

## PARAMETER ESTIMATION FOR MOVING AVERAGES WITH POSITIVE INNOVATIONS<sup>1</sup>

BY PAUL D. FEIGIN, MARIE F. KRATZ AND SIDNEY I. RESNICK

*Technion–Israel Institute of Technology, Université René Descartes,  
Paris V and Cornell University*

This paper continues the study of time series models generated by nonnegative innovations which was begun by Feigin and Resnick. We concentrate on moving average processes. Estimators for moving average coefficients are proposed and consistency and asymptotic distributions established for the case of an order-one moving average assuming either the right or the left tail of the innovation distribution is regularly varying. The rate of convergence can be superior to that of the Yule–Walker or maximum likelihood estimators.

**1. Introduction.** This paper continues the study of time series models generated by nonnegative innovations which was begun in Feigin and Resnick (1992, 1994). This program is motivated by the need to model teletraffic and hydrologic data sets where quantities such as holding times and stream flows are inherently positive and hence possibly unsuited to the usual time series methods which are based on Gaussian models. In Feigin and Resnick (1994), we showed how to estimate parameters of a pure autoregression using linear programming (lp) techniques. Assuming regular variation of either the left or the right innovation distribution tail, such lp estimators have a good rate of convergence which is frequently superior to those achieved by Yule–Walker or maximum likelihood estimators. The lp estimators can be used for model selection and for testing for independence [Feigin, Resnick and Stărică (1995)]. In this paper, we focus on estimation of moving average coefficients. This is a necessary step along the road to being able to estimate parameters in more general ARMA processes which combine both autoregressive and moving average components.

There is an abundance of heavy-tailed data coming out of teletraffic studies. Recent examples include Crovella and Bestavros (1995) and Cunha, Bestavros and Crovella (1995). Since ARMA modeling is firmly established as an effective tool for modeling finite variance data, it is natural to inquire if

---

Received January 1995; revised May 1996.

<sup>1</sup>Paul Feigin and Sidney Resnick were partially supported by U.S.–Israel Binational Science Foundation (BSF) Grant 92-00227. Sidney Resnick was also partially supported by NSF Grant DMS-94-00535 at Cornell University, which also supported the travel to Cornell by Marie Kratz in the fall of 1993.

*AMS 1991 subject classifications.* 60B10, 60F05, 60G55, 60E20, 26F10, 62M10.

*Key words and phrases.* Poisson processes, linear programming, autoregressive processes, moving average processes, parameter estimation, weak convergence, consistency, time series analysis.

ARMA models are a sufficiently flexible class for modeling infinite variance data. Affirmative evidence to date is scanty at best [Davis and Resnick (1996) and Resnick (1996)] but without efficient estimation techniques such as proposed here and in Feigin and Resnick (1994), it is not possible to decisively reject a linear model as an adequate description of the data. Note also that there are finite variance circumstances where the left innovation distribution tail is regularly varying and in this case lp estimation techniques applied to linear time series models can be quite successful. A striking example is lp analysis of the lynx data presented in Feigin, Resnick and Stărică (1995).

The process under consideration in this paper is the finite-order moving average of order  $q$ , denoted  $MA(q)$  and specified as follows. Let  $\{Z_t\}$  be an iid sequence of nonnegative random variables. For a positive integer  $q \geq 1$ , suppose we have parameters  $\theta_1, \dots, \theta_q$  such that  $\theta_i \geq 0$  for  $1 \leq i \leq q$ . The  $MA(q)$  process  $\{X_t\}$  is

$$(1.1) \quad X_t = Z_t + \sum_{i=1}^q \theta_i Z_{t-i}, \quad t = 0, \pm 1, \pm 2, \dots,$$

and we are interested in estimating  $\theta_1, \dots, \theta_q$ . It is convenient to be able to write (1.1) compactly and to achieve this we define the moving average polynomial

$$\Theta(z) = \sum_{i=0}^q \theta_i z^i,$$

where  $\theta_0 = 1$  and the backward shift operator  $B$  is defined symbolically by

$$BX_t = X_{t-1}, \quad BZ_t = Z_{t-1}.$$

With this notation we may write the  $MA(q)$  as

$$X_t = \Theta(B)Z_t, \quad t = 0, \pm 1, \pm 2, \dots$$

For a pure autoregressive process of order  $p$ , denoted by  $AR(p)$ , with positive innovations  $\{Z_t\}$  and with autoregressive coefficients  $\phi_1, \dots, \phi_p$ ,  $\phi_p \neq 0$ ,  $\sum_{i=1}^p \phi_i < 1$ , of the form

$$(1.2) \quad X_t = \sum_{k=1}^p \phi_k X_{t-k} + Z_t, \quad t = 0, \pm 1, \pm 2, \dots,$$

Feigin and Resnick (1994) defined the linear programming estimators  $\hat{\phi}$  based on observing  $X_1, \dots, X_n$  as

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

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

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

Such estimators can be motivated either as approximate maximum likelihood estimators or from the point of view of generalized martingale estimating equations. [See Feigin (1991) and Feigin and Resnick (1992, 1994).] Assuming regular variation conditions on either the left or the right tails of the innovations was sufficient to show that a limit distribution existed for  $\hat{\phi}$  and that rates of convergence were often superior to the Yule–Walker estimators. Furthermore, lp estimators frequently give some warning of model misspecification whereas Yule–Walker estimators do not [Feigin and Resnick (1996), Davis and Resnick (1996) and Resnick (1996)].

A natural approach to the estimation problem for moving averages is to see what results from the autoregressive case can be brought to bear and thus we assume the moving average in (1.1) is *invertible* which, according to Brockwell and Davis (1991), means  $\Theta(z) \neq 0$ ,  $|z| \leq 1$ . This allows us to write

$$\Pi(z) := \frac{1}{\Theta(z)} = \sum_{i=0}^{\infty} \pi_k z^k, \quad |z| \leq 1,$$

and we convert (1.1) into an infinite-order autoregression

$$\Pi(B)X_t = Z_t, \quad t = 0, \pm 1, \pm 2, \dots$$

If we now try to apply the lp estimators, we find we have a nice objective function but the constraints involve an infinite number of variables. If we truncate the constraints suitably, we should obtain an estimator with worthwhile properties. The precise definition of our estimator of the moving average coefficients in the MA( $q$ ) process is

$$(1.5) \quad \hat{\theta} := \arg \max_{D_n} \sum_{i=1}^q \theta_i,$$

where

$$(1.6) \quad D_n := \left\{ \theta: \left[ \sum_{i=0}^{2l} (I - \Theta(B))^i \right] X_i \geq 0, t = 2lq + 1, \dots, n \right\}$$

and  $l$  is the first integer such that  $2l \geq q$ . Further motivation and discussion of this estimator is the subject of Section 2.

Here is a precise statement of the assumptions which will allow discussion of the properties of our estimators. We need conditions which specify the model. 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  $U: [0, \infty) \mapsto (0, \infty)$  is regularly varying with exponent  $\rho \in \mathbb{R}$  if

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

CONDITION M (Model specification). The process  $\{X_t: t = 0, \pm 1, \pm 2, \dots\}$  satisfies (1.1):

$$(1.1) \quad X_t = Z_t + \sum_{i=1}^q \theta_i Z_{t-1}, \quad t = 0, \pm 1, \pm 2, \dots,$$

where  $\{Z_t\}$  is an independent and identically distributed sequence of nonnegative random variables and common distribution function  $F$ . The nonnegative coefficients  $\theta_1, \dots, \theta_q$  satisfy the invertibility condition that the moving average polynomial  $\Theta(z) = \sum_{i=0}^q \theta_i z^i$  has no roots in the unit disk  $\{z: |z| \leq 1\}$ .

CONDITION L (Left tail). The distribution  $F$  of the innovations  $Z_t$  satisfies, for some  $\alpha > 0$ :

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

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

CONDITION R (Right tail). The distribution  $F$  of the innovations  $Z_t$  satisfies, for some  $\alpha > 0$ :

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

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

Our results have as hypotheses M and either L or R. 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.

Section 2 discusses further motivation and properties of the mathematical programming estimator given in (1.6). Section 3 assumes Condition R and reviews the point process limit theory [Resnick (1987)] which underlies discussion of the limit distributions for  $\hat{\theta}$  carried out for the case  $q = 1$ . We show that for  $q = 1$ , if  $\theta^{(0)}$  is the true MA(1) parameter, then under Conditions M and R, our estimator  $\hat{\theta}$  is asymptotically Weibull distributed and (Theorem 3.1)

$$\lim_{n \rightarrow \infty} P[b_n(\hat{\theta} - \theta^{(0)}) \leq x] = 1 - \exp\{-cx^\alpha\}, \quad x > 0,$$

where

$$c = E|Z_1 + (\theta^{(0)})^3 Z_2|^{-\alpha}, \quad b_n = F^{-1}\left(1 - \frac{1}{n}\right).$$

Section 4 parallels Section 3 but assumes Condition L. When  $q = 1$ , under Conditions M and L,  $\hat{\theta}$  is also asymptotically Weibull distributed (Theorem 4.1):

$$\lim_{n \rightarrow \infty} P\left[a(\sqrt{n})^{-1}(\hat{\theta} - \theta^{(0)}) \leq x\right] = 1 - \exp\{-kx^{2\alpha}\}, \quad x > 0,$$

where

$$\begin{aligned} k &= (\theta^{(0)})^{-3\alpha} c(\alpha) E(|Z_1 - \theta^{(0)} Z_2|^{2\alpha} \mathbf{1}_{[Z_1 > Z_2]}), \\ c(\alpha) &= \int_0^1 (1-s)^\alpha \alpha s^{\alpha-1} ds, \\ a(n) &= F^{-1}\left(\frac{1}{n}\right). \end{aligned}$$

In Section 5 we give the results of a simulation study which investigates how well the limit distributions derived in Sections 3 and 4 for MA(1) estimation approximate the true finite-sample distributions. The conclusion is that in the right-tail case, the limit distribution provides an excellent approximation over a wide range of  $\theta^{(0)}$  values. In the left-tail case, this is true provided  $\theta^{(0)}$  is not too close to 1.

In Section 6 we present some concluding remarks which emphasize the point that, in contrast to the autoregressive case, the moving average estimators in the left-tail case suffer a performance degradation depending on the order  $q$  of the model; no such degradation is present under Condition R. Some future issues to be resolved are also considered.

**2. The parameter estimator for MA( $q$ ).** Assume we have the invertible model  $\{X_t\}$  specified by Condition M. Suppose the true value of the moving average coefficients is  $\theta^{(0)}$ . In inverted form, the model can be written as the AR( $\infty$ ) process

$$(2.1) \quad \Pi(B)X_t = Z_t, \quad t = 0, \pm 1, \pm 2, \dots,$$

where

$$\frac{1}{\Theta(z)} = \Pi(z), \quad |z| \leq 1.$$

For a finite-order autoregression (1.2), the linear programming estimator of autoregressive coefficients is given by (1.3) and (1.4). If in (1.2) we write as usual the autoregressive polynomial as

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

then the objective function in (1.3) can be written as  $1 - \Phi(1)$  and the constraints in (1.4) can be expressed as

$$(2.2) \quad \Phi(B) X_t \geq 0, \quad t = p + 1, \dots, n.$$

If we try to write down an analogous expression for the parameter estimators for the AR( $\infty$ ) process in (2.1), we obtain as the objective function

$$1 - \frac{1}{\Theta(1)} = \frac{\sum_{i=1}^q \theta_i}{1 + \sum_{i=1}^q \theta_i},$$

which is monotone in  $\sum_{i=1}^q \theta_i$ . So we try to maximize  $\sum_{i=1}^q \theta_i$ . For the constraints, (2.2) suggests the set of conditions

$$\Pi(B) X_t \geq 0, \quad t = 1, \dots, n.$$

A problem arises in that this constraint set requires knowledge of  $X_t, X_{t-1}, \dots$  with the index extending back to  $-\infty$ , and since we only have knowledge of  $X_1, \dots, X_n$  we must somehow truncate this constraint set.

A suggestion for how to construct a truncated set of constraints comes from symbolically expanding  $1/\Theta$ :

$$\frac{1}{\Theta(B)} = \frac{1}{I - (I - \Theta(B))} = \sum_{k=0}^{\infty} (I - \Theta(B))^k \approx \sum_{k=0}^{2l} (I - \Theta(B))^k,$$

where  $l \geq 1$  is an integer to be specified. Note that

$$\sum_{k=0}^{2l} (I - \Theta(B))^k \Theta(B) = I - (I - \Theta(B))^{2l+1} = I + Q(B)^{2l+1},$$

where  $Q(B) = \Theta(B) - I = \sum_{i=1}^q \theta_i B^i$ . Let  $\Theta^{(0)}(B) = \sum_{i=0}^q \theta_i^{(0)} B^i$  and  $Q_0(B) = \sum_{i=1}^q \theta_i^{(0)} B^i$  and thus

$$\begin{aligned} \sum_{k=0}^{2l} (I - \Theta^{(0)}(B))^k X_t &= \sum_{k=0}^{2l} (I - \Theta^{(0)}(B))^k \Theta^{(0)}(B) Z_t \\ &= (I + Q_0^{2l+1}(B)) Z_t \geq 0, \end{aligned}$$

since all  $\theta_i^{(0)}$ 's are assumed nonnegative. So, by truncating the series expansion for  $1/\Theta$  in a judicious manner, the truncated expansion is always positive at the true value of the parameter vector.

Thus our estimator is

$$(2.3) \quad \hat{\theta} = \arg \max_{D_n} \sum_{i=1}^q \eta_i,$$

where the constraint set is

$$(2.4) \quad D_n = \left\{ \boldsymbol{\eta} \in \mathbb{R}_+^q : \sum_{k=0}^{2l} \left( I - \sum_{i=0}^q \eta_i B^i \right)^k X_t \geq 0, \right. \\ \left. t = 2lq + 1, \dots, n; \eta_0 = 1, \sum_{i=0}^q \eta_i z^i \neq 0, |z| \leq 1 \right\}.$$

The choice of  $l$  suggested by the limit theory is to choose  $l$  to be the first integer such that  $2l \geq q$ .

*Change of variable.* We seek a limit distribution for  $q_n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})$ , where  $q_n$  is an appropriate scaling satisfying  $q_n \rightarrow \infty$ . It will turn out that, under Condition R, the right choice of  $q_n$  is

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

and, assuming Condition L and  $q = 1$ , the appropriate choice of  $q_n$  is

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

We observe that  $q_n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})$  satisfies

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

for all  $\boldsymbol{\eta} \in D_n$ . Let  $\boldsymbol{\delta} = q_n(\boldsymbol{\eta} - \boldsymbol{\theta}^{(0)})$  so that  $q_n^{-1} \boldsymbol{\delta} + \boldsymbol{\theta}^{(0)} = \boldsymbol{\eta}$ . Then  $q_n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})$  satisfies  $q_n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})' \mathbf{1} \geq \boldsymbol{\delta}' \mathbf{1}$  for all  $\boldsymbol{\delta}$  such that

$$1 + \sum_{i=1}^q \left( \frac{\delta_i}{q_n} + \theta_i^{(0)} \right) z^i \neq 0, \quad |z| \leq 1,$$

and

$$\sum_{k=0}^{2l} (-1)^k \left( Q_0(B) + \frac{\delta(B)}{q_n} \right)^k X_t \geq 0$$

for  $t = 2lq + 1, \dots, n$ , where  $\delta(B) = \sum_{i=1}^q \delta_i B^i$ . Thus

$$q_n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)}) = \arg \max_{\Lambda_n} \boldsymbol{\delta}' \mathbf{1},$$

where

$$\Lambda_n = \left\{ \boldsymbol{\delta} \in D_n : 1 + \sum_{i=1}^q \left( \frac{\delta_i}{q_n} + \theta_i^{(0)} \right) z^i \neq 0, |z| \leq 1, \right. \\ \left. \sum_{k=0}^{2l} (-1)^k \left( Q_0(B) + \frac{\delta(B)}{q_n} \right)^k X_t \geq 0, t = 2lq + 1, \dots, n \right\}.$$

In case  $q = 1$ , we also have  $l = 1$  and  $\Theta(B) = I + \theta B$  so  $I - \Theta(B) = -\theta B$  and the estimator is

$$\begin{aligned} \hat{\theta} &= \sup\{\eta \in [0, 1): X_t - \eta X_{t-1} + \eta^2 X_{t-2} \geq 0, t = 3, \dots, n\} \\ &= \bigwedge_{t=3}^n \sup\{\eta \in [0, 1): X_t - \eta X_{t-1} + \eta^2 X_{t-2} \geq 0\}. \end{aligned}$$

Also,

$$q_n(\hat{\theta} - \theta^{(0)}) = \arg \max_{\Lambda_n} \delta = \sup\{\delta \in \Lambda_n\},$$

where

$$\begin{aligned} \Lambda_n &= \left\{ \delta: 1 + \left( \frac{\delta}{q_n} + \theta^{(0)} \right) z \neq 0, |z| \leq 1, \right. \\ &\quad \left. X_t - \left( \theta^{(0)} + \frac{\delta}{q_n} \right) X_{t-1} + \left( \theta_0 + \frac{\delta}{q_n} \right)^2 X_{t-2} \geq 0, t = 3, \dots, n \right\}. \end{aligned}$$

Let  $\eta = \delta/q_n$  and recall that  $X_t = Z_t + \theta^{(0)}Z_{t-1}$ . If we set

$$\begin{aligned} (2.5) \quad A_t &= Z_{t-2} + \theta^{(0)}Z_{t-3}, \\ B_t &= -Z_{t-1} + \theta^{(0)}Z_{t-2} + 2(\theta^{(0)})^2 Z_{t-3}, \\ C_t &= Z_t + (\theta^{(0)})^3 Z_{t-3}, \end{aligned}$$

then

$$\Lambda_n = q_n\{-\theta^{(0)} \leq \eta < 1 - \theta^{(0)}: A_t\eta^2 + B_t\eta + C_t \geq 0, t = 3, \dots, n\}$$

and we find in case  $q = 1$  that, since  $0 \in \Lambda_n$ ,

$$(2.6) \quad q_n(\hat{\theta} - \theta^{(0)}) = q_n \bigwedge_{t=3}^n \sup\{0 \leq \eta < 1 - \theta^{(0)}: A_t\eta^2 + B_t\eta + C_t \geq 0\}.$$

So the limit distribution depends on the behavior of random parabolas and from extreme value theory we expect the limit distribution to be in the Weibull family [cf. Resnick (1987), pages 14 and 15].

To analyze the limit distribution in (2.6), we intend to proceed as follows. Denote the random parabola by

$$p_t(\eta) = A_t\eta^2 + B_t\eta + C_t.$$

Only those parabolas such that  $p_t(1 - \theta^{(0)}) < 0$  are of interest since if  $p_t(1 - \theta^{(0)}) \geq 0$ , then

$$\sup\{0 \leq \eta \leq 1 - \theta^{(0)}: p_t(\eta) \geq 0\} = 1 - \theta^{(0)},$$

which is an uninteresting contribution to the minimum in (2.6). Note that the condition  $p_t(1 - \theta^{(0)}) < 0$  also implies  $B_t < 0$  and that the discriminant of the quadratic is positive so that the two roots of the quadratic are real. [The product of the roots of  $p_t(\eta)$  is  $C_t/A_t \geq 0$  so that both roots have the same



sign. The sum of the roots is  $-B_t/2A_t$ . If  $p_t(1 - \theta^{(0)}) < 0$ , then the bigger root is positive which implies both roots are positive and hence  $B_t < 0$ .] Thus the smaller root  $r_t^-$  is the desired root and in (2.6)

$$q_n(\hat{\theta} - \theta^{(0)}) = q_n \bigwedge_{\substack{1 \leq t \leq n \\ p_t(1 - \theta^{(0)}) < 0}} r_t^-.$$

Now

$$\begin{aligned} r_t^- &= \frac{-B_t - \sqrt{B_t^2 - 4A_t C_t}}{2A_t} \\ (2.7) \quad &= \left( |B_t| - |B_t| \left( 1 - \frac{4A_t C_t}{B_t^2} \right)^{1/2} \right) / 2A_t \\ &\approx \frac{C_t}{|B_t|}, \end{aligned}$$

where the last step resulted from expanding the function  $(1 + x)^{1/2}$  and neglecting remainder terms. Thus the behavior of

$$q_n \bigwedge_{\substack{1 \leq t \leq n \\ p_t(1 - \theta^{(0)}) < 0}} r_t^-$$

should be determined by

$$q_n \bigwedge_{\substack{1 \leq t \leq n \\ p_t(1 - \theta^{(0)}) < 0}} C_t/|B_t|,$$

and the behavior of this quantity can be determined by using a point process argument which depends on whether Condition L or R is assumed.

The approximation in (2.7) can be justified by the following mechanism: Assume  $p_t(1 - \theta^{(0)}) < 0$ . Consider the roots of the lines

$$L_1: B_t \eta + C_t, \quad L_2: (A_t(1 - \theta^{(0)}) + B_t) \eta + C_t,$$

$L_2$  being the line passing through  $C_t$  and  $p_t(1 - \theta^{(0)})$ . The roots of the two lines are

$$x_{t1} = \frac{C_t}{|B_t|}, \quad x_{t2} = \frac{C_t}{|B_t| - (1 - \theta^{(0)}) A_t}$$

and

$$x_{t1} \leq r_t^- \leq x_{t2}.$$

If we know the limit behavior of the point process depending on  $\{x_{t1}\}$ , and if  $x_{t2}$  is sufficiently close to  $x_{t1}$ , then the sandwiched piece  $r_t^-$  will behave properly and give us the limit distribution.

Details are in the next two sections which assume Condition R and then Condition L.

**3. The limit distribution in the right-tail case for  $q = 1$ .** In this section we assume Conditions M and R hold. We assume we are dealing with MA(1) so that  $q = 1$ . The goal of this section is to present the limit distribution for  $\hat{\theta}$ . We will prove the following theorem.

**THEOREM 3.1.** *Suppose  $\{X_t\}$  is the MA(1) process given in (1.1) and that Conditions M and R hold. Suppose the true parameter is  $\theta^{(0)} \in (0, 1)$  and that  $F$  is continuous. Let  $q_n = b_n$  be the quantile function*

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

where  $F$  is the distribution of  $Z_1$ . The estimator  $\hat{\theta}$  given in Section 2 has a Weibull limit distribution. In  $[0, \infty)$ ,

$$(3.1) \quad b_n(\hat{\theta} - \theta^{(0)}) \Rightarrow \bigwedge_{k=1}^{\infty} \Gamma_k^{1/\alpha} \left( Y_k + (\theta^{(0)})^3 Y'_k \right),$$

where  $\{Y_k, Y'_k, k \geq 1\}$  are iid with common distribution  $F$  and

$$\Gamma_k = E_1 + \cdots + E_k, \quad k \geq 1,$$

is a sum of iid unit exponentially distributed random variables independent of  $\{(Y_k, Y'_k)\}$ . The limit distribution of  $\hat{\theta}$  is Weibull:

$$(3.2) \quad \lim_{n \rightarrow \infty} P \left[ b_n(\hat{\theta} - \theta^{(0)}) \leq x \right] = 1 - \exp\{-cx^\alpha\}, \quad x > 0,$$

where

$$c = E|Y_k + (\theta^{(0)})^3 Y'_k|^{-\alpha},$$

which is finite by the second statement of Condition R.

Before discussing the limit theory which leads to the asymptotic distribution of  $\hat{\theta}$ , we quickly review some facts about point processes.

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  $\rho(\cdot, \cdot)$ . Its properties are discussed, for example, in Resnick (1987), Section 3.4, or Kallenberg (1983). 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  $m_n(f) = \int_E f dm_n$ . The nonnegative continuous functions with compact support will be denoted  $C_K^+(E)$ .

A Poisson process on  $E$  with mean measure  $\mu$  will be denoted  $\text{PRM}(\mu)$ . Two examples of the space  $E$  that interest us are  $E = [0, \infty)$ , where compact sets are those closed sets bounded away from  $\infty$ , and  $E = [0, \infty]^p \setminus \{\mathbf{0}\}$ , where compact sets are closed subsets of  $[0, \infty]^p$  which are bounded away from  $\mathbf{0}$ . Other examples of the space  $E$  will be needed as well.

The fact that (3.1) implies (3.2) is a standard fact in extreme value theory. Since

$$\sum_{k=1}^{\infty} \varepsilon_{(\Gamma_k, Y_k + (\theta^{(0)})^3 Y'_k)}$$

is a Poisson process with mean measure  $du \times P[Y_1 + (\theta^{(0)})^3 Y'_1 \in dv]$  [Resnick (1987), page 135], if we set

$$\Lambda := \{(u, v) \in [0, \infty)^2 : u^{1/\alpha} v \leq x\},$$

then, for  $x > 0$ ,

$$\begin{aligned} & P \left[ \bigwedge_{k=1}^{\infty} \Gamma_k^{1/\alpha} (Y_k + (\theta^{(0)})^3 Y'_k) > x \right] \\ &= P \left[ \sum_{k=1}^{\infty} \varepsilon_{(\Gamma_k, Y_k + (\theta^{(0)})^3 Y'_k)}(\Lambda) = 0 \right] \\ &= \exp \left\{ - \iint_{\Lambda} du P[Y_1 + (\theta^{(0)})^3 Y'_1 \in dv] \right\} \\ &= \exp \left\{ - \int_{v=0}^{\infty} \int_{u \leq (x/v)^\alpha} du P[Y_1 + (\theta^{(0)})^3 Y'_1 \in dv] \right\} \\ &= \exp \left\{ -x^\alpha \int_0^{\infty} v^{-\alpha} P[Y_1 + (\theta^{(0)})^3 Y'_1 \in dv] \right\}, \end{aligned}$$

which is (3.2).

The main point process limit theorem which underlies our work in this section now follows. It is more general than we need for considering the asymptotic behavior of  $\hat{\theta}$  in the MA(1) case but is stated in full generality for application to future work. A simpler version is discussed as Proposition 3.1 of Davis and Resnick (1988).

**PROPOSITION 3.2.** *Suppose Conditions M and R hold and that  $P[Z_1 = 0] = 0$ . Define the measure*

$$\nu(dx) = \alpha x^{-\alpha-1} dx, \quad x > 0.$$

For any positive integer  $m$  we have

$$\begin{aligned} & nP \left[ Z_0 \in dy_0, \frac{Z_i}{b_n} \in dx_i, i = 1, \dots, m; Z_j \in dy_j, j = 1, \dots, m \right] \\ (3.3) \quad & \rightarrow_v F(dy_0) \sum_{i=1}^m \nu(dx_i) \varepsilon_x(dy_i) \prod_{\substack{j \neq i \\ 1 \leq j \leq m}} \varepsilon_0(dx_j) F(dy_j) \end{aligned}$$

in

$$(3.4) \quad E := [0, \infty] \times ([0, \infty]^m \setminus \{\mathbf{0}\}) \times [0, \infty]^m.$$

Furthermore, let  $E$  be defined as in (3.4) and for  $l = 1, \dots, m$  set

$$\mathbf{e}_l = (0, \dots, 1, \dots, 0) \in \mathbb{R}^m,$$

where the 1 appears in the  $l$ th coordinates. Suppose  $\{Y_{k,l}, Y'_{k,l}, k \geq 1, l \geq 1\}$  are iid with distribution  $F$ . Then

$$(3.5) \quad \begin{aligned} & \sum_{t=1}^n \mathcal{E}_{(Z_t, b_n^{-1}(Z_{t-i}, i=1, \dots, m), Z_{t-j}, j=1, \dots, m)} \\ & \Rightarrow \sum_{k=1}^{\infty} \mathcal{E}_{(Y_{k,1}, \Gamma_k^{-1/\alpha} \mathbf{e}_1, \infty, Y'_{k,1}, \dots, Y'_{k,m-1})} \\ & + \sum_{k=1}^{\infty} \mathcal{E}_{(Y_{k,2}, \Gamma_k^{-1/\alpha} \mathbf{e}_2, Y_{k,1}, \infty, Y'_{k,1}, \dots, Y'_{k,m-2})} + \dots \\ & + \sum_{k=1}^{\infty} \mathcal{E}_{(Y_{k,m}, \Gamma_k^{-1/\alpha} \mathbf{e}_m, Y_{k,m-1}, \dots, Y_{k,1}, \infty)} \end{aligned}$$

in  $M_p(E)$ .

PROOF. The proof of (3.3) is based on the following two simple results. For  $y, a, b \geq 0$ , since  $b_n \rightarrow \infty$ ,

$$nP[b_n^{-1}Z_t > y, Z_t \in [a, b]] \rightarrow \begin{cases} 0, & \text{if } b < \infty, \\ y^{-\alpha}, & \text{if } b = \infty, \end{cases}$$

and, for  $x > 0, y \geq 0$ ,

$$nP[b_n^{-1}Z_t \geq x, b_n^{-1}Z_{t-1} \geq y] \rightarrow \begin{cases} x^{-\alpha}, & \text{if } y = 0, \\ 0, & \text{if } y > 0, \end{cases}$$

and furthermore we have checked the vague convergence of

$$nP\left[\left(\frac{Z_t}{b_n}, Z_t\right) \in \cdot\right] \rightarrow_v \nu \times \varepsilon_\infty$$

as measures on  $(0, \infty] \times [0, \infty]$ . To prove (3.5), we let  $\rho$  be the vague metric on  $M_p(E)$ . Then we can show that

$$(3.6) \quad \rho\left(\sum_{t=1}^n \mathcal{E}_{(Z_t, b_n^{-1}(Z_{t-1}, Z_{t-2}, \dots, Z_{t-m}), Z_{t-1}, Z_{t-2}, \dots, Z_{t-m})}, \sum_{t=1}^n \sum_{i=1}^m \mathcal{E}_{(Z_t, b_n^{-1}Z_{t-i} \mathbf{e}_i, Z_{t-1}, Z_{t-2}, \dots, Z_{t-m})}\right) \rightarrow_P 0$$

and

$$(3.7) \quad \rho\left(\sum_{t=1}^n \sum_{i=1}^m \mathcal{E}_{(Z_t, b_n^{-1}Z_{t-i} \mathbf{e}_i, Z_{t-1}, Z_{t-2}, \dots, Z_{t-m})}, \sum_{t=1}^n \sum_{i=1}^m \mathcal{E}_{(Z_{t+i}, b_n^{-1}Z_t \mathbf{e}_i, Z_{t+i-1}, Z_{t+i-2}, \dots, Z_{t+i-m})}\right) \rightarrow_P 0.$$

The proof of (3.6) is almost exactly the same as that of Proposition 4.26 of Resnick (1987) and rests on the fact that, for any  $\delta > 0$ ,

$$I_n := \sum_{t=1}^n \varepsilon_{(Z_t, b_n^{-1}(Z_{t-1}, Z_{t-2}, \dots, Z_{t-m}), Z_{t-1}, Z_{t-2}, \dots, Z_{t-m})}$$

has the property that

$$\begin{aligned} EI_n & \left( \bigcup_{1 \leq i < j \leq m} \{(y_0, x_1, \dots, x_m, y_1, \dots, y_m) : x_i > \delta, x_j > \delta\} \right) \\ & \leq \binom{m}{2} n P[Z_1 > \delta b_n, Z_2 > \delta b_n], \end{aligned}$$

which tends to 0 as  $n \rightarrow \infty$  [cf. Feigin and Resnick (1992), (3.37)]. The proof of (3.7) is identical to that of (3.40) in Feigin and Resnick (1992). From (3.7) and (3.6) we see that, to prove (3.5), it suffices to show that the rightmost point process in (3.7) converges to the limit in (3.5). Toward showing this, we assert that

$$\begin{aligned} & \sum_{t=1}^n \varepsilon_{(Z_{t+m}, \dots, Z_{t+1}, b_n^{-1}Z_t, Z_t, Z_{t-1}, \dots, Z_{t-m+1})} \\ & \Rightarrow \sum_{k \geq 1} \varepsilon_{(Y_{k,m}, \dots, Y_{k,1}, \Gamma_k^{-1/\alpha}, \infty, Y'_{k,1}, \dots, Y'_{k,m-1})}, \end{aligned}$$

where  $\{Y'_{k,i}, Y_{k,j}, k \geq 1, 1 \leq i, j \leq m\}$  are iid random variables with the distribution of  $Z_1$  and independent of  $\{\Gamma_k\}$ . To see this, let

$$X_{n,t} = \left( Z_{t+m}, \dots, Z_{t+1}, \frac{Z_t}{b_n}, Z_t, Z_{t-1}, \dots, Z_{t-m+1} \right)$$

and observe that  $\{X_{n,t}, -\infty < t < \infty\}$  satisfies:

- (i)  $\{X_{n,t}, -\infty < t < \infty\}$  is stationary and  $2m$ -dependent.
- (ii) We have

$$\begin{aligned} nP[X_{n,1} \in (dz_m, \dots, dz_1, dx, dy, du_1, \dots, du_{m-1})] \\ \rightarrow_v \prod_{i=1}^m F(dz_i) \nu(dx) \varepsilon_\infty(dy) \prod_{j=1}^{m-1} F(du_j) \end{aligned}$$

on  $[0, \infty]^m \times (0, \infty] \times [0, \infty]^m := E'$ .

- (iii) For  $g \in C_K^+(E')$ ,

$$\lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{t=2}^{\lfloor n/k \rfloor} Eg(X_{n,1})g(X_{n,t}) = 0.$$

The desired result then follows from Theorem 2.1 in Davis and Resnick (1988).

By projecting, we then get in  $(M_p([0, \infty] \times (0, \infty] \times [0, \infty]^{m-1}))^m$  that

$$\begin{aligned} & \left( \sum_{t=1}^n \mathcal{E}_{(Z_{t+1}, b_n^{-1}Z_t, Z_t, \dots, Z_{t+1-m})}, \sum_{t=1}^n \mathcal{E}_{(Z_{t+2}, b_n^{-1}Z_t, Z_{t+1}, \dots, Z_{t+2-m})}, \dots, \right. \\ & \qquad \qquad \qquad \left. \sum_{t=1}^n \mathcal{E}_{(Z_{t+m}, b_n^{-1}Z_t, Z_{t+m-1}, \dots, Z_t)} \right) \\ & \Rightarrow \left( \sum_k \mathcal{E}_{(Y_{k,1}, \Gamma_k^{-1/\alpha}, \infty, Y'_{k,1}, \dots, Y'_{k,m-1})}, \sum_k \mathcal{E}_{(Y_{k,2}, \Gamma_k^{-1/\alpha}, Y_{k,1}, \infty, Y'_{k,1}, \dots, Y'_{k,m-2})}, \dots, \right. \\ & \qquad \qquad \qquad \left. \sum_k \mathcal{E}_{(Y_{k,m}, \Gamma_k^{-1/\alpha}, Y_{k,m-1}, \dots, Y_{k,1}, \infty)} \right). \end{aligned}$$

Hence, by using a mapping argument, we have

$$\begin{aligned} & \left( \sum_{t=1}^n \mathcal{E}_{(Z_{t+i}, b_n^{-1}Z_t \mathbf{e}_i, Z_{t+i-1}, Z_{t+i-2}, \dots, Z_{t+i-m})}; i = 1, \dots, m \right) \\ & \Rightarrow \left( \sum_k \mathcal{E}_{(Y_{k,i}, \Gamma_k^{-1/\alpha} \mathbf{e}_i, Y_{k,i-1}, \dots, Y_{k,1}, \infty, Y'_{k,1}, \dots, Y_{k,m-i})}; i = 1, \dots, m \right). \end{aligned}$$

Since addition is vaguely continuous we finally obtain that the rightmost point process in (3.7) converges weakly to the limit in (3.5) and therefore, by using (3.6) and (3.7), we deduce the result of the proposition.  $\square$

Before proceeding with the proof of Theorem 3.1 based on Proposition 3.2, we state some preliminaries. Suppose  $E' \subset E$  and give  $E'$  the relative topology inherited from  $E$ . The compact subsets of  $E'$  are those subsets  $K' \subset E'$  such that  $K'$  is compact when considered as a subset of  $E$ . To see this, suppose  $K'$  is compact in  $E'$ . Suppose  $O_\beta, \beta \in A$ , is an open covering of  $K'$  in  $E$ , so that  $K' \subset \bigcup_{\beta \in A} O_\beta$ . Since  $K' \subset E'$ , we also have

$$K' \subset \left( \bigcup_{\beta \in A} O_\beta \right) \cap E' = \bigcup_{\beta \in A} (O_\beta \cap E').$$

Since  $O_\beta \cap E'$  is open in  $E'$  and  $K'$  is compact in  $E'$ , we have

$$K' \subset \bigcup_{\beta \in I} (O_\beta \cap E') \subset \bigcup_{\beta \in I} O_\beta,$$

where  $I$  is a finite index set. Thus  $K'$  is compact in  $E$ . The converse is similar.

**PROPOSITION 3.3.** *Suppose  $E'$  is a measurable subset of  $E$  and give  $E'$  the relative topology inherited from  $E$ . For a set  $B \subset E'$  denote by  $\partial_{E'} B$  the boundary of  $B$  in  $E'$  and denote by  $\partial_E B$  the boundary of  $B$  in  $E$ .*

(a) Define

$$\hat{T}: M_p(E) \mapsto M_p(E')$$

by

$$\hat{T}m = m(\cdot \cap E').$$

If  $m \in M_p(E)$  and  $m(\partial_E E') = 0$ , then  $\hat{T}$  is continuous at  $m$  so that if  $m_n \rightarrow_v m$  in  $M_p(E)$ , then  $\hat{T}m_n \rightarrow_v \hat{T}m$  in  $M_p(E')$ .

(b) The same conclusion holds if we define  $\hat{T}$  the same way but consider it as a mapping

$$\hat{T}: M_p(E) \mapsto M_p(E).$$

(c) Conversely, suppose  $m_n \in M_p(E)$  for  $n \geq 0$  and that  $m_n \rightarrow_v m_0$  in  $M_p(E')$ . If

$$m_n((E')^c) = 0, \quad n \geq 0,$$

and  $m(\partial_E E') = 0$ , then  $m_n \rightarrow_v m_0$  in  $M_p(E)$  as well.

PROOF. (a) Suppose  $B \subset E'$  is relatively compact in  $E'$  and  $m(\partial_E B) = 0$ . It suffices to show  $m_n(B) \rightarrow m(B)$  [Resnick (1987), page 142]. Since  $m_n \rightarrow_v m$  in  $M_p(E)$ , it suffices to show that  $m(\partial_E B) = 0$ . One can readily check the inclusion

$$(3.8) \quad \partial_E B \subset \partial_{E'} B \cup \partial_E E'$$

for  $B \subset E'$ . Thus

$$m(\partial_E B) \leq m(\partial_{E'} B) + m(\partial_E E') = 0.$$

The proofs of (b) and (c) are very similar except one needs the inclusion

$$(3.9) \quad \partial_E (B \cap E') \subset (\partial_E B) \cap \overline{E'} \cup \partial_E E'. \quad \square$$

The following simple result allows us to discard components in a point process convergence result.

LEMMA 3.4. Suppose  $E_i$ ,  $i = 1, 2$ , are locally compact Hausdorff topological spaces and that  $E_2$  is compact. If  $m_n \in M_p(E_1 \times E_2)$  for  $n \geq 0$  and  $m_n \rightarrow_v m_0$  in  $M_p(E_1 \times E_2)$ , then

$$m_n(\cdot \times E_2) \rightarrow_v m_0(\cdot \times E_2)$$

in  $M_p(E_1)$ .

PROOF. Let  $f_1 \in C_K^+(E_1)$  and define  $f: E_1 \times E_2 \mapsto \mathbb{R}_+$  by  $f(x, y) = f_1(x)$ . The support of  $f$  is contained in  $K_1 \times E_2$ , where  $K_1$  is the compact support of  $f_1$ . Thus the support of  $f$  is compact and  $m_n(f) \rightarrow m_0(f)$  which translates to

$$\int_{E_1} f_1(x) m_n(dx \times E_2) \rightarrow \int_{E_1} f_1(x) m_0(dx \times E_2),$$

which is equivalent to the desired convergence in  $M_p(E_1)$ .  $\square$

We now proceed to prove Theorem 3.1. From Proposition 3.2 and Lemma 3.4 we have

$$\begin{aligned}
 N_n &:= \sum_{t=1}^n \mathcal{E}_{(Z_t, b_n^{-1}(Z_{t-1}, Z_{t-2}, Z_{t-3}), Z_{t-1}, Z_{t-2}, Z_{t-3})} \\
 &\Rightarrow N_\infty := \sum_{k=1}^\infty \mathcal{E}_{(Y_{k,1}, \Gamma_k^{-1/\alpha}, 0, 0, \infty, Y_{k,1}, Y_{k,2})} \\
 &\quad + \sum_{k=1}^\infty \mathcal{E}_{(Y_{k,2}, 0, \Gamma_k^{-1/\alpha}, 0, Y_{k,1}, \infty, Y_{k,1})} \\
 &\quad + \sum_{k=1}^\infty \mathcal{E}_{(Y_{k,3}, 0, 0, \Gamma_k^{-1/\alpha}, Y_{k,2}, Y_{k,1}, \infty)}
 \end{aligned}
 \tag{3.10}$$

in  $M_p(E)$  where, recall,  $E := [0, \infty] \times ([0, \infty]^3 \setminus \{\mathbf{0}\}) \times [0, \infty]^3$ .

Referring back to the end of Section 2 and (2.5), recall that we are interested only in the case where  $p_t(1 - \theta^{(0)}) < 0$ . However, it is initially easier to deal with the restriction of the point process convergence in (3.10) for the case that  $B_t < 0$  which is implied by  $p_t(1 - \theta^{(0)}) < 0$ . So we define the region  $[B < 0] \subset E$  by

$$\begin{aligned}
 [B < 0] &:= \left\{ (x_0, \dots, x_6) \in E : x_4 > \theta^{(0)}x_5 + 2(\theta^{(0)})^2 x_6, \right. \\
 &\quad \left. x_1 \neq \theta^{(0)}x_2 + 2(\theta^{(0)})^2 x_3 \right\} \\
 &= \left\{ (x_0, \dots, x_6) \in [0, \infty] \times ((0, \infty] \times [0, \infty])^2 : \right. \\
 &\quad \left. x_4 > \theta^{(0)}x_5 + 2(\theta^{(0)})^2 x_6, x_1 > \theta^{(0)}x_2 + 2(\theta^{(0)})^2 x_3 \right\} \\
 &\quad \cup \left\{ (x_0, \dots, x_6) \in [0, \infty] \times [0, \infty] \times ([0, \infty]^2 \setminus \{(0, 0)\}) \times (0, \infty] \right. \\
 &\quad \left. \times [0, \infty]^2 : x_4 > \theta^{(0)}x_5 + 2(\theta^{(0)})^2 x_6, x_1 < \theta^{(0)}x_2 + 2(\theta^{(0)})^2 x_3 \right\} \\
 &= [B < 0]_> \cup [B < 0]_<.
 \end{aligned}$$

Note that for  $b > 0, c > 0, b_1 > 0, b_4 > 0, a_j < \infty, j = 2, 3, 5, 6$ , the set

$$\begin{aligned}
 &\left\{ (x_0, \dots, x_6) \in [0, \infty] \times ((0, \infty] \times [0, \infty])^2 : \right. \\
 &\quad \left. x_4 > \theta^{(0)}x_5 + 2(\theta^{(0)})^2 x_6, x_1 > \theta^{(0)}x_2 + 2(\theta^{(0)})^2 x_3, \right. \\
 &\quad \left. x_i \geq b_i, i = 1, 4; x_j \leq a_j, j = 2, 3, 5, 6 \right\}
 \end{aligned}$$

is a compact subset of  $[0, \infty] \times ((0, \infty] \times [0, \infty])^2$ . Define the mappings

$$\begin{aligned}
 T_1(x_0, \dots, x_6) &:= x_0 + (\theta^{(0)})^3 x_6, \\
 T_2(x_0, \dots, x_6) &:= |-x_1 + \theta^{(0)}x_2 + 2(\theta^{(0)})^2 x_3|, \\
 T(x_0, \dots, x_6) &:= (T_1(x_0, \dots, x_6), T_2(x_0, \dots, x_6), x_4, x_5, x_6)
 \end{aligned}
 \tag{3.11}$$



and think of  $T$  as

$$\begin{aligned} T: [B < 0] &\mapsto E' := [0, \infty] \times (0, \infty) \\ &\quad \times \left\{ (x_4, x_5, x_6) \in [0, \infty]^3: x_4 > \theta^{(0)}x_5 + 2(\theta^{(0)})^2 x_6 \right\} \\ &= [0, \infty] \times (0, \infty) \times \left\{ (x_4, x_5, x_6) \in (0, \infty] \times [0, \infty)^2: \right. \\ &\quad \left. x_4 > \theta^{(0)}x_5 + 2(\theta^{(0)})^2 x_6 \right\}. \end{aligned}$$

Note that

$$\begin{aligned} T(Z_t, b_n^{-1}(Z_{t-1}, Z_{t-2}, Z_{t-3}), Z_{t-1}, Z_{t-2}, Z_{t-3}) \\ = \left( C_t, \frac{|B_t|}{b_n}, Z_{t-1}, Z_{t-2}, Z_{t-3} \right). \end{aligned}$$

We seek to show

$$\begin{aligned} (3.12) \quad &\sum_{t=1}^n \mathbf{1}_{[B_t < 0]} \mathcal{E}_{(C_t, b_n^{-1}|B_t|, Z_{t-1}, Z_{t-2}, Z_{t-3})} \\ &\Rightarrow N_\infty^\# := \sum_{k=1}^{\infty} \mathcal{E}_{(Y_{k,1} + (\theta^{(0)})^3 Y'_{k,2}, \Gamma_k^{-1/\alpha}, \infty, Y'_{k,1}, Y'_{k,2})} \end{aligned}$$

in  $M_p([0, \infty] \times (0, \infty) \times [0, \infty]^3)$ .

Since

$$P[N_\infty(\partial_E([B < 0])) = 0] = 1,$$

we obtain from Proposition 3.3 that if we take restrictions to  $[B < 0]$  in (3.10), we get

$$(3.13) \quad N_n|_{[B < 0]} \Rightarrow N_\infty|_{[B < 0]} = \sum_{k=1}^{\infty} \mathcal{E}_{(Y_{k,1}, \Gamma_k^{-1/\alpha}, 0, 0, \infty, Y'_{k,1}, Y'_{k,2})}$$

in  $M_p([B < 0])$ . It is tempting to try to apply Proposition 3.18, page 148, of Resnick (1987). To apply this theorem, we need to check that  $T$  is continuous (no problem) and that  $T^{-1}$  maps compact sets into compact sets (problem). This last compactness property fails, so to get around the problem, truncation of the domain is necessary. For  $M > 0$ , let

$$\begin{aligned} [B < 0]_M &= [B < 0]_> \cap \{(x_0, \dots, x_6): x_1 \geq M^{-1}, x_2 \vee x_3 \leq M\} \\ &\quad \cup [B < 0]_< \cap \{(x_0, \dots, x_6): x_1 \leq M, x_2 \vee x_3 \geq M^{-1}\}. \end{aligned}$$

From (3.13) we get by restriction

$$N_n|_{[B < 0]_M} \Rightarrow N_\infty|_{[B < 0]_M}$$

in  $M_p([B < 0]_M)$ . Considered as a mapping on  $[B < 0]_m$ ,  $T^{-1}$  maps compacta into compacta. For instance, for  $a > 0, b, b', c > 0, a_i < \infty, i = 5, 6$ ,

$$\begin{aligned} & T^{-1}\left([0, a] \times [b, \infty] \times \left\{(x_4, x_5, x_6): x_4 \geq b', x_i \leq a_i, i = 5, 6; \right. \right. \\ & \qquad \qquad \qquad \left. \left. x_4 \geq \theta^{(0)}x_5 + 2(\theta^{(0)})^2 x_6 + c\right\}\right) \\ &= \left\{(x_0, \dots, x_6) \in [0, \infty] \times ((0, \infty] \times [0, \infty)^2)^2: \right. \\ & \quad x_4 \geq b', x_i \leq a_i, i = 5, 6; x_4 \geq \theta^{(0)}x_5 + 2(\theta^{(0)})^2 x_6 + c, \\ & \quad \left. x_0 + (\theta^{(0)})^3 x_6 \leq a, x_2 \vee x_3 \leq M, x_1 \geq \theta^{(0)}x_2 + 2(\theta^{(0)})^2 x_3 + b\right\} \end{aligned}$$

is compact in  $[0, \infty] \times ((0, \infty] \times [0, \infty)^2)^2$  and hence in  $[B < 0]_M$ . Thus we conclude from Proposition 3.13, page 148, Resnick (1987) that

$$N_n|_{[B < 0]_M} \circ T^{-1} \Rightarrow N_\infty|_{[B < 0]_M} \circ T^{-1}$$

in  $M_p(E')$ . Part (c) of Proposition 3.3 then implies

$$\begin{aligned} N_{n, M} &:= \sum_{t=1}^n \mathbf{1}_{\{|B_t < 0\} \cap (\{Z_{t-1} > b_n M^{-1}, Z_{t-2} \vee Z_{t-3} \leq b_n M\} \cup \{Z_{t-1} \leq b_n M, Z_{t-2} \vee Z_{t-3} > M^{-1} b_n\})\}} \\ & \quad \times \mathcal{E}_{(C_t, b_n^{-1}|B_t|, Z_{t-1}, Z_{t-2}, Z_{t-3})} \\ &\Rightarrow N_\infty^\# \end{aligned}$$

in  $M_p([0, \infty] \times (0, \infty] \times [0, \infty]^3)$ . From Billingsley (1968), Theorem 4.2, to show (3.12), we need to verify that, for any  $\eta > 0$ ,

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} P[\rho(N_{n, M}, N_n) > \eta] = 0.$$

Let  $g \in C_K^+([0, \infty] \times (0, \infty] \times [0, \infty]^3)$  and suppose the support of  $g$  is in  $[0, \infty] \times [b, \infty] \times [0, \infty]^3$ . It suffices to show that, for all  $\eta > 0$ ,

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} P[|N_{n, M}(g) - N_n(g)| > \eta] = 0.$$

Now

$$\begin{aligned} & |N_{n, M}(g) - N_n(g)| \\ & \leq \sum_{t=1}^n \mathbf{1}_{\{|B_t < 0\} \cap (\{Z_{t-1} > b_n M^{-1}, Z_{t-2} \vee Z_{t-3} \leq b_n M\} \cup \{Z_{t-1} \leq b_n M, Z_{t-2} \vee Z_{t-3} > b_n M^{-1}\})^c\}} \\ & \quad \times g(C_t, b_n^{-1}|B_t|, Z_{t-1}, Z_{t-2}, Z_{t-3}) \end{aligned}$$

and so

$$\begin{aligned} & P[|N_{n, M}(g) - N_n(g)| > \eta] \\ & \leq nP\left\{[B_t < 0, |B_t|/b_n \geq b] \cap \left(\{Z_{t-1} > b_n M^{-1}, Z_{t-2} \vee Z_{t-3} \leq b_n M\} \right. \right. \\ & \qquad \qquad \qquad \left. \left. \cup \{Z_{t-1} \leq b_n M, Z_{t-2} \vee Z_{t-3} > b_n M^{-1}\}\right)^c\right\} \end{aligned}$$

and since  $B_t < 0$  and  $|B_t|/b_n \geq b$  imply  $Z_{t-1}/b_n > b$  we have the bound

$$\begin{aligned} &\leq nP\{[Z_{t-1}/b_n > b] \cap ([Z_{t-1} \leq b_n M^{-1}] \cup [Z_{t-2} \vee Z_{t-3} > b_n M])\} \\ &\leq nP[Z_{t-1} > b_n b, Z_{t-2} \vee Z_{t-3} > b_n M] \\ &= nP[Z_{t-1} > b_n b]P[Z_{t-2} \vee Z_{t-3} > b_n M]. \end{aligned}$$

As  $n \rightarrow \infty$  this last expression is asymptotic to

$$b^{-\alpha}P[Z_{t-2} \vee Z_{t-3} > b_n M],$$

which goes to 0 as  $n \rightarrow \infty$ .

Now corresponding to the condition  $p_t(1 - \theta^{(0)})$  define

$$\begin{aligned} NEG := \{ &(x_0, x_1, x_2, x_3, x_4) \in [0, \infty] \times (0, \infty] \times [0, \infty]^3: \\ &(x_3 + \theta^{(0)}x_4)(1 - \theta^{(0)})^2 + (-x_2 + \theta^{(0)}x_3 + 2(\theta^{(0)})^2 x_4)(1 - \theta^{(0)}) \\ &\quad + x_0 + (\theta^{(0)})^3 x_4 < 0\}. \end{aligned}$$

Since

$$P[N_\infty^\#(\partial(NEG)) = 0] = 1$$

and  $p_t(1 - \theta^{(0)}) < 0$  implies  $B_t < 0$ , we get from parts (a) and (c) of Proposition 3.3 that

$$\begin{aligned} (3.14) \quad &\sum_{t=1}^n \mathbf{1}_{[p_t(1-\theta^{(0)}) < 0]} \mathcal{E}_{(C_t, b_n^{-1}|B_t|, Z_{t-1}, Z_{t-2}, Z_{t-3})} \\ &\Rightarrow N_\infty^\# = \sum_{k=1}^{\infty} \mathcal{E}_{(Y_{k,1} + (\theta^{(0)})^3 Y_{k,2}, \Gamma_k^{-1/\alpha}, \infty, Y'_{k,1}, Y'_{k,2})}, \end{aligned}$$

and applying Lemma 3.4 gives

$$(3.15) \quad \sum_{t=1}^n \mathbf{1}_{[p_t(1-\theta^{(0)}) < 0]} \mathcal{E}_{(C_t, b_n^{-1}|B_t|)} \Rightarrow \sum_{k=1}^{\infty} \mathcal{E}_{(Y_{k,1} + (\theta^{(0)})^3 Y_{k,2}, \Gamma_k^{-1/\alpha})}$$

in  $M_p([0, \infty] \times (0, \infty])$ . Finally, another application of Proposition 3.3(a) gives the desired result

$$(3.16) \quad \sum_{t=1}^n \mathbf{1}_{[p_t(1-\theta^{(0)}) < 0]} \mathcal{E}_{(C_t, b_n^{-1}|B_t|)} \Rightarrow \sum_{k=1}^{\infty} \mathcal{E}_{(Y_{k,1} + (\theta^{(0)})^3 Y_{k,2}, \Gamma_k^{-1/\alpha})}$$

in  $M_p([0, \infty] \times (0, \infty])$ .

The next step is to show that we can take the ratio of the components in (3.16) and we show

$$\begin{aligned} (3.17) \quad \nu_{0,n} &:= \sum_{t=1}^n \mathbf{1}_{[p_t(1-\theta^{(0)}) < 0]} \mathcal{E}_{b_n C_t / |B_t|} \\ &\Rightarrow \nu_{0,\infty} := \sum_{k=1}^{\infty} \mathcal{E}_{\Gamma_k^{1/\alpha} (Y_{k,1} + (\theta^{(0)})^3 Y_{k,2})} \end{aligned}$$

in  $M_p([0, \infty))$ . The map

$$T_3: (u, v) \in [0, \infty) \times (0, \infty] \mapsto u/v \in [0, \infty),$$

although continuous, does not have the property that  $T_3^{-1}$  carries compacta into compacta and so a truncation of the domain must be done. For small  $\delta > 0$ , we restrict attention to  $[0, \delta^{-1}] \times [\delta, \infty]$  and then apply  $T_3$  to get, as  $n \rightarrow \infty$ ,

$$\begin{aligned} \nu_{\delta, n} &:= \sum_{t=1}^n \mathbf{1}_{[p_t(1-\theta^{(0)}) < 0, C_t \leq \delta^{-1}, b_n^{-1}|B_t| \geq \delta]} \mathcal{E} b_n C_t / |B_t| \\ &\Rightarrow \nu_{\delta, \infty} := \sum_k \mathbf{1}_{[Y_{k,1} + (\theta^{(0)})^3 Y'_{k,2} \leq \delta^{-1}, \Gamma_k^{-1/\alpha} \geq \delta]} \mathcal{E} \Gamma_k^{1/\alpha} (Y_{k,1} + (\theta^{(0)})^3 Y'_{k,2}) \end{aligned}$$

in  $M_p([0, \infty))$ . As  $\delta \rightarrow 0$  we have

$$\begin{aligned} \nu_{\delta, \infty} &:= \sum_k \mathbf{1}_{[Y_{k,1} + (\theta^{(0)})^3 Y'_{k,2} \leq \delta^{-1}, \Gamma_k^{-1/\alpha} \geq \delta]} \mathcal{E} \Gamma_k^{1/\alpha} (Y_{k,1} + (\theta^{(0)})^3 Y'_{k,2}) \\ &\Rightarrow \nu_{0, \infty} := \sum_k \mathcal{E} \Gamma_k^{1/\alpha} (Y_{k,1} + (\theta^{(0)})^3 Y'_{k,2}) \end{aligned}$$

in  $M_p([0, \infty))$ , which is the right-hand side of (3.17). To show (3.17), therefore, it suffices by Theorem 4.2 of Billingsley (1968) to prove that, for any  $\eta > 0$ ,

$$(3.18) \quad \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P[\rho(\nu_{\delta, n}, \nu_{0, n}) > \eta] = 0,$$

where  $\rho$  is the vague metric.

Let  $g \in C_K^+([0, \infty))$  and suppose the support of  $g$  is in  $[0, \alpha]$ . For (3.18) it suffices to show that

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P[|\nu_{\delta, n}(g) - \nu_{0, n}(g)| > \eta] = 0,$$

or, equivalently,

$$(3.19) \quad \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P \left[ \left| \sum_{t=1}^n g \left( b_n \frac{C_t}{|B_t|} \right) \mathbf{1}_{[|B_t|/b_n \geq \delta, C_t \leq \delta^{-1}, p_t(1-\theta^{(0)}) < 0]} - \sum_{t=1}^n g \left( b_n \frac{C_t}{|B_t|} \right) \mathbf{1}_{[p_t(1-\theta^{(0)}) < 0]} \right| > \eta \right] = 0.$$

The absolute value of the difference between the two sums is bounded by

$$\begin{aligned} &\sum_{t=1}^n g \left( b_n \frac{C_t}{|B_t|} \right) \mathbf{1}_{[p_t(1-\theta^{(0)}) < 0]} \mathbf{1}_{[|B_t| \geq \delta b_n, C_t \leq \delta^{-1}]^c} \\ &\leq \sum_{t=1}^n g \left( b_n \frac{C_t}{|B_t|} \right) \mathbf{1}_{[B_t < 0] \cap [ |B_t| \geq \delta b_n, C_t \leq \delta^{-1}]^c} \\ &\leq \sum_{t=1}^n g \left( b_n \frac{C_t}{|B_t|} \right) \mathbf{1}_{[B_t < 0, |B_t| \leq \delta b_n, C_t \leq \delta^{-1}]} \\ &\quad + \sum_{t=1}^n g \left( b_n \frac{C_t}{|B_t|} \right) \mathbf{1}_{[B_t < 0, C_t > \delta^{-1}]}. \end{aligned}$$

Taking into account the compact support of  $g$ , the probability in (3.19) is thus bounded by

$$\begin{aligned} & P\left\{\bigcup_{t=1}^n \left[ b_n \frac{C_t}{|B_t|} \leq a, B_t < 0, \frac{|B_t|}{b_n} \leq \delta, C_t \leq \delta^{-1} \right]\right\} \\ & + P\left\{\bigcup_{t=1}^n \left[ b_n \frac{C_t}{|B_t|} \leq a, B_t < 0, C_t > \delta^{-1} \right]\right\} \\ & \leq nP\left[ b_n \frac{C_t}{|B_t|} \leq a, B_t < 0, \frac{|B_t|}{b_n} \leq \delta, C_t \leq \delta^{-1} \right] \\ & \quad + nP\left[ b_n \frac{C_t}{|B_t|} \leq a, B_t < 0, C_t > \delta^{-1} \right] \\ & = I + II. \end{aligned}$$

Now  $II$  is bounded as follows:

$$II \leq nP\left[ \frac{\delta^{-1}}{a} \leq \frac{|B_t|}{b_n}, B_t < 0 \right],$$

and because  $B_t < 0$  implies  $Z_{t-1} > \theta^{(0)}Z_{t-2} + 2(\theta^{(0)})^3Z_{t-3}$ , the previous expression is dominated by

$$\begin{aligned} nP\left[ \frac{\delta^{-1}}{a} \leq \frac{Z_{t-1}}{b_n} \right] & \rightarrow \left( \frac{\delta^{-1}}{a} \right)^{-\alpha}, \quad n \rightarrow \infty, \\ & = \delta^\alpha a^\alpha \rightarrow 0, \quad \delta \rightarrow 0. \end{aligned}$$

To bound  $I$ , observe that for arbitrarily small  $\omega < \delta^{-1}$  we have

$$\begin{aligned} I & \leq nP\left[ b_n \frac{C_t}{|B_t|} \leq a, B_t < 0, \frac{|B_t|}{b_n} \leq \delta, C_t \leq \omega \right] \\ & \quad + nP\left[ b_n \frac{C_t}{|B_t|} \leq a, B_t < 0, C_t > \omega \right] \\ & = Ia + Ib. \end{aligned}$$

Now  $Ib$  has a double limit which is 0 by the argument that handled  $II$  so we concentrate on  $Ia$ . From the definition of  $C_t$  we have  $C_t \geq Z_t$  and  $C_t \geq (\theta^{(0)})^3Z_{t-3}$ . Recall that  $Z_{t-1} \geq \theta^{(0)}Z_{t-2} + 2(\theta^{(0)})^3Z_{t-3}$  when  $B_t < 0$ . Thus  $Ia$  is bounded by

$$\begin{aligned} Ia & \leq nP\left[ a^{-1} \leq (Z_t^{-1}) \frac{|B_t|}{b_n}, B_t < 0, C_t \leq \omega \right] \\ & \leq nP\left[ a^{-1} \leq (Z_t^{-1}) \frac{Z_{t-1}}{b_n}, (\theta^{(0)})^3Z_{t-3} \leq \omega \right] \\ & \rightarrow E(Z_t)^{-\alpha} a^\alpha P\left[ (\theta^{(0)})^3Z_{t-3} \leq \omega \right], \quad n \rightarrow \infty, \end{aligned}$$

since, by Condition R,  $E(Z_t)^{-\beta} < \infty$  for some  $\beta > \alpha$  and thus

$$\lim_{n \rightarrow \infty} \limsup_{\delta \rightarrow 0} I\alpha \leq (\text{constant}) P\left[(\theta^{(0)})^3 Z_{t-3} \leq \omega\right]$$

and since  $\omega$  can be picked arbitrarily small, the double limit must be 0 as desired. This completes the verification of (3.17).

From (3.13) a standard argument mapping the points of  $\nu_{0,n}$  into the minimum [see, e.g., Resnick (1987), page 214] yields

$$(3.20) \quad b_n \bigwedge_{\substack{1 \leq t \leq n \\ p_t(1-\theta^{(0)}) < 0}} \frac{C_t}{|B_t|} \Rightarrow \bigwedge_{k=1}^{\infty} \Gamma_k^{1/\alpha} \left( Y_{k,1} + (\theta^{(0)})^3 Y'_{k,2} \right).$$

The rest of the proof consists of showing that, in fact,  $b_n(\hat{\theta} - \theta^{(0)})$  has the same limit distribution as the limit random variable in (3.20).

To do this, recall the outline presented at the end of Section 2. When  $p_t(1 - \theta^{(0)}) < 0$ , we approximate the polynomial  $p_t(\eta)$  with two lines

$$L_1: B_t \eta + C_t, \quad L_2: (A_t(1 - \theta^{(0)}) + B_t) \eta + C_t,$$

which have roots

$$x_{t1} = \frac{C_t}{|B_t|}, \quad x_{t2} = \frac{C_t}{|B_t| - (1 - \theta^{(0)}) A_t}$$

and therefore the smallest root  $r_t^-$  of the random polynomial satisfies

$$(3.21) \quad x_{t1} \leq r_t^- \leq x_{t2}.$$

We know from (3.13) that

$$\sum_{t=1}^n \mathbf{1}_{[p_t(1-\theta^{(0)}) < 0]} \mathcal{E}_{b_n x_{t1}} \Rightarrow \nu_{0,\infty} := \sum_k \mathcal{E}_{\Gamma_k^{1/\alpha} (Y_{k,1} + (\theta^{(0)})^3 Y'_{k,2})}$$

in  $M_p([0, \infty))$ , and we now propose to show that the same holds true with  $x_{t2}$  replacing  $x_{t1}$  and we show

$$(3.22) \quad \rho \left( \sum_{t=1}^n \mathbf{1}_{[p_t(1-\theta^{(0)}) < 0]} \mathcal{E}_{b_n x_{t1}}, \sum_{t=1}^n \mathbf{1}_{[p_t(1-\theta^{(0)}) < 0]} \mathcal{E}_{b_n x_{t2}} \right) \rightarrow_P 0$$

as  $n \rightarrow \infty$ . Suppose  $f \in C_k^+([0, \infty))$  with support in  $[0, k]$ . It is enough to show, for any  $\eta$ ,

$$\lim_{n \rightarrow \infty} P \left[ \sum_{t=1}^n |f(b_n x_{t1}) - f(b_n x_{t2})| \mathbf{1}_{[p_t(1-\theta^{(0)}) < 0]} > \eta \right] = 0.$$

For any small  $\delta > 0$ , this probability is bounded by

$$\begin{aligned} & P \left[ \sum_{t=1}^n |f(b_n x_{t1}) - f(b_n x_{t2})| \mathbf{1}_{[B_t < 0]} > \eta \right] \\ & \leq P \left[ \sum_{t=1}^n |f(b_n x_{t1}) - f(b_n x_{t2})| \mathbf{1}_{[B_t < 0, b_n x_{t2} > k + \delta, b_n |x_{t2} - x_{t1}| < \delta]} > \eta/3 \right] \\ & \quad + P \left[ \sum_{t=1}^n |f(b_n x_{t1}) - f(b_n x_{t2})| \mathbf{1}_{[B_t < 0, b_n x_{t2} > k + \delta, b_n |x_{t2} - x_{t1}| \geq \delta]} > \eta/3 \right] \\ & \quad + P \left[ \sum_{t=1}^n |f(b_n x_{t1}) - f(b_n x_{t2})| \mathbf{1}_{[B_t < 0, b_n x_{t2} \leq k + \delta]} > \eta/3 \right] \\ & = I + II + III. \end{aligned}$$

Note that  $I$  is 0 because both arguments of  $f$  are outside the support of  $f$ .

For  $II$  we have the bounds

$$\begin{aligned} II & \leq P \left[ \sum_{t=1}^n f(b_n x_{t1}) \mathbf{1}_{[b_n x_{t1} \leq k, B_t < 0, |x_{t1} - x_{t2}| \geq \delta]} > \eta/3 \right] \\ & \leq P \left\{ \bigcup_{t=1}^n [b_n x_{t1} \leq k, B_t < 0, b_n |x_{t1} - x_{t2}| \geq \delta] \right\} \\ & \leq nP [b_n x_{t1} \leq k, B_t < 0, b_n |x_{t1} - x_{t2}| \geq \delta]. \end{aligned}$$

Now

$$b_n |x_{t1} - x_{t2}| = b_n x_{t1} \left| 1 - \frac{1}{1 - (1 - \theta^{(0)}) A_t / |B_t|} \right|$$

and on the set  $[b_n x_{t1} \leq k]$  this difference is bounded by

$$k \left| 1 - \frac{1}{1 - (1 - \theta^{(0)}) A_t / |B_t|} \right|.$$

Thus, for some  $\delta' > 0$ ,

$$\begin{aligned} & [B_t < 0, b_n x_{t1} \leq k, |x_{t1} - x_{t2}| \geq \delta] \\ & \subset \left[ B_t < 0, b_n x_{t1} \leq k, \frac{A_t}{|B_t|} > \delta' \right]. \end{aligned}$$

On  $[b_n x_{t1} \leq k]$ , we have

$$\frac{1}{|B_t|} \leq \frac{k}{b_n C_t},$$

which implies

$$\frac{A_t}{|B_t|} \leq k \frac{A_t}{b_n C_t}.$$

Thus, using the definitions of  $A_t, C_t$ , we have the bound for some new  $\delta''$ :

$$\begin{aligned} II &\leq nP \left[ B_t < 0, \frac{Z_{t-2} + \theta^{(0)}Z_{t-3}}{b_n(Z_t + (\theta^{(0)})^3 Z_{t-3})} > \delta'' \right] \\ &\leq nP \left[ B_t < 0, \frac{Z_{t-2} + \theta^{(0)}Z_{t-3}}{b_n Z_t} > \delta'', \frac{Z_{t-2} + \theta^{(0)}Z_{t-3}}{b_n(\theta^{(0)})^3 Z_{t-3}} > \delta'' \right]. \end{aligned}$$

On the set  $[B_t < 0]$ , we have  $(\theta^{(0)})^{-1}Z_{t-1} \geq Z_{t-2} + \theta^{(0)}Z_{t-3}$  and so

$$\begin{aligned} II &\leq nP \left[ \frac{Z_{t-1}}{b_n \theta^{(0)} Z_t} > \delta'' \frac{Z_{t-2} + \theta^{(0)}Z_{t-3}}{b_n(\theta^{(0)})^3 Z_{t-3}} > \delta'' \right] \\ &\leq nP \left[ \frac{Z_{t-1}}{b_n \theta^{(0)} Z_t} > \delta'' \right] P \left[ \frac{Z_{t-2} + \theta^{(0)}Z_{t-3}}{b_n(\theta^{(0)})^3 Z_{t-3}} > \delta'' \right] \\ &\sim (\text{const})(\delta'')^{-\alpha} P \left[ \frac{Z_{t-2} + \theta^{(0)}Z_{t-3}}{b_n(\theta^{(0)})^3 Z_{t-3}} > \delta'' \right] \end{aligned}$$

[by a result of Breiman (1965) which is applicable from the second part of Condition R]

$$\rightarrow 0, \quad n \rightarrow \infty,$$

which shows  $II \rightarrow 0$ .

Observe that what was proven in the treatment of  $II$  is that, for any constants  $k > 0$  and  $\delta > 0$ ,

$$(3.23) \quad \lim_{n \rightarrow \infty} nP[B_t < 0, b_n|x_{t1} - x_{t2}| > \delta, b_n x_{t1} \leq k] = 0.$$

For  $III$  we note that

$$\begin{aligned} &\sum_{t=1}^n |f(b_n x_{t1}) - f(b_n x_{t2})| \mathbf{1}_{[B_t < 0, b_n x_{t2} \leq k + \delta]} \\ &\leq \sum_{t=1}^n |f(b_n x_{t1}) - f(b_n x_{t2})| \mathbf{1}_{[B_t < 0, b_n x_{t2} \leq k + \delta, b_n|x_{t2} - x_{t1}| \leq \delta]} \\ &\quad + \sum_{t=1}^n |f(b_n x_{t1}) - f(b_n x_{t2})| \mathbf{1}_{[B_t < 0, b_n x_{t2} \leq k + \delta, b_n|x_{t2} - x_{t1}| > \delta]} \\ &= S_1 + S_2. \end{aligned}$$

So

$$III \leq P[S_1 > \eta/6] + P[S_2 > \eta/6] = IIIa + IIIb.$$

Letting  $\omega_f(\delta)$  be the modulus of continuity for  $f$ :

$$\omega_f(\delta) := \sup_{|x-y| \leq \delta} |f(x) - f(y)|,$$

we have for  $IIIa$  the bound

$$IIIa \leq P \left[ \omega_f(\delta) \sum_{t=1}^n \mathbf{1}_{[B_t < 0, b_n x_{t1} \leq k + 2\delta]} > \eta/6 \right],$$



and by the argument which showed (3.17), this probability converges to  $(n \rightarrow \infty)$

$$P[\omega_f(\delta)\nu_{0,\infty}([0, k + 2\delta]) > \eta/6],$$

which converges to 0 as  $\delta \rightarrow 0$  since  $\omega_f(\delta) \rightarrow 0$ .

For IIIb we have by (3.23) that

$$\begin{aligned} \text{IIIb} &\leq nP[B_t < 0, b_n x_{t_2} \leq k + \delta, b_n |x_{t_2} - x_{t_1}| > \delta, b_n x_{t_1} > 2k] + o(1) \\ &= o(1), \end{aligned}$$

since  $x_{t_1} < x_{t_2}$ .

The proof of Theorem 3.1 is concluded by recalling that, when  $p_t(1 - \theta^{(0)}) < 0$ ,

$$x_{t_1} \leq r_t^- \leq x_{t_2}$$

and

$$b_n(\hat{\theta} - \theta^{(0)}) = \bigwedge_{\substack{1 \leq t \leq n \\ p_t(1 - \theta^{(0)}) < 0}} b_n r_t^-.$$

Since

$$\bigwedge_{\substack{1 \leq t \leq n \\ p_t(1 - \theta^{(0)}) < 0}} b_n x_{t_1} \leq \bigwedge_{\substack{1 \leq t \leq n \\ p_t(1 - \theta^{(0)}) < 0}} b_n r_t^- \leq \bigwedge_{\substack{1 \leq t \leq n \\ p_t(1 - \theta^{(0)}) < 0}} b_n x_{t_2}$$

and the two extremes converge to the same weak limit, the minimum in the middle converges to the same weak limit. The proof of Theorem 3.1 is complete.

**4. The limit distribution in the left-tail case for  $q = 1$ .** In this section we assume Conditions M and L hold. We continue to assume the order is  $q = 1$  and we present the limit distribution for  $\hat{\theta}$ . We will discuss the following theorem whose proof parallels that of Theorem 3.1 for the right-tail case.

**THEOREM 4.1.** *Suppose  $\{X_t\}$  is the MA(1) process given in (1.1) and that Conditions M and L hold. Suppose the true parameter is  $\theta^{(0)} \in (0, 1)$  and that  $F$ , the distribution of  $Z_1$ , is continuous. Let  $a(n)$  be the quantile function*

$$a(n) = F^{\leftarrow}(1/n).$$

*Note that  $a(n) \rightarrow 0$ . The estimator  $\hat{\theta}$  given in Section 2 has a Weibull limit distribution. In  $[0, \infty)$ ,*

$$(4.1) \quad a(\sqrt{n})^{-1}(\hat{\theta} - \theta^{(0)}) \Rightarrow \frac{(\theta^{(0)})^{3/2}}{c(\alpha)^{1/(2\alpha)}} \bigwedge_{\substack{1 \leq k < \infty \\ Y_{k,1} > Y_{k,2}}} \frac{\Gamma_k^{1/2\alpha}}{|Y_{k,1} - \theta^{(0)}Y_{k,2}|},$$

where  $\{Y_{k,1}, Y_{k,2}, k \geq 1\}$  are iid with common distribution  $F$  and

$$\Gamma_k = E_1 + \cdots + E_k, \quad k \geq 1,$$

is a sum of iid unit exponentially distributed random variables. The constant  $c(\alpha)$  is defined by the beta integral

$$c(\alpha) = \int_0^1 (1-s)^\alpha \alpha s^{\alpha-1} ds.$$

The limit distribution of  $\hat{\theta}$  is Weibull:

$$(4.2) \quad \begin{aligned} \lim_{n \rightarrow \infty} P \left[ a(\sqrt{n})^{-1} (\hat{\theta} - \theta^{(0)}) \leq x \right] \\ = 1 - \exp\{-kx^{2\alpha}\}, \quad x > 0, \end{aligned}$$

where

$$k = (\theta^{(0)})^{-3\alpha} c(\alpha) E \left( |Y_{k,1} - \theta^{(0)} Y_{k,2}|^{2\alpha} \mathbf{1}_{\{Y_{k,1} > Y_{k,2}\}} \right),$$

which is finite by Condition L. The convergence rate is  $1/a(\sqrt{n})$ .

REMARK. In the right-tail case the convergence rate was  $b_n$  which up to a slowly varying multiplicative factor is of order  $n^{1/\alpha}$ . However, under Condition L, the convergence rate is only  $1/a(\sqrt{n})$  which up to a slowly varying multiplicative factor is of order  $n^{1/(2\alpha)}$ . The convergence rate is slowed by the presence of a moving average component.

The proof of Theorem 4.1 parallels that of Theorem 3.1 and is only outlined. Our plan of attack is to show first that  $\min\{C_t/B_t : 1 \leq t \leq n, p_t(1 - \theta^{(0)}) < 0\}$  has the limit distribution given in (4.2) and then we show that  $\hat{\theta}$  has, in fact, this limit distribution.

We begin with the following limit theorem which parallels Proposition 3.2. It is built on the observations that, for  $x > 0, y > 0$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \sqrt{n} P \left[ Z_t \leq x, \frac{Z_t}{a(\sqrt{n})} \leq y \right] &= y^\alpha, \\ \lim_{n \rightarrow \infty} \sqrt{n} P \left[ Z_t \geq x, \frac{Z_t}{a(\sqrt{n})} \leq y \right] &= 0 \end{aligned}$$

and

$$\begin{aligned} \lim_{n \rightarrow \infty} nP \left[ \frac{Z_t}{a(\sqrt{n})} \leq x, \frac{Z_{t-1}}{a(\sqrt{n})} \leq y \right] \\ = \lim_{n \rightarrow \infty} \sqrt{n} P \left[ \frac{Z_t}{a(\sqrt{n})} \leq x \right] \sqrt{n} P \left[ \frac{Z_{t-1}}{a(\sqrt{n})} \leq y \right] \\ = x^\alpha y^\alpha. \end{aligned}$$

PROPOSITION 4.2. Suppose Conditions M and L hold. Define the measure

$$\mu(dx) = \alpha x^{\alpha-1} dx, \quad x > 0.$$

We have

$$(4.3) \quad \begin{aligned} nP \left[ Z_{t-i} \in dy_i, i = 0, \dots, 3; \frac{Z_t}{a(\sqrt{n})} \in dx_0, \frac{Z_{t-3}}{a(\sqrt{n})} \in dx_3 \right] \\ \rightarrow_v \varepsilon_0(dy_0) F(dy_1) F(dy_2) \varepsilon_0(dy_3) \mu(dx_0) \mu(dx_3) \end{aligned}$$

in

$$(4.4) \quad E := \left\{ (y_0, y_1, y_2, y_3, x_0, x_3) \in [0, \infty]^4 \times ([0, \infty]^2) \right\}.$$

Furthermore, with  $E$  as defined in (4.4), we have in  $M_p(E)$  that

$$(4.5) \quad \begin{aligned} N_n &:= \sum_{t=1}^n \mathcal{E}_{(Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3}, Z_t/a(\sqrt{n}), Z_{t-3}/a(\sqrt{n}))} \\ &\Rightarrow N_\infty := \sum_k \mathcal{E}_{(0, Y_{k1}, Y_{k2}, 0, j_{k1}, j_{k2})}, \end{aligned}$$

where

$$\sum_k \mathcal{E}_{(j_{k1}, j_{k2})}$$

is PRM with mean measure  $\mu \times \mu$  on  $[0, \infty)^2$ .

Again, as in Section 3, we wish only to consider points corresponding to  $p_t(1 - \theta^{(0)}) < 0$  but because it is easier, we start by restricting attention to the part of the state space corresponding to  $B_t < 0$ . So we define

$$[B < 0] := \left\{ (x_0, \dots, x_5) \in E: x_1 > \theta^{(0)}x_2 + 2(\theta^{(0)})^2 x_3 \right\}.$$

Further we need the maps

$$(4.6) \quad \begin{aligned} T_1(x_0, \dots, x_5) &:= x_4 + (\theta^{(0)})^3 x_5, \\ T_2(x_0, \dots, x_5) &:= | -x_1 + \theta^{(0)}x_2 + 2(\theta^{(0)})^2 x_3 |, \\ T(x_0, \dots, x_5) &:= (x_0, x_1, x_2, x_3, T_1(x_0, \dots, x_5), T_2(x_0, \dots, x_5)), \end{aligned}$$

with domains and ranges

$$T_1: [B < 0] \mapsto [0, \infty), \quad T_2: [B < 0] \mapsto (0, \infty]$$

and

$$T: [B < 0] \mapsto E' := [0, \infty]^4 \times [0, \infty) \times (0, \infty].$$

Also,  $T$  is continuous and  $T^{-1}$  maps compact sets into compact sets since, for instance,

$$\begin{aligned} &T^{-1} \left( \prod_{i=0}^4 [0, a_i] \times [a_5, \infty] \right) \\ &= \left\{ (x_0, \dots, x_5) \in E: x_i \leq a_i, i = 0, \dots, 3; x_4 + (\theta^{(0)})^3 x_5 \leq a_4, \right. \\ &\quad \left. a_5 \leq x_1 - (\theta^{(0)}x_2 + 2(\theta^{(0)})^2 x_3) \right\}. \end{aligned}$$

From Propositions 4.2 and 3.3 we have

$$N_n|_{[B < 0]} \Rightarrow N_\infty|_{[B < 0]}$$

in  $M_p([B < 0])$ . Applying Proposition 3.18, page 148, of Resnick (1987) yields

$$N_n|_{[B < 0]} \circ T^{-1} \Rightarrow N_\infty|_{[B < 0]} \circ T^{-1}$$

in  $M_p(E')$ , where, remember,  $E' = [0, \infty]^4 \times [0, \infty) \times (0, \infty]$ . Written another way, this is

$$\begin{aligned} & \sum_{t=1}^n \mathbf{1}_{[B < 0]} \mathcal{E}(Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3}, C_t/a(\sqrt{n}), |B_t|) \\ & \Rightarrow \sum_k \mathbf{1}_{[Y_{k1} > \theta^{(0)}Y_{k2}]} \mathcal{E}(0, Y_{k1}, Y_{k2}, 0, j_{k1} + (\theta^{(0)})^3 j_{k2}, |-Y_{k1} + \theta^{(0)}Y_{k2}|) \end{aligned}$$

in  $M_p(E')$ .

Now define

$$\begin{aligned} NEG := & \left\{ (x_0, x_1, x_2, x_3, x_5) \in E' : (x_2 + \theta^{(0)}x_3)(1 - \theta^{(0)})^2 \right. \\ & \left. + (-x_1 + \theta^{(0)}x_2 + 2(\theta^{(0)})^2 x_3)(1 - \theta^{(0)}) + (x_0 + (\theta^{(0)})^3 x_3) < 0 \right\}. \end{aligned}$$

We get from parts (a) and (c) of Proposition 3.3 that

$$\begin{aligned} & \sum_{t=1}^n \mathbf{1}_{[p_t(1 - \theta^{(0)}) < 0]} \mathcal{E}(Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3}, C_t/a(\sqrt{n}), |B_t|) \\ & \Rightarrow N_\infty^\# = \sum_{k=1}^\infty \mathbf{1}_{[Y_{k1} > Y_{k2}]} \mathcal{E}(0, Y_{k1}, Y_{k2}, 0, j_{k1} + (\theta^{(0)})^3 j_{k2}, |-Y_{k1} + \theta^{(0)}Y_{k2}|). \end{aligned}$$

Note that in the indicator on the right-hand side, the condition  $[Y_{k1} > Y_{k2}]$  is equivalent to the condition

$$Y_{k2}(1 - \theta^{(0)})^2 + (-Y_{k1} + \theta^{(0)}Y_{k2})(1 - \theta^{(0)}) < 0.$$

Applying Lemma 3.4 yields

$$\begin{aligned} & \sum_{t=1}^n \mathbf{1}_{[p_t(1 - \theta^{(0)}) < 0]} \mathcal{E}(C_t/a(\sqrt{n}), |B_t|) \\ & \Rightarrow N_\infty^\# = \sum_{k=1}^\infty \mathbf{1}_{[Y_{k1} > Y_{k2}]} \mathcal{E}(j_{k1} + (\theta^{(0)})^3 j_{k2}, |-Y_{k1} + \theta^{(0)}Y_{k2}|). \end{aligned}$$

After an argument that verifies division between the two components is permitted, we get

$$\begin{aligned} (4.7) \quad & \sum_{t=1}^n \mathbf{1}_{[p_t(1 - \theta^{(0)}) < 0]} \mathcal{E}(C_t/a(\sqrt{n})|B_t|) \\ & \Rightarrow \sum_k \mathbf{1}_{[Y_{k1} > Y_{k2}]} \mathcal{E}(j_{k1} + \theta^{(0)}j_{k2})/|-Y_{k1} + \theta^{(0)}Y_{k2}| \end{aligned}$$

in  $M_p((0, \infty))$ . Now one finishes the derivation with a comparison argument as in Section 3.

The form of the limit in (4.1) is based on the fact that

$$\sum_k \mathcal{E}(j_{k1} + (\theta^{(0)})^3 j_{k2})$$

is PRM with a mean measure of  $[0, x]$  equal to  $(\theta^{(0)})^{-3}c(\alpha)x^{2\alpha}$  and so is

$$\sum_{k=1}^{\infty} \mathcal{E}_{((\theta^{(0)})^{3/2}c(\alpha))^{-1/(2\alpha)}\Gamma_k^{1/(2\alpha)}}.$$

The form of the Weibull limit is gotten from the usual argument that the minimum of the points is greater than  $x$  iff the point process has no points in  $[0, x]$ .

**5. Simulation study.** There is often a problem with applying limit theorems based on weak convergence results that are driven by maxima or minima. In particular, the rates of convergence to the limit distribution may be very slow [see, e.g., de Haan and Resnick (1996)]. In order to verify the usefulness of our limit results we simulated MA(1) processes for both the left- and the right-tail cases and investigated the closeness of the empirical distribution of the estimator and the theoretical limit distribution. We note that this study deals only with the issue of the closeness of the approximation, and does not deal with further practical issues such as estimating  $\alpha$ , the parameter of regular variation, which would be required in an analysis of real data.

Each simulation study involved 200 replications, and time series of lengths  $n = 250, 500$  and  $1000$  were considered. The innovations  $Z_t$  were Pareto with  $\alpha = 1$  for the right-tail case ( $P[Z_t > x] = x^{-1}; x \geq 1$ ), and unit exponential for the left-tail case ( $P[Z_t \leq x] = 1 - \exp\{-x\}; x \geq 0$ ).

Three values of  $\theta^{(0)}$  (0.2, 0.5 and 0.8) were used for each case. A QQ-plot of the distribution of the normalized estimates  $q_n(\hat{\theta} - \theta^{(0)})$  versus that of the limit distribution was used. The value of  $q_n$  is  $n$  in the right-tail case and  $-1/\log(1 - 1/\sqrt{n}) \sim \sqrt{n}$  in the left-tail case.

In the right-tail case, from Theorem 3.1, the limit distribution is exponential with scale parameter given by

$$\begin{aligned} c(\theta) &= E\left((Z_1 + \theta^3 Z_2)^{-1}\right) \\ &= \frac{1}{3} \left( 1 - \theta^{-3} - \theta^{-6} \log(1 + \theta^3) - \theta^3 \log\left(\frac{1 + \theta^3}{\theta^3}\right) \right). \end{aligned}$$

In the left-tail case the limit distribution, according to Theorem 4.1, is Weibull with form:

$$1 - \exp(-k(\theta)x^2),$$

and

$$k(\theta) = (4 - 4\theta + \theta^2)/(8\theta^3).$$

We see from Figure 1 for the right-tail case that the simulated and theoretical distributions match up very well for all sample sizes and for all values of  $\theta^{(0)}$  that were investigated. In other words, in this case, confidence statements based on the limiting distributions will provide good approximations to the true ones even for sample sizes of 250.

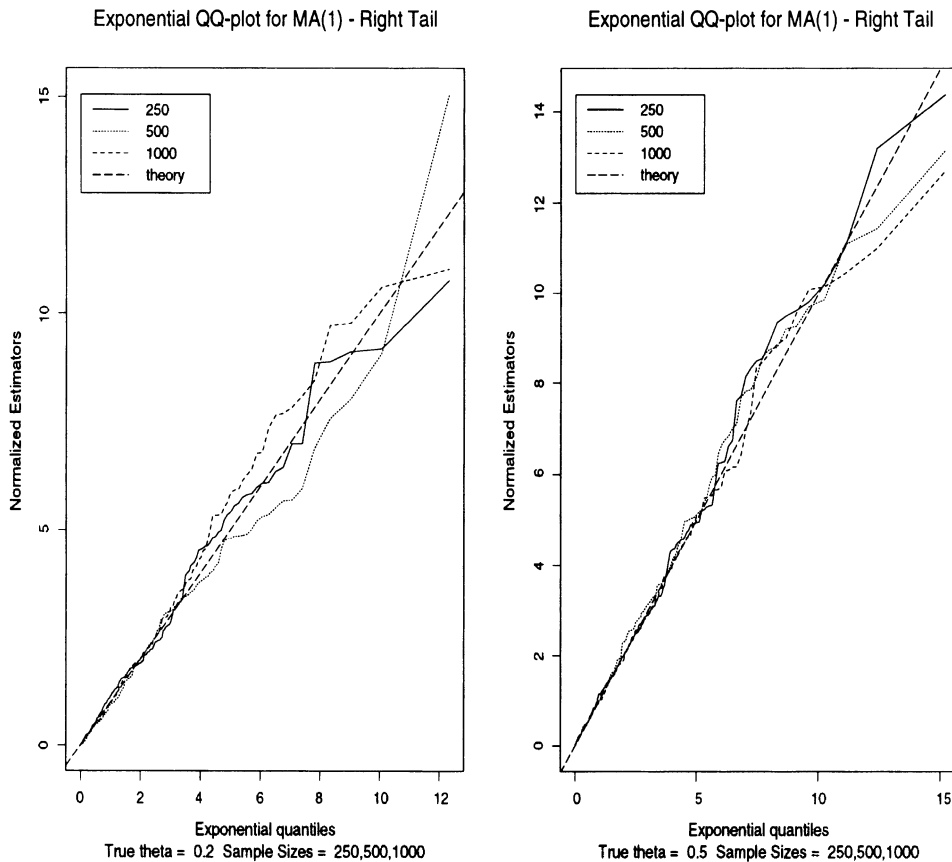


FIG. 1.

For the left-tail case, we find good approximations to the limiting distribution for  $\theta^{(0)} = 0.2$  and  $\theta^{(0)} = 0.5$ , but a poorer one for  $\theta^{(0)} = 0.8$ , for the sample sizes considered.

**6. Concluding remarks.** It is noteworthy that, in contrast to the autoregressive case, the moving average estimators in the left-tail case suffer a performance degradation depending on the order  $q$  of the model; no such degradation is present under Condition R. From the results of Section 4 we see that the convergence rate for the estimator of the MA(1) parameter is  $1/\alpha(\sqrt{n})$  which is a regularly varying function of index  $1/(2\alpha)$ . Contrast this to the convergence rate of the lp estimators in the autoregressive case which is regularly varying of index  $1/\alpha$ . We anticipate that the convergence rate in the left-tail case for MA( $q$ ) parameters will have index  $1/(2\lceil q/2 \rceil \alpha)$ , where  $\lceil q/2 \rceil$  is the first integer greater than or equal to  $q/2$ . Thus, under Condition L, a sharp penalty is paid for using models which have moving average

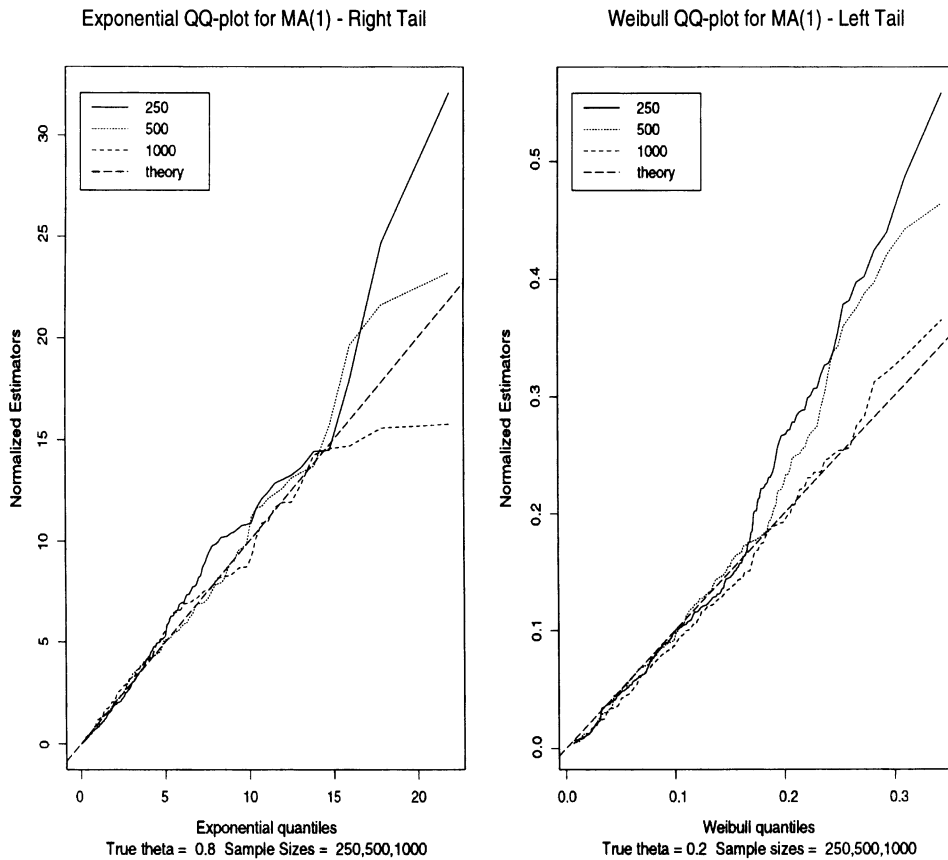


FIG. 1. (continued)

components and the penalty increases as the order of the model increases. This is in contrast to results under Condition R and to the results found for lp estimators for autoregressive parameters.

How does our linear programming estimator compare with other estimators? Under condition L, the rate of convergence of our MA(1) estimator is  $1/a(\sqrt{n})$ . Because  $a(\cdot)$  is regularly varying with index  $1/\alpha$ , the rate of convergence [neglecting the slowly varying component of  $a(\cdot)$ ] is of the order  $n^{1/(2\alpha)}$ . When Condition L is applicable one could also consider using the Gaussian maximum likelihood estimator which has a rate of convergence  $\sqrt{n}$ . Our estimator will be superior provided  $1/(2\alpha) > 1/2$ ; that is, when  $\alpha < 1$ . The validity of the condition  $\alpha < 1$  can be statistically checked by using, for example, the Hill estimator. See Resnick and Stărică (1995).

When Condition R holds, Davis and Resnick's (1986) method of moments estimator for  $\theta$  based on the sample correlation function can be used. Such estimators, because they are based on products of the innovations variables,

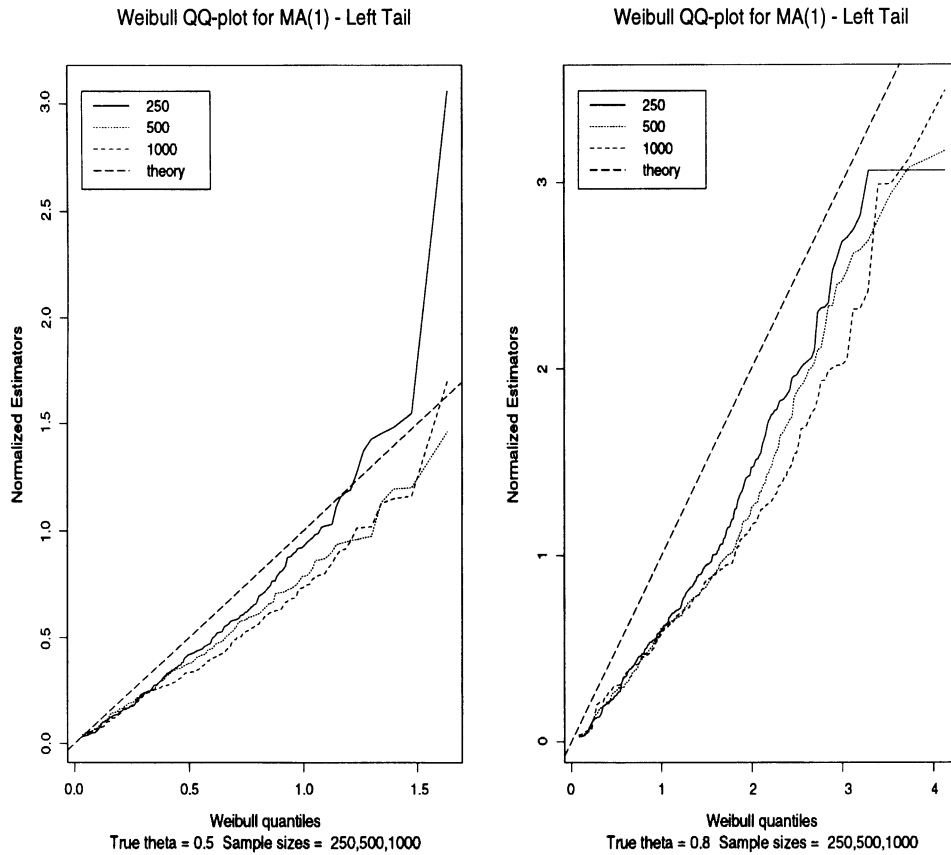


FIG. 1. (continued)

will have a rate of convergence which is inferior to our linear programming estimator. For example, if the innovations have Pareto tails, the rate of convergence of the Davis and Resnick (1986) estimator will be  $(n/\log n)^{1/\alpha}$ , whereas our linear programming estimator has a rate of convergence  $n^{1/\alpha}$ . Thus, for example, when  $\alpha = 1$  and the sample size is 5000 (a not uncommon sample size in teletraffic studies), the rate of convergence of the moment estimator is about 8 times slower than the linear programming estimator. A Gaussian maximum likelihood estimator can be expected to have the same rate of convergence as the method of moments estimator.

The important challenge now is to extend these results from the MA(1) case to more general moving average processes and then on to the general ARMA model. We anticipate that for ARMA models the rate of convergence under Condition L of the lp estimators will suffer depending on the order of the moving average component. The theoretical challenge in dealing with higher-order MA models or general ARMA models arises from the fact that



the estimators are not representable explicitly but are solutions of mathematical programming problems. Thus general weak convergence techniques involving random sets representing feasible regions for the mathematical program will be needed. These techniques will be similar to but harder than those used by Feigin and Resnick (1994). As for practical implementation of numerical methods to compute the estimated values, there is an emerging field of *semidefinite programming* which offers general computational tools that will be relevant. See, for example, Vandenberghe and Boyd (1996).

There is some evidence that many data sets exhibiting heavy tails cannot be fit by ARMA models, perhaps because of nonlinearities. See Davis and Resnick (1996) and Resnick (1996). However, one cannot appreciate the potential shortcomings of classical Box–Jenkins modeling when applied to the new context of heavy tails without a reasonably complete set of estimation and fitting techniques and thus present efforts to estimate coefficients in linear models are crucial for progress in finding acceptable models for heavy-tailed data.

**Acknowledgment.** Marie Kratz acknowledges with gratitude the hospitality of Cornell's School of Operations Research and Industrial Engineering during the fall of 1993, 1994 and 1995.

## REFERENCES

- BILLINGSLEY, P. (1968). *Convergence of Probability Measure*. Wiley, New York.
- BREIMAN, L. (1965). On some limit theorems similar to the arc-sin law. *Theory Probab. Appl.* **10** 323–331.
- BROCKWELL, P. and DAVIS, R. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.
- CROVELLA, M. and BESTAVROS, A. (1992). Explaining world wide web traffic self-similarity. Technical Report 95-015, Dept. Computer Science, Boston Univ.
- CUNHA, C., BESTAVROS, A. and CROVELLA, M. (1995). Characteristics of www client-based traces. Preprint, Dept. Computer Science, Boston Univ.
- DAVIS, R. and RESNICK, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14** 533–558.
- DAVIS, R. and RESNICK, S. (1988). Extremes of moving averages of random variables from the domain of attraction of the double exponential distribution. *Stochastic Processes Appl.* **30** 41–68.
- DAVIS, R. and RESNICK, S. (1996). Limit theory for bilinear processes with heavy tailed noise. *Ann. Applied Probab.* **6** 1191–1210.
- FEIGIN, P. (1991). Inference and martingale estimating equations for stochastic processes on a semigroup. In *Statistical Inference in Stochastic Processes* (N. U. Prabhu and I. V. Basawa, eds.) **2**. Dekker, New York.
- FEIGIN, P. and RESNICK, S. (1992). Estimation for autoregressive processes with positive innovations. *Stochastic Models* **8** 479–498.
- FEIGIN, P. and RESNICK, S. (1994). Limit distributions for linear programming time series estimators. *Stochastic Process. Appl.* **51** 135–166.
- FEIGIN, P. and RESNICK, S. (1996). Pitfalls of fitting autoregressive models with emphasis on heavy-tailed time series. Technical Report 1163 ORIE, Cornell Univ.
- FEIGIN, P., RESNICK, S. and STÁRICÁ, C. (1995). Testing for independence in heavy tailed and positive innovation time series. *Stochastic Models* **11** 587–612.

- HAAN, L. DE and RESNICK, S. (1996). Second-order regular variation and rates of convergence in extreme-value theory. *Ann. Probab.* **24** 97–124.
- KALLENBERG, O. (1993). *Random Measures*, 3rd ed. Akademie, Berlin.
- RESNICK, S. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.
- RESNICK, S. (1996). Why non-linearities can ruin the heavy tailed modeler's day. Technical Report 1157, ORIE, Cornell Univ.
- RESNICK, S. and STĂRICĂ, C. (1995). Consistency of Hill's estimator for dependent data. *J. Appl. Probab.* **32** 139–167.
- VANDENBERGHE, L. and BOYD, S. (1996). Semidefinite programming. *SIAM Rev.* **38** 49–95.

P. D. FEIGIN  
FACULTY OF INDUSTRIAL ENGINEERING  
AND MANAGEMENT  
TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY  
HAIFA 32000  
ISRAEL  
E-MAIL: paulf@ie.technion.ac.il

M. F. KRATZ  
UNIVERSITÉ RENÉ DESCARTES  
PARIS V U.F.R. DE MATHÉMATIQUES  
ET INFORMATIQUE 45  
RUE DES SAINTS-PÈRES 75270  
PARIS CEDEX 06  
FRANCE  
E-MAIL: kratz@math-info.univ-paris5.fr

S. I. RESNICK  
CORNELL UNIVERSITY  
SCHOOL OF OPERATIONS RESEARCH  
AND INDUSTRIAL ENGINEERING  
RHODES HALL  
ITHACA, NEW YORK 14853  
E-MAIL: sid@orie.cornell.edu