Stochastic Processes and their Applications (1988) 47-59
North-Holland

SIMPLE CONSISTENT ESTIMATION OF THE COEFFICIENTS OF A LINEAR FILTER

P.J. BROCKWELL and R.A. DAVIS

Department of Statistics, Colorado State University, Fort Collins, Colorado 80523, USA

Received 9 March 1987
Revised 22 September 1987

A simple procedure is proposed for estimating the coefficients $\{\psi_j\}$ from observations of the linear process $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, $t = 1, 2, \ldots$. The method is based on the representation of X_t in terms of the innovations, $X_n - \hat{X}_n$, $n = 1, \ldots, t$, where $\hat{X}_n$ is the best mean square predictor of X_n in span $\{X_1, \ldots, X_{n-1}\}$. The asymptotic distribution of the sequence of estimators is derived and its applications to inference for ARMA processes are discussed.

AMS 1980 Subject Classifications: 62M20, 60G15, 60G10.

linear process * ARMA process * linear filter * innovations * preliminary estimation * identification * asymptotic distribution

1. Introduction

We are concerned in this paper with the development of a simple consistent technique for estimating the coefficients $\{\psi_j\}$ of the linear process,

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \ldots, \tag{1.1}$$

where $\{Z_t\}$ is an i.i.d. sequence of random variables such that $EZ_t = 0$, $EZ_t^2 = \sigma^2$ and $EZ_t^4 < \infty$. We assume that $\psi_0 = 1$, $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $\psi(z) := \sum_{j=0}^{\infty} \psi_j z^j$ is non-zero for all $z \in \mathbb{C}$ such that $|z| \leq 1$. These conditions ensure that $\{Z_t\}$ has the representation

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t = 0, \pm 1, \ldots, \tag{1.2}$$

where $\pi_0 = 1$ and $\pi(z) := \sum_{j=0}^{\infty} \pi_j z^j = 1/\psi(z)$, $|z| \leq 1$.

The estimation procedure is based on the *innovations representation* of X_t, viz.

$$X_t = \sum_{j=0}^{t-1} \theta_{t-1,j} (X_{t-j} - \hat{X}_{t-j}), \quad t = 1, 2, \ldots, \tag{1.3}$$

where $\hat{X}_1 = 0$ and $\hat{X}_t$, $t \geq 2$, is the best linear predictor of X_t based on $\{X_1, \ldots, X_{t-1}\}$. The coefficients θ_{nj}, $0 \leq j \leq n$, $n = 0, 1, 2, \ldots$, can be found recursively from the following proposition (or equivalently from the Cholesky factorization of the covariance matrix; see Rissanen and Barbosa (1969)).

Research supported by NSF Grant No. MCS 8501763.

0304-4149/88/$3.50

Proposition 1.1. *Every zero-mean second order process* $\{X_t\}$ *has a representation of the form* (1.3). *If* $E(X_iX_j)=\kappa(i,j)$, *and if* $[\kappa(i,j)]_{i,j=1}^{n}$ *is non-singular for every n, then the coefficients* θ_{nj} *and the mean squared errors* $v_n=E(X_{n+1}-\hat{X}_{n+1})^2$ *are determined by the recursion relations,*

$$\theta_{n,n-k}=v_k^{-1}\Big(\kappa(n+1,k+1)-\sum_{0\le j<k}\theta_{k,k-j}\theta_{n,n-j}v_j\Big),\quad k=0,\ldots,n, \tag{1.4}$$

where $\theta_{n0}=1$, $n=0,1,2,\ldots$, *and* $v_0=\kappa(1,1)$. (*The equations* (1.5) *can easily be solved recursively for* θ_{11}, v_1; θ_{22}, θ_{21}, v_2; θ_{33}, θ_{32}, θ_{31}, v_3; ...)

Remarks. 1. For the process (1.1), the coefficients θ_{nj} appearing in the representation (1.3) are independent of σ^2 and can be found from Proposition 1.1 with $\kappa(i,j)=\sum_{r=0}^{\infty}\psi_r\psi_{r+|i-j|}$.

2. It will be shown in Section 2 that for the process (1.1), $\theta_{mj}\to\psi_j$ as $m\to\infty$ for each fixed *j*. This suggests the use of an estimator of θ_{mj} to estimate ψ_j. Given the sample $X_1,\ldots,X_n$, an obvious estimator of the vector $\boldsymbol{\theta}_m:=(\theta_{m1},\ldots,\theta_{mm})'$ for $m<n$ is $\hat{\boldsymbol{\theta}}_m:=(\hat{\theta}_{m1},\ldots,\hat{\theta}_{mm})'$, whose components are found by applying Proposition 1.1 to the *sample* covariances,

$$\hat{\kappa}(i,j)=\hat{\gamma}(i-j):=n^{-1}\sum_{r=1}^{n-|i-j|}X_rX_{r+|i-j|},\quad i,j=1,\ldots,m+1.$$

Provided *m* is chosen to depend on the sample size *n* in such a way that $m(n)\to\infty$, $m(n)=o(n^{1/3})$ and $n^{1/2}\sum_{j>m(n)}|\pi_j|\to 0$ as $n\to\infty$, we show in Section 2 that

$$n^{1/2}(\hat{\theta}_{m1}-\psi_1,\ldots,\hat{\theta}_{mm}-\psi_m,0,0,\ldots)\Rightarrow N(\mathbf{0},\Sigma),$$

where $N(\mathbf{0},\Sigma)$, denotes a zero-mean Gaussian sequence with covariance matrix,

$$\Sigma=\left[\sum_{k=0}^{\min(i,j)-1}\psi_k\psi_{k+|i-j|}\right]_{i,j=1}^{\infty}. \tag{1.5}$$

This result implies in particular that $(\hat{\theta}_{m1},\ldots,\hat{\theta}_{mm},0,0,\ldots)$ is consistent for $(\psi_1,\psi_2,\ldots)$. Notice that the calculation of the estimators $\hat{\theta}_{mj}$ from (1.4), with the covariances $\kappa(i,j)$ replaced by $\hat{\kappa}(i,j)$, requires *no* matrix inversions.

3. The explicit expression (1.5) for the asymptotic covariances of $n^{1/2}(\hat{\theta}_{mi}-\psi_i)$ and $n^{1/2}(\hat{\theta}_{mj}-\psi_j)$ makes the results outlined in Remark 2 particularly valuable for the identification of moving average models and preliminary estimation of the coefficients. This is because the asymptotic 95 per cent confidence bounds,

$$\hat{\theta}_{mj}\pm 1.96\left(n^{-1}\sum_{k=0}^{j-1}\hat{\theta}_{mk}^2\right)^{1/2},$$

for ψ_j, have the correct asymptotic coverage probability, regardless of the order of the moving average representation of the process. If on the other hand we use the sample autocorrelations $\hat{\rho}(j)$ to identify a moving average model, the situation is complicated by the fact that the asymptotic distribution of $\hat{\rho}(j)$ depends on the order of the underlying model. An illustration of the use of the estimators $\hat{\theta}_{mj}$ in analyzing the International Airline Passenger Data (Box and Jenkins (1976)) is given in Section 3.

4. The estimation procedure we have proposed in Remark 2 is closely analogous to the use of Yule-Walker estimators for the coefficients π_j in the AR(∞) representation (1.2) of the process $\{X_t\}$. The Yule-Walker estimators of $(\pi_1, \pi_2, \ldots)$ are the vectors $-(\hat{\phi}_{m1}, \hat{\phi}_{m2}, \ldots, \hat{\phi}_{mm}, 0, 0, \ldots)$, $m = 1, 2, \ldots$, where $\hat{\boldsymbol{\phi}}_m := (\hat{\phi}_{m1}, \ldots, \hat{\phi}_{mm})'$ satisfies

$$\hat{\Gamma}_m \hat{\boldsymbol{\phi}}_m = \hat{\boldsymbol{\gamma}}_m, \tag{1.6}$$

$\hat{\Gamma}_m = [\hat{\gamma}(i-j)]_{i,j=1}^m$, $\hat{\boldsymbol{\gamma}}_m = (\hat{\gamma}(1), \ldots, \hat{\gamma}(m))'$ and $\hat{\gamma}(j)$ is the sample autocovariance at lag j. The equations (1.6) can be solved recursively, for $m = 1, 2, \ldots$, using the Durbin-Levinson algorithm. Under the assumption that m depends on the sample size n in such a way that $m(n) \to \infty$, $m(n) = o(n^{1/3})$ and $n^{1/2} \sum_{j > m(n)} |\pi_j| \to 0$ as $n \to \infty$, Bhansali (1978), using results of Berk (1974), has shown that

$$n^{1/2}(\hat{\phi}_{m1} + \pi_1, \ldots, \hat{\phi}_{mm} + \pi_m, 0, 0, \ldots) \Rightarrow N(\mathbf{0}, \Lambda),$$

where $N(\mathbf{0}, \Lambda)$, denotes a zero-mean Gaussian sequence with covariance matrix,

$$\Lambda = \left[\sum_{k=0}^{\min(i,j)-1} \pi_k \pi_{k+|i-j|} \right]_{i,j=1}^{\infty}. \tag{1.7}$$

This result implies in particular that $-(\hat{\phi}_{m1}, \ldots, \hat{\phi}_{mm}, 0, 0, \ldots)$ is consistent for $(\pi_1, \pi_2, \ldots)$. The similarity between (1.5) and (1.7) is quite striking. There is a duality between the determination of $\hat{\boldsymbol{\theta}}_m$ from (1.4) as described in Remark 2 and the determination of $\hat{\boldsymbol{\phi}}_m$ be means of the Durbin-Levinson algorithm applied to (1.6). Neither technique requires any matrix inversion. The choice between them will clearly depend on whether primary interest is in the MA(∞) or AR(∞) representation of X_t.

5. The estimators $\hat{\theta}_{mj}$ of ψ_j, $j = 1, \ldots, p+q$, can also be used to find preliminary estimators of the coefficients $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$, in the ARMA($p, q$) model,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{1.8}$$

where $\{Z_t\}$ is white noise and $(1 - \phi_1 z - \cdots - \phi_p z^p)$ $(1 + \theta_1 z + \cdots + \theta_q z^q) \neq 0$ for $|z| \leq 1$. We simply solve the equations,

$$\hat{\theta}_{mj} = \hat{\theta}_j + \sum_{i=1}^{\min(j,p)} \hat{\phi}_i \hat{\theta}_{m,j-i}, \quad j = 1, \ldots, p+q, \tag{1.9}$$

which are obtained from the corresponding equations,

$$\psi_j = \theta_j + \sum_{i=1}^{\min(j,p)} \phi_i \psi_{j-i}, \tag{1.10}$$

relating $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$, to the coefficients ψ_j in the representation (1.1) of the process defined by (1.8). The asymptotic distribution of the preliminary estimators $\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots \hat{\theta}_q$, can easily be found from that of $\hat{\theta}_{m1}, \ldots, \hat{\theta}_{m,p+q}$ using the relation (1.9) (see Brockwell and Davis (1987b)). This technique is simply the dual of the more customary procedure (see Fuller (1976)) of estimating π_j using $-\hat{\phi}_{m,j}$ as described in Remark 4, then finding estimates for $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ by solving the analogue of (1.9), namely,

$$\hat{\phi}_{mj} = \hat{\phi}_j - \sum_{i=1}^{\min(j,q)} \hat{\theta}_i \hat{\phi}_{m,j-i}, \quad j = 1, \ldots, p+q. \tag{1.11}$$

If (1.8) is the true model then both (1.9) and (1.11) will give causal invertible coefficient estimates asymptotically as $n \to \infty$ but not necessarily for finite n. If p is small equations (1.9) are trivial to solve, while if q is small equations (1.11) are trivial to solve.

6. For the ARMA(p, q) process defined by (1.8) the innovation representation (1.4) can be reexpressed as

$$X_t = \begin{cases} \sum_{j=0}^{t-1} \theta_{t-1,j}(X_{t-j} - \hat{X}_{t-j}), & t = 1, \ldots, \max(p, q), \\ \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=0}^{q} \theta_{t-1,j}(X_{t-j} - \hat{X}_{t-j}), & t > \max(p, q), \end{cases} \tag{1.12}$$

where $\theta_{nj}, 0 \leq j \leq n, n = 0, 1, 2, \ldots,$ now denote the coefficients obtained when Proposition 1.1 is applied to the covariance function of the process,

$$W_t = \begin{cases} X_t, & t = 1, \ldots, \max(p, q) \\ X_t - \sum_{j=1}^{p} \phi_j X_{t-j}, & t > \max(p, q) \end{cases}$$

The advantage of the representation (1.12) is that the last of the sums involves only $(q+1)$ terms instead of t terms as in the sum on the right of (1.3). The one-step predictors $\hat{X}_t$ are obtained from (1.12) by suppressing the summands with $j = 0$. From $\hat{X}_t$, the Gaussian likelihood of $(X_1, \ldots, X_n)$ is easily computed as

$$L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2; X_1, \ldots, X_n) = (2\pi)^{-n/2}(v_0 \cdots v_{n-1})^{-1/2} \exp\left\{-\tfrac{1}{2} \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{v_{j-1}}\right\},$$

where the mean squared errors $v_j, j = 0, \ldots, n-1$, are found by applying Proposition 1.1 to the process $\{W_t\}$. This is the method of likelihood calculation used by Ansley (1979). The maximum likelihood estimators of $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ and σ^2 are found by numerical maximization of L. Although these are asymptotically more efficient than the estimators of Remark 4, the latter provide good preliminary estimates with which to initialize the search and thereby accelerate the maximization of the likelihood. For estimating the coefficient θ of an MA(1) process, the efficiencies of the preliminary estimator relative to the maximum likelihood estimator are 0.94, 0.75 and 0.44 for $\theta = 0.25$, 0.50 and 0.75 respectively. The corresponding relative efficiencies for the moment estimator of θ are 0.82, 0.37 and 0.06. For higher order MA processes, analytic efficiency comparisons are difficult owing to the difficulty in computing the asymptotic covariance matrix of the moment estimators. However simulations indicate that the preliminary estimation procedure of Remark 4 is in fact substantially more efficient than moment estimation as in the case $q=1$.

7. Proposition 1.1 has a multivariate generalization (see Brockwell and Davis (1987a), p. 412). However, when it is applied to the multivariate analogue of (1.1), the coefficient matrices Θ_{nj} of the innovations will depend on the covariance matrix of the multivariate white noise sequence $\{Z_t\}$.

2. Properties of the estimators

Our data will consist of observations $X_1, \ldots, X_n$ of the process defined by (1.1). From now on $\theta_{mj}, j=0,1,\ldots,m$, and v_m, $m=0,1,2,\ldots$, will denote the coefficients and one-step mean square prediction errors obtained by applying the recursions (1.4) to the true covariances $\kappa(i,j)$ of the process (1.1). Similarly $\hat{\theta}_{mj}$ and $\hat{v}_m$ will denote the corresponding quantities obtained by applying the recursions to the *sample* covariances defined in Remark 2. Defining the coefficients $\pi_j, j=0,1,\ldots,$ as in (1.2) we have

$$\begin{aligned}\sigma^2=\operatorname{Var}(Z_{m+1})&=E(X_{m+1}+\sum_{j=1}^{\infty}\pi_j X_{m+1-j})^2\leq v_m\\&\leq E(X_{m+1}+\sum_{j=1}^{m}\pi_j X_{m+1-j})^2\\&=E(Z_{m+1}-\sum_{j>m}\pi_j X_{m+1-j})^2\\&\leq\sigma^2+\left(\sum_{j>m}|\pi_j|\right)^2\gamma(0).\end{aligned}\tag{2.1}$$

Moreover from (1.1) and (1.3) we have

$$\theta_{mk}=v_{m-k}^{-1}E[X_{m+1}(X_{m+1-k}-\hat{X}_{m+1-k})]$$

and

$$\psi_k=\sigma^{-2}E[X_{m+1}Z_{m+1-k}].$$

Using these relations with (2.1) we find that

$$\begin{aligned}|\theta_{mk}-\psi_k|^2&\leq\gamma(0)E[v_{m-k}^{-1}(X_{m+1-k}-\hat{X}_{m+1-k})-\sigma^{-2}Z_{m+1-k}]^2\\&=\gamma(0)[\sigma^{-2}-v_{m-k}^{-1}]\\&\leq\gamma^2(0)\sigma^{-4}\left(\sum_{j>m-k}|\pi_j|\right)^2\\&\to 0\quad\text{as } m\to\infty.\end{aligned}\tag{2.2}$$

The last result suggests the possibility of using an estimator of θ_{mk} in order to estimate ψ_k. We therefore consider the vectors $\hat{\boldsymbol{\theta}}_m$ defined in Remark 2 of Section 1. The following theorem, which is a dual of Bhansali's result in Remark 4 of Section 1, gives the asymptotic distribution of $\hat{\boldsymbol{\theta}}_m$.

Theorem 2.1. *Let $\{X_t\}$ be the linear process defined by* (1.1) *and let* $\{m(n), n = 1, 2, \ldots\}$ *be a sequence of integers such that as $n \to \infty$,*

$$(i)\ m < n, m \to \infty \text{ and } m = o(n^{1/3})$$

and

$$(ii)\ n^{1/2} \sum_{j>m} |\pi_j| \to 0.$$

Then, in $\mathbb{R}^\infty$,

$$n^{1/2}(\hat{\theta}_{m1} - \psi_1, \ldots, \hat{\theta}_{mm} - \psi_m, 0, 0, \ldots) \Rightarrow N(\mathbf{0}, \Sigma),$$

where $N(\mathbf{0}, \Sigma)$ denotes a zero-mean Gaussian sequence with covariance matrix Σ, defined in (1.5). *It follows in particular that*

$$n^{1/2}(\hat{\theta}_{mj} - \psi_j) \Rightarrow N(0, \sum_{k=0}^{j-1} \psi_k^2),$$

where $\psi_0 = 1$. □

Before proving Theorem 2.1 we need some preliminary results.

Proposition 2.1. Under the conditions of Theorem 2.1, we have in $\mathbb{R}^\infty$,

$$n^{1/2}(\hat{\phi}_{m1} - \phi_{m1}, \hat{\phi}_{m2} - \phi_{m2}, \ldots, \hat{\phi}_{mm} - \phi_{mm}, 0, 0, \ldots) \Rightarrow N(\mathbf{0}, \Lambda),$$

where Λ is defined in (1.7).

Proof. For $m < n$, define $\Gamma_m = [\gamma(i-j)]_{i,j=1}^m$ where $\gamma(h) = E(X_{t+h}X_t)$ is the autocovariance at lag h of the process $\{X_t\}$. Note that the eigenvalues of Γ_m are bounded below by $L = 2\pi \min_\lambda f(\lambda) > 0$ and above by $U = 2\pi \max_\lambda f(\lambda) < \infty$ where $f(\lambda) = |\psi(e^{-i\lambda})|^2\sigma^2/(2\pi)$ is the spectral density of $\{X_t\}$ (see e.g. Brockwell and Davis (1987, p. 132).

In view of Remark 4 of Section 1 it suffices to show that

$$n^{1/2}(\phi_{mi} + \pi_i) \to 0, \quad i = 1, 2, \ldots. \tag{2.3}$$

In order to show this, recall from the definitions of $\hat{X}_{m+1}$ and $\boldsymbol{\phi}_m$ that

$$\hat{X}_{m+1} = P_m X_{m+1} = \phi_{m1}X_m + \cdots + \phi_{mm}X_1, \tag{2.4}$$

where P_m denotes projection onto the span of $X_1, \ldots, X_m$ in L^2. Setting $\boldsymbol{\pi}_m = (\pi_1, \ldots, \pi_m)'$, we find from (1.2), (2.4) and the orthogonality of Z_{m+1} and $\hat{X}_{m+1}$, that

$$\begin{aligned}
\|\boldsymbol{\pi}_m + \boldsymbol{\phi}_m\|^2 &= \sum_{j=1}^{m} (\pi_j + \phi_{mj})^2 \\
&\leq L^{-1}(\boldsymbol{\pi}_m + \boldsymbol{\phi}_m)'\Gamma_m(\boldsymbol{\pi}_m + \boldsymbol{\phi}_m) \\
&= L^{-1}\operatorname{Var}\left(\sum_{j=1}^{m} (\pi_j + \phi_{mj})X_{m+1-j}\right) \\
&= L^{-1}\operatorname{Var}\Big(Z_{m+1} - (X_{m+1} - \hat{X}_{m+1}) - \sum_{j>m} \pi_j X_{m+1-j}\Big) \\
&\leq L^{-1}2\left(\left(\sum_{j>m} |\pi_j|\right)^2 \gamma(0) + (v_m - \sigma^2)\right), \\
&\leq 4L^{-1}\gamma(0)\left(\sum_{j>m} |\pi_j|\right)^2,
\end{aligned}$$

where the last inequality follows from (2.1). The required result (2.3) now follows from assumption (ii) of Theorem 2.1. □

Next recall that $\hat{X}_{m+1}$ has the two representations,

$$\hat{X}_{m+1} = \sum_{j=1}^{m} \theta_{mj}(X_{m+1-j} - \hat{X}_{m+1-j})$$

and

$$\hat{X}_{m+1} = \sum_{j=1}^{m} \phi_{mj} X_{m+1-j} = \sum_{j=1}^{m} \phi_{mj} \sum_{k=0}^{m-j} \theta_{m-j,k}(X_{m+1-j-k} - \hat{X}_{m+1-j-k}),$$

where $\theta_{i0} = 1$. Identifying the coefficients of $(X_{m+1-j} - \hat{X}_{m+1-j})$ we find that

$$\begin{bmatrix} \theta_{m1} \\ \vdots \\ \theta_{mk} \end{bmatrix} = R_{mk} \begin{bmatrix} \phi_{m1} \\ \vdots \\ \phi_{mk} \end{bmatrix}, \quad 1 \leq k \leq m, \tag{2.5}$$

where

$$R_{mk} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ \theta_{m-1,1} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \theta_{m-1,k-2} & \theta_{m-2,k-3} & \cdots & 1 & 0 \\ \theta_{m-1,k-1} & \theta_{m-2,k-2} & \cdots & \theta_{m-k+1,1} & 1 \end{bmatrix}. \tag{2.6}$$

Moreover, because of the way in which the estimators $\hat{\theta}_{mj}$ and $\hat{\phi}_{mj}$ are defined, we also have

$$\begin{bmatrix} \hat{\theta}_{m1} \\ \vdots \\ \hat{\theta}_{mk} \end{bmatrix} = \hat{R}_{mk} \begin{bmatrix} \hat{\phi}_{m1} \\ \vdots \\ \hat{\phi}_{mk} \end{bmatrix}, \quad 1 \leq k \leq m, \tag{2.7}$$

where $\hat{R}_{mk}$ is defined as in (2.6) with $\hat{\theta}_{ij}$ replacing θ_{ij} for each i and j. Now if $\{m(n)\}$ satisfies the conditions of Theorem 2.1, then since $\hat{\phi}_{m1} \xrightarrow{P} -\pi_1$, we have $\hat{\theta}_{m1} = \hat{\phi}_{m1} \xrightarrow{P} -\pi_1 = \psi_1$. Similarly $\hat{\phi}_{m2} \xrightarrow{P} -\pi_2$, so that $\hat{\theta}_{m2} = \hat{\theta}_{m-1,1}\hat{\phi}_{m1} + \hat{\phi}_{m2} \xrightarrow{P} -\psi_1\pi_1 - \pi_2 = \psi_2$. Repetition of this argument gives $\hat{\theta}_{mj} \xrightarrow{P} \psi_j$, for $j = 1, 2, \ldots, k$, and hence

$$\hat{R}_{mk} \xrightarrow{P} R_k \quad \text{as } m \to \infty, \tag{2.8}$$

where

$$R_k = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ \psi_1 & 1 & \cdots & 0 & 0 \\ \psi_2 & \psi_1 & \ddots & \vdots & \vdots \\ \vdots & \vdots & & 1 & 0 \\ \psi_{k-1} & \psi_{k-2} & \cdots & \psi_1 & 1 \end{bmatrix}.$$

Proof of Theorem 2.1. For a fixed positive integer k, define $\boldsymbol{\theta} := (\theta_{m1}, \ldots, \theta_{mk})'$ and $\boldsymbol{\phi} := (\phi_{m1}, \ldots, \phi_{mk})'$, where we have suppressed the dependence of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ on m. Define also the corresponding estimators $\hat{\boldsymbol{\theta}} := (\hat{\theta}_{m1}, \ldots, \hat{\theta}_{mk})'$ and $\hat{\boldsymbol{\phi}} := (\hat{\phi}_{m1}, \ldots, \hat{\phi}_{mk})'$. Using (2.5) and (2.6) we can write

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \hat{R}_{mk}\hat{\boldsymbol{\phi}} - R_{mk}\boldsymbol{\phi},$$

i.e.

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \hat{R}_{mk}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) + (\hat{R}_{mk} - R_{mk})\boldsymbol{\phi}. \tag{2.9}$$

The second term of (2.9) can be decomposed further as

$$(\hat{R}_{mk} - R_{mk})\boldsymbol{\phi} = (\hat{R}_{mk} - \hat{R}^*_{mk})\boldsymbol{\phi} + (\hat{R}^*_{mk} - R^*_{mk})\boldsymbol{\phi} + (R^*_{mk} - R_{mk})\boldsymbol{\phi} \tag{2.10}$$

where

$$R^*_{mk} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ \theta_{m1} & 1 & \cdots & 0 & 0 \\ \theta_{m2} & \theta_{m1} & \ddots & \vdots & \vdots \\ \vdots & \vdots & & 1 & 0 \\ \theta_{m,k-1} & \theta_{m,k-2} & \cdots & \theta_{m1} & 1 \end{bmatrix}$$

and $\hat{R}^*_{mk}$ is the corresponding matrix obtained by replacing θ_{ij} with $\hat{\theta}_{ij}$ for each i and j. The next step in the argument is to show that

$$\hat{R}_{mk} - \hat{R}^*_{mk} = o_p(n^{1/2}) \quad \text{and} \quad R_{mk} - R^*_{mk} = o(n^{1/2}),$$

where $o(n^{1/2})$ means that each component is $o(n^{1/2})$. We thus need to show that

$$n^{1/2}(\hat{\theta}_{mj} - \hat{\theta}_{m-i,j}) \xrightarrow{P} 0 \quad \text{and} \quad n^{1/2}(\theta_{mj} - \theta_{m-i,j}) \to 0 \tag{2.11}$$

for $i = 1, 2, \ldots, k-1$, and $j = 1, 2, \ldots, k$. It is easy to show, by arguments we have used earlier, that

$$n^{1/2}(\hat{\phi}_{mj} - \hat{\phi}_{m-i,j}) \xrightarrow{P} 0 \quad \text{and} \quad n^{1/2}(\phi_{mj} - \phi_{m-i,j}) \to 0 \tag{2.12}$$

and, since $\hat{\theta}_{m1}-\hat{\theta}_{m-i,1}=\hat{\phi}_{m1}-\hat{\phi}_{m-i,1}$ and $\theta_{m1}-\theta_{m-i,1}=\phi_{m1}-\phi_{m-i,1}$, this establishes (2.11) with $j=1$. The cases $j=2,3,\ldots,k$ follow iteratively using (2.5), (2.7) (2.12) and the arguments used to derive (2.8).

Now from (2.9), (2.10) and (2.11) it follows that

$$\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}=\hat{R}_{mk}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi})+(\hat{R}^*_{mk}-R^*_{mk})\boldsymbol{\phi}+o_p(n^{1/2}). \qquad (2.13)$$

Inspection of the middle term on the right side of (2.13) shows that it can be rewritten in the form,

$$(\hat{R}^*_{mk}-R^*_{mk})\boldsymbol{\phi}=\begin{bmatrix}0 & 0 & \cdots & 0 & 0\\ \phi_{m1} & 0 & \cdots & 0 & 0\\ \phi_{m2} & \phi_{m1} & \ddots & \vdots & \vdots\\ \vdots & \vdots & & 0 & 0\\ \phi_{m,k-1} & \phi_{m,k-2} & \cdots & \phi_{m1} & 0\end{bmatrix}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}),$$

so that

$$\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}=A_m^{-1}\hat{R}_{mk}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi})+o_p(n^{1/2}),$$

where

$$A_m=\begin{bmatrix}1 & 0 & \cdots & 0 & 0\\ -\phi_{m1} & 1 & \cdots & 0 & 0\\ -\phi_{m2} & -\phi_{m1} & \ddots & \vdots & \vdots\\ \vdots & \vdots & & 1 & 0\\ -\phi_{m,k-1} & -\phi_{m,k-2} & \cdots & -\phi_{m1} & 1\end{bmatrix}\to A=\begin{bmatrix}1 & 0 & \cdots & 0 & 0\\ \pi_1 & 1 & \cdots & 0 & 0\\ \pi_2 & \pi_1 & \ddots & \vdots & \vdots\\ \vdots & \vdots & & 1 & 0\\ \pi_{k-1} & \pi_{k-2} & \cdots & \pi_1 & 1\end{bmatrix}.$$

Hence, using (2.8) together with Bhansali's result in Remark 4 of Section 1, we find that

$$n^{1/2}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \Rightarrow N(\mathbf{0},V),$$

where

$$V=A^{-1}R_k\Lambda_{k\times k}R_k'(A')^{-1}. \qquad (2.14)$$

and $\Lambda_{k\times k}$ denotes the top left $k\times k$ truncation of the matrix Λ defined in (1.7). From (1.7) it is clear that

$$\Lambda_{k\times k}=AA'$$

and, since $\pi(z)\psi(z)=1$ for $|z|\leq 1$, we also have

$$R_k A=I_k,$$

where I_k is the $k\times k$ identiy matrix. Consequently

$$\begin{aligned}V&=A^{-1}R_kAA'R_k'(A')^{-1}\\&=R_kR_k'\\&=\textstyle\sum_{k\times k},\end{aligned}$$

where $\sum_{k\times k}$ is the top left truncation of the matrix $\sum$ defined in (1.5). We have thus established that

$$n^{1/2}(\hat{\theta}_{m1}-\theta_{m1},\ldots,\hat{\theta}_{mm}-\theta_{mm},0,0,\ldots) \Rightarrow N(\mathbf{0},\textstyle\sum)$$

in $\mathbb{R}^{\infty}$, since the finite-dimensional distributions converge. To complete the proof of Theorem 2.1, we need only show that

$$n^{1/2}(\theta_{mi}-\psi_i)\to 0, \quad i=1,2,\ldots,$$

as $n\to\infty$. But this follows from (2.2) and the assumptions on the sequence $\{m(n)\}$. □

3. Preliminary estimation for MA(q) processes

Let $\{X_t\}$ be the MA(q) process,

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

where $\{Z_t\}$ is an i.i.d. sequence of random variables such that $EZ_t=0$, $EZ_t^2=\sigma^2$ and $EZ_t^4<\infty$, and assume that $\theta(z):=1+\theta_1 z+\cdots+\theta_q z^q\neq 0$ for $|z|\leq 1$. The vector $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_q)'$ can then be estimated by $\hat{\boldsymbol{\theta}}=(\hat{\theta}_{m1},\ldots,\hat{\theta}_{mq})'$. which by Theorem 2.1 is asymptotically normal with mean $\boldsymbol{\theta}$ and the covariance matrix whose (i,j) element is equal to $n^{-1}\sum_{k=0}^{\min(i,j)-1}\theta_k\theta_{k+|i-j|}$ (where $\theta_0:=1$ and $\theta_j:=0$ for $j>q$). Moreover for *any* fixed j (possibly greater than q) the asymptotic distribution of θ_{mj} is asymptotically normal with mean θ_j and variance $n^{-1}\sum_{k=0}^{j-1}\theta_{mk}^2$, *regardless of the value of* q. Inspection of the asymptotic 95 percent confidence bounds,

$$\hat{\theta}_{mj}\pm 1.96\left(n^{-1}\sum_{k=0}^{j-1}\hat{\theta}_{mk}^2\right)^{1/2}, \tag{3.1}$$

for $\theta_j, j=1,2,\ldots$, therefore provides a means for deciding which of the coefficients $\theta_1,\theta_2,\ldots$ are different from zero, and thus for estimating the order q of the underlying process. The vector $\hat{\boldsymbol{\theta}}$ can be computed extremely rapidly and it has reasonable efficiency relative to the maximum likelihood estimator. It is also substantially more efficient than some other commonly-used preliminary estimators, such as those derived by equating theoretical and sample autocovariances at lags $0,\ldots,q$ (see Brockwell and Davis (1987b)). We demonstrate the use of the technique in the following example by applying it to the International Airline Passenger Data of Box and Jenkins (1976), p. 531.

Example 3.1. Let $\{Y_t, t=1,\ldots,144\}$ denote the International Airline Passenger Data. As in the analysis of Box and Jenkins, the data was first transformed by taking natural logarithms, $L_t=\ln Y_t$, then applying the operator $(1-B)(1-B^{12})$ to produce a new series, stationary in appearance, and with rapidly decaying sample autocorrelation function. If we write

$$X_t = (1-B)(1-B^{12})L_{t+13}, \quad t=1,\ldots,131,$$

then the sample autocorrelation function of X_t suggests a moving average model with zero coefficients for lags greater than 23. (Box and Jenkins fitted a multiplicative moving average model of order 13.)

The graphs of $\hat{\theta}_{mj}$, $1 \le j \le 30$, and the bounds $\pm 1.96(n^{-1}\sum_{k=0}^{j-1}\hat{\theta}_{mk}^2)^{1/2}$ are shown in Fig. 1 for $m=30$ and $m=50$. In view of (3.1) a value of $\hat{\theta}_{mj}$ outside the bounds suggests that the corresponding coefficient θ_j is non-zero. The graphs thus suggest the model,

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_3 Z_{t-3} + \theta_{12} Z_{t-12} + \theta_{23} Z_{t-23}, \tag{3.2}$$

where $\{Z_t\}$ is white noise. At the same time they provide us with the preliminary estimates $\hat{\theta}_j = \hat{\theta}_{30,j}$, $j=1,3,12,23$, where

$$\hat{\theta}_1 = -0.357, \quad \hat{\theta}_3 = -0.158, \quad \hat{\theta}_{12} = -0.479 \quad \text{and} \quad \hat{\theta}_{23} = 0.254. \tag{3.3}$$

(There is very little difference between the values of $\hat{\theta}_{mj}$ for $30 \le m \le 50$.)

Using the maximum likelihood technique described in Remark 6 of Section 1, estimates of the parameters $\theta_1, \theta_3, \theta_{12}$ and θ_{23} were then obtained, using as initial values in the optimization the preliminary estimates found in the preceding paragraph. The maximum likelihood model was found to be

$$X_t = Z_t - 0.372Z_{t-1} - 0.214Z_{t-3} - 0.537Z_{t-12} + 0.232Z_{t-\ }, \tag{3.4}$$

where $\{Z_t\}$ is white noise with variance 0.00123. The Akaike information criterion for this model has the value, AIC $= -861.757$.

The model for $\{X_t\}$ fitted by Box and Jenkins was

$$X_t = (1-0.396B)(1-0.614B^{12})Z_t, \qquad \{Z_t\} \sim WN(0, 0.00134). \tag{3.5}$$

Although this model has two fewer parameters than (3.4), it gives a higher AIC value, viz. AIC $= -856.247$. The sample autocorrelation function of the residuals from the model (3.5) is compatible with that of white noise insofar as it passes the portmanteau test. There is however a rather large value, 0.219, at lag 23, which is well outside the 0.95 bounds, $\pm 1.96/\sqrt{131} = \pm 0.171$. Maximum likelihood fitting was also carried out for the more general model,

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_3 Z_{t-3} + \theta_{12} Z_{t-12} + \theta_{13} Z_{t-13} + \theta_{23} Z_{t-23},$$

and the for models obtained by setting subsets of these parameters equal to zero. However the best such model on the basis of the AIC criterion was found to be (3.4), the one which was first suggested by our identification technique. Notice also that the preliminary estimated values (3.3) are quite close to the maximum likelihood estimates in (3.4).

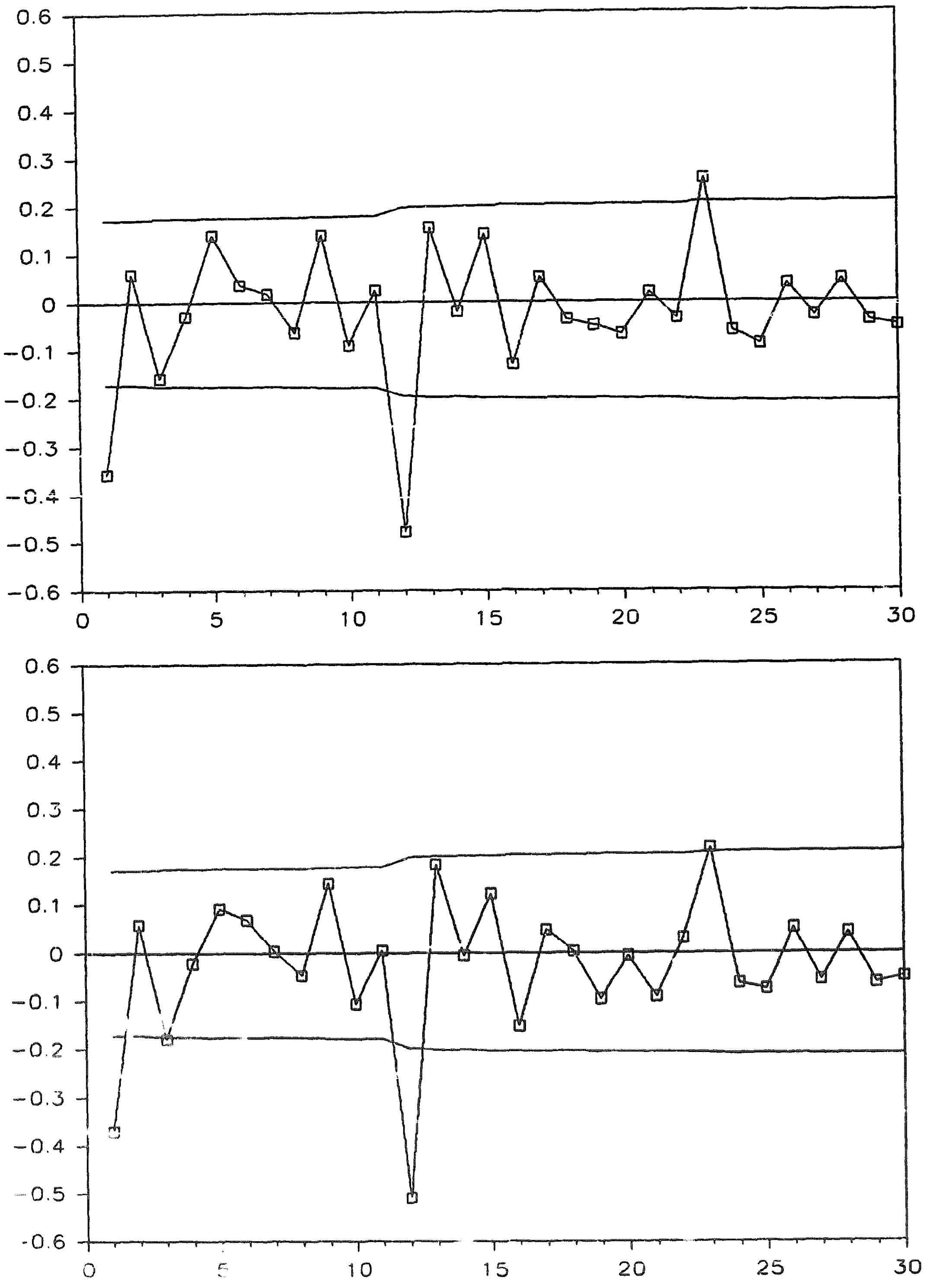

Fig. 1. Graphs of $\theta_{30,j}$ and $\theta_{50,j}$, $j = 1, \ldots, 30$.

References

C.F. Ansley, An algorithm for the exact likelihood of a mixed autoregressive-moving average process, Biometrika 66 (1979) 59–65.

K.N. Berk, Consistent autoregressive spectral estimates, Ann. Statist. 2 (1974) 489–501.

R.J. Bhansali, Linear prediction by autoregressive model fitting in the time domain, Ann. Statist. 6 (1978) 224–231.

P. Billingsley, Convergence of Probability Measures (John Wiley & Sons Inc., New York, (1968))

G.E.P. Box and G.M. Jenkins, Time Series Analysis, Forecasting and Control, 2nd edition (Holden Day, San Francisco, 1976).

P.J. Brockwell and R.A. Davis, Time Series: Theory and Methods (Springer-Verlag, New York, 1987a).

P.J. Brockwell and R.A. Davis, Applications of innovation representations in time series analysis, in: J.N. Srivastava, ed., Graybill Festschrift Volume (North-Holland, Amsterdam, 1987b).

W.A. Fuller, Introduction to Statistical Time Series (John Wiley & Sons Inc., New York, 1976).

J. Kissanen and L. Barbosa, Properties of infinite covariance matrices and stability of optimum predictors, Information Sci. 1 (1969) 221–236.