Stochastic Processes and their Applications 19 (1985) 281-296
North-Holland

LINEAR PREDICTION OF ARMA PROCESSES WITH INFINITE VARIANCE

Daren B. H. CLINE*
Department of Statistics, Texas A&M University, College Station, TX 77483, USA

Peter J. BROCKWELL
Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA

Received 20 December 1983
Revised 29 October 1984

In order to predict unobserved values of a linear process with infinite variance, we introduce a linear predictor which minimizes the dispersion (suitably defined) of the error distribution. When the linear process is driven by symmetric stable white noise this predictor minimizes the scale parameter of the error distribution. In the more general case when the driving white noise process has regularly varying tails with index α, the predictor minimizes the size of the error tail probabilities. The procedure can be interpreted also as minimizing an appropriately defined l_α-distance between the predictor and the random variable to be predicted. We derive explicitly the best linear predictor of X_{n+1} in terms of $X_1, \ldots, X_n$ for the process ARMA(1, 1) and for the process AR(p). For higher order processes general analytic expressions are cumbersome, but we indicate how predictors can be determined numerically.

ARMA process * regular variation * stable process

1. Introduction

We shall be concerned in this paper with prediction of the causal stationary solution $\{X_n\}$ of the ARMA(p, q) equations,

$$X_n - \phi_1 X_{n-1} - \cdots - \phi_p X_{n-p} = W_n + \theta_1 W_{n-1} + \cdots + \theta_q W_{n-q} \tag{1.1}$$

where $\{W_n\}_{n=-\infty}^{\infty}$ is an independently and identically distributed (iid) sequence of random variables, and the polynomials $\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ satisfy the condition

$$\Phi(z)\theta(z) \neq 0 \text{ for all } z \in C \text{ such that } |z| \leq 1. \tag{1.2}$$

It will be assumed throughout that there exists $\alpha > 0$ such that

$$\lim_{t\to\infty} \frac{P[|W_n| > xt]}{P[|W_n| > t]} = x^{-\alpha} \quad \text{for all } x > 0. \tag{1.3}$$

Research partially supported by the National Science Foundation Grant MCS-8202335.
* Work performed in partial fulfillment of Ph.D. requirements at Colorado State University.

The distribution of W_n is then said to have regularly varying tails and the parameter α is called the tail index. If $\alpha < 2$ the variance of W_n does not exist.

A straightforward argument shows that (1.1) has a unique stationary causal solution, namely

$$X_n = \sum_{j=0}^{\infty} \pi_j W_{n-j}, \tag{1.4}$$

where the coefficients $\{\pi_j\}_{j=0}^{\infty}$ are uniquely determined as the coefficients in the power series expansion

$$\sum_{j=0}^{\infty} \pi_j z^j = \Theta(z)/\Phi(z), \quad |z| \leq 1.$$

The argument for this basically follows the classical argument for the finite variance case, with minor adjustments. See, e.g. Yohai and Maronna (1977) for an alternative approach. In addition,

$$X_n - \sum_{j=1}^{\infty} \psi_j X_{n-j} = W_n, \tag{1.5}$$

where

$$1 - \psi_1 z - \psi_2 z^2 - \cdots = \Phi(z)/\Theta(z), \quad |z| \leq 1.$$

Our aim is to predict the values $X_{n+1}, X_{n+2}, \ldots$ in terms of the observed values of $X_1, \ldots, X_n$. We shall restrict attention to linear predictors. The predictor $\hat{Y}$ of a random variable Y will thus have the form

$$\hat{Y} = \boldsymbol{a}' \boldsymbol{X}_n$$

where $\boldsymbol{a}' = (a_1, \ldots, a_n)$ and $\boldsymbol{X}_n = (X_n, X_{n-1}, \ldots, X_1)$.

For ARMA(p, q) processes in which the white noise sequence $\{W_n\}$ has finite variance, predictors are usually determined by minimizing the expected squared error $E(Y - \hat{Y})^2$ (see for example Fuller (1976) and Box and Jenkins (1976)). If the process is Gaussian this procedure also minimizes the probabilities of large deviations $P(|Y - \hat{Y}| > K)$ for every $K > 0$. For processes with infinite variance however, an alternative criterion for selection of a best predictor is needed. Alternative approaches which have been suggested include minimization of the expected absolute error and the pseudo-spectral technique of Cambanis and Soltani (1982). Most criteria however are complicated to use, are of limited applicability and require precise knowledge of the distribution of W_n. This contrasts sharply (but not surprisingly) with the elegant Hilbert-space theory of minimum mean squared error prediction which is applicable when $EW_n^2 < \infty$.

It would be extremely useful, in the infinite variance case, to have a predictor which is reasonably simple to compute, which does not require full knowledge of the distribution of W_n and which (in a sense to be specified) minimizes the prob-

abilities of large prediction errors. In this paper we discuss such a predictor, based on the natural criterion of minimizing error 'dispersion' where dispersion is defined by (1.6) below. This criterion was introduced by Stuck (1978) who used it with considerable success to solve Kalman filtering problems associated with symmetric stable sequences. Blattberg and Sargent (1971), as well as others, have used the dispersion criterion in regression models with stable errors. In the special case when W_n has a symmetric stable distribution (i.e. $E\exp(itW_n) = \exp(-c|t|^\alpha)$, $t \in \mathbb{R}$, $0 < \alpha < 2$), the minimization is equivalent to minimization of the scale parameter of the error distribution. Thus if $\{W_n\}$ is iid symmetric stable with index α and if $\sum_{j=-\infty}^{\infty} |\rho_j|^\alpha < \infty$, then $Y = \sum_{j=-\infty}^{\infty} \rho_j W_j$ is also symmetric stable and in fact,

$$Y \overset{d}{=} \left(\sum_{j=-\infty}^{\infty} |\rho_j|^\alpha \right)^{1/\alpha} W_1.$$

Stuck therefore defines the dispersion of Y (relative to that of W_1) as

$$\operatorname{disp}(Y) = \sum_{j=-\infty}^{\infty} |\rho_j|^\alpha. \tag{1.6}$$

We shall adopt the definition (1.6) for all random variables of the form $Y = \sum_{j=-\infty}^{\infty} \rho_j W_j$, so long as the iid sequence $\{W_n\}$ satisfies (1.3).

The ARMA process $\{X_n\}$ can be expressed (using (1.4)) as the moving average $X_n = \sum_{j=-\infty}^{\infty} \pi_{n-j} W_j$ with $\pi_j = 0$, $j < 0$. Hence $\operatorname{disp}(X_n) = \sum_{j=0}^{\infty} |\pi_j|^\alpha$. If $Y = \sum_{j=-\infty}^{\infty} \rho_j W_j$ then we define the minimum error dispersion linear predictor of Y (based on $X_1, \ldots, X_n$) to be the linear combination $\hat{Y} = a_1 X_n + \cdots + a_n X_1 = \boldsymbol{a}' \boldsymbol{X}_n$ which minimizes

$$\operatorname{disp}(\hat{Y} - Y) = \sum_{j=-\infty}^{\infty} |\rho_j - (a_1 \pi_{n-j} + \cdots + a_n \pi_{1-j})|^\alpha. \tag{1.7}$$

In the special case where $Y = X_{n+k}$ we minimize

$$\operatorname{disp}(\hat{X}_{n+k} - X_{n+k}) = \sum_{j=0}^{k-1} |\pi_j|^\alpha + \sum_{j=k}^{\infty} |\pi_j - (a_1 \pi_{j-k} + \cdots + a_n \pi_{j+1-n-k})|^\alpha. \tag{1.8}$$

For a linear process driven by symmetric stable noise, the prediction error for any linear predictor also has symmetric stable distribution. The minimum dispersion prediction error has the distribution with the smallest scale and hence is optimal. The procedure is easily extended to more general linear processes, since it requires only knowledge of the coefficients of the process and of the tail index α of the noise distribution. Furthermore, by using the following theorem due to Cline (1983) we can relate dispersion to the probability of large error values. A corollary of this is that among linear predictors, the minimum dispersion predictor is optimal in the sense that it minimizes the probability of large prediction errors.

Theorem 1.1. *Suppose* $\{W_j\}$ *are independent and identically distributed and satisfy* (1.3) *and suppose* $Y = \sum_{j=-\infty}^{\infty} \rho_j W_j$ *where* $\sum_{j=-\infty}^{\infty} |\rho_j|^\delta < \infty$ *for some* $\delta < \min(1, \alpha)$. *Then* Y *exists almost surely (is absolutely convergent) and*

$$\lim_{t\to\infty} \frac{P[|Y| > t]}{P[|W_1| > t]} = \operatorname{disp}(Y) = \sum_{j=-\infty}^{\infty} |\rho_j|^\alpha. \qquad \square$$

Since the coefficients $\{\pi_j\}$ in the representation (1.4) are geometrically decreasing in magnitude, this theorem indicates that $\operatorname{disp}(\hat{X}_{n+k} - X_{n+k})$ is roughly proportional to the probability of a large prediction error.

We see from (1.7) that minimization of $\operatorname{disp}(\hat{Y} - Y)$ is equivalent to minimization of a suitably defined l_α-distance between $\hat{Y}$ and Y on the linear space generated by $\{W_n\}$. In the case $\alpha = 2$, $\hat{Y} = P_n Y$ where P_n is the orthogonal projection from the L^2 space $\overline{\operatorname{span}}\{W_n, n \in \mathbb{Z}\}$ into $\operatorname{span}\{X_n, \ldots, X_1\}$, the space generated by linear combinations of $X_n, \ldots, X_1$. With $\alpha < 2$, we can still define an operator so that $\hat{Y} = P_n Y$ but it is not necessarily unique if $\alpha \leq 1$.

We shall see in Sections 2 and 3 that minimum dispersion linear predictors can be found quite explicitly for autoregressive processes and for the mixed ARMA(1, 1) process. In both cases, the prediction operator, P_n is unique and linear on $\operatorname{span}\{X_1, X_2, \ldots\}$. For higher order moving average and mixed processes, however, one cannot always give a single general expression which is acceptable for all values of the parameters. For particular values, determination of the predictor is straightforward. Section 4 discusses the higher order processes.

2. Linear prediction with the infinite past; the AR(p) process

We begin with the simple but important problem of finding an optimal predictor for X_{n+k}, $k \geq 1$, of the form $\sum_{j=1}^{\infty} a_j X_{n+1-j}$, i.e., a linear predictor based on the infinite past. The practical importance of this predictor lies in the fact that for large n its truncation $\sum_{j=1}^{n} a_j X_{n+1-j}$ is approximately minimum dispersion optimal for predicting X_{n+k} on the basis of $X_n, \ldots, X_1$. If $\{X_n\}$ is a pure autoregressive process (AR(p)) and if $n \geq p$, the truncation will in fact be optimal. The results of this section will be seen to be almost identical to the corresponding results for least squares prediction of a finite variance process. First we establish a useful lemma. The sequences $\{\pi_j\}$ and $\{\psi_j\}$ are as in (1.4) and (1.5), respectively.

Lemma 2.1. *Fix* $\delta < \min(1, \alpha)$. *For the* ARMA(p, q) *process* $\{X_n\}$ *satisfying* (1.1)–(1.3) *let* S_* *be the class of random variables of the form*

$$\sum_{j=n+1}^{\infty} \rho_j W_j + \sum_{j=1}^{\infty} \nu_j X_{n+1-j}$$

where

$$\sum_{j=n+1}^{\infty} |\rho_j|^{\delta} < \infty \quad \text{and} \quad \sum_{j=1}^{\infty} |\nu_j|^{\delta} < \infty. \tag{2.1}$$

Then for each $Y \in S_*$, *the set*

$$P_\infty Y = \left\{ \sum_{j=1}^{\infty} a_j X_{n+1-j} : \operatorname{disp}\left(Y - \sum_{j=1}^{\infty} \nu_j X_{n+1-j} \right) \text{ is minimum} \right\}$$

consists of exactly one element. For $Y = \sum_{j=n+1}^{\infty} \rho_j W_1 + \sum_{j=1}^{\infty} \nu_j X_{n+1-j}$, *this element is* $Y^* = \sum_{j=1}^{\infty} \nu_j X_{n+1-j}$. *Furthermore the mapping* $Y \to Y^*$ *is linear on* S_*.

Proof. By Cline (1983), the condition in (2.1) guarantees that each element in S_* exists and has a finite dispersion. Now,

$$\begin{aligned} \operatorname{disp}\left(Y - \sum_{j=1}^{\infty} a_j X_{n+1-j} \right) &= \operatorname{disp}\left(\sum_{j=n+1}^{\infty} \rho_j W_j + \sum_{j=1}^{\infty} (\nu_j - a_j) X_{n+1-j} \right) \\ &= \operatorname{disp}\left(\sum_{j=n+1}^{\infty} \rho_j W_j + \sum_{j=1}^{\infty} \left(\sum_{i=1}^{j} (\nu_i - a_i) \pi_{j-i} \right) W_{n+1-j} \right) \\ &= \sum_{j=n+1}^{\infty} |\rho_j|^{\alpha} + \sum_{j=1}^{\infty} \left| \sum_{i=1}^{j} (\nu_i - a_i) \pi_{j-1} \right|^{\alpha} \geqslant \sum_{j=n+1}^{\infty} |\rho_j|^{\alpha}. \end{aligned} \tag{2.2}$$

Equality holds in (2.2) if and only if $a_j = \nu_j$. Thus the unique element of $P_\infty Y$ is $Y^* = \sum_{j=1}^{\infty} \nu_j X_{n+1-j}$ as asserted. The linearity of the mapping $Y \to Y^*$ is apparent from the form of Y and the form of Y^*. □

We remark that for symmetric stable processes with $\alpha > 1$, we have $Y^* = E[Y \mid X_n, X_{n-1}, X_{n-2}, \ldots]$ (Cambanis and Miller (1981)).

Theorem 2.2. *For the* ARMA(p, q) *process there exists a unique minimum dispersion linear predictor* X^*_{n+k} *for* X_{n+k}, $k \geqslant 1$, *based on the infinite past* $X_n, X_{n-1}, \ldots$. *This predictor satisfies the recursive relationship*

$$X^*_{n+k} = \sum_{j=1}^{k-1} \psi_j X^*_{n+k-j} + \sum_{j=k}^{\infty} \psi_j X_{n+k-j}. \tag{2.2}$$

Proof. For each $k \geqslant 1$,

$$X_{n+k} = W_{n+k} + \sum_{j=1}^{k-1} \psi_j X_{n+k-j} + \sum_{j=k}^{\infty} \psi_j X_{n+1-j}.$$

(The second term is taken to be zero when $k = 1$.) It follows by induction that $X_{n+k} \in S_*$ and thus there exists a linear predictor X^*_{n+k}. Furthermore, by the linearity

of the prediction mapping.

$$X^*_{n+k} = W^*_{n+k} + \sum_{j=1}^{k-1} \psi_j X^*_{n+k-j} + \left(\sum_{j=k}^{\infty} \psi_j X_{n+k-j} \right)^*.$$

Clearly $W^*_{n+k} = 0$ and $(\sum_{j=k}^{\infty} \psi_j X_{n-k+j})^* = \sum_{j=k}^{\infty} \psi_j X_{n+k-j}$ so that we have the result (2.2). □

In practice, of course, one will usually have only the data $X_n, \ldots, X_1$. For any $Y \in S_*$ one can use the 'truncated' predictor $Y^*(n) = \sum_{j=1}^{n} \nu_j X_{n+1-j}$, where $\{\nu_j\}_{j=1}^{\infty}$ is defined as in Lemma 2.1. Though the dispersion is not minimum, we have from (2.2)

$$\begin{aligned} &\operatorname{disp}(Y - Y^*(n)) - \operatorname{disp}(Y - Y^*) \\ &= \sum_{j=n+1}^{\infty} |\rho_j|^\alpha + \sum_{j=n+1}^{\infty} \left| \sum_{i=n+1}^{k} \nu_i \pi_{j-i} \right|^\alpha - \sum_{j=n+1}^{\infty} |\rho_j|^\alpha \\ &= \sum_{j=n+1}^{\infty} \left| \sum_{i=n+1}^{j} \nu_i \pi_{j-i} \right|^\alpha. \end{aligned}$$

In particular, if $Y = X_{n+1}$, then

$$\operatorname{disp}(X_{n+1} - X^*_{n+1}) = 1$$

and

$$\operatorname{disp}(X_{n+1} - X^*_{n+1}(n)) = 1 + \sum_{j=k+1}^{\infty} \left| \sum_{i=n+1}^{\infty} \psi_i \pi_{j-i} \right|^\alpha$$

so that for large n the truncation predictor is nearly optimal.

The truncated predictor is in fact optimal when the process is purely autoregressive and n is large enough, that is when $\{X_n\}$ satisfies

$$X_n = \phi_1 X_{n-1} + \cdots + \phi_p X_{n-p} + W_n \tag{2.3}$$

and $n \geq p$. As always, $\{W_n\}$ is an iid sequence of random variables satisfying (1.3). Assumption (1.2) reduces to $(1 - \phi_1 z - \cdots - \phi_p z^p) \neq 0$ for all $|z| \leq 1$. We state the results for the autoregressive process in a lemma and corollary which are proved in a fashion identical to the previous lemma and theorem. Recall that $\boldsymbol{X}_n = (X_n, X_{n-1}, \ldots, X_1)$.

Lemma 2.3. *Let S_* be the class of random variables of the form $Y = Z + \boldsymbol{\nu}' \boldsymbol{X}_n$ for some $\boldsymbol{\nu} \in \mathbb{R}^n$ and $Z = \sum_{j=n+1}^{\infty} \rho_j W_j$ such that Z exists. Then for each $Y \in S_*$, the set $P_n Y = \{\boldsymbol{a}' \boldsymbol{X}_n : \operatorname{disp}(Y - \boldsymbol{a}' \boldsymbol{X}_n)$ is minimum$\}$ consists of exactly one variable. For $Y = Z + \boldsymbol{\nu}' \boldsymbol{X}_n$, this unique variable is $\hat{Y} = \boldsymbol{\nu}' \boldsymbol{X}_n$. Furthermore, the mapping $Y \to \hat{Y}$ is linear on S_*.* □

Corollary 2.4. *For the process* (2.3), *provided* $n \geq p$, *there exists a unique minimum dispersion linear predictor* $\hat{X}_{n+k}$ *for* X_{n+k} $(k \geq 1)$ *in terms of* $X_1, \ldots, X_n$. *This predictor satisfies the recursive relationship*

$$\hat{X}_{n+k} = \phi_1 \hat{X}_{n+k-1} + \cdots + \phi_p \hat{X}_{n+k-p} \tag{2.4}$$

with initial conditions $\hat{X}_j = X_j$ *for* $1 \leq j \leq n$. □

Remarks. 1. The minimum dispersion predictor $\hat{X}_{n+k}$ is exactly the same as the least squares predictor $\tilde{X}_{n+k}$ for an autoregressive process. This is not the case for more general ARMA processes.

2. The residuals $W_{n+1}, W_{n+2}, \ldots$ are predicted with zeroes, and for $p < j \leq n$, then

$$W_j = X_j - \phi_1 X_{j-1} - \cdots - \phi_p X_{j-p},$$

but the linearity principle does not apply to $W_1, \ldots, W_p$. In fact, if $\alpha \leq 1$, the set $P_n W_j = \{\boldsymbol{a}' \boldsymbol{X}_n : \operatorname{disp}(W_j - \boldsymbol{a}' \boldsymbol{X}_n)$ is minimum$\}$ may not consist of exactly one element for $j \leq p$.

3. Prediction of the ARMA(1, 1) process

In this section we are concerned with the stationary process $\{X_n\}$ defined by

$$X_n - \phi X_{n-1} = W_n + \theta W_{n-1} \tag{3.1}$$

where $|\phi| < 1$ and $|\theta| < 1$ and $\{W_n\}$ are iid, satisfying (1.3). We find it necessary to distinguish between the cases $\alpha \leq 1$ and $\alpha > 1$. For both cases, however, we shall need the following lemma.

Lemma 3.1. *If* $a > 0$ *and* $\alpha > 0$, *then* $h(x) = a|x|^\alpha + |x - b|^\alpha$ *has its minimum value at* x_m, *where*

$$x_m = \begin{cases} b & \text{if } \alpha \leq 1,\ a \leq 1, \\ 0 & \text{if } \alpha \leq 1,\ a > 1, \\ \dfrac{b}{1 + a^{1/\alpha - 1}} & \text{if } \alpha > 1, \end{cases}$$

and x_m *is unique if* $a \neq 1$ *or* $\alpha > 1$.

The minimum value of h *is*

$$h(x_m) = \begin{cases} |b|^\alpha \min(1, a) & \text{if } \alpha \leq 1, \\ a|b|^\alpha (1 + a^{1/\alpha - 1})^{1-\alpha} & \text{if } \alpha > 1. \end{cases}$$

Proof. Define the function $[x]^\gamma = \operatorname{sgn}(x)|x|^\gamma$. Suppose $b > 0$. Then for $x \neq 0$, $x \neq b$

$$h'(x) = \alpha(a[x]^{\alpha-1} + [x - b]^{\alpha-1}), \qquad h''(x) = \alpha(\alpha - 1)(a|x|^{\alpha-2} + |x - b|^{\alpha-2}).$$

So for $x < 0$, $h'(x) < 0$ and for $x > b$, $h'(x) > 0$. Thus h is minimized in $[0, b]$.

If $\alpha \leq 1$, then $h''(x) \leq 0$, so the minimum must be either at 0 or at b. It is easy to see that $h(b) \leq h(0)$ if and only if $a \leq 1$.

If $\alpha > 1$, then h' is continuous on $[0, b]$ and h'' is nonnegative. Thus $h'(x_m) = 0$ gives us the point of minimum. On $[0, b]$, $h'(x) = \alpha(ax^{\alpha-1} - (b-x)^{\alpha-1})$, so that $x_m = b(1 + a^{1/\alpha-1})^{-1}$. Also

$$h(x_m) = a\left(\frac{b}{1+a^{1/\alpha-1}}\right)^{\alpha} + \left(\frac{ba^{1/\alpha-1}}{1+a^{1/\alpha-1}}\right)^{\alpha} = \frac{ab}{(1+a^{1/\alpha-1})^{\alpha-1}}.$$

The proof is similar if $b < 0$. □

Theorem 3.2.(i) For the ARMA(1, 1) process (3.1) there is a unique minimum dispersion predictor $\hat{X}_{n+k} = \boldsymbol{a}'\boldsymbol{X}_n$ for X_{n+k}, except when $\alpha \leq 1$ and $|\phi + \theta|^{\alpha} = 1 - |\phi|^{\alpha}$, in which case a minimum dispersion predictor exists but is not unique.

(ii) If $\alpha \leq 1$ the error dispersion is minimized when the coefficient vector $\boldsymbol{a}$ satisfies

$$\begin{aligned} a_j &= (\phi+\theta)(-\theta)^{j-1}\phi^{k-1}, \quad 1 \leq j \leq n-1, \\ a_n &= \begin{cases} (\phi+\theta)(-\theta)^{n-1}\phi^{k-1} & \text{if } |\phi+\theta|^{\alpha} \leq 1-|\phi|^{\alpha}, \\ \phi^{k}(-\theta)^{n-1} & \text{if } |\phi+\theta|^{\alpha} \geq 1-|\phi|^{\alpha}, \end{cases} \end{aligned} \tag{3.2}$$

and the corresponding minimum dispersion is

$$1 + |\phi+\theta|^{\alpha}\frac{1-|\phi|^{\alpha(k-1)}}{1-|\phi|^{\alpha}} + |\phi|^{\alpha(k-1)}|\theta|^{n\alpha}\min\left(1, \frac{|\phi+\theta|^{\alpha}}{1-|\phi|^{\alpha}}\right).$$

(iii) If $\alpha > 1$ the error dispersion is minimized when the coefficient vector $\boldsymbol{a}$ satisfies

$$a_j = \phi^{k-1}(-\theta)^{j-1}\frac{(\phi+\theta)(1-\eta+\xi) - \xi\eta^{n-j}(\eta\phi+\theta)}{1-\eta+\xi(1-\eta^{n})}, \quad 1 \leq j \leq n, \tag{3.3}$$

where

$$\eta := |\theta|^{\alpha/(\alpha-1)} \quad \text{and} \quad \xi := \left(\frac{|\phi+\theta|^{\alpha}}{1-|\phi|^{\alpha}}\right)^{1/(\alpha-1)}.$$

The corresponding minimum dispersion is

$$1 + \xi^{\alpha-1}(1-|\phi|^{\alpha(k-1)}) + \left(\frac{\xi\eta^{n}(1-\eta)}{1-\eta+\xi(1-\eta^{n})}\right)^{\alpha-1}.$$

Proof. Since $|\phi| < 1$, we have

$$X_j = W_j + (\phi+\theta)\sum_{k=1}^{\infty}\phi^{k-1}W_{j-k} \quad \text{for all } j.$$

If $\boldsymbol{m} \in \mathbb{R}^n$ and if we define $m_0 = -\phi^{k-1}$, then from the previous equation we can write for $k \geq 1$,

$$\boldsymbol{m}'\boldsymbol{X}_n - X_{n+k} = -W_{n+k} - \sum_{j=1}^{k-1} (\phi+\theta)\phi^{k-j-1} W_{n+j}$$
$$+ \sum_{j=1}^{n} \left[m_j + (\phi+\theta) \sum_{i=0}^{j-1} m_j \phi^{j-i-1} \right] W_{n+1-j}$$
$$+ (\phi+\theta)\left(\sum_{i=0}^{n} m_i \phi^{n-i} \right) \sum_{j=0}^{\infty} \phi^j W_{-j}.$$

Consequently, the dispersion is

$$\operatorname{disp}(\boldsymbol{m}'\boldsymbol{X}_n - X_{n+k}) = 1 + |\phi+\theta|^\alpha \frac{1-|\phi|^{\alpha(k-1)}}{1-|\phi|^\alpha}$$
$$+ \sum_{j=1}^{n} |c_j + \theta c_{j-1}|^\alpha + \frac{|\phi+\theta|^\alpha}{1-|\phi|^\alpha} |c_n|^\alpha \tag{3.4}$$

where $c_j = \sum_{i=0}^{j} m_i \phi^{j-i}$, $j \geq 0$ (and $m_j = c_j - \phi c_{j-1}$, $j \geq 1$). It suffices now to minimize

$$h(\boldsymbol{c}) = \sum_{j=1}^{n} |c_j + \theta c_{j-1}|^\alpha + \frac{|\phi+\theta|^\alpha}{1-|\phi|^\alpha} |c_n|^\alpha \tag{3.5}$$

and this will be done recursively, minimizing first with respect to c_n, then c_{n-1} and so on.

There are three cases to consider: (a) $\alpha \leq 1$ and $|\phi+\theta|^\alpha \leq 1-|\phi|^\alpha$, (b) $\alpha \leq 1$ and $|\phi+\theta|^\alpha > 1-|\phi|^\alpha$, and (c) $\alpha > 1$. We consider these in turn.

(a) By Lemma 3.1, for fixed $c_{n-1}, \ldots, c_1$, $h(\boldsymbol{c})$ is minimized by choosing $c_n = -\theta c_{n-1}$. Under this condition (3.5) becomes

$$\min_{c_n} h(\boldsymbol{c}) = \sum_{j=1}^{n-1} |c_j + \theta c_{j-1}|^\alpha + |\theta|^\alpha \frac{|\phi+\theta|^\alpha}{1-|\phi|^\alpha} |c_{n-1}|^\alpha.$$

Since $|\theta| < 1$ (and hence $|\theta|^\alpha |\phi+\theta|^\alpha < 1-|\phi|^\alpha$), then $h(\boldsymbol{c})$ is minimized further by choosing $c_{n-1} = -\theta c_{n-2}$, again using Lemma 3.1. The resulting value for $h(\boldsymbol{c})$ will have a similar form so that continuing recursively, we can choose $c_j = -\theta c_{j-1}$, $1 \leq j \leq n$. Since $c_0 = m_0 = -\phi^{k-1}$, we find that $c_j = -(-\theta)^j \phi^{k-1}$ and the minimizing vector $\boldsymbol{m}$ is $\boldsymbol{a}$, as given in (3.2). The minimum value of $\operatorname{disp}(X_{n+k} - \boldsymbol{a}'\boldsymbol{X}_n)$ is

$$1 + |\phi+\theta|^\alpha \frac{1-|\phi|^{\alpha(k-1)}}{1-|\phi|^\alpha} + \min h(\boldsymbol{c})$$
$$= 1 + |\phi+\theta|^\alpha \frac{1-|\phi|^{\alpha(k-1)}}{1-|\phi|^\alpha} + |\theta|^{n\alpha} |\phi|^{\alpha(k-1)} \frac{|\phi+\theta|^\alpha}{1-|\theta|^\alpha}.$$

(b) If $\alpha \leq 1$ and $|\phi+\theta|^\alpha > 1-|\phi|^\alpha$, the argument is the same except that first we choose $c_n = 0$, according to Lemma 3.1, to minimize (3.5). In this case

$$\min_{c_n} h(\boldsymbol{c}) = \sum_{j=1}^{n-1} |c_j + \theta c_{j-1}|^\alpha + |\theta|^\alpha |c_{n-1}|^\alpha. \tag{3.6}$$

Since $|\theta| < 1$, then (3.5) is further minimized by setting $c_j = -\theta c_{j-1}$, $1 \leq j \leq n-1$, as done previously. Again using $c_0 = -\phi^{k-1}$ and $m_j = c_j - \theta c_{j-1}$, we find that $\boldsymbol{a}$, as defined in (3.2), is the minimizing value of $\boldsymbol{m}$. The minimum error dispersion is

$$1 + |\phi+\theta|^\alpha \frac{1-|\phi|^{\alpha(k-1)}}{1-|\phi|^\alpha} + \min h(\boldsymbol{c})$$

$$= 1 + |\phi+\theta|^\alpha \frac{1-|\phi|^{\alpha(k-1)}}{1-|\phi|^\alpha} + |\theta|^{n\alpha} |\phi|^{\alpha(k-1)}.$$

(c) In the case $\alpha > 1$ we find from Lemma 3.1 that

$$c_n = -\theta c_{n-1}(1+\xi)^{-1}$$

with the corresponding minimum value,

$$\min_{c_n} h(\boldsymbol{c}) = \sum_{j=1}^{n-1} |c_j + \theta c_{j-1}|^\alpha + \left(\frac{\eta\xi}{1+\xi}\right)^{\alpha-1} |c_{n-1}|^\alpha.$$

This is further minimized when

$$c_{n-1} = -\theta c_{n-2}\left(1 + \frac{\eta\xi}{1+\xi}\right)^{-1},$$

and the corresponding minimum is

$$\min_{c_n, c_{n-1}} h(\boldsymbol{c}) = \sum_{j=1}^{n-2} |c_j + \theta c_{j-1}|^\alpha + \left(\frac{\eta^2\xi}{1+\xi+\eta\xi}\right)^{\alpha-1} |c_{n-2}|^\alpha.$$

Continuing the stepwise minimization we find that

$$c_j = -\theta c_{j-1} \frac{1-\eta+\xi(1-\eta^{n-j})}{1-\eta+\xi(1-\eta^{n-j+1})}.$$

Since $c_0 = -\phi^{k-1}$, we deduce that

$$c_j = -\phi^{k-1}(-\theta)^j \frac{1-\eta+\xi(1-\eta^{n-j})}{1-\eta+\xi(1-\eta^n)}. \tag{3.7}$$

From this, and the relations $a_j = c_j - \phi c_{j-1}$, we find that $\boldsymbol{a}$ as defined by (3.3) is the unique vector minimizing $\mathrm{disp}(X_{n+k} - \boldsymbol{a}'\boldsymbol{X}_n)$. The minimum error dispersion, from

the expressions (3.4), (3.5) and (3.7), is

$$\begin{aligned}\operatorname{disp}(X_{n+k} - \boldsymbol{a}'\boldsymbol{X}_n) &= 1 + \xi^{\alpha-1}(1 - |\phi|^{\alpha(k-1)}) + \min h(\boldsymbol{c}) \\ &= 1 + \xi^{\alpha-1}(1 - |\phi|^{\alpha(k-1)}) + \left(\frac{\xi\eta^n(1-\eta)}{1-\eta+\xi(1-\eta^n)}\right)^{\alpha-1}.\end{aligned}$$

In all three cases (a), (b) and (c), Lemma 3.1 guarantees the uniqueness of the optimal $\boldsymbol{a}$, except when $\alpha \leq 1$ and $|\phi+\theta|^\alpha = 1 - |\phi|^\alpha$, in which case the final coefficient a_n may be chosen as either of the two expressions in (3.2). □

Remarks. 1. For an AR(1) process the minimum dispersion predictor of X_{n+k} is $\hat{X}_{n+k} = \phi^k X_n$, $k \geq 1$, $n \geq 1$, and the corresponding error dispersion is

$$\frac{1 - |\phi|^{\alpha k}}{1 - |\phi|^\alpha}.$$

2. For the MA(1) process $X_n = W_n + \theta W_{n-1}$, the optimal predictor of X_{n+k}, $k \geq 2$, is $\hat{X}_{n+k} = 0$ with error dispersion $1 + |\theta|^\alpha$. For $k = 1$ the optimal predictor $\hat{X}_{n+1}$ is obtained by choosing

$$a_j = -(-\theta)^j, \qquad j = 1, \ldots, n \quad \text{if } \alpha \leq 1,$$

and

$$a_j = -(-\theta)^j \frac{1 - \eta^{n+1-j}}{1 - \eta^{n+1}}, \qquad j = 1, \ldots, n \quad \text{if } \alpha > 1,$$

where $\eta = |\theta|^{\alpha/(\alpha-1)}$. The error dispersion of $\hat{X}_{n+1}$ is

$$\begin{aligned} &1 + |\theta|^{(n+1)\alpha} && \text{if } \alpha \leq 1, \\ &1 + |\theta|^{(n+1)\alpha}\left(\frac{1-\eta}{1-\eta^n}\right)^{\alpha-1} && \text{if } \alpha > 1. \end{aligned}$$

3. In the case $\alpha \leq 1$, although the predictor may not be unique, it can be specified in such a way that the mapping $Y \to \hat{Y}$ is linear on $\operatorname{span}\{X_1, X_2, \ldots\}$. To see this we need only observe that for each $j \geq 1$

$$X_j = W_j + (\phi+\theta)\sum_{i=1}^{j-1} \phi^{i-1} W_{j-i} + \phi^{j-1}(\phi+\theta) W_0^*$$

where $W_0^* = \sum_{i=0}^{\infty} \phi^i W_{-i}$, and to apply Theorem 3.5 of Cline (1983). In particular this allows us to write

$$\hat{X}_{n+k} = \hat{W}_{n+k} + \theta\hat{W}_{n+k-1} + \phi\hat{X}_{n+k-1} = \phi\hat{X}_{n+k-1} = \phi^{k-1}\hat{X}_{n+1}$$

in agreement with Theorem 3.2.

4. In the case $\alpha > 1$ we again have a partial linearity property for the minimum dispersion prediction operator. Thus if

$$Y = l_1 X_{n+1} + \cdots + l_k X_{n+k},$$

then

$$\hat{Y} = l_1\hat{X}_{n+1} + \cdots + l_k\hat{X}_{n+k} = (l_1 + \phi l_2 + \cdots + \phi^{k-1}l_k)\hat{X}_{n+1}.$$

This can be established by minimizing $h(\boldsymbol{c})$ in (3.5), now subject to $c_0 = -(l_1 + \phi l_2 + \cdots + \phi^{k-1}l_k)$.

5. Predictors can easily be computed recursively. Defining $\hat{X}_j(k) = P_k X_j$, the best predictor of X_j based on $X_1, \ldots, X_k$, $j > k$, we find that

(a) *For* $\alpha \leq 1$,

$$\hat{X}_{n+1}(n) = \phi X_n + \theta(\hat{X}_n(n-1) - X_n),$$

with

$$\hat{X}_2(1) = \begin{cases} (\phi+\theta)X_1 & \text{if } |\phi+\theta|^\alpha \leq 1 - |\phi|^\alpha, \\ \phi X_1 & \text{otherwise.} \end{cases}$$

(b) *For* $\alpha > 1$,

$$X_{n+1}(n) = \phi X_n + \phi\frac{\nu_{n-1}}{\nu_n}(\hat{X}_n(n-1) - X_n),$$

with

$$\hat{X}_1(0) = 0 \quad \text{and} \quad \nu_n = 1 + \eta + \xi(1 - \eta^n).$$

The linearity properties and recursion formulae do not extend to higher order ARMA models.

6. Minimizing (3.5) with $\alpha = 2$ gives the least squares predictor for X_{n+k}, namely $\tilde{X}_{n+k} = b'\boldsymbol{X}_n$ where

$$b_j = \phi^{k-1}(-\theta)^{j-1}(\phi+\theta)\left(\frac{1 - \theta\rho\theta^{2(n-j)}}{1 - \rho^2\theta^{2n}}\right), \quad \rho = \frac{\phi+\theta}{1+\phi\theta},$$

and the error dispersion of this predictor, for any α, is

$$1 + |\phi+\theta|^\alpha \frac{1 - |\phi|^{\alpha(k-1)}}{1 - |\phi|^\alpha} + \left|\frac{\theta^n}{1-\rho^2\theta^{2n}}\right|^\alpha \times \left(|\rho^2(1-\theta^2)|^\alpha \frac{1 - |\theta|^{n\alpha}}{1 - |\theta|^\alpha} + \frac{|(\phi+\theta)(1-\rho^2)|^\alpha}{1 - |\phi|^\alpha}\right).$$

The least squares predictor $\tilde{X}_{n+1}(n)$ is recursively calculated from

$$\tilde{X}_{n+1}(n) = \phi X_n + \theta\frac{1 - \rho^2\theta^{2(n-1)}}{1 - \rho^2\theta^{2n}}(\tilde{X}_n(n-1) - X_n), \quad \tilde{X}_1(0) = 0.$$

(See also Brockwell and Davis (1983) for a general discussion of least squares prediction.)

4. Prediction for the MA(q) and ARMA(p, q) models

Assume the process $\{X_n\}$ satisfies $X_n = W_n + \theta_1 W_{n-1} + \cdots + \theta_q W_{n-q}$ where $(1 + \theta_1 z + \cdots + \theta_q z^q) \neq 0$ for complex $|z| \leq 1$. In order to predict X_{n+1}, we need to minimize

$$\operatorname{disp}(X_{n+1} - \boldsymbol{a}'\boldsymbol{X}_n) = 1 + \sum_{j=1}^{n+q} |a_j + a_{j-1}\theta_1 + \cdots + a_{j-q}\theta_q|^\alpha \tag{4.1}$$

where $a_0 = -1$ and $a_j = 0$ for $j < 0$ or $j > n$. According to Cline (1983, Theorem 3.4), when $\alpha \leq 1$ it suffices to consider only vectors $\boldsymbol{a} \in \mathbb{R}^n$ which satisfy

$$a_j + a_{j-1}\theta_1 + \cdots + a_{j-q}\theta_q = 0 \tag{4.2}$$

for at least n of the $n+q$ equations, $1 \leq j \leq n+q$. The set of choices is thus limited to $\binom{n+q}{q}$ possibilities. In Theorem 3.2 we have already established which choice is optimal for the MA(1) model. Exactly one choice was the best for all values of θ_1 in the parameter space, $|\theta_1| < 1$. If $q > 1$, however, the optimal formula depends on the particular region of the parameter space. We look specifically at the MA(2) model.

Lemma 4.1. *Suppose* $\{X_n\}$ *is an* MA(2) *process with* $\alpha \leq 1$. *Define* z_1, z_2 *to be the solutions to* $(z^2 + \theta_1 z + \theta_2) = 0$, *and*

$$S_j = \begin{cases} \dfrac{z_1^j - z_2^j}{z_1 - z_2} & \text{if } z_1 \neq z_2, \\ j z_1^{j-1} & \text{if } z_1 = z_2. \end{cases}$$

Then the minimum dispersion predictor for X_{n+1} *lies in the set of the* $\binom{n+2}{2}$ *choices for* $\boldsymbol{a}'\boldsymbol{X}_n$ *where* $1 \leq j_1 < j_2 \leq n+2$ *and*

$$a_j = \begin{cases} -S_{j+1} & \text{if } 1 \leq j < j_2, \\ -\theta_2^{j-j_1+1} \dfrac{S_{j_2-j-1} S_{j_1}}{S_{j_2-j_1}} & \text{if } j_1 \leq j < j_2, \\ 0 & \text{if } j_2 \leq j \leq n+2. \end{cases}$$

The dispersion of the prediction error is

$$\operatorname{disp}(X_{n+1} - \boldsymbol{a}'\boldsymbol{X}_n) = 1 + \left|\frac{S_{j_2}}{S_{j_2-j_1}}\right|^\alpha + \left|\theta_2^{j_2-j_1} \frac{S_{j_1}}{S_{j_2-j_1}}\right|. \tag{4.3}$$

Proof. We recognize that the S_j's can be determined recursively by $S_{j+2} + \theta_1 S_{j+1} + \theta_2 S_j = 0$ and that in fact $S_{j+1}S_{k+1} - \theta_2 S_j S_k = S_{j+k+1}$. From these we can easily verify that (using $a_0 = -1$, $a_{-1} = a_{n+1} = a_{n+2} = 0$ and fixing j_1, j_2)

$$a_j + \theta_1 a_{j-1} + \theta_2 a_{j-2} = \begin{cases} 0 & \text{for } j \neq j_1, j \neq j_2, \\ -\dfrac{S_{j_2}}{S_{j_2-j_1}} & \text{for } j = j_1, \\ -\theta_2^{j_2-j_1} \dfrac{S_{j_1}}{S_{j_2-j_1}} & \text{for } j = j_2. \end{cases}$$

Thus for each of the $\binom{n+2}{n}$ choices of j_1 and j_2, $\boldsymbol{a}$ satisfies n of the $n+2$ equations in (4.2). It follows that a minimum dispersion predictor is of this form. The error dispersion is

$$\begin{aligned}\operatorname{disp}(X_{n+1}-\boldsymbol{a}'\boldsymbol{X}_n) &= 1+\sum_{j=1}^{n+2}|a_j+\theta_1 a_{j-1}+\theta_2 a_{j-2}|^{\alpha}\\ &= 1+\left|\frac{S_{j_2}}{S_{j_2-j_1}}\right|^{\alpha}+\left|\theta_2^{j_2-j_1}\frac{S_{j_1}}{S_{j_2-j_1}}\right|^{\alpha}.\end{aligned}$$

To actually determine the predictor, we need to minimize (4.3) over the $\binom{n+2}{n}$ possible choices of j_1, j_2. □

Now for a general MA(q) process with $\alpha>1$, the minimum dispersion predictor is obtained by finding $\boldsymbol{a}\in\mathbb{R}^n$ to satisfy (using the notation $[x]^{\alpha-1}=\operatorname{sgn}(x)|x|^{\alpha-1}$),

$$\begin{aligned}&[a_j+\theta_1 a_{j-1}+\cdots+\theta_q a_{j-q}]^{\alpha-1}\\ &\quad+\theta_1[a_{j+1}+\theta_1 a_j+\cdots+\theta_q a_{j-q+1}]^{\alpha-1}\\ &\quad+\cdots+\theta_q[a_{j+q}+\cdots+\theta_q a_j]^{\alpha-1}=0,\quad 1\leq j\leq n.\end{aligned}$$

This can be accomplished recursively in the following manner: Let Π and $\boldsymbol{\rho}$ respectively be the $n\times(n+q)$ matrix and $1\times(n+q)$ vector,

$$\Pi=\begin{bmatrix}1 & \theta_1 & \theta_2 & \cdots & \theta_q & 0 & \cdots & 0\\ 0 & 1 & \theta_1 & & \cdots & \theta_q & \cdots & 0\\ \vdots & & \ddots & & & & & \\ & & & 1 & \theta_1 & \cdots & & \theta_q\end{bmatrix}$$

and

$$\boldsymbol{\rho}=[\theta_1\theta_2\cdots\theta_q 0\cdots 0].$$

Set $\boldsymbol{a}_0=(\Pi\Pi')^{-1}\Pi\boldsymbol{\rho}$. ($\boldsymbol{a}_0'\boldsymbol{X}_n$ is the least squares predictor.) Next define $l_j(\boldsymbol{a})=[\boldsymbol{a}'\boldsymbol{\Pi}_j-\rho_j]^{\alpha-1}=[-\theta_j+a_1\theta_{j-1}+\cdots+a_q\theta_{j-q}]^{\alpha-1}$, $1\leq j\leq n+q$. The recursion is then given by

$$\boldsymbol{a}_{k+1}=\boldsymbol{a}_k-(\Pi\Pi')^{-1}\Pi\boldsymbol{l}(\boldsymbol{a}_k).$$

The ARMA(1, q) process can be handled similarly to the MA(q) process. Let $\pi_j=\phi^j+\phi^{j-1}\theta_1+\cdots+\phi\theta_{j-1}+\theta_j$. The process can then be expressed as

$$X_n=W_n+\pi_1 W_{n-1}+\cdots+\pi_{n+q-2}W_{2-q}+\frac{\pi_{n+q-1}}{(1-|\phi|^{\alpha})^{1/\alpha}}W^*,\quad n\geq 1,$$

where

$$W^*=(1-|\phi|^{\alpha})^{1/\alpha}\sum_{j=0}^{\infty}\phi^j W_{1-q-j}\overset{\mathrm{d}}{=}W_{1-q}$$

and is independent of $W_{2-q}, W_{3-q}, \ldots$. To predict X_{n+1} we need to minimize

$$\operatorname{disp}(X_{n+1} - \boldsymbol{a}'\boldsymbol{X}_n) = 1 + \sum_{j=1}^{n+q-1} |a_j + \pi_1 a_{j-1} + \cdots + \pi_j a_0|^\alpha + \frac{|a_{n+q} + \cdots + \pi_{n+q} a_0|^\alpha}{1 - |\phi|^\alpha}$$

$$= 1 + \sum_{j=1}^{n+q-1} |c_j + \theta_1 c_{j-1} + \cdots + \theta_q c_{j-q}|^\alpha + \frac{|\pi_q c_n|^\alpha}{1 - |\phi|^\alpha},$$

where

$$c_j = \sum_{k=0}^{j} \phi^k z_{j-k}, \; a_j = c_j - \phi c_{j-1}, \quad j \geq 0$$

$$c_j = 0, \qquad j < 0.$$

Except for the last term this (as a function of $\boldsymbol{c}$) is similar to (4.1). The minimization is thus done with respect to $\boldsymbol{c}$ and then $\boldsymbol{a}$ is obtained from $\boldsymbol{c}$.

The ARMA(1, q) minimization involves a finite sum. This is not true for the more general ARMA(p, q) process (1.1). To deal with this process we define a sequence $\{c_j\}$ satisfying $c_j = 0$, $j < 0$, and $a_j = c_j - \phi_1 c_{j-1} - \cdots - \phi_p c_{j-p}$, $j \geq 0$. Then X_{n+1} is predicted by minimizing

$$\operatorname{disp}(X_{n+1} - \boldsymbol{a}'\boldsymbol{X}_n) = 1 + \sum_{j=1}^{\infty} |c_j + \theta_1 c_{j-1} + \cdots + \theta_q c_{j-q}|^\alpha$$

$$= 1 + \sum_{j=1}^{n} |c_j + \theta_1 c_{j-1} + \cdots + \theta_q c_{j-q}|^\alpha + \sum_{j=1}^{\infty} |\sigma_{j1} c_n + \cdots + \sigma_{jn} c_1|^\alpha$$

where $\sigma_{j1}, \ldots, \sigma_{jn}$ satisfy $c_{n+j} + \theta_1 c_{n+j-1} + \cdots + \theta_q c_{n+j-q} = \sigma_{j1} c_n + \cdots + \sigma_{jn} c_1$ ($n > \max(p, q+1)$). The sum can be truncated after an appropriate number of terms to facilitate the minimization. Alternatively, we can use the truncated predictor described in Section 2, $X^*_{n+1}(n) = \sum_{j=1}^{n} \psi_j X_{n+1-j}$, which will be close to optimal for large n.

Acknowledgements

The authors are indebted to Richard A. Davis and to Sidney I. Resnick for their helpful comments, to the National Science Foundation for its support and to the Universities of British Columbia and Kuwait for additional support. We are also indebted to a referee for suggesting a number of improvements in presentation of the results.

References

R. Blattberg and T. Sargent, Regression with non-Gaussian stable disturbances: Some sampling results, Econometrica 39 (1971) 501-510.

G.E.P. Box and G.M. Jenkins, Time Series Analysis: Forecasting and Control (Holden-Day, San Francisco, 1976).

P.J. Brockwell and R.A. Davis, Recursive prediction and exact likelihood determination for Gaussian processes, Tech. Rpt. 65, Department Statistics, Colorado State University, Fort Collins (1983).

S. Cambanis and G. Miller, Linear problems in pth order and stable processes, Siam J. Appl. Math. 41 (1981) 43-69.

S. Cambanis and A.R. Soltani, Prediction of stable processes: Spectral and moving average representations, Tech. Rpt. 11, Center for Stochastic Processes, University No. Carolina, Chapel Hill, 1982.

D.B.H. Cline, Infinite series of random variables with regularly varying tails, Technical Report, No. 83-24, University of British Columbia, 1983.

W.A. Fuller, Introduction to Statistical Times Series (Wiley, New York, 1976).

E.J. Hannan and M. Kanter, Autoregressive processes with infinite variance, J. Appl. Prob. 14 (1977) 411-415.

M. Kanter and W.L. Steiger, Regression and autoregression with infinite variance, Adv. Appl. Prob. 6 (1974) 768-783.

I. Singer, Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces (Springer-Verlag, Berlin, 1970).

B.W. Stuck, Minimum error dispersion linear filtering of scalar symmetric stable processes, IEEE Trans. Aut. Cont. 23 (1978) 507-509.

V.T. Yohai and R.A. Maronna, Asymptotic behavior of least-squares estimates for autoregressive processes with infinite variance, Ann. Statist. 5 (1977) 554-560.