F .

(2) *Condition S (stationarity)*: The coefficients ϕ_1, \dots, ϕ_p satisfy the stationarity condition that the autoregressive polynomial $\Phi(z) \equiv 1 - \sum_1^p \phi_i z^i$ has no roots in the unit disk $\{z: |z| \leq 1\}$. Furthermore we assume $\Phi(1) > 0$; i.e., we require

$$\sum_{i=1}^p \phi_i < 1. \quad (2.2)$$

(3) *Condition L (left tail)*: The distribution F of the innovations Z_t satisfies, for some $\alpha > 0$

$$\lim_{s \downarrow 0} \frac{F(sx)}{F(s)} = x^\alpha \quad \text{for all } x > 0;$$

$$E(Z_t^\beta) = \int_0^\infty u^\beta F(du) < \infty \quad \text{for some } \beta > \alpha.$$

(4) *Condition R (right tail)*: The distribution F of the innovations Z_t satisfies, for some $\alpha > 0$

$$\lim_{s \rightarrow \infty} \frac{1 - F(sx)}{1 - F(s)} = x^{-\alpha} \quad \text{for all } x > 0;$$

$$E(Z_t^{-\beta}) = \int_0^\infty u^{-\beta} F(du) < \infty \quad \text{for some } \beta > \alpha.$$

Our results have as hypotheses M , S , and either L or R . In Feigin and Resnick (1992), we assumed ϕ_1, \dots, ϕ_p non-negative as would be suitable for the majority of modelling applications of data which are inherently non-negative, such as stream flows, interarrival times, teletraffic applications such as video conference scenes. We remarked that the non-negativity of the ϕ 's was sufficient to guarantee that (2.2) holds. However, we have discovered that our results can be proven under the weaker assumption (2.2) and so we drop the assumption that the ϕ 's are non-negative in the hope that the added flexibility will prove useful in fitting models to data.

For additional remarks concerning these conditions see Feigin and Resnick (1992). However, note that condition L is rather mild. It is satisfied if a density f of F exists which is continuous at 0 and with $f(0) > 0$. In this case $\alpha = 1$. Other common cases where condition L holds are the Weibull distributions of the form $F(x) = 1 - \exp\{-x^\alpha\}$ where $F(x) \sim x^\alpha$, as $x \downarrow 0$ and the gamma densities $f(x) = ce^{-x}x^{r-1}$, $r > 0$, $x > 0$ so that $f(x) \sim cx^{r-1}$ as $x \downarrow 0$ and therefore the associated gamma distribution function satisfies $F(x) \sim cr^{-1}x^r$, as $x \downarrow 0$. Examples of distributions satisfying condition R include positive stable densities and the Pareto density.

We now review the form of our estimator given by (1.3). We suppose the true autoregressive parameter is $\phi^{(0)} = (\phi_1^{(0)}, \dots, \phi_p^{(0)})'$. Our estimator is obtained by solving the linear program to get

$$\hat{\phi}^{(n)} = \arg \max_{\delta \in D_n} \delta' \mathbf{1} \quad (1.3)$$

where $\mathbf{1}' = (1, \dots, 1)$ and where the feasible region D_n is defined as

$$D_n = \left\{ \delta \in \mathbb{R}^p: X_t - \sum_{i=1}^p \delta_i X_{t-i} \geq 0, t = 1, \dots, n \right\}. \quad (1.4)$$

We now state our main result, which gives a limit distribution for the linear programming estimator in (1.3). We use “ \Rightarrow ” to denote weak convergence of random elements. For a monotone function U , we denote the left continuous inverse by U^\leftarrow so that

$$U^\leftarrow(y) = \inf\{t: U(t) \geq y\}.$$

Theorem 2.1. *Suppose M and S hold.*

(a) *If condition L holds, define*

$$a_n = F^\leftarrow(1/n).$$

Then

$$a_n^{-1}(\hat{\phi}^{(n)} - \phi^{(0)}) \Rightarrow U$$

where U is non-degenerate and

$$U \stackrel{d}{=} \arg \max_{\delta \in A} \delta' \mathbf{1}$$

where

$$A = \{\delta \in \mathbb{R}^p: \delta' \mathbf{1} \geq -1, \delta' \mathbf{u}_k \leq 1, k \geq 1\}. \quad (2.3)$$

The points $\{\mathbf{u}_k\}$ are specified as follows. Let $\{E_j, j \geq 1\}$ be i.i.d. unit exponential random variables and define

$$\Gamma_k = E_1 + \dots + E_k.$$

Let $\{Y_j, j \geq 1\}$ be i.i.d. \mathbb{R}^p valued random vectors with

$$Y_1 \stackrel{d}{=} (X_p, \dots, X_1)$$

and suppose $\{Y_j, j \geq 1\}$ are independent of $\{\Gamma_k\}$. Then

$$\mathbf{u}_k = \Gamma_k^{-1/2} Y_k.$$

(b) *If condition R holds, define b_n by*

$$b_n = \left(\frac{1}{1 - F} \right)^\leftarrow(n).$$

Define for $|z| \leq 1$

$$C(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{1}{\Phi(z)}.$$

Then

$$b_n(\hat{\phi}^{(n)} - \phi^{(0)}) = O_p(1)$$

so the rate of convergence of $\hat{\phi}^{(n)}$ to $\phi^{(0)}$ is b_n . Furthermore, if for any $p-1$ non-negative and distinct indices $\{l_1, \dots, l_{p-1}\}$, the set of p vectors

$$\{\mathbf{1}, (c_{l_j}, c_{l_j-1}, \dots, c_{l_j-p+1}); 1 \leq j \leq p-1\} \quad (2.4)$$

is linearly independent, then

$$b_n(\hat{\phi}^{(n)} - \phi^{(0)}) \Rightarrow V$$

where V is non-degenerate,

$$V \stackrel{d}{=} \arg \max_{\delta \in A} \delta' \mathbf{1},$$

and

$$A = \{\delta \in \mathbb{R}^p: \delta' \mathbf{1} \geq -1, \delta' \mathbf{v}_k \leq 1, k \geq 1\}. \quad (2.5)$$

The points $\{\mathbf{v}_k\}$ are specified as follows. Let $\{E_j, j \geq 1\}$ be i.i.d. unit exponential random variables and define

$$\Gamma_k = E_1 + \dots + E_k.$$

Let $\{Y_{kl}, k \geq 1, l \geq 0\}$ be a doubly infinite array of i.i.d. random variables with the distribution F . Then

$$\mathbf{v}_l = \left(\bigvee_{k=1}^x \Gamma_k^{-1/x} Y_{kl}^{-1} \right) (c_{l-1}, \dots, c_{l-p})' = V_{l-1}(c_{l-1}, \dots, c_{l-p}), \quad l = 1, 2, \dots$$

If we set q_n equal to either a_n^{-1} or b_n as appropriate, then q_n is regularly varying with index $1/x$. So the rate of convergence will be faster than the \sqrt{n} rate provided that $\alpha < 2$. For example, for the Weibull distribution where $F(x) \sim x^\alpha$, $x \downarrow 0$, we have $a_n \sim n^{-1/\alpha}$ and for the gamma density where $F(x) \sim x^r/\Gamma(r+1)$ we obtain

$$a_n \sim \left(\frac{\Gamma(r+1)}{n} \right)^{1/r}.$$

In the right tailed case, for example for the Pareto distribution function where $1 - F(x) \sim x^{-\alpha}$ as $x \rightarrow \infty$, we have $b_n \sim n^{1/\alpha}$.

Note, if condition R holds with $\alpha < 2$, the variance of Z_1 is infinite.

3. Convergence of feasible regions and solutions

In this section we consider the continuity properties of two mappings: one from point measures to feasible regions; and the second from feasible regions to optimal solutions. These continuity results are required to derive the asymptotic distribution of our estimators from the weak convergence of the relevant random point measures.

We also investigate conditions on the limit process that will ensure the continuity — namely, boundedness of the feasible region in \mathbb{R}^p and uniqueness of the solution for the limit point measure.

For a locally compact, Hausdorff topological space E , we let $M_p(E)$ be the space of Radon point measures on E . This means $m \in M_p(E)$ is of the form

$$m = \sum_{i=1}^{\infty} \varepsilon_{x_i}$$

where $x_i \in E$ are the point masses of m and where

$$\varepsilon_x(A) = \begin{cases} 1 & \text{if } x \in A; \\ 0 & \text{if } x \notin A. \end{cases}$$

We emphasize that we assume that all measures in $M_p(E)$ are Radon which means that for any $m \in M_p(E)$ and any compact $K \subset E$, $m(K) < \infty$. On the space $M_p(E)$ we use the vague metric $d(\cdot, \cdot)$. Its properties are discussed in Resnick (1987, Section 3.4). Note that a sequence of measures $m_n \in M_p(E)$ converge vaguely to $m_0 \in M_p(E)$ if for any continuous function $f: E \rightarrow [0, \infty)$ with compact support we have $m_n(f) \rightarrow m_0(f)$ where

$$m_n(f) = \int_E f d m_n.$$

We will be considering point measures $\mu \in M_p(E)$ where $E = [-\infty, \infty]^p \setminus \{0\}$ and for which the compact sets are of the form U^c where $U \ni 0$ is an open set. Then $\mu(U^c) < \infty$ for any such U . In particular, given μ , we can order its points u_i in decreasing order of magnitude $\|u_1\| \geq \|u_2\| \geq \|u_3\| \geq \dots$.

We now define, for any Radon point measure $\mu \in M_p([-\infty, \infty]^p \setminus \{0\})$ of the form

$$\mu = \sum_i \varepsilon_{u_i}, \quad u_i \in \mathbb{R}^p,$$

the closed set $A(\mu)$ as follows

$$\begin{aligned} A(\mu) &= \{\delta \in \mathbb{R}^p: \delta' \mathbf{1} \geq -1, \delta' u_i \leq 1; i = 1, 2, \dots\} \\ &= \{\delta \in \mathbb{R}^p: \delta' u_i \leq 1; i = 0, 1, 2, \dots\} \end{aligned} \quad (3.1)$$

where we have set $u_0 = -(1, 1, \dots, 1)'$. We also define the set of points of μ together with u_0 , $P(\mu) = \{u_0, u_1, \dots\}$ and $C(\mu) = \text{conv}(P(\mu))$ — the *convex hull* of $P(\mu)$.

We need to define a distance, the Hausdorff distance, between two compact subsets A and B of \mathbb{R}^p

$$\rho(A, B) = \inf\{\eta: \sup_{y \in B} \inf_{x \in A} \|x - y\| \leq \eta; \sup_{x \in A} \inf_{y \in B} \|x - y\| \leq \eta\};$$

or equivalently

$$\rho(A, B) = \inf\{\eta: A \subset B^\eta; B \subset A^\eta\}$$

where A^η denotes the η -swelling of A

$$A^\eta = \{y \in \mathbb{R}^p: \inf_{x \in A} \|x - y\| < \eta\}.$$

In the above $\|\cdot\|$ denotes the usual Euclidean metric on \mathbb{R}^p .

Suppose that $\{\mu_n\}$ is a sequence of Radon point measures in $M_p(E)$ that converges to μ in the vague metric: $d(\mu_n, \mu) \rightarrow 0$ as $n \rightarrow \infty$. Our goal is to show that if $A(\mu)$ is compact then $\rho(A(\mu_n), A(\mu)) \rightarrow 0$ also.

We proceed through a series of lemmas which use some concepts from convex analysis. In particular we define A^* as the *polar* set of $A \subset \mathbb{R}^p$

$$A^* = \{x \in \mathbb{R}^p: x'a \leq 1, \text{ for all } a \in A\}.$$

We thus note that $A(\mu) = (P(\mu))^* = (C(\mu))^*$. We also write $\text{int}(A)$ for the interior of a set A .

Lemma 3.1. Suppose $\mu \in M_p([-\infty, \infty]^p \setminus \{\mathbf{0}\})$. $A(\mu)$ is bounded in \mathbb{R}^p if $\mathbf{0} \in \text{int}(C(\mu))$.

Proof. Since $\mathbf{0} \in \text{int}(C(\mu))$ and $C(\mu)$ is a bounded convex set in \mathbb{R}^p , the result is immediate from Theorem 23.4 of Lay (1982). \square

Lemma 3.2. Suppose $\mu \in M_p([-\infty, \infty]^p \setminus \{\mathbf{0}\})$. If $A(\mu)$ is bounded in \mathbb{R}^p then there exists a finite K such that

$$A(\mu) = P_K^* = C_K^*$$

where $P_K = \{\mathbf{u}_i; i = 0, 1, \dots, K\}$ and $C_K = \text{conv}(P_K)$.

Proof. By the boundedness of $A(\mu)$ in \mathbb{R}^p , there exists $M > 1$ such that

$$A(\mu) \subset S_p(M) \equiv \{\delta \in \mathbb{R}^p: \|\delta\| \leq M\}.$$

Let

$$A^{(M)} = \left\{ \delta \in \mathbb{R}^p: \delta' \mathbf{u}_k \leq 1, \text{ for all } k \geq 0 \text{ such that } \|\mathbf{u}_k\| > \frac{1}{2M} \right\}$$

and it is clear that $A(\mu) \subset A^{(M)}$. If there exists

$$\delta \in A^{(M)} \setminus A(\mu),$$

then for all \mathbf{u}_k such that $\|\mathbf{u}_k\| > (2M)^{-1}$, we have $\delta' \mathbf{u}_k \leq 1$ and also there must exist \mathbf{u}_s such that $\|\mathbf{u}_s\| \leq (2M)^{-1}$ and $\delta' \mathbf{u}_s > 1$. Thus, from this last property

$$1 < \delta' \mathbf{u}_s \leq \|\delta\| \cdot \|\mathbf{u}_s\| \leq (2M)^{-1} \|\delta\|$$

and therefore

$$\|\delta\| > 2M.$$

Define

$$\tilde{\delta} = \frac{2M}{\|\delta\|} \delta.$$

Then $\tilde{\delta} \in A^{(M)}$. Indeed for any $\|\mathbf{u}_k\| > (2M)^{-1}$, we have $\delta' \mathbf{u}_k \leq 1$, since $\delta \in A^{(M)}$, and therefore

$$\tilde{\delta}' \mathbf{u}_k = \frac{2M}{\|\delta\|} \delta' \mathbf{u}_k \leq \frac{2M}{\|\delta\|} < 1.$$

We now show that $\tilde{\delta} \in A(\mu)$. Indeed, since $\delta \in A^{(M)}$, we only need show that for \mathbf{u}_s satisfying $\|\mathbf{u}_s\| < (2M)^{-1}$, $\tilde{\delta}' \mathbf{u}_s \leq 1$. This fact follows simply

$$\begin{aligned} \tilde{\delta}' \mathbf{u}_s &= \frac{2M}{\|\delta\|} \delta' \mathbf{u}_s \leq \frac{2M}{\|\delta\|} |\delta' \mathbf{u}_s| \\ &\leq \frac{2M}{\|\delta\|} \|\delta\| \cdot \|\mathbf{u}_s\| \\ &= \frac{2M}{\|\delta\|} \|\delta\| \frac{1}{2M} = 1. \end{aligned}$$

We thus conclude that

$$\tilde{\delta} \in A(\mu).$$

However $\|\tilde{\delta}\| = 2M$ so $\tilde{\delta} \notin A(\mu) \subset S_p(M)$ which yields a contradiction.

Now $K = \max\{k: \|\mathbf{u}_k\| > (2M)^{-1}\}$ is finite by the compactness of $\{\mathbf{x} \in \mathbb{R}^p: \|\mathbf{x}\| > (2M)^{-1}\}$ in $[-\infty, \infty]^p \setminus \{\mathbf{0}\}$, and satisfies the claim of the lemma. \square

For the set C_K derived above we also have

Lemma 3.3. *If $\mu \in M_p([-\infty, \infty]^p \setminus \{\mathbf{0}\})$ and $A(\mu)$ is bounded then $\mathbf{0} \in C_K$ and $(C_K^*)^* = C_K$.*

Proof. Suppose that $\mathbf{0} \notin C_K$. Then by the separating hyperplane theorem we have that there exists a vector $\mathbf{v} \in \mathbb{R}^p$ such that $\mathbf{x}' \mathbf{v} > 1$ for all $\mathbf{x} \in C_K$. This implies that $(-c)\mathbf{v} \in A(\mu)$ for any $c > 0$ which contradicts that $A(\mu)$ is bounded. Thus $\mathbf{0} \in C_K$. Since C_K is the convex hull of a finite number of points it is closed. For a closed convex set which contains the origin, we have $(C_K^*)^* = C_K$ from Theorem 23.5 of Lay (1982). \square

Proposition 3.4. *If $\mu \in M_p([-\infty, \infty]^p \setminus \{\mathbf{0}\})$, the following are equivalent*

- (a) $\Lambda(\mu)$ is bounded in \mathbb{R}^p ;
- (b) there exists an integer $K < \infty$ such that $0 \in \text{int}(C_K)$;
- (c) for all $\mathbf{0} \neq a \in \mathbb{R}^p$, $a' \mathbf{1} \geq 0 \Rightarrow a' \mathbf{u}_i > 0$ for some $i = 1, 2, 3, \dots$

Proof. (a) \Rightarrow (b): It is clear that $0 \in \text{int}(\Lambda(\mu))$, since P_K is a bounded set in \mathbb{R}^p . From Theorem 23.4 of Lay (1982) we have that $(\Lambda(\mu))^*$ is bounded and contains 0 in its interior, since so does $\Lambda(\mu)$. However, from Lemma 3.3 we have that $(\Lambda(\mu))^* = (C_K^*)^* = C_K$.

(b) \Rightarrow (a): This follows from Lemma 3.1 since we get $\Lambda(\sum_{k=0}^K \varepsilon_{\mathbf{u}_k})$ bounded and therefore so is

$$\Lambda\left(\sum_{k=0}^{\infty} \varepsilon_{\mathbf{u}_k}\right) = \Lambda(\mu).$$

(a) \Leftrightarrow (c): Follows from the following equivalences

- there exists $a \in \mathbb{R}^p$ such that $a' \mathbf{u}_i \leq 0$ for all $i \geq 0$
- \Leftrightarrow there exists $a \in \mathbb{R}^p$ such that $(ca)' \mathbf{u}_i \leq 0$ for all $i \geq 0$ and for all $c > 0$
- \Leftrightarrow there exists $a \in \mathbb{R}^p$ such that $(ca) \in \Lambda(\mu)$ for all $c > 0$
- $\Leftrightarrow \Lambda(\mu)$ unbounded.

From the definition that $\mathbf{u}_0 = -\mathbf{1}$ the result follows. \square

Theorem 3.5. *If a sequence of Radon point measures μ_n converges to μ vaguely in $M_p([-\infty, \infty]^p \setminus \{\mathbf{0}\})$, and $\Lambda(\mu)$ is bounded in \mathbb{R}^p , then $\Lambda(\mu_n)$ converges to $\Lambda(\mu)$ in the Hausdorff metric.*

Proof. Take $v = \mu_n$ for some n sufficiently large such that $d(v, \mu) < \varepsilon$. We will show that under the conditions of the theorem $\rho(\Lambda(v), \Lambda(\mu)) < \eta(\varepsilon)$, where $\eta(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$. Let

$$v = \sum_{i=1}^{\infty} \varepsilon_{\mathbf{v}_i};$$

and for convenience in defining $\Lambda(v)$, we define $\mathbf{v}_0 = -\mathbf{1}$.

Since $\Lambda(\mu)$ is bounded we can take $1 < M < \infty$ such that $\Lambda(\mu) \subset S_p(M)$.

If $\Lambda(\mu)$ is bounded then from Proposition 3.4 there exists finite $K = \max\{k; \|\mathbf{u}_k\| > (2M)^{-1}\} > 0$ such that $\text{conv}(\{\mathbf{u}_k, 0 \leq k \leq K\})$ is bounded with $\mathbf{0}$ in its interior. If $d(\mu, v)$ is sufficiently small, then also $\text{conv}(\{\mathbf{v}_k, 0 \leq k \leq K\})$ is bounded with $\mathbf{0}$ in its interior; cf. Davis, Mulrow, Resnick, 1988. Thus $\Lambda(v)$ is also bounded. (We may have to rearrange the ordering of the points \mathbf{v}_i of v .)

We may suppose for the rest of the proof that

$$\max_{\delta \in \Lambda(\mu) \cup \Lambda(v)} \|\delta\| \leq M$$

and that $d(\mu, \nu)$ is so small that

$$\mu \left\{ \mathbf{x} \in \mathbb{R}^p: \|\mathbf{x}\| \geq \frac{1}{2M} \right\} = \nu \left\{ \mathbf{x} \in \mathbb{R}^p: \|\mathbf{x}\| \geq \frac{1}{2M} \right\}$$

and that if $\|\mathbf{u}_k\| > (2M)^{-1}$ there is $\|\mathbf{v}_k\| > (2M)^{-1}$ and $\tau = \tau(\varepsilon)$ such that

$$\|\mathbf{v}_k - \mathbf{u}_k\| < \tau(\varepsilon)$$

where $\tau(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$. Then if

$$\delta \in \Lambda(\mu) = \left\{ \delta \in \mathbb{R}^p: \delta' \mathbf{u}_k \leq 1, \text{ for all } k \geq 0 \text{ such that } \|\mathbf{u}_k\| > \frac{1}{2M} \right\}$$

and $\|\mathbf{u}_k\| \wedge \|\mathbf{v}_k\| > (2M)^{-1}$ we have

$$\begin{aligned} \delta' \mathbf{v}_k &= \delta' \mathbf{u}_k + \delta'(\mathbf{v}_k - \mathbf{u}_k) \\ &\leq 1 + \|\delta\| \|\mathbf{v}_k - \mathbf{u}_k\| \\ &\leq 1 + M\tau, \end{aligned}$$

whence

$$(1 + \tau M)^{-1} \delta' \mathbf{v}_k \leq 1.$$

Thus we conclude that if $\delta \in \Lambda(\mu)$ then also

$$(1 + \tau M)^{-1} \delta \in \Lambda(\nu).$$

However

$$\|\delta - (1 + \tau M)^{-1} \delta\| = \frac{M\tau}{1 + M\tau} \|\delta\| \leq M^2\tau.$$

Therefore we get the set inclusion

$$\Lambda(\mu) \subset \Lambda(\nu)^{M^2\tau(\varepsilon)}$$

where $\Lambda(\nu)^\eta$ is the η -swelling of $\Lambda(\nu)$; that is, the set of points at distance less than η from $\Lambda(\nu)$. An appeal to symmetry finishes the proof. \square

We now turn to conditions for the uniqueness of the solution to

$$\max\{\delta' \mathbf{1}: \delta' \mathbf{u}_k \leq 1; k \geq 1\} = \max\{\delta' \mathbf{1}: \delta \in \Lambda(\mu)\}. \quad (3.2)$$

Suppose $P_L = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L\}$ is the finite set of extreme points of C_K — they actually determine $\Lambda(\mu)$.

Lemma 3.6. *Suppose $\Lambda(\mu)$ is bounded in \mathbb{R}^p . The solution to (3.2) is unique if for all $\mathbf{a} \in \mathbb{R}^p$ such that $\mathbf{a}' \mathbf{1} = 0$, and for all subsets $\{\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{p-1}}\}$ of $p - 1$ points from P_L , $\mathbf{a}' \mathbf{x}_{i_j} \neq 0$ for some $1 \leq j \leq p - 1$.*

Proof. The condition ensures that there is no edge of the polyhedron P_L^* which lies in a plane $\{\delta: \delta' \mathbf{1} = \text{constant}\}$. If there were such an edge, then any points on it would be solutions of (3.2). \square

For the limiting random point measure, this last condition is rather hard to check. We will check the sufficient condition.

Corollary 3.7. *If for all collections of $(p-1)$ points $\{\mathbf{u}_{k_1}, \dots, \mathbf{u}_{k_{p-1}}\}$ of μ the determinant $\det[\mathbf{u}_{k_1} \cdots \mathbf{u}_{k_{p-1}} \mathbf{1}] \neq 0$, then (3.2) has a unique solution.*

Finally we turn to the question of continuity of the mappings

$$A \rightarrow \arg \max \{\delta' \mathbf{1} : \delta \in A\}$$

with respect to the Hausdorff metric on compact sets A and the ordinary Euclidean distance on $\delta \in \mathbb{R}^p$.

Theorem 3.8. *Suppose $A_n \rightarrow A$ in the Hausdorff metric and A is bounded in \mathbb{R}^p and has a unique solution \mathbf{w} :*

$$\mathbf{w} = \arg \max \{\delta' \mathbf{1} : \delta \in A\}.$$

Then for any sequence of solutions

$$\mathbf{w}_n = \arg \max \{\delta' \mathbf{1} : \delta \in A_n\},$$

we have $\mathbf{w}_n \rightarrow \mathbf{w}$.

Proof. Since $A_n \rightarrow A$ and A is bounded, there exist M and N such that

$$A \cup A_n \subset S_p(M) \quad \text{for all } n > N.$$

($S_p(M)$ is a p -dimensional sphere of radius M .) For each $n > N$, take any optimal solution \mathbf{w}_n of $\max \{\delta' \mathbf{1} : \delta \in A_n\}$ and since the sequence $\{\mathbf{w}_n, n > N\} \subset S_p(M)$ we have a convergent subsequence $\{n'\}$ for which $\mathbf{w}_{n'} \rightarrow \mathbf{w}_\infty$. Since $A_n \rightarrow A$, we have from the definition of the Hausdorff metric that $\mathbf{w}_\infty \in A$.

Since

$$\mathbf{w} \in A \subset A_n^{\rho(A, A_n)}$$

(where ρ is the Hausdorff metric), there exists $\delta_n \in A_n$ such that

$$\|\mathbf{w} - \delta_n\| \leq \rho(A, A_n).$$

So we have for any n ,

$$\begin{aligned} \mathbf{w}' \mathbf{1} &= \max_{A_n} \delta' \mathbf{1} \leq \delta_n' \mathbf{1} + \sqrt{p} \rho(A_n, A) \\ &\leq \max_{A_n} \delta' \mathbf{1} + \sqrt{p} \rho(A_n, A) \\ &= \mathbf{w}_n' \mathbf{1} + \sqrt{p} \rho(A_n, A). \end{aligned}$$

Therefore, for any n' ,

$$\begin{aligned} w'_\infty \mathbf{1} &\geq w'_{n'} \mathbf{1} - \sqrt{p} \|w_{n'} - w_\infty\| \\ &\geq w' \mathbf{1} - \sqrt{p} \rho(A_{n'}, A) - \sqrt{p} \|w_{n'} - w_\infty\|, \end{aligned}$$

and letting $n' \rightarrow \infty$ yields

$$w'_\infty \mathbf{1} \geq w' \mathbf{1}.$$

Since $\max_A \delta' \mathbf{1}$ is uniquely achieved by w we get $w_\infty = w$. So every subsequential limit of $\{w_n\}$ equals w and thus $w_n \rightarrow w$ as required. \square

4. Background results and weak convergence

In this section we review some background needed to smooth the way for the probabilistic derivations which follow in subsequent sections.

MA(∞) representations: Suppose $\{Z_n\}$ are i.i.d. non-negative random variables with common distribution F . Assuming that $\sum_{j=0}^{\infty} c_j Z_j$ converges in some suitable sense, consider an infinite order one sided moving average process (MA(∞)) of the form

$$X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}, \quad -\infty < t < \infty.$$

For us, the most useful form of convergence of infinite series is almost sure convergence. Note, that the $\{Z_j\}$ sequence, being i.i.d., is also stationary and ergodic. Consequently, by Breiman, 1968, Proposition 6.31, we have that the MA(∞) process $\{X_t\}$ is stationary and ergodic.

The particular case of interest is where $\{X_t\}$ is an autoregressive process of order p . Under conditions M, S and either L or R, we can show readily that $\sum_{j=0}^{\infty} c_j Z_j$ converges almost surely and if we set

$$C(z) = \sum_{j=0}^{\infty} c_j z^j, \quad |z| \leq 1,$$

then

$$C(z) = \frac{1}{\Phi(z)}$$

where

$$\Phi(z) = 1 - \sum_{j=1}^p \phi_j z^j$$

is the autoregressive polynomial. As is well-known, the quantities c_j decrease in absolute value to zero geometrically fast. (See Brockwell and Davis, 1991, Section 3.3.)

Convergence of the series, then follows readily by, say, the technique of Yohai and Maronna, 1977. Since

$$C(z)\Phi(z) = 1$$

we obtain

$$c_0 = 1 \quad c_l = \sum_{i=1}^p \phi_i c_{l-i}, \quad l = 1, 2, \dots \quad (4.1)$$

In vector notation, we write

$$\sigma_l = (c_l, c_{l-1}, \dots, c_{l-p+1})'$$

($c_j = 0$, for $j < 0$), and then

$$\sigma_l = \Phi \sigma_{l-1} = \Phi^l \sigma_0 = \Phi^l e_1$$

where

$$e_1 = (1, 0, \dots, 0)'$$

and

$$\Phi = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \dots & \phi_p \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

Note that Φ is non-singular provided $\phi_p \neq 0$, which we always assume to guarantee the order of the autoregression is p .

Right tail behavior of the distribution of X_t : We will need the following fact about the tail behaviour of X_t under conditions M, S and R. For $x > 0$ we write

$$P[|X_t| > x] \leq P\left[\sum_{j=0}^{\infty} |c_j| Z_j > x\right]$$

and from a result of Cline (1983) (see also the account in Resnick, 1987, page 227)

$$\lim_{x \rightarrow \infty} \frac{P\left[\sum_{j=0}^{\infty} |c_j| Z_j > x\right]}{P[Z_1 > x]} = \sum_{j=0}^{\infty} |c_j|^{\alpha}. \quad (4.2)$$

Also, it is not hard to show

$$\lim_{x \rightarrow \infty} \frac{P[X_t > x]}{P[Z_1 > x]} = \sum_{j: c_j > 0} c_j^{\alpha}.$$

This comes from decomposing

$$X_t = \sum_{j:c_j > 0} c_j Z_{t-j} + \sum_{j:c_j \leq 0} c_j Z_{t-j} =: X_t^+ + X_t^-,$$

as two independent terms and using dominated convergence in

$$P[X_t > x] = \int_0^\infty P[X_t^+ > x + y] P[X_t^- \in -dy],$$

together with

$$\lim_{x \rightarrow \infty} \frac{P[X_t^+ > x + y]}{P[Z_t > x]} = \lim_{x \rightarrow \infty} \frac{P[X_t^+ > x + y]}{P[Z_t > x + y]} \frac{P[Z_t > x + y]}{P[Z_t > x]} = \sum_{j:c_j > 0} c_j^\alpha \cdot 1$$

using (4.2) and condition R.

Left tail behavior of X_t under condition L, is more complex and does not follow the pattern of this result. Davis and Resnick (1991) show that if condition L is satisfied, finite positive linear combinations of Z 's satisfy condition L. However, with some regularity assumptions imposed, an infinite linear combination will have different tail behavior.

Ultimate positivity of $\sum_{i=1}^n X_i$: The motivation for the form of our estimator in (1.3) rested on

$$\sum_{t=1}^n X_t \approx \sum_{t=0}^{n-1} X_t$$

and the fact that $\sum_{t=1}^n X_t$ was non-negative for all sufficiently large n . Both requirements are met if $\sum_{t=1}^n X_t \rightarrow \infty$. This is discussed in the next lemma.

Lemma 4.1. *Under the conditions M, S and either R or L,*

$$\sum_{t=1}^n X_t \xrightarrow{P} \infty.$$

(If $\phi_i \geq 0$ for all $i = 1, \dots, p$ then the convergence is almost sure.)

Proof. Using the MA(∞) representation we write

$$\begin{aligned} \sum_{t=1}^n X_t &= \sum_{t=1}^n \sum_{j=0}^{\infty} c_j Z_{t-j} \\ &= \sum_{t=1}^n \sum_{i=-\infty}^t c_{t-i} Z_i \\ &= \sum_{i=-\infty}^n \left(\sum_{t=i \vee 1}^n c_{t-i} \right) Z_i \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=-\infty}^0 \left(\sum_{t=1}^n c_{t-i} \right) Z_i + \sum_{i=1}^n \left(\sum_{t=i}^n c_{t-i} \right) Z_i \\
&= \sum_{i=-\infty}^0 \left(\sum_{v=1-i}^{n-i} c_v \right) Z_i + \sum_{i=1}^n \left(\sum_{v=0}^{n-i} c_v \right) Z_i \\
&=: A_n + B_n.
\end{aligned}$$

Now

$$C(1) = \frac{1}{\Phi(1)} = \frac{1}{1 - \sum_{i=1}^p \phi_i},$$

so we may pick $\varepsilon > 0$ and K such that for $k \geq K$ we have

$$\sum_{v=0}^k c_v \geq \varepsilon.$$

Then for large n

$$B_n = \sum_{i=1}^{n-K} \left(\sum_{v=0}^{n-i} c_v \right) Z_i + \sum_{i=n-K+1}^n \left(\sum_{v=0}^{n-i} c_v \right) Z_i =: B_n(1) + B_n(2).$$

Now we have

$$B_n(1) \geq \varepsilon \sum_{i=1}^{n-K} Z_i \rightarrow \infty$$

almost surely. Also,

$$|B_n(2)| \leq \sum_{i=n-K+1}^n \left(\sum_{v=0}^{K-1} |c_v| \right) Z_i,$$

which is stochastically bounded and hence

$$\frac{B_n(2)}{B_n(1)} \xrightarrow{P} 0.$$

For A_n we have

$$|A_n| \leq \sum_{i=0}^{\infty} \sum_{v=1+i}^{n+i} |c_v| Z_{-i}.$$

For an autoregressive process, we have for $\xi > 1$ and some non-negative integer r

$$|c_v| = O(v^r \xi^{-v})$$

(Brockwell and Davis, 1991, Section 3.3). Therefore

$$|A_n| \leq \sum_{i=0}^{\infty} \sum_{v=1+i}^{\infty} |c_v| Z_{-i} < \infty \quad \text{a.s.}$$

and the desired result follows. \square

Continuity: We will use several times the following result about the continuity of linear combinations of (X_t, \dots, X_{t-p+1}) .

Lemma 4.2. *Suppose conditions M and S hold. Set $X_t = (X_t, \dots, X_{t-p+1})'$. For any $0 \neq a \in \mathbb{R}^p$, $a'X_t$ has a continuous distribution.*

Proof. Direct calculation verifies the expansion

$$a'X_t \stackrel{d}{=} \sum_{l=0}^{\infty} (a'\sigma_l)Z_l, \quad (4.3)$$

so that if we can show that $a'\sigma_l \neq 0$ for infinitely many indices l , the result will follow from Davis and Rosenblatt (1991). To get a contradiction, suppose that $a'\sigma_l = 0$ for all $l \geq L$. Recall $\sigma_l = \Phi^l e_1$ and Φ is non-singular. The matrix with columns $e_1, \Phi e_1, \dots, \Phi^{p-1} e_1$ is upper triangular with 1's on the main diagonal and is therefore invertible. It follows that $e_1, \Phi e_1, \dots, \Phi^{p-1} e_1$ is a basis of \mathbb{R}^p . Since Φ is non-singular, we conclude that $\{\Phi^L e_1, \Phi^{L+1} e_1, \dots, \Phi^{L+p-1} e_1\}$ is also a basis of \mathbb{R}^p . However, it is therefore not possible that for $0 \neq a \in \mathbb{R}^p$, $a'\Phi^l e_1 = 0$ for all $l = L, L+1, \dots, L+p-1$. Hence $a'\sigma_l \neq 0$ infinitely often and we have the desired result that $a'X_t$ has a continuous distribution for $a \neq 0$. \square

Point processes and weak convergence: Our limit theory is based on weak convergence of point processes (Resnick, 1987) rather than central limit theory as is typical of the usual L_2 treatment of Yule–Walker estimators. We now discuss the necessary limit theory, some of which is based on Feigin and Resnick (1992), which built on the work of Davis and McCormick (1989). This limit theory will underlie our efforts to obtain an asymptotic distribution for our estimator of the autoregressive coefficients.

For a nice topological space E , we consider weak convergence of random elements of $M_p(E)$, the space of Radon point measures on E . A Poisson process on E with mean measure μ will be denoted $\text{PRM}(\mu)$. Particular examples of the space E that interest us are $E = [0, \infty)$, where compact sets are those closed sets bounded away from ∞ and $E = [-\infty, \infty] \setminus \{0\}$, where compact sets are closed subsets of $[-\infty, \infty]^p$ which are bounded away from 0 .

Here are the results we will need. We state them as a theorem for ready reference.

Theorem 4.3. (a) *Suppose conditions M , S and L hold. Set*

$$a_n = F^{*n}(1/n).$$

Then in $M_p([0, \infty))$ we have

$$\sum_{t=1}^n \varepsilon_{Z_t/a_n} \Rightarrow \sum_k \varepsilon_{j_k}$$

where the point process on the right hand side is Poisson with mean measure $\alpha x^{\alpha-1} dx$. Furthermore in $M_p([-\infty, \infty]^p \setminus \{\mathbf{0}\})$

$$\sum_{t=1}^n \mathcal{E}_{a_n(X_{t-1}/Z_t, \dots, X_{t-p}/Z_t)} \Rightarrow \sum_k \mathcal{E}_{(Y_{kt}/j_k, \dots, Y_{kp}/j_k)} \quad (4.4)$$

where

$$(Y_{k1}, \dots, Y_{kp}) \stackrel{d}{=} (X_p, \dots, X_1)$$

and the vectors $\mathbf{Y}_k = (Y_{k1}, \dots, Y_{kp})$ are i.i.d. and independent of $\{j_k\}$.

(b) Suppose conditions M , S and R hold. Set

$$b_n = \left(\frac{1}{1-F} \right)^{\leftarrow} (n).$$

Then in $M_p((0, \infty])$ we have

$$\sum_{t=1}^n \mathcal{E}_{Z_t/b_n} \Rightarrow \sum_k \mathcal{E}_{j_k}$$

where the point process on the right hand side is Poisson with mean measure $\alpha x^{\alpha-1} dx$. Furthermore in $M_p([-\infty, \infty]^p \setminus \{\mathbf{0}\})$

$$\sum_{t=1}^n \mathcal{E}_{(b_n^{-1} Z_{t-1}, \dots, X_{t-p})} \Rightarrow \sum_k \sum_{l=1}^{\infty} \mathcal{E}_{j_k(c_{l-1} Y_{kt}, \dots, c_{l-p} Y_{kt})} = \sum_k \sum_{l=1}^{\infty} \mathcal{E}_{j_k Y_{kl}^{-1} \sigma_{l-1}} \quad (4.5)$$

where for $j < 0$ we set $c_j = 0$ and for $j \geq 0$ the c_j are the coefficients in the infinite moving average expansion $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$. $\{Y_{kl}\}$ is a doubly infinite array of i.i.d. random variables with distribution equal to that of Z_1 and independent of $\{j_k\}$.

Proof. (a) The proof is almost exactly the same as the derivation of the convergence result (3.22) in Feigin and Resnick (1992). Examining this derivation shows that the assumption $\phi_i \geq 0$, $i = 1, \dots, p$ is unnecessary and that the method of proof applies when we only assume $\sum_{i=1}^p \phi_i < 1$.

(b) As in the derivation of (3.54) of Feigin and Resnick (1992) we get

$$\sum_{t=1}^n \mathcal{E}_{(b_n^{-1}(X_{t-1}, \dots, X_{t-p}), Z_t)} \Rightarrow \sum_k \sum_{l=1}^{\infty} \mathcal{E}_{j_k(c_{l-1}, \dots, c_{l-p}), Y_{kl}} \quad (4.6)$$

in $M_p([-\infty, \infty]^p \setminus \{\mathbf{0}\}) \times [0, \infty)$. Set

$$\xi_n = b_n^{-1}(X_{t-1}, \dots, X_{t-p})'$$

and recall the notation

$$\sigma_l = (c_l, c_{l-1}, \dots, c_{l-p+1})'.$$

We need to show

$$\mu_n := \sum_{t=1}^n \mathcal{E}_{Z_t^{-1} \xi_n} \Rightarrow \mu_\infty := \sum_k \sum_{l=1}^{\infty} \mathcal{E}_{Y_{kl}^{-1} j_k \sigma_{l-1}}. \quad (4.7)$$

For a vector \mathbf{x} write

$$|\mathbf{x}| = \bigvee_{i=1}^p |x_i|$$

and define the compact region in $([-\infty, \infty]^p \setminus \{\mathbf{0}\}) \times [0, \infty)$

$$S_M = \{\mathbf{x} \in [-\infty, \infty]^p \setminus \{\mathbf{0}\} : |\mathbf{x}| \geq M^{-1}\} \times [0, M].$$

Applying the map

$$(\mathbf{x}, z) \rightarrow z^{-1} \mathbf{x}$$

from $([-\infty, \infty]^p \setminus \{\mathbf{0}\}) \times [0, \infty) \rightarrow [-\infty, \infty]^p \setminus \{\mathbf{0}\}$ and Proposition 3.18 of Resnick (1987), we obtain

$$\mu_n^{(M)} := \sum_{t=1}^n \varepsilon_{Z_t^{-1} \xi_{nt}} I_{[(\xi_{nt}, Z_t) \in S_M]} \Rightarrow \mu_\infty^{(M)} := \sum_k \sum_l \varepsilon_{Y_{kl}^{-1} j_k \sigma_{l-1}} I_{[(j_k \sigma_{l-1}, Y_{kl}) \in S_M]}.$$

Since as $M \rightarrow \infty$ we have

$$\sum_k \sum_l \varepsilon_{Y_{kl}^{-1} j_k \sigma_{l-1}} I_{[(j_k \sigma_{l-1}, Y_{kl}) \in S_M]} \Rightarrow \sum_k \sum_l \varepsilon_{Y_{kl}^{-1} j_k \sigma_{l-1}} =: \mu_\infty$$

the desired result will follow from Billingsley, 1968, Theorem 4.2 if we show

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} P[|\mu_n^{(M)}(f) - \mu_n(f)| > 2\eta] = 0$$

for f bounded, continuous on $(-\infty, \infty)^p \setminus \{\mathbf{0}\}$.

Suppose the support of f is contained in $\{\mathbf{x} : |\mathbf{x}| \geq c\}$. Then

$$\begin{aligned} P[|\mu_n^{(M)}(f) - \mu_n(f)| > 2\eta] &\leq P\left[\sum_{t=1}^n f(Z_t^{-1} \xi_{nt}) I_{[|\xi_{nt}| \leq M^{-1}]} > \eta\right] \\ &\quad + P\left[\sum_{t=1}^n f(Z_t^{-1} \xi_{nt}) I_{[|Z_{it}| \geq M]} > \eta\right] \\ &=: A + B \end{aligned}$$

Now for B we have

$$\begin{aligned} B &\leq P\left[\bigcup_{t=1}^n [Z_t^{-1} \xi_{nt} > c, Z_t > M]\right] \\ &\leq nP[|\xi_{nt}| > cM] \\ &= nP\left[\bigvee_{i=1}^p b_n^{-1} |X_{t-i}| > cM\right] \\ &\leq npP[b_n^{-1} |X_t| > cM] \end{aligned}$$

and (4.2) provides a bound on the tail behavior of the series, as $n \rightarrow \infty$. Therefore,

$$\limsup_{n \rightarrow \infty} B \leq \text{const } M^{-\alpha} \rightarrow 0$$

as $M \rightarrow 0$.

For A we have

$$\begin{aligned} P \left[\sum_{t=1}^n f(Z_t^{-1} \xi_{nt}) I_{[|\xi_{nt}| \leq M^{-1}]} > \eta \right] &\leq P \left\{ \bigcup_{t=1}^n [|\xi_{nt}/Z_t| > c, |\xi_{nt}| \leq M^{-1}] \right\} \\ &\leq nP[|\xi_{n1}/Z_1| > c, |\xi_{n1}| \leq M^{-1}] \\ &\leq nP[Z_1^{-1}|\xi_{n1}| > c, Z_1^{-1} > Mc] \\ &\leq nP \left[Z_1^{-1} \bigvee_{i=1}^n b_n^{-1} |X_{1-i}| > c, Z_1^{-1} > Mc \right] \\ &\leq npP[Z_1^{-1} b_n^{-1} |X_0| > c, Z_1^{-1} > Mc] \\ &\leq npP[Z_1^{-1} I_{[Z_1^{-1} > Mc]} b_n^{-1} |X_0| > c] \\ &\leq npP \left[Z_1^{-1} I_{[Z_1^{-1} > Mc]} b_n^{-1} \sum_{j=0}^{\infty} |c_j| Z_{-j} > c \right] \\ &\sim \text{const } E(Z_1^{-\alpha} I_{[Z_1^{-1} > Mc]}), \end{aligned}$$

as $n \rightarrow \infty$ by (4.2) and a result of Breiman (1965). From the fact that $EZ_1^{-\beta} < \infty$ for $\beta > \alpha$, we get the forgoing bound converging to 0 as $M \rightarrow \infty$. \square

Corollary 4.4. *Suppose conditions M, S and R hold and define*

$$\mu = \sum_k \sum_{l=1}^{\infty} e_{j_k} Y_{kl}^{-1} \sigma_{l-1}$$

and

$$\mu' = \sum_{l=1}^{\infty} e_{V_l} \sigma_{l-1}$$

where

$$V_l = \bigvee_{k=1}^{\infty} j_k Y_{kl}^{-1}.$$

Then V_l is finite for each l and

$$A(\mu) = A(\mu').$$

Proof. We first check V_l is finite. Suppose $\{Y_k\}$ is i.i.d., independent of $\{j_k\}$ and having common distribution F , the distribution of Z_1 . Then

$$\sum_k e_{(j_k, Y_k)}$$

is Poisson with mean measure of $dx dy$ equal to $\alpha x^{-\alpha-1} dx F(dy)$ (Proposition 3.8, Resnick, 1987). Therefore, for any $M > 0$

$$\begin{aligned} E \sum_k \varepsilon_{j_k/Y_k}((M, \infty]) &= \int_0^\infty (My)^{-\alpha} F(dy) \\ &= M^{-\alpha} \int_0^\infty y^{-\alpha} F(dy) \end{aligned}$$

which is finite from the moment condition in condition R.

To verify the characterization of $\Lambda(\mu)$ note that for each l , the direction σ_{l-1} is fixed and is, in fact, non-random. Therefore, we see that

$$\delta'(V_l \sigma_{l-1}) \leq 1$$

iff

$$\delta'(j_k Y_{kl}^{-1} \sigma_{l-1}) \leq 1 \quad \text{for all } k$$

and the result follows. \square

5. Compactness of $\Lambda(\mu)$

Here we prove two theorems which guarantee that for the limit PRM

$$\mu = \sum_{k=1}^{\infty} \varepsilon_{u_k}, \quad (5.1)$$

the set $\Lambda(\mu)$ is almost surely bounded in \mathbb{R}^p . We deal separately with the left and right tail cases.

In the left tail case, that is under conditions M, S and L, we have that the points $u_k \in \mathbb{R}^p$ of the limit PRM have the form

$$u_k = j_k^{-1} (Y_{k1}, \dots, Y_{kp})' = j_k^{-1} Y_k, \quad k = 1, 2, \dots \quad (5.2)$$

where the Y_k are i.i.d. with distribution that of $X_{t-1} = (X_{t-1}, \dots, X_{t-p})'$; and the j_k are points of an independent PRM on $[0, \infty)$ with mean measure $\alpha x^{\alpha-1} dx$.

Theorem 5.1. *Under conditions M, S and L, the set $\Lambda(\mu)$ defined in (3.1) is bounded almost surely for μ defined by (5.1) and (5.2).*

Proof. From Proposition 3.4 we need to show that for all $a \in \mathbb{R}^p$ satisfying $a'1 \geq 0$ there exists $k \geq 1$ such that $a'u_k > 0$, or equivalently from (5.2), such that $a'Y_k > 0$. Since $\{Y_k\}$ form an i.i.d. sequence all we need show is that $P[a'Y_k > 0] = P[a'X_t > 0] > 0$.

Case 1: $E(Z_t) < \infty$. From condition M and the fact that $\sum |c_j| < \infty$ we can conclude that $m = E(X_t)$ is well-defined (see Section 4). Moreover, from the conditions M and S we have

$$m = \frac{E(Z_1)}{1 - \sum_{i=1}^p \phi_i} > 0;$$

and

$$E(a'X_t) = ma'1. \quad (5.3)$$

If $a'1 > 0$, then we conclude $E(a'X_t) > 0$ and so $P[a'X_t > 0] > 0$.

If $a'1 = 0$, we suppose by way of contradiction that $P[a'X_t > 0] = 0$. Then from (5.3) we must have that $a'X_t = 0$ almost surely. However, from Lemma 4.2, we know $a'X_t$ has a continuous distribution and hence $P[a'X_t = 0] = 0$ yielding a contradiction.

Case 2: $E(Z_t) = \infty$. Suppose that $a'1 > 0$ and in order to obtain a contradiction, that $P[a'X_t \geq 0] = 0$. Then almost surely, for any $n \geq 1$,

$$\begin{aligned} 0 &\geq \sum_{t=1}^n \sum_{i=1}^p a_i X_{t-i+1} \\ &= \sum_{i=1}^p a_i \sum_{t=1}^n X_{t-i+1} = \sum_{i=1}^p a_i \sum_{t=2-i}^{n-i+1} X_t \\ &= \sum_{i=1}^p a_i \left(\sum_{t=1}^n X_t - \sum_{t=n-i+2}^n X_t + \sum_{t=2-i}^0 X_t \right) \\ &= A + B + C. \end{aligned}$$

Since $\sum_{i=1}^p a_i > 0$, we have that $A \xrightarrow{P} \infty$ from the results in Section 4. For B we have

$$\begin{aligned} |B| &\leq p \bigvee_{i=1}^p |a_i| \sum_{t=n-p+2}^n |X_t| \stackrel{d}{=} p \bigvee_{i=1}^p |a_i| \sum_{t=2}^p |X_t| \\ &\leq p \bigvee_{i=1}^p |a_i| \sum_{t=1}^p |X_t| \end{aligned}$$

and thus $|B|$ is stochastically bounded. A similar argument shows $|C|$ is stochastically bounded and therefore,

$$0 \geq A + B + C \xrightarrow{P} \infty;$$

which is a contradiction. We conclude

$$a'1 > 0 \text{ implies } P \left[\sum_{i=1}^p a_i X_{t-i+1} \leq 0 \right] < 1. \quad (5.4)$$

Now we deal with the case $a_1 \neq 0$. We may suppose $a_1 = \pm 1$ since if $a_1 = 0$ we may move to the first non-zero entry in the list a_1, \dots, a_p and use this one in the role that a_1 plays in the following argument.

For the case $a_1 = 1$ we have

$$\sum_{i=1}^p a_i X_{t-i+1} = X_t + \sum_{i=2}^p a_i X_{t-i+1}$$

and by (2.1) this is

$$= \sum_{i=1}^p v_i X_{t-i} + Z_t,$$

where

$$v_i = \phi_i + a_{i+1}, \quad i = 2, \dots, p-1,$$

$$v_p = \phi_p.$$

Thus

$$\begin{aligned} P\left[\sum_{i=1}^p a_i X_{t-i+1} > 0\right] &= P\left[\sum_{i=1}^p v_i X_{t-i} + Z_t > 0\right] \\ &= P\left[Z_t > \sum_{i=1}^p -v_i X_{t-i}\right] \end{aligned}$$

where the two variables on either side of the inequality are independent. For any $r > 0$, this probability is bounded below by

$$\geq P[Z_t > r] P\left[\sum_{i=1}^p -v_i X_{t-i} < r\right].$$

Since EZ_1 is assumed infinite, no matter what value of r is used, we have $P[Z_t > r] > 0$. We may pick r large enough that $P[\sum_{i=1}^p -v_i X_{t-i} < r] > 0$ and then the desired conclusion

$$P\left[\sum_{i=1}^p a_i X_{t-i+1} > 0\right] > 0$$

follows.

The last case to consider is when $a_1 = -1$. In this case

$$\begin{aligned} \sum_{i=1}^p a_i X_{t-i+1} &= -X_t + \sum_{i=2}^p a_i X_{t-i+1} \\ &= -\sum_{i=1}^p \phi_i X_{t-i} - Z_t + \sum_{i=1}^{p-1} a_{i+1} X_{t-i} \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{p-1} (a_{i+1} - \phi_i) X_{t-i} - \phi_p X_{t-p} - Z_t \\
&= \sum_{i=1}^p v_i X_{t-i} - Z_t.
\end{aligned}$$

In this case we have

$$\mathbf{v}'\mathbf{1} = \sum_{i=2}^p a_i - \sum_{i=1}^p \phi_i = 1 - \sum_{i=1}^p \phi_i > 0.$$

So

$$\begin{aligned}
\Pr \left[\sum_{i=1}^p a_i X_{t-i+1} > 0 \right] &= P \left[\sum_{i=1}^p v_i X_{t-i} > Z_t \right] \\
&\geq \Pr[Z_t < r] \Pr \left[\sum_{i=1}^p v_i X_{t-i} > r \right].
\end{aligned} \tag{5.5}$$

The first factor is positive for every $r > 0$ because zero is the left endpoint of F , the distribution of Z_1 . Since $\mathbf{v}'\mathbf{1} > 0$, the positivity of the second factor in (5.5) follows for some $r > 0$ from (5.4). \square

In the right tail case, that is under conditions M, S and R, we have that the points $\mathbf{v}_l \in \mathbb{R}^p$ have the form:

$$\mathbf{v}_l = V_l(c_{l-1}, \dots, c_{l-p})' = V_l \sigma_{l-1}, \quad l = 1, 2, \dots \tag{5.6}$$

where the V_l variables are described in Section 4, the c_l are the coefficients from the $\text{MA}(\infty)$ representation. We use Proposition 3.4 in the following.

Theorem 5.2. *Under conditions M, S and R, the set $A(\mu)$ defined in (3.1) is bounded almost surely for μ defined by (5.1) and (5.6).*

Proof. Take $\mathbf{a} \in \mathbb{R}^p$ such that $\mathbf{a}'\mathbf{1} > 0$. Suppose that $\mathbf{a}'\sigma_l \leq 0$ for all $l \geq 0$. Then

$$0 \geq \mathbf{a}' \left(\sum_{l=0}^{\infty} \sigma_l \right) = \mathbf{a}' \mathbf{1} \left(1 - \sum_{i=1}^p \phi_i \right)^{-1} > 0 \tag{5.7}$$

by assumption on \mathbf{a} and condition S — hence a contradiction.

Suppose now that $\mathbf{a}'\mathbf{1} = 0$. If $\mathbf{a}'\sigma_l \leq 0$ for all $l \geq 0$ then from (5.7) $\mathbf{a}'\sigma_l = 0$ for all $l = 0, 1, 2, \dots$. If we let $j = \min\{i : a_i \neq 0\}$ then $\mathbf{a}'\sigma_j = 0 \Rightarrow c_0 = 0$ and we readily see that $c_l = 0$ for all $l \geq 0$. This will of course again lead to a degenerate process for the X_n , which is in contradiction to the assumptions on Z_t . \square

6. Uniqueness of solutions

Under the assumptions M, S and L, we will have uniqueness of the solution to (3.2) if we can show that with probability zero the determinant

$$D \equiv \det[Y_1 \ Y_2 \ \cdots \ Y_{p-1} \mathbf{1}] = 0. \quad (6.1)$$

Here the Y_k i.i.d. copies of X_{t-1} — see (5.2).

Theorem 6.1. *Assume M, S and L hold. Then for the limit PRM of (5.1), (5.2) the linear program (3.2) has a unique solution with probability 1.*

Proof. From Lemma 4.2, $a'Y_k$ has a continuous distribution for any $0 \neq a \in \mathbb{R}^p$. For (6.1) to hold we require that

$$D = \sum_{j=1}^p Y_{1,j} q_j(Y_2, \dots, Y_{p-1}) = 0$$

for some functions q_j . By independence, we can condition on Y_2, \dots, Y_{p-1} and conclude from the continuity of $a'Y_1$ that $P[D = 0 \mid Y_2, \dots, Y_{p-1}] = 0$ a.s. Integrating yields $P[D = 0] = 0$.

Thus, from Corollary 3.7 we conclude that the solution of (3.2) is unique almost surely. \square

For the right tail case under conditions M, S and R, from (5.6) we see that the conditions for uniqueness depend on the (non-random) directions σ_l . Indeed uniqueness will follow from Corollary 3.7 as follows.

Theorem 6.2. *Assume conditions M, S and R hold. If, for any collection of $p - 1$ non-negative and distinct indices $\{l_1, \dots, l_{p-1}\}$ the set of vectors $\{\mathbf{1}, \sigma_{l_1}, \dots, \sigma_{l_{p-1}}\}$ is linearly independent then the solution of (3.2) is unique.*

Proof. This result is basically a restatement of Corollary 3.7. \square

The linear independence condition of Theorem 6.2 can fail. For example, suppose $p = 2$. Then linear independence requires $c_l \neq c_{l-1}$ for every l . However, if $\phi_1(1 - \phi_1) = \phi_2$, it is readily checked from (4.1) that $c_1 = c_2$. Furthermore

$$\begin{aligned} 1 - \phi_1 - \phi_2 &= 1 - \phi_1 - \phi_1(1 - \phi_1) \\ &= (1 - \phi_1)^2 > 0, \end{aligned}$$

assuming $\phi_1 \neq 1$.

7. Proofs of main results

We are now in a position to prove our main Theorem 2.1. We will treat parts (a) and (b) separately.

Proof of Theorem 2.1(a). Under conditions M, S and L we have, from Theorem 4.3(a) that $\mu_n \Rightarrow \mu$ where μ_n and μ are defined in (4.4). Moreover, it is easy to show that if

$$\hat{\phi}^{(n)} = \arg \max \{ \delta' \mathbf{1} : \delta' X_{t-1} \leq X_t; t = 1, \dots, n \}$$

then

$$\begin{aligned} a_n^{-1}(\hat{\phi}^{(n)} - \phi^{(0)}) &= \arg \max \{ \delta' \mathbf{1} : \delta' (a_n Z_t^{-1} X_{t-1}) \leq 1; t = 1, \dots, n \} \\ &= \arg \max \{ \delta' \mathbf{1} : \delta \in A(\mu_n) \}. \end{aligned} \quad (7.1)$$

Note that $A(\mu_n)$ has the extra constraint $\delta' \mathbf{1} \geq -1$, but it has no effect on the linear program since $\delta = \mathbf{0}$ is always a feasible point and hence this constraint is never active.

By Theorem 5.1 $A(\mu)$ is almost surely bounded in \mathbb{R}^p and by Theorem 6.1 it has almost surely a unique solution. Now we may employ the continuity theorems 3.5 and 3.8, together with the continuous mapping theorem, to conclude that

$$\arg \max \{ \delta' \mathbf{1} : \delta \in A(\mu_n) \} \Rightarrow \arg \max \{ \delta' \mathbf{1} : \delta \in A(\mu) \}.$$

This last result, given the relationship (7.1), is just a paraphrasing of the claim of part (a) of Theorem 2.1 and so the proof is completed. \square

Proof of Theorem 2.1(b). Under conditions M, S and R we have again that $\mu_n \Rightarrow \mu$ where μ_n and μ are defined in (4.5) and (4.7). Moreover, we have that $A(\mu) = A(\mu')$, for μ' as defined in (4.8), and from Theorem 5.1 we have that $A(\mu')$ is bounded in \mathbb{R}^p .

Similarly to the proof of part (a), we have

$$b_n(\hat{\phi}^{(n)} - \phi^{(0)}) = \arg \max \{ \delta' \mathbf{1} : \delta \in A(\mu_n) \}$$

and therefore $b_n \|\hat{\phi}^{(n)} - \phi^{(0)}\| \leq \max \{ \|\delta\| : \delta \in A(\mu_n) \}$.

However the mapping $Q: \mathbf{K} \rightarrow [0, \infty) - \mathbf{K}$ is the collection of compact sets in \mathbb{R}^p — given by

$$Q(A) = \max \{ \|\mathbf{x}\| : \mathbf{x} \in A \}$$

is continuous with respect to the Hausdorff metric on \mathbf{K} . Theorem 3.5 ensures that $A(\mu_n) \Rightarrow A(\mu')$ and so

$$Q(A(\mu_n)) \Rightarrow Q(A(\mu')),$$

where the latter is a finite random variable. Hence

$$\begin{aligned} P[b_n \|\hat{\phi}^{(n)} - \phi^{(0)}\| > x] &\leq P[Q(A(\mu_n)) > x] \\ &\rightarrow P[Q(A(\mu')) > x] \end{aligned}$$

and so we conclude that

$$b_n \|\hat{\phi}^{(n)} - \phi^{(0)}\| = O_p(1).$$

If, moreover, the linear independence condition of the statement of the theorem holds, (see (2.4)) then by Corollary 3.7,

$$V = \arg \max \{ \delta' \mathbf{1} : \delta \in A(\mu') \}$$

is unique and so we can again apply the extra continuity result of Theorem 3.8 to conclude that

$$b_n \left(\hat{\phi}^{(n)} - \phi^{(0)} \right) \Rightarrow V$$

just as we did in the proof of part (a). This completes the proof of Theorem 2.1(b). \square

8. Concluding remarks

Comparison with Yule–Walker estimates: Under conditions M, S and R, Davis and Resnick (1986) have given results about the asymptotic distribution of the Yule–Walker estimators. These results arose as a by product of the study of the limit distribution theory for the sample correlation function. Define

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X})$$

and

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

The Yule–Walker estimator is defined by

$$\hat{\phi}_{YW}^{(n)} = \hat{R}_p^{-1} \hat{\rho}_p$$

where

$$\hat{\rho}_p = (\hat{\rho}(1), \dots, \hat{\rho}(p))'$$

and

$$\hat{R}_p = (\hat{\rho}(i-j); i, j = 1, \dots, p).$$

Furthermore, define

$$b_n = \left(\frac{1}{P[Z_1 > \cdot]} \right)^{\leftarrow} (n)$$

and

$$\tilde{b}_n = \left(\frac{1}{P[Z_1 Z_2 > \cdot]} \right)^{\leftarrow} (n).$$

Suppose $EZ_1^\alpha = \infty$. Sample results from Davis and Resnick (1986) include the following. If $\alpha \in (1, 2)$ then

$$\frac{b_n^2}{\tilde{b}_n} (\hat{\phi}_{YW}^{(n)} - \phi^{(0)}) \Rightarrow DY = D(Y_1, \dots, Y_p)'$$

If $\alpha < 1$ then

$$n(\hat{\phi}_{YW}^{(n)} - \phi^{(0)}) \Rightarrow DS = D(S_1, \dots, S_p)'$$

Here D is a matrix of partial derivatives and Y and S are vectors of jointly stable random variables.

Suppose we compare this with the rate of convergence b_n of our estimator given in Theorem 2.1(b). If $\alpha \in (1, 2)$, then

$$\frac{\tilde{b}_n^{-1} b_n^2}{b_n} = \frac{b_n}{\tilde{b}_n} \rightarrow 0$$

(by (3.5) in Davis and Resnick, 1986) and so our estimator improves on the rate of convergence of the Yule–Walker estimator, albeit by a slowly varying factor. Similarly if $\alpha < 1$, then our rate b_n is regularly varying of index $1/\alpha > 1$ and hence approaches infinity faster than n .

Statistical implications: Under conditions M, S and L a limit distribution for our estimator ensues. While it is doubtful that the distribution of the limiting random variable U in Theorem 2.1(a) can be explicitly computed, the structure of the random variable U provides a recipe that can be simulated. However, the limit distribution depends on the unknown parameter α and the unknown distribution of the vector (X_1, \dots, X_p) . So the limit result in its present form is not suitable for the construction of a confidence region. We intend to explore ways of using random normalizations to yield limit distributions which contain no unknown quantities.

Under conditions M, S and R, we have shown our estimator is consistent and we have given the order of convergence. However, we have not established that a limit distribution exists without a linear independence condition. We expect that the limit distribution will usually exist but the extent to which this condition restricts applicability remains to be seen.

We hope to explore further the statistical nature of our results. We also hope to extend our results to ARMA models with a moving average component and to fractionally differenced models. One approach to estimating coefficients in an ARMA(p, q) or fractionally differenced model is to extend our results to the $p = \infty$ case and thereby look at estimation of the coefficients in AR(∞) models. Good criteria for model selection is also an open issue.

We would like to be able to jointly estimate α and ϕ . Under condition L, estimation of α based on (X_0, \dots, X_n) is complicated by the fact that the behavior of the distribution of X_1 at the left endpoint may be nothing like the behavior of the distribution of Z_1 at the left endpoint (Davis and Resnick (1991)).

Li and McCleod (1988) have fitted some ARMA models with gamma and log normal innovations to the sunspots and lynx data sets. We plan to try our estimators on these data sets as well as doing extensive simulations to see how our estimators work in practice.

References

- J. Andel, Non-negative autoregressive processes, *J. Time Ser. Anal.* 10 (1989) 1–11.
- P. Billingsley, *Convergence of Probability Measures* (Wiley, New York, 1968).
- L. Breiman, On some limit theorems similar to the arc-sin law, *Theor. Probab. Appl.* 10 (1965) 323–331.
- L. Breiman, *Probability* (Addison Wesley, Reading, MA, 1968).
- P. Brockwell and R. Davis, *Time Series: Theory and Methods*, (Springer-Verlag, New York, 2nd ed., 1991).
- D. Cline, Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data, Thesis, Dept of Statistics, Colorado State University (Ft. Collins CO 80521 USA, 1983).
- R. Davis and W. McCormick, Estimation for first-order autoregressive processes with positive or bounded innovations, *Stochastic Process. Appl.* 31 (1989) 237–250.
- R. Davis, E. Mulrow and S. Resnick, The convex hull of a random sample in R^2 , *Stochastic Models* 3 (1987) 1–27.
- R. Davis and S. Resnick, Limit theory for the sample covariance and correlation functions of moving averages, *Ann. Statist.* 14 (1986) 533–558.
- R. Davis and S. Resnick, Extremes of moving averages of random variables with finite endpoint, *Ann. Probab.* 19 (1991) 312–328.
- R. Davis and M. Rosenblatt, Parameter estimation for some time series models without continuity, *Statist. Probab. Lett.* 11 (1991) 515–521.
- P. Feigin, Inference and martingale estimating equations for stochastic processes on a semigroup, *J. Statist. Planning Inference* 39 (1994) 239–254.
- P. Feigin and S. Resnick, Estimation for autoregressive processes with positive innovations, *Stochastic Models* 8 (1992) 479–498.
- S. Lay, *Convex Sets and their Applications* (Wiley, New York, 1982).
- W.K. Li and A.I. McLeod, ARMA modelling with non-Gaussian innovations, *J. Time Ser. Anal.* 9 (1988) 155–168.
- S. Resnick, *Extreme Values, Regular Variation, and Point Processes* (Springer, New York, 1987).
- V. Yohai and R. Maronna, Asymptotic behavior of least-squares estimates for autoregressive processes with infinite variances, *Ann. Statist.* 5 (1977) 554–560.