SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

TECHNICAL REPORT NO. 800

May 1988

BASIC PROPERTIES AND PREDICTION OF MAX–ARMA PROCESSES

by

Richard A. Davis[1]
Colorado State University and
U. of California, San Diego

Sidney I. Resnick[2]
Cornell University

[1]Research supported by NSF Grant MCS–8501763.

[2]Research supported by NSF Grant MCS–8501763 and the Mathematical Sciences Institute at Cornell University.

Basic Properties and Prediction of Max-ARMA Processes

Richard A. Davis*
Colorado State University and U. of California, San Diego

and

Sidney I. Resnick†
Cornell University

Abstract

A max-autoregressive moving average (MARMA(p,q)) process $\{X_t\}$ satisfies the recursion

$$X_t = \phi_1 X_{t-1} \vee \cdots \vee \phi_p X_{t-p} \vee Z_t \vee \theta_1 Z_{t-1} \vee \cdots \vee \theta_q Z_{t-q}$$

for all t where $\phi_i, \theta_j \geq 0$, and $\{Z_t\}$ is iid with common distribution function $\Phi_{1,\sigma}(x) := \exp\{-\sigma x^{-1}\}$ for $x \geq 0$, $\sigma > 0$. Such processes have finite dimensional distributions which are max-stable and hence are examples of max-stable processes. We provide necessary and sufficient conditions for the existence of a stationary solution to the MARMA recursion and we examine the reducibility of the process to a MARMA(p', q') with $p' < p$ or $q' < q$. After introducing a natural metric between two jointly max-stable random variables, we consider the prediction problem for MARMA processes. Assuming that $X_1, \ldots, X_n$ have been observed, we restrict our class of predictors to be max-linear, ie. of the form $\vee_1^n b_i X_i$, and find $b_1, \ldots, b_n$ to minimize the distance between this predictor and X_{n+k} for $k \geq 1$. The optimality criterion is designed to minimize the probability of large errors and is similar in spirit to the dispersion criterion adopted in Cline and Brockwell [*Stoch. Process. Appl.* 19(1985):281-296] for the prediction of ARMA processes with stable noise. Most of our results remain valid for the case when the distribution of Z_1 is only in the domain of attraction of $\Phi_{1,\sigma}$. In addition, we give a naive estimation procedure for the ϕ's and the θ's which, with probability one, identifies the true parameter values exactly for n sufficiency large.

*Research supported by NSF Grant MCS-8501763.

†Research supported by NSF Grant MCS-8501763 and the Mathematical Sciences Institute at Cornell University.

AMS 1980 Subject Classification: 60G10, 60G25, 62M20.

Key Words and Phrases. Max-stable distribution, max-stable processes, prediction, causality, domain of attraction of an extreme value distribution, ARMA processes.

1. Introduction

A time series which is stationary in appearance and exhibits large peaks or sudden bursts of outlying observations is a potential candidate for modelling as an ARMA process with heavy tailed noise. Examples of such data can be found in Stuck and Kleiner (1974), who considered telephone signals, and Fama (1965), who modelled stock market prices. In this paper we explore the class of max-autoregressive moving average (MARMA) processes, which share many of the characteristics of heavy-tailed ARMA processes. As such, MARMA processes offer an alternative class of models for modelling stationary data with outlying observations.

A stationary process $\{X_n\}$ is a MARMA(p,q) process if it satisfies the MARMA recursion, namely:

$$X_n = \phi_1 X_{n-1} \vee \cdots \vee \phi_p X_{n-p} \vee Z_n \vee \theta_1 Z_{n-1} \vee \cdots \vee \theta_q Z_{n-q}$$

for all n where $\phi_i, \theta_j \geq 0$, $1 \leq i \leq p$, $1 \leq j \leq q$, and $\{Z_n\}$ is iid with common distribution function $\Phi_{1,\sigma}(x) := \exp\{-\sigma x^{-1}\}$. As will be evident from the discussion of Section 2, such processes have finite dimensional distributions which are max-stable (cf. Resnick, 1987) and in fact MARMA processes are examples of general max-stable processes considered by de Haan (1984), de Haan and Pickands (1986), Balkema and de Haan (1988). MARMA processes constitute a relatively simple parametric family with $p + q + 1$ parameters. The finite dimensional distributions of a MARMA process possess much structure that can be used for prediction (Section 4), estimation (Section 5), and model fitting.

Section 2 deals with some foundational issues such as when the defining equation for a MARMA has a stationary solution. This coincides with the property that the MARMA process can be written as a max-moving average (MMA) of order ∞ or synonymously a max-linear process of the form $X_n = \bigvee_{i=0}^{\infty} \psi_i Z_{n-i}$. This property we call *causality* and we show how the coefficients $\{\psi_j\}$ can be computed by means of a recursive procedure. We are also interested in when the MARMA process is reducible; i.e. when the process can be written as the solution of a MARMA(p', q') recursion with either $p' < p$ or $q' < q$.

Section 3 considers a natural metric between two jointly max-stable random variables and in Section 4 we give a procedure which produces predictors which are optimal in the sense of minimizing the distance (in the sense of Section 3) between a max-linear function of the past and the future value one needs to predict. So for instance given that $X_1, \ldots, X_n$ have been observed, we restrict our class of predictors to be max-linear, ie of the form $\bigvee_{i=1}^{n} b_i X_i$ and find $b_1, \ldots, b_n$ to minimize the distance between this predictor and X_{n+k} for $k \geq 1$. The optimality criterion is designed to minimize the probability of large errors (cf. Cline and Brockwell, 1985) and thus is different in spirit to the usual Hilbert space approach to prediction. We can exhibit explicitly the optimal predictor for a variety of examples and the form of the predictor seems natural in most cases.

To get a feel for the sample paths of MARMA(p,q) processes, we have simulated 250 observations from each of the following models (see Figures 1, 3, and 5),

$$X_t = .7X_{t-1} \vee Z_t,$$
$$X_t = Z_t \vee .8Z_{t-1},$$
$$X_t = .5X_{t-1} \vee .3X_{t-2} \vee Z_t \vee .6Z_{t-1},$$

where $Z_t \sim \Phi_{1,1}$. Figure 2 contains the plot of 250 observations from the AR(1) process $X_t = .7X_{t-1} + Z_t$ where the noise variables are identical to those used in generating the plot in Figure 1. Observe that the MARMA(1,0) graph is slightly smoother than the AR(1) graph but the corresponding peaks in the two plots are practically the same. Figure 4 contains a realization of the MA(1) process, $X_t = Z_t + .8Z_{t-1}$, which again was generated from the same innovations used to construct the MARMA(0,1) process in Figure 3. This time the two plots are essentially indistinguishable from one another. As an alternative to comparing ARMA and MARMA, one can imagine a MARMA model as an alternative to a shot noise model. For example, in Figure 1, the process descends exponentially from each spike and then fluctuates around *small* values until the next major spike is encountered. By adjusting the order and parameter values of the process, the frequency and width of the spikes can be altered.

In Section 5, we give a naive estimation procedure for the $\phi's$ and $\theta's$ which, with probability one, identifies the parameter values exactly for n sufficiently large. This procedure is applied to the MARMA examples displayed in Figures 1, 3, and 5, as well as to the AR(1) process in Figure 2.

It is our contention that the class of max-stable processes possesses many desirable and elegant mathematical properties. We intend in future work to pursue questions of parameter estimation and model fitting and it is our hope that, with the inclusion of some form of observational noise, MARMA processes will be useful in fitting data which exhibit sudden large jumps. It is entirely possible that the MARMA class will model the small values of such data badly. However since many applications are concerned with whether large values exceed specified limits, if a MARMA process is a good model for large values of the data we may have an adequate model for the intended purpose and in this case our prediction criterion–minimizing large differences between predicted and true values–is quite appropriate.

Our results are more widely applicable than may at first be apparent. Although we have assumed $Z_1 \sim \Phi_{1,\sigma}$, this is not essential. If instead, Z_1 has the *heavy tailed* extreme value distribution $\exp\{\sigma x^{-\alpha}\}$, then one can transform back to the $\alpha = 1$ case by taking $1/\alpha$ powers in the MARMA recursion. More generally we hope to examine estimation techniques for long tailed stationary data which allow instantaneous transformations of the data in such a way that the transformed data can be reasonably modelled by the MARMA model with $\Phi_{1,\sigma}$ marginals. Furthermore, most of our results remain valid for the case when the distribution of Z_1 is in the domain of attraction of $\Phi_{1,\sigma}$. These instances are pointed out in the sequel when applicable.

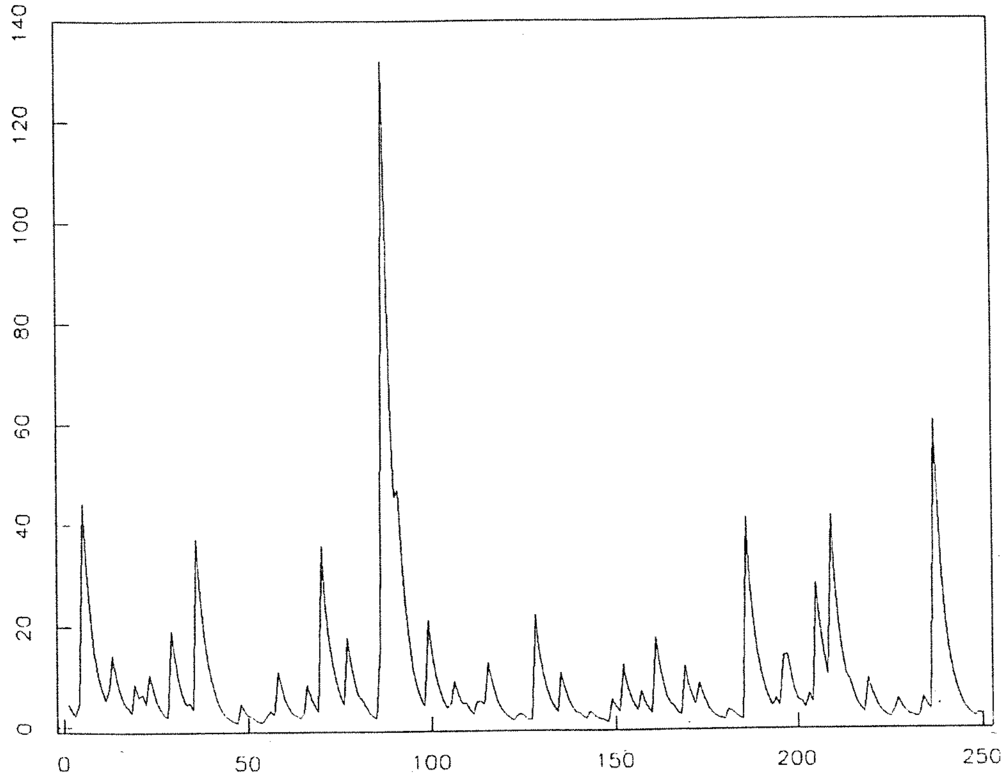

Figure 1. 250 observations of the process $X_t = .7X_{t-1} \vee Z_t$.

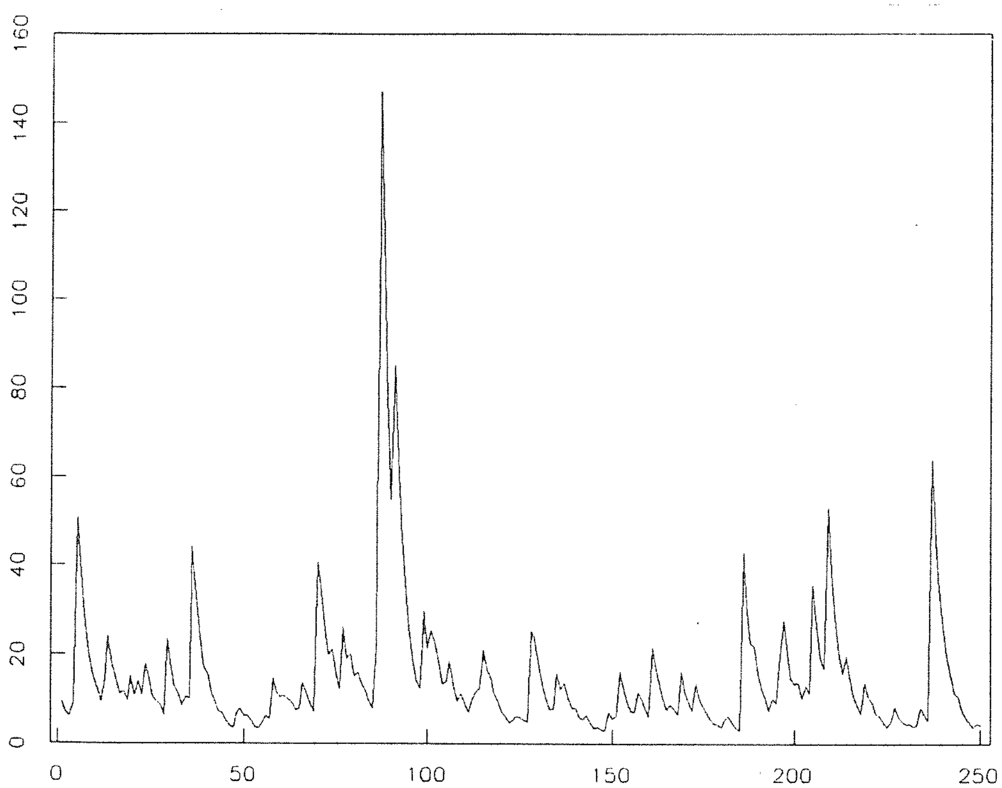

Figure 2. 250 observations of the process $X_t = .7X_{t-1} + Z_t$ where $\{Z_t\}$ is the same innovation sequence as in Figure 1.

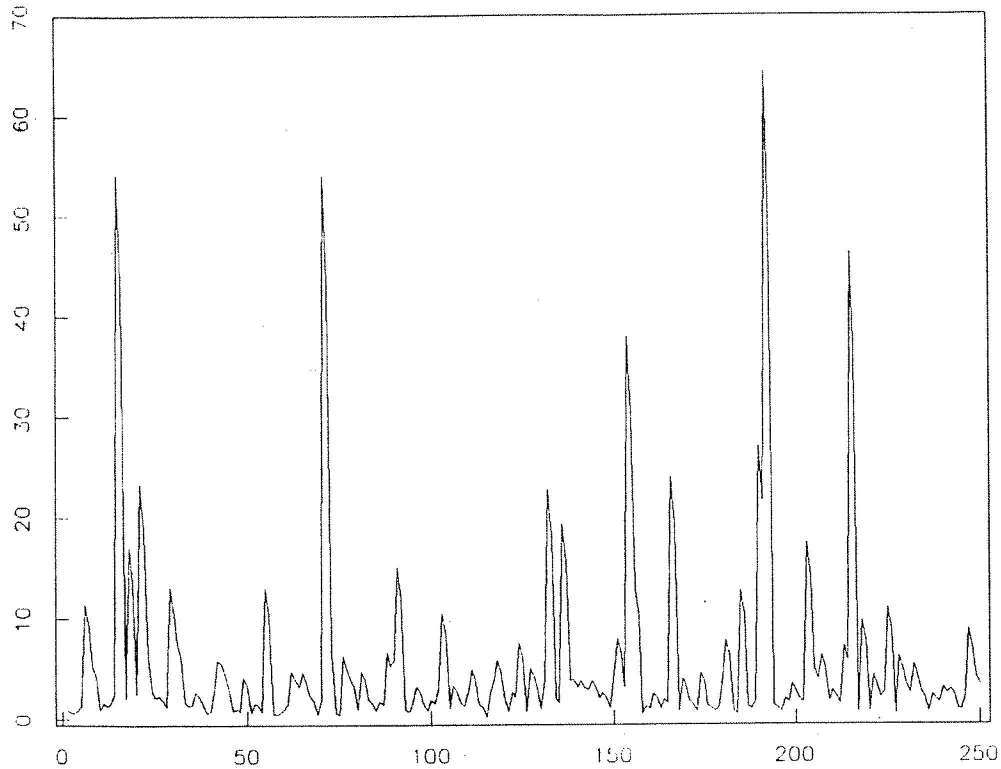

Figure 3. 250 observations of the process $X_t = Z_t \vee .8Z_{t-1}$.

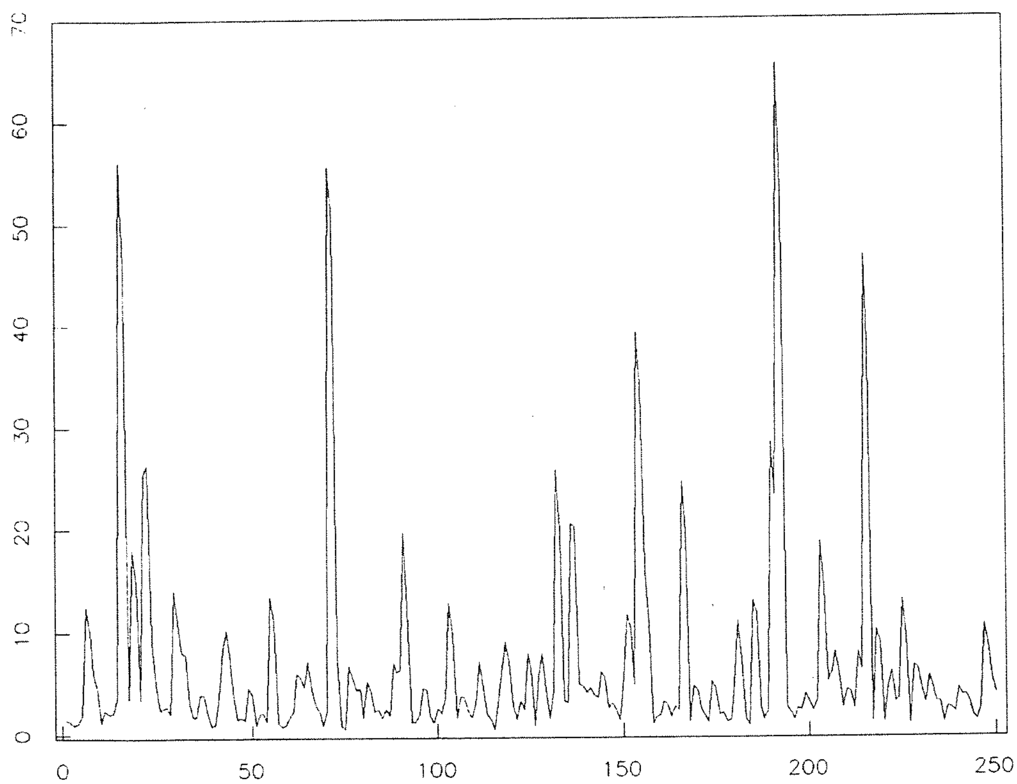

Figure 4. 250 observations of the process $X_t = Z_t + .8Z_{t-1}$ where $\{Z_t\}$ is the same innovation sequence as in Figure 3.

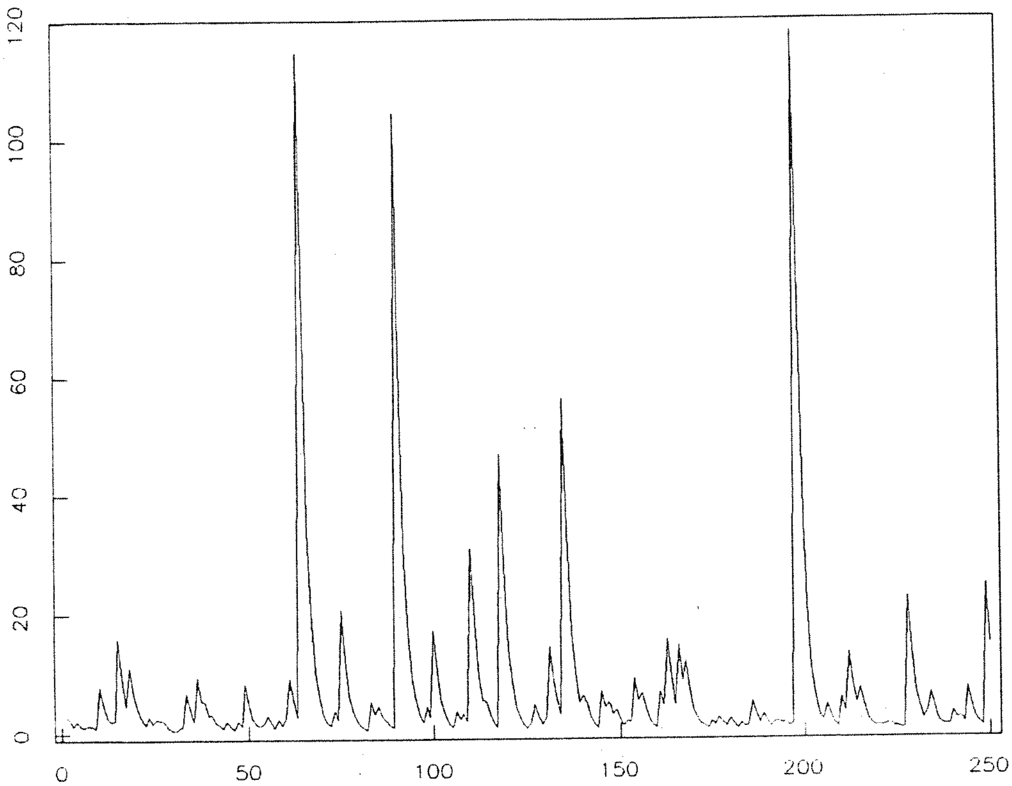

Figure 5. 250 observations of the process $X_t = .5X_{t-1} \vee .3X_{t-2} \vee Z_t \vee .6Z_{t-1}$.

2. Causality and reducibility.

Consider a stationary process $\{X_n\}$ which satisfies the max-ARMA(p,q) (or MARMA(p,q) for short) recursion

$$X_n = \phi_1 X_{n-1} \vee \cdots \vee \phi_p X_{n-p} \vee Z_n \vee \theta_1 Z_{n-1} \vee \cdots \vee \theta_q Z_{n-q}, \qquad n = 0, \pm 1, \pm 2, \ldots \tag{2.1}$$

for $\phi_i,\, \theta_j \geq 0,\ 1 \leq i \leq p,\ 1 \leq j \leq q$, and $\{Z_n\}$ iid with common distribution function $\Phi_{1,\sigma}(x) := \exp\{-\sigma x^{-1}\},\ x \geq 0,\ \sigma > 0$. Such a process is called *causal* if there exist constants $\psi_j \geq 0,\ j \geq 0$ such that $\bigvee_{j=0}^{\infty} \psi_j Z_j < \infty$ a.s. and

$$X_n = \bigvee_{j=0}^{\infty} \psi_j Z_{n-j}. \tag{2.2}$$

In analogy with traditional time series models, a process having representation (2.2) will be called *max-linear*. We first note the following easy but useful facts about the existence of max-linear processes.

PROPOSITION 2.1. *If $\{Z_n\}$ is iid with $P[Z_1 \leq x] = \Phi_{1,\sigma}(x)$ then*

(a) $\bigvee_{j=0}^{\infty} \psi_j Z_j < \infty$ a.s. *for* $\psi_j \geq 0$ *if and only if* $\sum_{j=0}^{\infty} \psi_j < \infty$.

(b) *Define* $\bar{Z}_n = \bigvee_{j=0}^{q} \theta_j Z_{n-j}$ *where* $\theta_0 = 1,\ \theta_j \geq 0$. *Then* $\bigvee_{j=0}^{\infty} \alpha_j \bar{Z}_{n-j} < \infty$ a.s. *for* $\alpha_j \geq 0$

if and only if

$$\sum_{j=0}^{\infty} \alpha_j < \infty.$$

In particular, $\{\bigvee_{j=0}^{\infty} \alpha_j \bar{Z}_{n-j},\ n = 0, \pm 1, \dots\}$ *has representation (2.2) with*

$$\psi_j = \bigvee_{k=0}^{j \wedge q} \alpha_{j-k}\theta_k.$$

PROOF: The proof of (a) follows readily since by the Kolomogorov 0-1 law $P[\bigvee_{j=0}^{\infty} \psi_j Z_j < \infty] = 0$ or 1. This probability is one if and only if for some $x > 0$

$$0 < P[\bigvee_{j=0}^{\infty} \psi_j Z_j \le x] = \prod_{j=0}^{\infty} \Phi_{1,\sigma}(\psi_j^{-1}x) = \exp\{-\sigma x^{-1} \sum_{j=0}^{\infty} \psi_j\}$$

However $\exp\{-\sigma x^{-1} \sum_{j=0}^{\infty} \psi_j\} > 0$ if and only if $\sum_{j=0}^{\infty} \psi_j < \infty$.

For (b) it is convenient to define $\theta_k = 0$, if $k > q$, then

$$\begin{aligned}\bigvee_{j=0}^{\infty} \alpha_j \bar{Z}_{n-j} &= \bigvee_{j=0}^{\infty} (\alpha_j \bigvee_{k=0}^{\infty} \theta_k Z_{n-j-k}) \\ &= \bigvee_{j=0}^{\infty} (\alpha_j \bigvee_{k=j}^{\infty} \theta_{k-j} Z_{n-k}) = \bigvee_{k=0}^{\infty} (\bigvee_{j=0}^{k} \alpha_j \theta_{k-j}) Z_{n-k} \\ &= \bigvee_{k=0}^{\infty} \psi_k Z_{n-k}.\end{aligned}$$

Since $\sum_j \psi_j < \infty$ if and only if $\sum_j \alpha_j < \infty$ the result now follows from (a). ∎

REMARK 1: There is a parallel result for the case when it is only assumed that the df F of Z_1 is in the domain of attraction of $\Phi_{1,\sigma}$. Using Potter's inequalities (see for example Resnick (1987), p.23), the counterparts of (a) and (b) of Proposition 2.1 can be combined into the following statement (see Cline (1983) and Hsing (1986)). If $F(0-) = 0$ and $F \in \mathcal{D}(\Phi_{1,\sigma})$, then for $\psi_j \ge 0$,

$$\bigvee_{j=0}^{\infty} \psi_j Z_j \begin{cases} < \infty & a.s. \quad \text{if } \sum_{j=0}^{\infty} \psi_j^{\delta} < \infty \text{ for some } \delta < 1, \\ = \infty & a.s. \quad \text{if } \sum_{j=0}^{\infty} \psi_j^{\delta} = \infty \text{ for some } \delta > 1, \end{cases}$$

and in the former case, we have

$$\lim_{x \to \infty} \frac{P[\bigvee_{j=0}^{\infty} \psi_j Z_j > x]}{P[Z_j > x]} = \sum_{j=0}^{\infty} \psi_j. \qquad ∎$$

In the remainder of this section we assume the df F of Z_1 is in fact equal to $\Phi_{1,\sigma}$. However by virtue of the preceeding remark, the following results will also be valid for the case $F(0-) = 0$ and $F \in \mathcal{D}(\Phi_{1,\sigma})$.

To investigate causality of the solution to the MARMA recursions, suppose a solution of (2.1) exists of the form $X_n = \bigvee_{j=0}^{\infty} \alpha_j \bar{Z}_{n-j}$ where again $\bar{Z}_n = \bigvee_{i=0}^{q} \theta_i Z_{n-i}$ and $\sum_j \alpha_j < \infty$. Then plugging this expression into both sides of (2.1) and setting $\alpha_j = 0$ for $j < 0$, we obtain

$$\begin{aligned}\bigvee_{j=0}^{\infty} \alpha_j \bar{Z}_{n-j} &= (\phi_1 \bigvee_{j=1}^{\infty} \alpha_{j-1} \bar{Z}_{n-j}) \vee (\phi_2 \bigvee_{j=1}^{\infty} \alpha_{j-2} \bar{Z}_{n-j}) \vee (\phi_p \bigvee_{j=1}^{\infty} \alpha_{j-p} \bar{Z}_{n-j}) \vee \bar{Z}_n \\ &= \bar{Z}_n \vee (\bigvee_{j=1}^{\infty} (\phi_1 \alpha_{j-1} \vee \phi_2 \alpha_{j-2} \vee \cdots \vee \phi_p \alpha_{j-p}) \bar{Z}_{n-j}).\end{aligned}$$

This leads to a solution of (2.1) provided $\{\alpha_j\}$ satisfies the recursions

$$\begin{aligned}&\alpha_0 = 1 \\ &\alpha_j = \phi_1 \alpha_{j-1} \vee \phi_2 \alpha_{j-2} \vee \cdots \vee \phi_p \alpha_{j-p}, \quad j \geq 1\end{aligned} \tag{2.3}$$

($\alpha_j = 0$ for $j < 0$) and $\sum_j \alpha_j < \infty$. If $\phi^* := \bigvee_{j=1}^{p} \phi_j < 1$, then $\alpha_j \leq (\phi^*)^j$ and hence by Proposition 2.1, the process $\bigvee_{j=0}^{\infty} \alpha_j \bar{Z}_{n-j}$ is one such stationary solution to (2.1). As the following propositon demonstrates this is the unique stationary solution to (2.1).

Proposition 2.2. *There is a solution to the equations (2.1) if and only if* $\phi^* := \bigvee_{j=1}^{p} \phi_j < 1$. *If* $\phi^* < 1$, *then there is a unique stationary solution given by*

$$X_n = \sum_{j=0}^{\infty} \alpha_j \bar{Z}_{n-j} = \bigvee_{j=0}^{\infty} \psi_j Z_{n-j} \tag{2.4}$$

where

$$\psi_j = \bigvee_{k=0}^{j \wedge q} \alpha_{j-k} \theta_k, \tag{2.5}$$

$\theta_0 = 1$ *and* $\{\alpha_j\}$ *satisfies the recursions (2.3). In addition, the* ψ_j *satisfy the recursion*

$$\psi_j = \phi_1 \psi_{j-1} \vee \cdots \vee \phi_p \psi_{j-p}$$

for $j \geq p$.

Proof: We have already seen that there is a stationary solution to (2.1) if $\phi^* < 1$. Conversely suppose $\phi^* = \phi_r \geq 1$ for some $1 \leq r \leq p$, then for any $k \geq 1$

$$\begin{aligned}X_n &\geq X_{n-r} \vee \bar{Z}_n \\ &\geq X_{n-2r} \vee \bar{Z}_n \vee \bar{Z}_{n-r} \\ &\geq \bar{Z}_n \vee \bar{Z}_{n-r} \vee \cdots \vee \bar{Z}_{n-kr}.\end{aligned}$$

Since, by Proposition 2.1, $\bigvee_{j=1}^{k} \bar{Z}_{n-jr} \to \infty$ a.s. as $k \to \infty$, we conclude that $X_n = \infty$ a.s., a contradiction. Consequently, ϕ^* must be less than one.

As for uniqueness, suppose $\phi^* < 1$ and let $\{Y_n\}$ be a strictly stationary solution to (2.1), ie.

$$Y_n = \phi_1 Y_{n-1} \vee \cdots \vee \phi_p Y_{n-p} \vee \bar{Z}_n.$$

Iterating this equation back k time lags, we may write

$$Y_n = \bigvee_{j=0}^{k} \alpha_j \bar{Z}_{n-j} \vee (a_{k,1} Y_{n-k-1} \vee a_{k,2} Y_{n-k-2} \vee \cdots \vee a_{k,p} Y_{n-k-p}) \tag{2.6}$$

where the constants $a_{k,j}$ satisfy $0 \le a_{k,j} \le (\phi^*)^{[k/p]}$, $j = 1, \ldots, p$, $[x]$ = integer part of x, and $\{\alpha_j\}$ is given by (2.3). Thus, since $\{Y_n\}$ is stationary,

$$a_{k,1} Y_{n-k-1} \vee a_{k,2} Y_{n-k-2} \vee \cdots \vee a_{k,p} Y_{n-k-p} \xrightarrow{p} 0$$

as $k \to \infty$ so that by letting $k \to \infty$ in (2.6) we obtain

$$Y_n = \bigvee_{j=0}^{\infty} \alpha_j \bar{Z}_{n-j}.$$

This is the unique stationary solution with causal representation as specified in the statement of the proposition.

Finally the relation $\psi_j = \phi_1 \psi_{j-1} \vee \cdots \vee \phi_p \psi_{j-p}$ for $j \ge p$ follows directly from (2.3) and (2.5). ∎

REMARK 2: The θ_j play no role in determining whether (2.1) has a stationary solution. ∎

REMARK 3: Proposition 2.2 does not assert that there is a unique solution to (2.1) if $\phi^* < 1$; only that there is a unique stationary solution. To construct nonstationary solutions to (2.1), let $\{X_n\}$ be the stationary solution as specified by (2.4) and suppose $\{Y_n\}$ is any solution to the homogeneous *max-linear* difference equation

$$Y_n = \phi_1 Y_{n-1} \vee \cdots \vee \phi_p Y_{n-p}.$$

For example, if $p = 1$ then $Y_n = \phi_1^n Y_0$. The process $\{X_n \vee Y_n\}$ is then easily seen to be a (nonstationary) solution to (2.1). The stationary solution is always the minimal solution of (2.1). ∎

If we write $\phi(x) = \bigvee_{i=1}^{p} \phi_i x^i$, $x > 0$ and define B to be the usual backwards shift operator, $BX_n = X_{n-1}$, we may represent (2.1) symbolically as

$$X_n = \phi(B) X_n \vee \bar{Z}_n. \tag{2.7}$$

If $\phi^* < 1$, the solution (2.4) can be seen as follows: Iterate (2.7)

$$\begin{aligned} X_n &= \phi(B)(\phi(B) X_n \vee \bar{Z}_n) \vee \bar{Z}_n \\ &= \phi^2(B) X_n \vee \phi(B) \bar{Z}_n \vee \bar{Z}_n \\ &= \ldots = \phi^{k+1}(B) X_n \vee \bigvee_{j=0}^{k} \phi^j(B) \bar{Z}_n \end{aligned}$$

where $\phi^0(B)$ is the identity. The solution is

$$X_n = (\bigvee_{j=0}^{\infty} \phi^j(B))\bar{Z}_n.$$

Now reading off the coefficient of x^j in

$$\bigvee_{k=0}^{\infty}(\bigvee_{i=1}^{p} \phi_i x^i)^k$$

we have an explicit expression for the α_j in terms of $\phi_1, \ldots, \phi_p$ given by

$$\alpha_0 = 1$$

$$\begin{aligned} \alpha_j &= \bigvee_{k=1}^{\infty} \bigvee_{i_1+\cdots+i_k=j} \phi_{i_1}\cdots\phi_{i_k} \\ &= \bigvee_{k=j/p}^{j} \bigvee_{i_1+\cdots+i_k=j} \phi_{i_1}\cdots\phi_{i_k} \\ &= \bigvee_{k=j/p}^{j} \bigvee_{\substack{n_1+\cdots+n_p=k \\ n_1+2n_2+\cdots+pn_p=j}} \prod_{i=1}^{p} \phi_i^{n_i}. \end{aligned} \tag{2.8}$$

It is a simple matter to check that this choice of α_j, is indeed the solution to the recursions (2.3).

We next consider when a process $\{X_n\}$ satisfying (2.1) is *reducible*in the AR component; i.e. when we can write

$$X_n = \bigvee_{i=1}^{p'} \phi_i X_{n-i} \vee \bar{Z}_n$$

for $p' < p$. In what follows, the dependence of α_j in (2.3) on $\boldsymbol{\phi}_p = (\phi_1, \ldots, \phi_p)$ is sometimes emphasized by writing $\alpha_j = \alpha_j(\boldsymbol{\phi}_p)$. Observe that with this notation $\alpha_j(\phi_1, \ldots, \phi_{p-1}, 0)$ are the corresponding coefficients in the causal representation (2.4) of a MARMA(p-1,q) process.

PROPOSITION 2.3. *Suppose $\{X_n\}$ is the stationary MARMA(p,q) process satisfying (2.1) with $p \geq 2$. Then the order of the autoregression is reducible to $p-1$ if and only if*

$$\alpha_p(\phi_1, \ldots, \phi_{p-1}, 0) = \phi_1\alpha_{p-1} \vee \cdots \vee \phi_{p-1}\alpha_1 \geq \phi_p. \tag{2.9}$$

By (2.8), this in turn is equivalent to

$$\bigvee_{k=2}^{p} \bigvee_{\substack{n_1+\cdots+n_{p-1}=k \\ n_1+2n_2+\cdots+(p-1)n_{p-1}=p}} \prod_{i=1}^{p-1} \phi_i^{n_i} \geq \phi_p.$$

PROOF: From the causal representation of $\{X_n\}$, it is clear that the process is reducible if and only if

$$\alpha_j(\boldsymbol{\phi}_p) = \alpha_j(\phi_1, \dots, \phi_{p-1}, 0) \qquad j = 0, 1, \dots .$$

Setting $j = p$ the necessity of (2.9) is immediate from the recursions (2.3).

Conversely suppose $\phi_1\alpha_{p-1} \vee \cdots \vee \phi_{p-1}\alpha_1 \geq \phi_p$, and we prove $\{\alpha_j(\phi_1, \dots, \alpha_{p-1}, 0)\}$ satisfies (2.3). Because $\alpha_j(\boldsymbol{\phi}_p) = \alpha_j = \alpha_j(\phi_1, \dots, \phi_{p-1}, 0) = \phi_1\alpha_{j-1} \vee \cdots \vee \phi_j\alpha_0$ for $j = 0, 1, \dots, p-1$, it thus suffices to show that for $h \geq 0$

$$\alpha_{p+h} = \phi_1\alpha_{p+h-1} \vee \cdots \vee \phi_{p-1}\alpha_{h+1}$$

or that

$$\phi_1\alpha_{p+h-1} \vee \cdots \vee \phi_{p-1}\alpha_{h+1} \geq \phi_p\alpha_h. \tag{2.10}$$

For $h = 0$, this follows by assumption so now assume the validity of (2.10) for $0 \leq h \leq n$. Then since $\alpha_j \geq \phi_p\alpha_{j-p}$. we have with $h = n + 1$

$$\begin{aligned}
\phi_1\alpha_{p+n} \vee \cdots \vee \phi_{p-1}\alpha_{n+2} &\geq \phi_1\phi_p\alpha_n \vee \cdots \vee \phi_{p-1}\phi_p\alpha_{n+2-p} \\
&= \phi_p(\phi_1\alpha_n \vee \cdots \vee \phi_{p-1}\alpha_{n+2-p}) \\
&= \phi_p\alpha_{n+1} \qquad \text{(by induction).}
\end{aligned}$$

This finishes the proof. ∎

An outgrowth of the argument just presented is that if $\{X_n\}$ satisfies (2.1), the coefficient ϕ_p is not necessarily uniquely determined. For example, any ϕ_p satisfying the inequality (2.9) gives rise to the same MARMA(p-1,q) defining equations. We may also consider the analogous questions relative to the other autoregressive coefficients. The coefficient ϕ_i where $2 \leq i \leq p$ is not uniquely determined whenever $\phi_i X_{n-i} \leq \phi_j X_{n-j}$ for some $j < i$. In this case, using the same argument as above, the MARMA(p,q) equations (2.1) may be reduced to

$$X_n = \phi_1 X_{n-1} \vee \cdots \vee \phi_{i-1}X_{n-i+1} \vee \phi_{i+1}X_{n-i-1} \vee \phi_p X_{n-p}\bar{Z}_n$$

for some $2 \leq i \leq p$ if and only if

$$\phi_1\alpha_{i-1} \vee \cdots \vee \phi_{i-1}\alpha_1 \geq \phi_i \tag{2.11}$$

or equivalently if

$$\bigvee_{k=1}^{i} \bigvee_{\substack{n_1+\cdots+n_{i-1}=k \\ n_1+2n_2+\cdots+(i-1)n_{i-1}=i}} \prod_{j=1}^{i-1} \phi_j^{n_j} \geq \phi_i.$$

This is exactly the condition that a MAR(i) process with parameter values $\phi_1, \ldots, \phi_i$ can be reduced to an MAR(i-1) process.

Examples: MAR(1). We have

$$X_n = \phi X_{n-1} \vee Z_n$$

and if $\phi < 1$

$$X_n = \bigvee_{j=0}^{\infty} \phi^j Z_{n-j},$$

A solution does not exist if $\phi \geq 1$.

MAR(2). We have

$$X_n = \phi_1 X_{n-1} \vee \phi_2 X_{n-2} \vee Z_n.$$

Assume $\phi_1 \vee \phi_2 < 1$. The coefficients $\{\psi_j\}$ in Proposition 2.2 are

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 &= \phi_1 \\ \psi_2 &= \phi_1^2 \vee \phi_2 \\ \psi_3 &= \phi_1^3 \vee \phi_1 \phi_2 \end{aligned}$$

and so on. The process is reducible if and only if $\phi_1^2 \geq \phi_2$ and if $\phi_2 > \phi_1^2$, then the ψ_j become

$$\psi_k = \begin{cases} \phi_1 \phi_2^{[k/2]}, & \text{if k is odd} \\ \phi_2^{k/2}, & \text{if k is even.} \end{cases}$$

MAR(3). We have

$$X_n = \phi_1 X_{n-1} \vee \phi_2 X_{n-2} \vee \phi_3 X_{n-3} \vee Z_n$$

and if $\phi_1 \vee \phi_2 \vee \phi_3 < 1$, the coefficients $\psi_j^{'}$s are

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 &= \phi_1 \\ \psi_2 &= \phi_1^2 \vee \phi_2 \\ \psi_3 &= \phi_1^3 \vee \phi_1 \phi_2 \vee \phi_3 \\ \psi_j &= \phi_1 \psi_{j-1} \vee \phi_2 \psi_{j-2} \vee \phi_3 \psi_{j-3} \qquad \text{for } j \geq 3. \end{aligned}$$

From (2.11), the coefficent ϕ_2 may be replaced by zero if and only if $\phi_1^2 \geq \phi_2$ and the process is reducible if and only if $\phi_1^3 \vee \phi_1 \phi_2 \geq \phi_3$.

The reducibility of the moving average component of a MARMA(p,q) process can also be addressed. It is clear that θ_i, for some $1 \le i \le q$, is not uniquely determined in (2.1) whenever $\theta_i Z_{n-i} \le \phi_1 X_{n-1} \vee \cdots \vee \phi_p X_{n-p}$. A necessary and sufficient condition for this to occur is that

$$\bigvee_{k=0}^{i-1} \alpha_{i-k}\theta_k \ge \theta_i, \tag{2.12}$$

where α_j satisfies the recursions (2.3). If the equations (2.1) do not depend on θ_i, then ψ_i in (2.5) must be independent of θ_i which is exactly (2.12). On the other hand if (2.12) holds, then we have from Proposition 2.2

$$\begin{aligned} X_n = \bigvee_{j=1}^{p} \phi_j X_{n-j} \vee \bigvee_{j=0}^{q} \theta_j Z_{n-j} &= \bigvee_{j=1}^{p} \phi_j X_{n-j} \vee \bigvee_{j=0}^{q} \theta_j Z_{n-j} \vee \psi_i Z_{n-i} \\ &= \bigvee_{j=1}^{p} \phi_j X_{n-j} \vee \bigvee_{\substack{j=0\\ j\neq i}}^{q} \theta_j Z_{n-j} \vee (\bigvee_{k=0}^{i-1} \alpha_{i-k}\theta_k) Z_{n-i} \\ &= \bigvee_{j=1}^{p} \phi_j X_{n-j} \vee \bigvee_{\substack{j=0\\ j\neq i}}^{q} \theta_j Z_{n-j} \end{aligned}$$

which does not depend on θ_i as asserted.

3. A metric for max-stable random variables.

In this section, we introduce a metric for max-stable random variables which will be the basis for prediction in the next section. Suppose (X,Y) has a max-stable distribution with marginals Φ_{1,σ_X} and Φ_{1,σ_Y}, respectively, which, by Proposition 5.11 in Resnick (1987), implies there exist non-negative integrable functions f_X and f_Y such that

$$\int_0^1 f_X(s)\,ds = \sigma_X, \quad \int_0^1 f_Y(s)\,ds = \sigma_Y \tag{3.1}$$

and

$$P[X \le x, Y \le y] = \exp\{-\int_0^1 \frac{f_X(s)}{x} \vee \frac{f_Y(s)}{y}\,ds\}. \tag{3.2}$$

There is an obvious extension of this relation to more than two max-stable rv's.

We now define the distance between X and Y by

$$d(X,Y) = \int_0^1 |f_X(s) - f_Y(s)|\,ds. \tag{3.3}$$

Since f_X and f_Y are not uniquely determined in the representation (3.2) (see de Haan and Pickands (1986)) it is not immediately obvious that $d(\cdot,\cdot)$ is a well defined function. To see that $d(\cdot,\cdot)$ is indeed well defined, observe that from (3.2)

$$P[X \vee Y \le x] = \Phi_{1,\sigma_{X\vee Y}}(x), \tag{3.4}$$

where $\sigma_{X \vee Y} = \int_0^1 f_X(s) \vee f_Y(s)\,ds$. Now if $Z \sim \Phi_{1,\sigma}$ then Z^{-1} is exponential with mean σ^{-1} and thus from the identity, $|a-b| = 2(a \vee b) - a - b$, we see that

$$\begin{aligned} \int_0^1 |f_X(s) - f_Y(s)|\,ds &= 2\sigma_{X \vee Y} - \sigma_X - \sigma_Y \\ &= 2(E(X \vee Y)^{-1})^{-1} - (EX^{-1})^{-1} - (EY^{-1})^{-1} \end{aligned} \tag{3.5}$$

and so $d(\cdot,\cdot)$ is unambiguously defined in terms of the moments $(X \vee Y)^{-1}$, X^{-1}, and Y^{-1}.

It is clear from (3.3) that $d(\cdot,\cdot)$ satisfies the triangle inequality; namely if (X, Y, Z) is max-stable, then

$$d(X,Z) \leq d(X,Y) + d(Y,Z).$$

In addition, setting $x = y$ in (3.2), we see that $X = Y$ a.s. if and only if $\sigma_{X \vee Y} = \sigma_X = \sigma_Y$ and hence from (3.5)

$$d(X,Y) = 0 \quad \text{if and only if} \quad X = Y \ a.s.$$

Thus $d(\cdot,\cdot)$ is a metric when applied to any collection of jointly max-stable random variables with one dimensional marginals $\Phi_{1,\sigma}$.

Using the metric d, a measure of dependence between a pair of max-stable random variables X and Y can be defined by

$$\rho(X,Y) = 1 - \frac{1}{2} d\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right).$$

This measure of dependence was originally proposed by Tiago de Oliveira (1962) and has the following properties:

(1) $0 \leq \rho(X,Y) \leq 1$.

(2) $\rho(X,Y) = 0$ if and only if X and Y are independent.

(3) $\rho(aX, bY) = \rho(X,Y)$ for all positive constants a and b.

(4) $\rho(X,Y) = 1$ if and only if $X = cY$ a.s. for some positive constant c.

Properties (1), (3), and (4) are immediate from the definitions of ρ and d. As for (2), it is easy to see from (3.2) that X and Y are independent if and only if $\int_0^1 \frac{f_X(s)}{x} \vee \frac{f_Y(s)}{y}\,ds = \int_0^1 \frac{f_X(s)}{x}\,ds + \int_0^1 \frac{f_Y(s)}{y}\,ds$ for all x and y which is equivalent to f_X and f_Y being supported on disjoint sets a.e. (de Haan (1984), de Haan and Pickands (1986), and p.292 of Resnick (1987)). This in turn is equivalent to

$$d\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) = \int_0^1 \left|\frac{f_X(s)}{\sigma_X} - \frac{f_Y(s)}{\sigma_Y}\right| ds = 2$$

which establishes (2).

As the following theorem indicates, the metric d may be roughly interpreted as the asymptotic scale of $|X - Y|$. This is analogous to the dispersion metric for random variables which are jointly (sum) stable (see Cline and Brockwell (1985) and Section 4, Remarks 1 and 4).

THEOREM 3.1. *Suppose (X,Y) is max-stable with joint distribution given by (3.2). Then*

$$xP[|X-Y|>x] \to d(X,Y)$$

as $x \to \infty$.

COROLLARY 3.2. *Suppose the distribution of the nonnegative random vector (U,V) is in the domain of attraction of the max-stable distribution of (X,Y) with normalizing constants (a_n, a_n), ie.,*

$$P^n[a_n^{-1}U \le x,\, a_n^{-1}V \le y] \to P[X \le x,\, Y \le y] = \exp\{-\int_0^1 \frac{f_X(s)}{x} \vee \frac{f_Y(s)}{y}\, ds\}$$

Then

$$nP[|U-V|>a_n] \to d(X,Y) = \int_0^1 |f_X(s) - f_Y(s)|\, ds.$$

PROOF OF COROLLARY: By assumption we have

$$nP[a_n^{-1}(U,V) \in \cdot] \xrightarrow{v} \nu(\cdot)$$

where $\xrightarrow{v}$ denotes vague convergence on the space $[0,\infty]^2 \setminus \{(0,0)\}$ and ν is the exponent measure of the distribution of (X,Y) (see Chapter 5, Resnick (1987)). Since the set $\{(x,y) : |x-y|>1\}$ is bounded away from the origin and $\nu(\partial\{(x,y) : |x-y|>1\}) = 0$, it follows from the vague convergence that

$$nP[a_n^{-1}|U-V|>1] \to \nu\{(x,y) : |x-y|>1\}.$$

But also

$$xP[x^{-1}|X-Y|>1] \to \nu\{(x,y) : |x-y|>1\}$$

so that by the theorem,

$$d(X,Y) = \nu\{(x,y) : |x-y|>1\}$$

which proves the corollary. ∎

The proof of Theorem 3.1 is based on a representation for (X,Y) in terms of a Poisson process (cf. de Haan (1984) and p.268 of Resnick (1987)). Let $\{\Gamma_n, n \ge 1\}$ be the points of a homogeneous Poisson process with rate 1 so that

$$\Gamma_n = E_1 + \ldots + E_n$$

where $\{E_i\}$ is a sequence of iid unit exponentials. Suppose $\{U_i\}$ is a sequence of iid uniform $(0,1)$ rv's, independent of $\{\Gamma_n\}$. Then

$$(X,Y) \stackrel{d}{=} (\bigvee_{i=1}^{\infty} \frac{f_X(U_i)}{\Gamma_i}, \bigvee_{i=1}^{\infty} \frac{f_Y(U_i)}{\Gamma_i}).$$

To verify this representation, note that $\{(U_i,\Gamma_i),\ i \geq 1\}$ are the points of a Poisson process on $[0,1] \times [0,\infty)$ with intensity measure $\mu(du, ds) = 1_{[0,1]}(u)\,du \times 1_{[0,\infty)}(s)\,ds$ (see Proposition 3.8 in Resnick (1987)). Hence

$$\begin{aligned} P[\bigvee_{i=1}^{\infty} \frac{f_X(U_i)}{\Gamma_i} \leq x,\ \bigvee_{i=1}^{\infty} \frac{f_Y(U_i)}{\Gamma_i} \leq y] \\ &= P[\{(U_i,\Gamma_i), i \geq 1\} \cap \{(u,s) : \frac{f_X(u)}{s} > x \text{ or } \frac{f_Y(u)}{s} > y\} = \phi] \\ &= \exp\{-\mu\{(u,s) : f_X(u) > sx \text{ or } f_Y(u) > sy\}\} \\ &= \exp\{-\mu\{(u,s) : 0 \leq s < \frac{f_X(u)}{x} \vee \frac{f_Y(u)}{y}\} \\ &= \exp\{-\int_0^1 \frac{f_X(u)}{x} \vee \frac{f_Y(u)}{y}\, du\} \end{aligned}$$

as asserted.

Proof of Theorem 3.1: Observe first that for $x > 0$,

$$\bigcap_{i=1}^{\infty}\{\frac{f_X(U_i)}{\Gamma_i} \leq \frac{f_Y(U_i)}{\Gamma_i} + x\} \subseteq \{\bigvee_{i=1}^{\infty} \frac{f_X(U_i)}{\Gamma_i} \leq \bigvee_{i=1}^{\infty} \frac{f_Y(U_i)}{\Gamma_i} + x\}$$

and therefore

$$\begin{aligned} P[X > Y + x] &\leq P[\bigcup_{i=1}^{\infty}\{\frac{f_X(U_i)}{\Gamma_i} > \frac{f_Y(U_i)}{\Gamma_i} + x\}] \\ &= 1 - P[\{(U_i,\Gamma_i),\ i \geq 1\} \cap \{(u,s) : s^{-1} f_X(u) > s^{-1} f_Y(u) + x\} = \phi] \\ &= 1 - \exp\{-\mu\{(u,s) : 0 \leq s < x^{-1}(f_X(u) - f_Y(u))\}\} \\ &= 1 - \exp\{-x^{-1} \int_{[f_X > f_Y]} (f_X(u) - f_Y(u))\, du\} \\ &\leq x^{-1} \int_{[f_X > f_Y]} (f_X(u) - f_Y(u))\, du \end{aligned}$$

Thus we conclude,

$$\limsup_{x \to \infty} xP[|X - Y| > x] \leq \int_0^1 |f_X(u) - f_Y(u)|\, du = d(X,Y). \tag{3.6}$$

For the reverse inequality, observe that

$$\begin{aligned} P[\bigvee_{i=1}^{\infty} \frac{f_X(U_i)}{\Gamma_i} > \bigvee_{i=1}^{\infty} \frac{f_Y(U_i)}{\Gamma_i} + x] \\ \geq P[\frac{f_X(U_1)}{\Gamma_1} > \frac{f_Y(U_1)}{\Gamma_1} + x,\ \bigvee_{i=2}^{\infty} \frac{f_X(U_i)}{\Gamma_i} \leq \frac{f_X(U_1)}{\Gamma_1},\ \bigvee_{i=2}^{\infty} \frac{f_Y(U_i)}{\Gamma_i} \leq \frac{f_Y(U_1)}{\Gamma_1}]. \end{aligned} \tag{3.7}$$

To evaluate this probability, note the first event is $\{x^{-1}(f_X(U_1) - f_Y(U_1)) > \Gamma_1\}$ and so by conditioning on U_1 and Γ_1, the probability on the right hand side of (3.7) is

$$(3.8)\quad \int_{[f_X > f_Y]} \left(\int_{[0, x^{-1}(f_X(u) - f_Y(u))]} e^{-s} P[\bigvee_{i=2}^{\infty} \frac{f_X(U_i)}{s + \Gamma_{i-1}} \le \frac{f_X(u)}{s}, \bigvee_{i=2}^{\infty} \frac{f_Y(U_i)}{s + \Gamma_{i-1}} \le \frac{f_Y(u)}{s}] \, ds \right) du.$$

Now for $y_1, y_2 > 0$,

$$\begin{aligned} P[\bigvee_{i=2}^{\infty} \frac{f_X(U_i)}{s + \Gamma_{i-1}} \le y_1, \bigvee_{i=2}^{\infty} \frac{f_Y(U_i)}{s + \Gamma_{i-1}} \le y_2] & \\ &= P[\bigvee_{i=1}^{\infty} \frac{f_X(U_i)}{s + \Gamma_i} \le y_1, \bigvee_{i=1}^{\infty} \frac{f_Y(U_i)}{s + \Gamma_i} \le y_2] \\ &= P[\{(U_i, \Gamma_i), i \ge 1\} \cap \{(v,t) : \frac{f_X(v)}{s+t} > y_1 \text{ or } \frac{f_Y(v)}{s+t} > y_2\} = \phi] \\ &= \exp\{-\mu\{(v,t) : 0 \le t < (y_1^{-1} f_X(v) \vee y_2^{-1} f_Y(v) - s)\}\} \\ (3.9) \qquad &= \exp\{-\int_0^1 (\frac{f_X(v)}{y_1} \vee \frac{f_Y(v)}{y_2} - s)^+ \, dv\}. \end{aligned}$$

If we replace (y_1, y_2) by $(f_X(u)/s, f_Y(u)/s)$ as dictated by (3.8), then (3.9) becomes

$$\exp\{-sI(u)\}, \text{ where } I(u) = \int_0^1 (\frac{f_X(v)}{f_X(u)} \vee \frac{f_Y(v)}{f_Y(u)} - 1)^+ \, dv.$$

Using this in (3.8), we obtain

$$\begin{aligned} xP[X - Y > x] &\ge x \int_{[f_X > f_Y]} (\int_{[0, x^{-1}(f_X(u) - f_Y(u))]} \exp\{-s(1 + I(u))\} \, ds) \, du \\ &= \int_{[f_X > f_2]} x(\frac{1 - \exp\{-x^{-1}(f_X(u) - f_Y(u))(1 + I(u)\}}{1 + I(u)}) \, du, \end{aligned}$$

so that by Fatou's lemma,

$$\liminf_{x \to \infty} xP[X - Y > x] \ge \int_{[f_X > f_Y]} (f_X(u) - f_Y(u)) \, du.$$

It thus follows

$$\liminf_{x \to \infty} xP[|X + Y| > x] \ge \int_0^1 |f_X(u) - f_Y(u)| \, du = d(X, Y)$$

which together with (3.6) proves the theorem. ∎

4. Prediction of max-linear processes.

In this section, we consider prediction of an observation from a max-linear process in terms of the infinite past of the process, and then specialize to prediction based on the finite past for MARMA(p,q) processes. The predictor will be restricted to the class of max-linear combinations of past observations and will be chosen so as to minimize the distance (in terms of the metric $d(\cdot,\cdot)$ of Section 3) between the observation and the predictor. This prediction criterion is similar in spirit to the dispersion criterion adopted by Cline and Brockwell (1985) for the prediction of linear processes with infinite variance. The minimum dispersion predictor has the property of minimizing the probabilities of large prediction errors. As we shall see in Remarks 1 and 4 below, the same interpretation is also valid for our criterion. Before embarking on the formal prediction set-up, we first record some elementary properties of max-linear processes.

PROPOSITION 4.1. *Let $\{Z_n\}$ be iid with common distribution function $\Phi_{1,\sigma}(x)$, $\sigma > 0$. Define*

$$X = \bigvee_{j=0}^{\infty} \alpha_j Z_j, \; Y = \bigvee_{j=0}^{\infty} \beta_j Z_j$$

where $\alpha_j \geq 0$, $\beta_j \geq 0$ with $\sum_j \alpha_j < \infty$, $\sum_j \beta_j < \infty$. Then the distribution (X,Y) has representation (3.2) (ie. (X,Y) is max-stable) with

$$f_X(s) = \sigma \sum_{j=0}^{\infty} \alpha_j 2^{j+1} 1_{(2^{-j-1}, 2^{-j}]}(s) \text{ and } f_Y(s) = \sigma \sum_{j=0}^{\infty} \beta_j 2^{j+1} 1_{(2^{-j-1}, 2^{-j}]}(s), \tag{4.1}$$

and

$$d(X,Y) = \sigma \sum_{j=0}^{\infty} |\alpha_j - \beta_j|. \tag{4.2}$$

Moreover,

$$X = Y \text{ a.s. if and only if } \alpha_j = \beta_j \text{ for all } j.$$

PROOF: We have

$$\begin{aligned} P[X \leq x, Y \leq y] &= P[\bigcap_{j=0}^{\infty}(\{\alpha_j Z_j \leq x\} \cap \{\beta_j Z_j \leq y\})] \\ &= \prod_{j=0}^{\infty} P[Z_j \leq \frac{x}{\alpha_j} \wedge \frac{y}{\beta_j}] \\ &= \exp\{-\sigma \sum_{j=0}^{\infty} \frac{\alpha_j}{x} \vee \frac{\beta_j}{y}\}. \end{aligned}$$

But with f_X and f_Y as defined by (4.1)

$$\int_0^1 \frac{f_X(s)}{x} \vee \frac{f_Y(s)}{y} \, ds = \sigma \sum_{j=0}^{\infty} \frac{\alpha_j}{x} \vee \frac{\beta_j}{y}$$

and hence the distribution of (X,Y) has the desired representation (3.2). Equation (4.2) is now immediate from the definition of d.

As for the last assertion of the proposition, note that $d(X,Y) = \sigma \sum_{j=0}^{\infty} |\alpha_j - \beta_j| = 0$ iff $\alpha_j = \beta_j$ for all j and since d is a metric, this is true iff $X = Y$ a.s. ∎

REMARK 1: With X and Y defined as in Proposition 4.1, we have from Theorem 3.1 and the above proposition (also see Remark 4 below)

$$\lim_{x\to\infty} \frac{P[|X - Y| > x]}{P[|Z_1| > x]} = \sigma^{-1} d(X,Y) = \sum_{j=0}^{\infty} |\alpha_j - \beta_j|.$$

Now to establish the connection between this metric and the dispersion metric discussed in Cline and Brockwell (1985), suppose

$$U = \sum_{j=0}^{\infty} \alpha_j W_j, \ V = \sum_{j=0}^{\infty} \beta_j W_j$$

where $\{W_j\}$ is iid and W_1 belongs to the domain of attraction of a stable random variable with index α, $0 < \alpha < 2$. Then under suitable summability conditions on $\{\alpha_j\}$and $\{\beta_j\}$, we have from a theorem of Cline (1983),

$$\lim_{x\to\infty} \frac{P[|U - V| > x]}{P[|W_1| > x]} = \sum_{j=0}^{\infty} |\alpha_j - \beta_j|^{\alpha} =: dispersion(U - V).$$

Consequently, our prediction criterion for max-linear processes is the natural analogue of the dispersion metric in the stable linear process setting. ∎

To set up the prediction problem for max-linear processes, let $\{X_t\}$ be the (causal) max-linear process

$$X_t = \bigvee_{j=0}^{\infty} \psi_j Z_{t-j} \tag{4.3}$$

where $\{Z_t\}$ is iid with $Z_1 \sim \Phi_{1,\sigma}$, $\psi_j \geq 0$, and $\sum_j \psi_j < \infty$. Let $\mathcal{H}$ be the class of random variables

$$\mathcal{H} = \{ \bigvee_{j=-\infty}^{\infty} \alpha_j Z_j : \ \alpha_j \geq 0, \sum_j \alpha_j < \infty \}.$$

Fix $n \geq 1$ and for each $Y \in \mathcal{H}$, define the set

$$P_n^* Y = \{ \bigvee_{j=1}^{\infty} b_j X_{n+1-j} : \ d(Y, \bigvee_{j=1}^{\infty} b_j X_{n+1-j}) \text{ is minimum} \}.$$

The following proposition, which parallels Lemma 2.1 in Cline and Brockwell (1985), shows that on a subset of $\mathcal{H}$, P_n^* is uniquely determined (i.e. consists of one element) and is max-linear.

PROPOSITION 4.2. *Let $\{X_t\}$ be as specified in (4.3) above.*
(i) If $Y = \bigvee_{j=-\infty}^{\infty} \alpha_j Z_j \in \mathcal{H}$, then

$$P_n^* Y = P_n^* (\bigvee_{j=-\infty}^{n} \alpha_j Z_j).$$

(ii) Let S_n^* be the subset of $\mathcal{H}$ consisting of the rv's

$$Y = \bigvee_{j=n+1}^{\infty} \alpha_j Z_j \vee \bigvee_{j=1}^{\infty} \beta_j X_{n+1-j} \tag{4.4}$$

where $\beta_j \geq 1$, $\sum_j \beta_j < \infty$. Then for each $Y \in S_n^*$, $P_n^* Y$ is uniquely determined and

$$P_n^*(Y) = \bigvee_{j=1}^{\infty} \beta_j X_{n+1-j}$$

with error of prediction $d(Y, P_n^* Y) = \sigma \sum_{j=n+1}^{\infty} \alpha_j$. Moreover, the mapping $Y \to P_n^* Y$ is max-linear on S_n^* in the sense that $P_n^*(aY \vee bZ) = (aP_n^* Y) \vee (bP_n^* Z)$ for all $a,\ b > 0$.

PROOF: For $b_j \geq 1$, $\sum_j b_j < \infty$, we have

$$\begin{aligned}
\bigvee_{j=1}^{\infty} b_j X_{n+1-j} &= \bigvee_{j=1}^{\infty} b_j \bigvee_{k=0}^{\infty} \psi_k Z_{n+1-j-k} \\
&= \bigvee_{j=1}^{\infty} b_j \bigvee_{i=j}^{\infty} \psi_{i-j} Z_{n+1-i} \\
&= \bigvee_{i=1}^{\infty} (\bigvee_{j=1}^{i} b_j \psi_{i-j}) Z_{n+1-i} \\
&= \bigvee_{i=1}^{\infty} a_i(\mathbf{b}) Z_{n+1-i}
\end{aligned}$$

where

$$a_i(\mathbf{b}) = \bigvee_{j=1}^{i} b_j \psi_{i-j}. \tag{4.5}$$

Now applying Proposition 4.1 with $Y = \bigvee_{j=-\infty}^{\infty} \alpha_j Z_j$, we have

$$\begin{aligned}
d(Y, \bigvee_{j=1}^{\infty} b_j X_{n+1-j}) &= d(\bigvee_{j=n+1}^{\infty} \alpha_j Z_j \vee \bigvee_{i=1}^{\infty} \alpha_{n+1-i} Z_{n+1-i}, \bigvee_{i=1}^{\infty} a_i(\mathbf{b}) Z_{n+1-i}) \\
&= \sigma(\sum_{j=n+1}^{\infty} \alpha_j + \sum_{i=1}^{\infty} |\alpha_{n+1-i} - a_i(\mathbf{b})|) \\
&= \sigma \sum_{j=n+1}^{\infty} \alpha_j + d(\bigvee_{j=-\infty}^{n} \alpha_j Z_j, \bigvee_{j=1}^{\infty} b_j X_{n+1-j})
\end{aligned}$$

from which (i) is now immediate.

As for (ii), suppose Y is given by (4.4). From the calculation above with $\alpha_{n+1-i} = a_i(\boldsymbol{\beta})$, $i \geq 1$, where $a_i(\cdot)$ is given by (4.5), we have

$$d(Y, \bigvee_{j=1}^{\infty} b_j X_{n+1-j}) = \sigma(\sum_{j=n+1}^{\infty} \alpha_j + \sum_{i=1}^{\infty} |a_i(\boldsymbol{\beta}) - a_i(\mathrm{b})|$$
$$\geq \sigma(\sum_{j=n+1}^{\infty} \alpha_j).$$

Equality holds if and only if $a_i(\boldsymbol{\beta}) = a_i(\mathrm{b})$, $i = 1, 2, \ldots$ in which case, by appealing to Proposition 4.1 once again, we conclude that there is a unique element of $P_n^* Y$ which is given by $\bigvee_{i=1}^{\infty} a_i(\boldsymbol{\beta}) Z_{n+1-i} = \bigvee_{j=1}^{\infty} \beta_j X_{n+1-j}$ as asserted. It is also clear that $d(Y, P_n^* Y) = \sigma \sum_{j=n+1}^{\infty} \alpha_j$.

The max-linearity is immediate from the form of Y and $P_n^* Y$. ∎

REMARK 2: If $Y \notin S_n^*$, then $P_n^* Y$ need not be unique (see the MARMA(1,1) example below). Also if $Y \in \mathcal{H}$ is independent of $X_n, X_{n-1}, \ldots$, then $P_n^* Y = 0$. ∎

REMARK 3: Proposition 4.2 remains valid if we replace S_n^* by

$$S_n = \{Y = \bigvee_{j=n+1}^{\infty} \alpha_j Z_j \vee \bigvee_{j=1}^{n} \beta_j X_{n+1-j} : \ \alpha_j \geq 0, \beta_j \geq 0, \sum_j \alpha_j < \infty\}$$

and $P_n^* Y$ by

$$P_n Y = \{\bigvee_{j=1}^{n} b_j X_{n+1-j} : d(Y, \bigvee_{j=1}^{n} b_j X_{n+1-j}) \text{ is minimum}\}.$$

In particular,

$$P_n(\bigvee_{j=-\infty}^{\infty} \alpha_j Z_j) = P_n(\bigvee_{j=-\infty}^{n} \alpha_j Z_j)$$

and

$$P_n(\bigvee_{j=n+1}^{\infty} \alpha_j Z_j \vee \bigvee_{j=1}^{n} \beta_j X_{n+1-j}) = \bigvee_{j=1}^{n} \beta_j X_{n+1-j}. \qquad ∎$$

REMARK 4: As in Section 2 we can extend the above results to the case when the distribution F of Z_1 satisfies $F(0-) = 0$ and

$$F^n(a_n x) \to \Phi_{1,\sigma}(x).$$

In particular let $\tilde{\mathcal{H}}$ be the class or rv's

$$\tilde{\mathcal{H}} = \{ \bigvee_{j=-\infty}^{\infty} \alpha_j Z_j : \ \alpha_j \geq 0 \text{ and } \sum_j \alpha_j^{\delta} < \infty \text{ for some } \delta < 1\}$$

and define the distance between the two rv's $U = \bigvee_j \alpha_j Z_j,\ V = \bigvee_j \beta_j Z_j \in \tilde{\mathcal{H}}$ by

$$d(U,V) := \sigma \sum_{j=-\infty}^{\infty} |\alpha_j - \beta_j|.$$

Then d is a metric on $\tilde{\mathcal{H}}$ and since (U,V) belongs to the domian of attraction of the max-stable distribution described in Proposition 4.1, we have by Corollary 3.2 and Proposition 4.1

$$nP[|U-V| > a_n] \to d(U,V)$$

or, equivalently,

$$\frac{P[|U-V| > x]}{P[Z > x]} \to \sigma^{-1} d(U,V). \quad \blacksquare$$

EXAMPLES:

MAR(p). Suppose $\{X_t\}$ is the MAR(p) process satisfying

$$X_t = \phi_1 X_{t-1} \vee \cdots \vee \phi_p X_{t-p} \vee Z_t.$$

and write $\hat{X}_{n+k} = P_n X_{n+k}$ for $k \geq 1$. Then using max-linearity, we have that for $n \geq p$, this predictor satisfies the recursion

$$\hat{X}_{n+k} = \phi_1 \hat{X}_{n+k-1} \vee \cdots \vee \phi_p \hat{X}_{n+k-p}$$

with initial conditions $\hat{X}_j = X_j,\ j = 1, \ldots, n$. The error of prediction is

$$d(X_{n+k}, \hat{X}_{n+k}) = \sigma(\sum_{j=0}^{k-1} \psi_j)$$

where $\{\psi_j\}$ are the coefficients in the causal representation of $\{X_t\}$ (see Proposition 2.2). As expected

$$d(X_{n+k}, \hat{X}_{n+k}) \to \sigma_{X_1} = \sigma(\sum_{j=0}^{\infty} \psi_j)$$

as $k \to \infty$.

MARMA(1,1). In this case, let $\{X_t\}$ be the causal MARMA(1,1) process satisfying the recursion

$$X_t = \phi X_{t-1} \vee Z_t \vee \theta Z_{t-1}$$

where $0 \leq \phi < 1$. We further assume that $\phi < \theta$ since otherwise $\{X_t\}$ is reducible to an MAR(1) (see (2.12)). Because Z_{n+1} is independent of $X_n, \ldots, X_1$, it follows from causality and Remark 3 that

$$\hat{X}_{n+1} := P_n(\phi X_n \vee Z_{n+1} \vee \theta Z_n) = P_n(\phi X_n \vee \theta Z_n).$$

However, P_n is not necessarily max-linear on this expression since $\phi X_n \vee \theta Z_n \notin S_n$. To compute $P_n(\phi X_n \vee \theta Z_n)$, write

$$X_n = \bigvee_{j=0}^{\infty} \psi_j Z_{n-j}$$

where $\{\psi_j\}$ is as specified in Proposition 2.2, namely

$$\psi_0 = 1 \text{ and } \psi_j = \theta\phi^{j-1} \quad \text{for } j \geq 1.$$

Now with $a_i(\mathbf{b}) = \bigvee_{j=1}^{i\wedge n} b_j \psi_{i-j}$, we have by Proposition 4.1,

$$\begin{aligned} d(\phi X_n \vee \theta Z_n, \bigvee_{j=1}^{n} b_j X_{n+1-j}) &= d((\phi \vee \theta) Z_n \vee \bigvee_{i=2}^{\infty} (\phi\psi_{i-1} Z_{n+1-i}), \bigvee_{i=1}^{\infty} a_i(\mathbf{b}) Z_{n+1-i}) \\ &= \sigma(|\theta - b_1| + \sum_{i=2}^{\infty} |\phi\psi_{i-1} - a_i(\mathbf{b})|) \\ &=: m(\mathbf{b}). \end{aligned}$$

If $\phi \leq b_1$, then $a_i(\mathbf{b}) \geq \phi\psi_{i-1}$ and hence

$$\begin{aligned} m(\mathbf{b}) &\geq \sigma(|\theta - b_1| + \sum_{i=2}^{\infty} \psi_{i-1}(b_1 - \phi)) \\ &= \sigma(|\theta - b_1| + \frac{\theta}{1-\phi}(b_1 - \phi)) \\ &= m(b_1, 0, \ldots, 0) \\ &\geq \begin{cases} m(\theta, 0, \ldots, 0) & \text{if } \theta < 1 - \phi \\ m(\phi, 0, \ldots, 0) & \text{if } \theta > 1 - \phi. \end{cases} \end{aligned}$$

On the other hand if $b_1 < \phi$, then

$$\begin{aligned} m(\mathbf{b}) &\geq \sigma(\theta - b_1) \\ &\geq \sigma(\theta - \phi) = m(\phi, 0, \ldots, 0). \end{aligned}$$

Combining both cases, we conclude that

$$P_n X_{n+1} = \begin{cases} \theta X_n & \text{if } \phi + \theta < 1 \\ \phi X_n & \text{if } \phi + \theta > 1. \end{cases}$$

For the case $\phi + \theta = 1$, the predictor is not unique and in fact can be chosen as $b_1 X_n$, for $\phi \leq b_1 \leq \theta$. For prediction of further lags ahead, we have by iterating backwards and using Remark 3

$$P_n X_{n+h} = \phi^{h-1} P_n X_{n+1}$$

for $h \geq 1$ with prediction error given by

$$d(X_{n+h}, P_n X_{n+h}) = d(\phi^{h-1} X_{n+1} \vee \bigvee_{j=1}^{h-1} \psi_j Z_{n+h-j}, \phi^{h-1} P_n X_{n+1})$$

$$= \sigma \sum_{j=1}^{h-1} \psi_j + \phi^{h-1} d(X_{n+1}, P_n X_{n+1})$$

where $d(X_{n+1}, P_n X_{n+1})$ is $\sigma(1 + (\theta - \phi))$ if $\phi + \theta \leq 1$ and is equal to $\sigma(1 + \theta(\theta - \phi)/(1 - \phi))$ otherwise.

Finally, we note that as long as $\phi + \theta \neq 1$, then the operator P_n is indeed max-linear for the term $\phi X_n \vee \theta Z_n$. To see this, clearly $P_n X_n = X_n$ and an argument similar to the one given above yields

$$P_n(Z_n) = \begin{cases} X_n & \text{if } \phi + \theta < 1 \\ 0 & \text{if } \phi + \theta > 1. \end{cases}$$

Therefore $P_n(\phi X_n \vee \theta Z_n) = \phi P_n X_n \vee \theta P_n Z_n$ as asserted.

5. Estimation

In this section we make some brief remarks concerning the estimation of the parameter vector $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$ of the MARMA(p,q) process

$$X_t = \phi_1 X_{t-1} \vee \cdots \vee \phi_p X_{t-p} \vee Z_t \vee \theta_1 Z_{t-1} \vee \cdots \theta_q Z_{t-q}$$

based on the observations $X_1, \dots, X_n$. We first estimate ϕ_j. Observe that

$$\frac{X_t}{X_{t-j}} \geq \phi_j \qquad \text{a.s.}$$

with equality holding if and only if

$$X_t = \phi_j X_{t-j}.$$

If ϕ_j is identifiable, or in the sense of Section 2, uniquely determined, then $P[X_t = \phi_j X_{t-j}] > 0$. Consequently, by the ergodic theorem, $P[X_t = \phi_j X_{t-j}, \text{ i.o.}] = 1$ and hence with probability one

$$\hat{\phi}_j := \bigwedge_{t=j+1}^{n} \frac{X_t}{X_{t-j}} = \phi_j$$

for n sufficiently large.

Now to estimate the $\theta_j' s$, assume for simplicity that $q = 1$ and that $\phi_1 < \theta_1$ (see (2.12)). Let A be the event

$$(5.1) \qquad A = \big[Z_t \geq \theta_1^{-1}[(\phi_1 \vee \phi_2) X_{t-1} \vee \cdots \vee (\phi_{p-1} \vee \phi_p) X_{t+1-p} \vee \phi_p X_{t-p}] \vee (\theta_1 Z_{t-1}) \vee Z_{t+1}\big]$$

if $\theta_1 \leq 1$, otherwise replace the leading coefficient θ_1^{-1} on the right hand side by 1. Since the rv's on both sides of the inequality in (5.1) are independent and finite a.s., it follows that $P[A] > 0$. Moreover on A we have

$$Z_t = X_t$$

and since $\theta_1 > \phi_1$

$$\theta_1 Z_t \geq \phi_1 X_t \vee \phi_2 X_{t-1} \vee \cdots \vee \phi_p X_{t+1-p}$$

which implies $X_{t+1} = \theta_1 Z_t$. Thus, $P[X_t = Z_t,\, X_{t+1} = \theta_1 X_t] > 0$ so that by the ergodic theorem

$$P[X_{t+1} = \theta_1 X_t, \text{ i.o.}] = 1. \tag{5.2}$$

Having already correctly identified $\phi_1, \ldots, \phi_p$, we consider those t for which

$$X_t \neq \phi_1 X_{t-1} \vee \cdots \vee \phi_p X_{t-p} \qquad \text{and} \qquad X_{t+1} \neq \phi_1 X_t \vee \cdots \vee \phi_p X_{t+1-p}. \tag{5.3}$$

For such t, we search for equality between any two ratios, X_{t+1}/X_t, and then estimate θ_1 to be their common value. Because of (5.2), it follows that with probability one, this procedure correctly identifies θ_1 for n large.

This estimation procedure, when applied to the simulated MARMA processes displayed in Figures 1, 3, and 5, correctly identified all of the parameters. For the MARMA(2,1) process, (5.3) held for 164 $t's$ of which 71 had ratios $\frac{X_{t+1}}{X_t} = .6 = \theta_1$. We also found that for the AR(1) process, $X_t = .7X_{t-1} + Z_t$ displayed in Figure 2, $\hat{\phi}_1 = \wedge_{t=1}^{249} \frac{X_{t+1}}{X_t} = .7063$. This is not too surprising since this estimator often performs quite well for AR(1) process with positive innovations (eg. Davis and McCormick (1988)).

It is unreasonable to expect any nonsimulated data to follow a MARMA model exactly. Nevertheless, the estimation procedure described above for the $\phi_j's$ may have desirable properties for a wider class of processes which will hopefully include a more realistic model such as a MARMA with observational noise. These questions and related issues will be addressed in future work.

REFERENCES

Balkema, A.A. and Haan, L. de (1988). Almost sure continuity of stable moving average processes with index less than one. *Ann. Probability, 16,* 334–343.

Cline, D.B.H. (1983). Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data. Ph.D. Dissertation, Dept. of Statistics, Colorado State University, Fort Collins, CO, 80523, USA.

Cline, D.B.H. and Brockwell, P.J. (1985). Linear prediction of ARMA processes with infinite variance. *Stoch. Processes and their Applications, 19,* 281–296.

Davis, R.A. and McCormick, W.P. (1988). Estimation for first–order autoregressive processes with positive or bounded innovations. Preprint.

Haan, L. de (1984). A spectral representation of max-stable processes. *Ann. Probability, 12,* 1194–1204.

Haan, L. de and Pickands, J. (1986). Stationary min-stable stochastic processes. *Probability Theory and Related Fields, 72,* 477–492.

Fama, E. (1965). Behavior of stock market prices. *J. Business, Univ. of Chicago, 38,* 34–105.

Hsing, T. (1986). Extreme value theory for the supremum of rv's with regularly varying tail probabilities. *Stoch. Processes and their Applications, 22,* 51–57.

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

Stuck, B.W. and Kleiner, B. (1974). A statistical analysis of telephone noise. *The Bell System Technical Journal, 53,* 1263– 1320.

Tiago de Oliveira, J. (1962/63). Structure theory of bivariate extremes: Extensions. *Estudos de Math. Estat. Econom., 7,* 165–195.