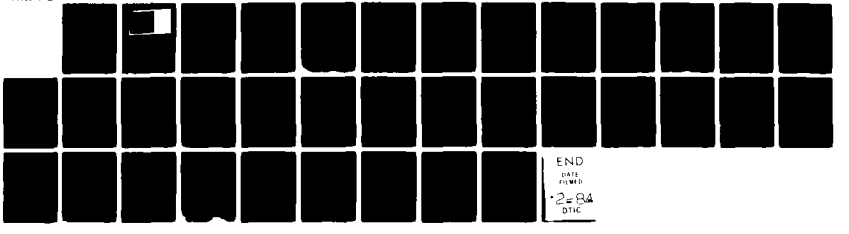
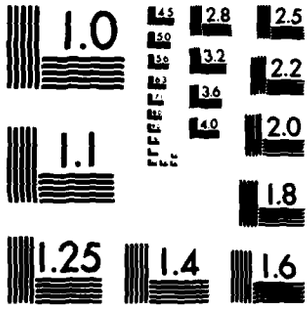
AD-A136 426 A UNIFIED APPROACH TO ARMA (AUTOREGRESSIVE-MOVING AVERAGE) MODEL IDENTIFI..(U) WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER G T WILSON ET AL. NOV 83 MRC-TSR-2592 DAAG29-80-C-0041 F/G 12/1 NL

1/1

UNCLASSIFIED



END
DATE FILMED
2-84
DTIC



MRC Technical Summary Report #2592

A UNIFIED APPROACH TO ARMA MODEL IDENTIFICATION AND PRELIMINARY ESTIMATION

G. Tunnicliffe Wilson and D. Piccolo

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

November 1983

(Received August 1, 1983)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

A UNIFIED APPROACH TO ARMA MODEL IDENTIFICATION AND PRELIMINARY ESTIMATION

G. Tunnicliffe Wilson and D. Piccolo*

Technical Summary Report #2592
November 1983

ABSTRACT

This paper reviews several different methods for identifying the orders of autoregressive-moving average models for time series data. The case is made that these have a common basis, and that a unified approach may be found in the analysis of a matrix G, defined to be the covariance matrix of forecast values.

The estimation of this matrix is considered, emphasis being placed on the use of high order autoregression to approximate the predictor coefficients. Statistical procedures are proposed for analysing G, and identifying the model orders.

A simulation example and three sets of real data are used to illustrate the procedure, which appears to be very useful as a tool for order identification and preliminary model estimation.

AMS (MOS) Subject Classifications: 62M10, 65U05

Key Words: ARMA model identification; preliminary estimation; Yule-Walker equations; Durbin-Levinson algorithm; prediction spaces; Choleski factorization

Work Unit Number 4 (Statistics and Probability)

*Partially sponsored by NATO Senior Fellowship Scheme, CNR, Italy.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

The prediction of a sequence of autocorrelated observations is generally facilitated by the use of autoregressive-moving average (ARMA) models. These represent the observations in terms of simple recurrence relations. The numbers of terms in these equations (the orders) and the coefficients (or parameters) need to be estimated from the observations. This is a complex problem which generally requires nonlinear optimisation for a range of possible orders.

Considerable effort has been devoted in recent years, to methods for obtaining rapid and moderately efficient, though necessarily approximate, solutions to this problem. The paper reviews the work in this area, and proposes a unified procedure which relies on detecting the singularity of a certain matrix. This is the covariance matrix of the predictions, and it may be estimated by fitting a (linear) high order autoregressive model to the data.

Examples indicate that this approach is remarkably successful in providing good preliminary estimates of the ARMA model orders and parameters.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

A UNIFIED APPROACH TO ARMA MODEL IDENTIFICATION AND PRELIMINARY ESTIMATION

G. Tunnicliffe Wilson and D. Piccolo*

1. INTRODUCTION

Since the work by Box and Jenkins (1970), henceforth referred to as BJ, there has been a renewed and continuing interest in improved methods of characterising the orders, and obtaining good preliminary estimates for the parameters of ARMA models. Following the notation of BJ let us therefore consider a zero-mean stationary times series x_t, which is believed to follow an ARMA (p,q) model:

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} . \qquad (1.1)$$

In theory, the orders p,q and parameters ϕ,θ can be determined by maximum likelihood estimation of the parameters for increasing values of the orders. This is time consuming, because nonlinear estimation methods are necessary whenever $q > 0$. In this case, quick techniques for identifying model orders and supplying preliminary parameter estimates are valuable even if they are not fully efficient, and on occasions might be inadequate. These identification methods are usually based on relatively rapid and direct computations deriving from sample autocorrelations and linear regression. The aim of this paper is to put the case that many of the procedures which have been advocated since the appearance of BJ, and also many of the earlier methods - particularly those due to Durbin (1959, 1960) - may be viewed as having essentially the same basis. This leads to a practical procedure which is very similar in its motivation to the implementation of the work of Durbin (1960) as presented recently by Hannan and Rissanen (1982) - henceforth referred to as HR. We believe however that is has some advantages, and may appeal to the practitioner of time series analysis. We end the paper with a discussion of how a simple regression may furnish estimates of ARMA model parameters which are very close to the maximum likelihood estimates.

*Partially sponsored by NATO Senior Fellowship Scheme, CNR, Italy.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

2. PROPERTIES OF THE MODEL

Readers of BJ will be familiar with three very similar equations which follow directly from the model (1.1), each of which also characterises the ARMA model. The first of these is (BJ p. 75)

$$\rho_j = \phi_1\rho_{j-1} + \ldots + \phi_p\rho_{j-p} \quad \text{for} \quad j > q \tag{2.1}$$

where ρ_k is the autocorrelation function (acf) of x_t.

The pattern of the acf for $j > q - p$ is a mixture of damped exponential and sinusoidal terms. Inspection of the sample acf with a view to recognising such a pattern is an important part of ARMA modelling. Increasingly, there is a desire for automatic recognition of this pattern by detecting for which values of p and q the linear relationship (2.1) holds.

The second equation is (BJ p. 139)

$$\hat{x}_n(j) = \phi_1\hat{x}_n(j-1) + \ldots + \phi_p\hat{x}_n(j-p) \quad \text{for} \quad j > q \tag{2.2}$$

where $\hat{x}_n(j)$ is, for fixed forecast origin n and increasing lead time j, the forecast function of the series, i.e.

$$\hat{x}_n(j) = E(x_{n+j}|P) \tag{2.3}$$

where P is a set of past variables. Although the past is usually taken to be the whole of the set $x_n, x_{n-1}, x_{n-2} \ldots$ the equation still holds if P is any subset of these. In a sense, (2.2) is more fundamental than (2.1) because by taking $P = \{x_n\}$, we get $\hat{x}_n(j) = \rho_j x_n$ by simple regression, and substituting this into (2.2) leads to (2.1).

In recent years, following very much on the work of Akaike (1974), the structure of the predictor space of random variables defined by the forecast function, has been seen to have an important role in characterising the ARMA model. See also Mazzali (1982) for work which is closely related to our own use of this concept. For an ARMA (p,q) model, (2.2) implies that this space has a finite dimension even if the past P is infinite, because for $j > q$, $\hat{x}_n(j)$ is a linear function of previous values. Define G to be the prediction covariance matrix of $\hat{x}_n(j)$ for $j = 1 \ldots K$, where K is some relatively large (and for the time being arbitrary) upper limit on the forecast function length. We

shall see how most methods of ARMA model identification are closely linked with the structure of G. The value of G will depend on the selection of variables in the past, P. Let us now choose these to be $x_n, x_{n-1} \cdots x_{n-M+1}$. Then partitioning the variables

$$x = x_{n+K} \cdots x_{n+1}, x_n \cdots x_{n-M+1} = x_F, x_P$$

into the future set x_F and past set x_P, we have as their joint covariance matrix

$$\Gamma = \begin{pmatrix} \Gamma_{FF} & \Gamma_{FP} \\ \Gamma_{PF} & \Gamma_{PP} \end{pmatrix} \tag{2.4}$$

where Γ has elements $\Gamma_{ij} = \sigma_x^2 \rho_{i-j}$, with $\sigma_x^2 = \text{Var } x_t$. In particular,

$$\Gamma_{PF} = \Gamma_{FP}' = \sigma_x^2 \begin{pmatrix} \rho_K & \rho_{K-1} & \cdots & \rho_1 \\ \rho_{K+1} & \rho_K & \cdots & \rho_2 \\ \vdots & & & \vdots \\ \rho_{M-1} & \rho_{M-2} & \cdots & 1 \end{pmatrix} \tag{2.5}$$

and

$$\Gamma_{PP} = \sigma_x^2 \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{M-1} \\ \rho_1 & 1 & \cdots & \rho_{M-2} \\ \vdots & & & \vdots \\ \rho_{M-1} & \rho_{M-2} & \cdots & 1 \end{pmatrix} \tag{2.6}$$

Also Γ_{FF} is the same as Γ_{PP} but with M replaced by K.

Using the standard result that $\hat{x}_F = E(x_F | x_P) = \Gamma_{FP}\Gamma_{PP}^{-1}x_P$, we obtain

$$G = E(\hat{x}_F \hat{x}_F') = \Gamma_{FP}\Gamma_{PP}^{-1}\Gamma_{PF} \tag{2.7}$$

The <u>third</u> equation to which we draw attention is (BJ p. 156)

$$\psi_j = \phi_1\psi_{j-1} + \cdots + \phi_p\psi_{j-p} \quad \text{for } j > q \tag{2.8}$$

where ψ_j is the coefficient of B^j in the expansion of $\psi(B) = \theta(B)/\phi(B)$, which arises in the representation $x_t = \psi(B)a_t$.

To see how this relates to the previous equation, recall that

$$\hat{x}_n(j) = \psi_j a_n + \psi_{j+1}a_{n+1} + \cdots$$

where now P is the infinite past (M = ∞). Define further the partitioned matrix

$$\Psi = \begin{pmatrix} \Psi_F \\ \Psi_P \end{pmatrix} = \left(\begin{array}{ccccc} 1 & & & & \\ \psi_1 & 1 & & & \\ \vdots & \vdots & & & \\ \psi_{K-1} & \psi_{K-2} & \cdots & 1 & \\ \hline \psi_K & \psi_{K-1} & \cdots & \psi_1 & \\ \psi_{K+1} & \psi_K & & \psi_2 & \\ \vdots & \vdots & & \vdots & \end{array} \right) \tag{2.9}$$

and vector

$$a' = (a_F, a_P)' = (a_{n+k} \cdots a_{n+1} \mid a_n \; a_{n-1} \cdots)' .$$

Then $x_F = \Psi' a$ and $\hat{x}_F = \Psi_P' a_P$, so that

$$\Gamma_{FF} = \Psi'\Psi\sigma_a^2 = (\Psi_F'\Psi_F + \Psi_P'\Psi_P)\sigma_a^2 \tag{2.10}$$

and

$$G = E(\hat{x}_F \hat{x}_F') = \Psi_P'\Psi_P\sigma_a^2 \tag{2.11}$$

$$= \Gamma_{FF} - \Psi_F'\Psi_F\sigma_a^2 . \tag{2.12}$$

Finally, defining the vector of length max(p,q) + 1,

$$\phi = \begin{pmatrix} 1 \\ -\phi_1 \\ \vdots \\ -\phi_p \end{pmatrix} \text{ if } p > q, \text{ or } \begin{pmatrix} 1 \\ -\phi_1 \\ \vdots \\ -\phi_p \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ if } p < q \tag{2.13}$$

we note that the three equations (2.1), (2.2) and (2.8) to which we have attached importance, may be presented for K = max(p,q) + 1 as

$$\Gamma_{PF}\phi = 0 \tag{2.14}$$

$$E(\hat{x}_F \hat{x}_F')\phi = G\phi = 0 \tag{2.15}$$

$$\Psi_p\phi = 0 \ . \tag{2.16}$$

Estimation of the matrices in these equations, and detection of the order p at which the implied rank deficiencies occur in (2.14)-(2.16), serves both to identify $\max(p,q)$ and to provide preliminary estimates of ϕ. We now examine how previous methods relate to this. For present convenience we shall use p in the place of $\max(p,q)$, and extend one or the other of ϕ_j, θ_j up to order p with zeros. We assume that one of ϕ_p or θ_p is non-zero and that $\phi(B)$ and $\theta(B)$ have no common factors. We shall refer to this as the assumption that $p = q$.

3. ARRAY METHODS

Béguin, Gourieroux and Monfort (1980) proposed a corner method in which an array of determinants is scanned for a set of zeros. This is closely related to the preliminary estimation procedure given in the appendix p. 499 of BJ, where a shifted set of Yule-Walker equations is solved for $\phi_1 \ldots \phi_p$. The matrix in these equations is that used by Beguin et al., and when both p and q exceed their true values the matrix is singular, and the determinant zero. In the case $p = q$ this matrix is simply Γ_{PF} with $M = K = p$. Gray, Kelley and McIntire (1978) also use an array, and a statistic which is closely related to the above determinant, being in fact a ratio of two determinants. They look for a change from a relatively constant pattern of statistics, to an undefined or erratic set when p and q become too large. Woodward and Gray (1981) relate this statistic to the methods of BJ.

Again, Glasbey (1982) proposes a related test statistic which similarly becomes undefined when p and q become too large.

The problems with these methods derives from the fact that when sample correlations are used to estimate the matrix Γ_{PF}, it is unlikely to be exactly singular and it is difficult to derive statistical tests for this hypothesis. We suggest that it should be more efficient to look for rank deficiency in Γ_{PF} when $K = p = 1$ and M is large.

This requires combining the information in the rows of Γ_{PF}, and this is precisely what is achieved in the matrix G.

4. CANONICAL CORRELATION ANALYSIS

Investigation of the prediction space structure by analysis of the correlation between the future (F) and the past (P) was advocated by Akaike (1976). He chose to use the tool of canonical correlation analysis. This recognises that (in the case $p = q$) the linear combination of future values:

$$x_{n+j} - \phi_1 x_{n+j-1} - \cdots - \phi_p x_{n+j-p} \quad ; \quad \text{for} \quad j > p$$

is uncorrelated with any set of variables in the past. The procedure is to choose a sufficiently large but fixed number M of values in the past, and an increasing number $K = 1,2\ldots$ values in the future until a zero canonical correlation is found. At this point $K = p + 1$, and the coefficients in the canonical factor, after scaling, supply the values of $\phi_1 \ldots \phi_p$. All that is required is an estimate of the covariance matrix Γ partitioned as in (2.4). This is then submitted to a standard routine. It is shown in Anderson (1958) p. 296 that the analysis seeks for eigenvalues λ_i of a reduced matrix system which in our notation is

$$(G - \lambda^2 \Gamma_{FF})v = 0 \ . \tag{4.1}$$

Because the hypothesis is that the smallest eigenvalue is zero, we are again in a situation of detecting a singularity of G.

5. REGRESSION USING ESTIMATED INNOVATIONS

The ARMA model (1.1) appears like a regression equation except that the lagged variables a_{t-j} on the RHS are unobserved. Durbin (1960) suggested that estimates of these terms may be obtained by first fitting a suitably high order regression to x_t, i.e.

$$x_t = \pi_1 x_{t-1} + \cdots + \pi_M x_{t-M} + \alpha_t \tag{5.1}$$

In theory, and in the limit $M \to \infty$, these coefficients are the π weights in the expansion of $\pi(B) = \phi(B)/\theta(B)$, and the residuals α_t are the model innovations a_t. In practice, sample correlations are used with the Levinson-Durbin algorithm to estimate the

coefficients, which should then be referred to as $w_{M,1} \ldots w_{M,M}$, and the residuals α_t should be distinguished from a_t.

A regression is then carried out following (1.1), using both past values of x_t and α_t

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} - \theta_1 \alpha_{t-1} - \ldots - \theta_q \alpha_{t-q} + e_t \qquad (5.2)$$

where e_t is the error in this regression. As a method of parameter estimation this is very useful, giving in theory consistent, though not fully efficient estimates. As a method of selecting the orders, the regression can be carried out for increasing p and q until no further appreciable reduction in residual sum of squares is obtained. HR show how this can be done in a consistent manner by minimizing an appropriate criterion.

Tsay and Tiao (1983a) use a similar regression to (5.2) using residuals from a regression such as (5.1), but the residuals they use are from equations of differing orders. They use in effect the minimum number M of past values in (5.1), to ensure consistency of the estimates of ϕ. The estimates of θ in (5.2) are in any case only consistent in the limit $M \to \infty$. Tsay and Tiao are not primarily concerned with estimation of θ and do not even refer to the coefficients of the residuals as moving average parameters. They do however use ordinary least squares in all their regressions, which allows the method to be used in the case of non-stationary ARMA processes.

Also recently, Young, Jakeman and McMurtries (1980) have reported methods of order identification in transfer function modelling which use the trace of the covariance matrix of the parameter estimates as a criterion. When this trace becomes large, the model is taken to be over parameterised, i.e. the orders are too large. Young has reported to us the results of applying this method to ARMA modelling by fitting a transfer function between α_t and x_t, i.e.

$$\alpha_t = \frac{\phi(B)}{\theta(B)} x_t + u_t \,, \qquad (5.3)$$

where α_t are the residuals already defined by a regression such as (5.1). An introduction to the instrumental variable estimation methods used, is given by Young (1974). In essence, they are in this context equivalent to estimating the parameters in the linear regression

$$x_t^* - \alpha_t^* = \phi_1 x_{t-1}^* + \ldots + \phi_p x_{t-p}^* - \theta_1 \alpha_{t-1}^* - \ldots - \theta_q \alpha_{t-q}^* + f_t \qquad (5.4)$$

where f_t is the error, and x_t^*, α_t^* are filtered versions of x_t, α_t, i.e.

$$x_t^* = L(B)x_t, \quad \alpha_t^* = L(B)\alpha_t .$$

From (5.3) the ideal choice for $L(B)$ is $1/\theta(B)$ where the true (but unknown) value of $\theta(B)$ should be used. The first stage is to set $L(B) = 1$, with the intention of using the estimates of $\theta(B)$ from this stage, to define $L(B)$ in a further stage. For the present we take $L(B) = 1$, giving $x_t^* = x_t$ and $\alpha_t^* = \alpha_t$, so that (5.4) differs from (5.2) only in the presence of α_t on the LHS.

We now deal with this point, and relate to previous sections, by noting that α_t is by construction orthogonal to $x_{t-1} \ldots x_{t-p}$, and provided M is sufficiently large, also to $\alpha_{t-1} \ldots \alpha_{t-q}$. Its presence in (5.4) does not therefore affect the estimates of ϕ and θ - it may even be included in the regression on the RHS. The estimated residual variance σ_f^2 is however smaller than σ_e^2 by the amount σ_α^2;

$$\sigma_f^2 = \sigma_e^2 - \sigma_\alpha^2 .$$

Thus as p and q are increased to obtain a better fit, $\sigma_f^2 \to 0$ as $\sigma_e^2 \to \sigma_\alpha^2$. It is of interest also to compare HR who use a bivariate autoregression on (x_t, α_t) to obtain the parameters in (5.2). They monitor the determinant of the bivariate error covariance matrix which may be seen to be approximately

$$\begin{vmatrix} \sigma_e^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 \end{vmatrix} = \sigma_\alpha^2(\sigma_e^2 - \sigma_\alpha^2) = \sigma_\alpha^2 \sigma_f^2 \qquad (5.5)$$

They also note that this may approach zero.

Considering now the case $p = q$ in (5.2), we point out that because $\alpha_t \ldots \alpha_{t-p}$ are orthogonal, it is a simple matter to orthogonalise $x_t \ldots x_{t-p}$ with respect to these variables. In theory, in the limit as $M \to \infty$, $\alpha_t \to a_t$. The result is a corrected set of variables given by the regression coefficients ψ_j of x_t on past values a_{t-j}:

$$\left.\begin{aligned}
\tilde{x}_t &= x_t - (a_t + \psi_1 a_{t-1} + \ldots + \psi_p a_{t-p}) \\
\tilde{x}_{t-1} &= x_{t-1} - (a_{t-1} - \psi_1 a_{t-2} + \ldots + \psi_{p-1} a_{t-p}) \\
&\vdots \\
\tilde{x}_{t-p} &= x_{t-p} - a_{t-p}
\end{aligned}\right\} \qquad (5.6)$$

These corrected variables are simply $\hat{x}_{t-p-1}(j)$ for $j = p+1 \ldots 1$. Substituting from (5.6) into the regression (5.2) then results in a partial regression equation in the forecast function only:

$$\tilde{x}_t = \phi_1 \tilde{x}_{t-1} + \ldots + \phi_p \tilde{x}_{t-p} + e_t \, . \qquad (5.7)$$

The coefficients ψ_j may be obtained by inversion of $\pi(B)$, i.e.

$$\psi_j = \pi_1 \psi_{j-1} + \ldots + \pi_{j-1} \psi_1 + \pi_j \, ; \quad j = 1,2,\ldots \qquad (5.8)$$

and the parameter values θ in (5.2) recovered from the substitution by

$$\theta_j = -(\psi_j - \phi_1 \psi_{j-1} - \ldots - \phi_{j-1} \psi_1 - \phi_j)$$

or

$$\theta = \begin{pmatrix} 1 \\ -\theta_1 \\ \vdots \\ -\theta_p \end{pmatrix} = \Psi_F \phi \, . \qquad (5.9)$$

Our main point is that the regression (5.7) depends only upon the covariance matrix G of $\tilde{x}_t \ldots \tilde{x}_{t-p}$ and therefore relates to the procedures in the foregoing sections. If p is increased so that G, with $K = p + 1$ becomes effectively singular (according to some statistical criterion) then p is the order selected. In the procedure of HR this is equivalent to a zero determinant in (5.5). In the procedure of Young the trace criterion is proportional to the trace of the inverse of the covariance matrix of the regressions, i.e.

$$\operatorname{tr} \begin{pmatrix} \Gamma_{FF} & \sigma_a^2 \Psi_F' \\ \sigma_a^2 \Psi_F & \sigma_a^2 I \end{pmatrix}^{-1}$$

The determinant of this matrix is the same as $\det G$, and the trace of the inverse will be dominated by the reciprocal of the smallest eigenvalue of G. Near singularity of G is therefore associated with large trace values.

With the matrix G figuring so prominently in the methods we have reviewed, we now consider how best to estimate it from a data sample, and what statistical analysis to apply in its investigation.

6. ESTIMATION OF THE PREDICTION COVARIANCE MATRIX G

Given a finite sample $x_1 \ldots x_N$ it is natural to use the sample autocorrelations $r_k = c_k/s_x^2$ where $s_x^2 = c_0$ and for $k > 0$

$$c_k = N^{-1} \sum_{1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \ . \tag{6.1}$$

Substituting r_k for ρ_k in Γ as defined and used in (2.4)-(2.6) then ensures that Γ is positive definite. So also then is G as defined in (2.7). A selection (considered later) of the length M of past values, and K of future values is required when (2.7) is used. If M is large the matrix inversion in (2.7) can be avoided by using the algorithm in Durbin (1960) for $k = 1 \ldots L$, where $L = K + M - 1$. This starts with $\tau_0 = 1$ and proceeds by the equations

$$\left.\begin{aligned} \pi_{k,k} &= (r_k - \pi_{k-1,1} r_{k-1} - \cdots - \pi_{k-1,k-1} r_1)/\tau_{k-1} \\ \pi_{k,j} &= \pi_{k-1,j} - \pi_{k,k}\pi_{k-1,k-j} \ , \quad j = 1 \ldots k-1 \\ \tau_k &= \tau_{k-1}(1 - \pi_{k,k}^2) \ . \end{aligned}\right\} \tag{6.2}$$

Setting

$$U = \begin{pmatrix} 1 & -\pi_{L,1} & \cdots & & -\pi_{L,K-1} \\ & 1 & & -\pi_{L-1,1} & \cdots & -\pi_{L-1,K-2} \\ & & \ddots & & \ddots & \\ & 0 & & & 1 & -\pi_{M+1,1} \\ & & & & & 1 \end{pmatrix} \tag{6.3}$$

and

$$D = \begin{pmatrix} \tau_L & & & 0 \\ & \tau_{L-1} & & \\ & & \ddots & \\ 0 & & & \tau_M \end{pmatrix} \qquad (6.4)$$

the value of G defined by (2.7) is precisely

$$G = \Gamma_{FF} - \sigma_x^2 SDS' , \qquad (6.5)$$

where $S = U^{-1}$ is economically obtained because U is triangular.

We shall call this the <u>direct estimate</u> of G. Cybenko (1980) provides assurance that the above procedure is numerically well conditioned.

An alternative is to stop the recursions in (6.2) at some value of $L = M$ say, and (dropping the first suffix L) to view the resulting values $\pi_1 \ldots \pi_M$ as estimates of the true π-weights of the ARMA model. Then using (5.8) supplies estimates of $\phi_1 \ldots \phi_K$ from which Ψ_F and G are constructed as in (2.9) and (2.12). The estimate of σ_a^2 is $\sigma_a^2 = c_0 \tau_M$ - no bias correction being made. The method of construction ensures that the operator $\pi(B)^{-1} = \phi(B) = 1 + \phi_1 B + \phi_2 B^2 + \ldots$ converges, and that both (2.10) and (2.11) hold. This alternative differs very little from (6.5), being simply a matter of replacing $\pi_{i,j}$ by $\pi_{M,j}$ and τ_i by τ_M throughout (6.3) and (6.4). Besides reducing computation, we shall discover other advantages of this estimate of G which we shall call the <u>AR(M) estimate</u>.

We have also investigated the use of spectral estimates of G. Assuming that the spectrum of the data is estimated using a classical lag window w_k applied to the sample acf, with a truncation lag M', estimates of the coefficients π_k and ϕ_k may be obtained by spectral factorization, and their properties have been studied by Bhansali (1974, 1977). It is possible to implement this by the Cramér-Wold factorization. Given the windowed acf $w_k r_k$, $k = 1 \ldots M'$, this supplies the coefficients $\theta_1 \ldots \theta_{M'}$ in a

high order MA(M') model. Then $\psi_k = -\theta_k$. The factorization can be done rapidly and accurately using algorithms such as that given by Laurie (1982). The matrix G is again constructed using (2.12), the elements of Γ_{FF} being the windowed autocovariances, and σ_a^2 the innovation variance from the MA(M') model. We shall call this the <u>MA(M') estimate</u> of G.

The foregoing estimates of G all suffer from the limitations on the accuracy of r_k as an estimate of ρ_k when the series borders on being non-stationary in the sense of ARIMA models. Tsay and Tiao have paid particular attention to this situation and recommend that the construction of G be based on the true least squares regression equations of the future upon the past. Thus in our notation their estimate of Γ would be given by

$$\Gamma_{ij} = (N - L)^{-1} \sum_{t=1+L}^{1+N} x_{t-i} x_{t-j} \; ; \quad i,j = 1 \ldots L$$

where here $L = K + M$. We call this the <u>regression estimate</u> of G.

In the following proposals for how G should be analysed, we shall assume that the AR(M) or MA(M') estimate is used.

7. ANALYSIS OF THE PREDICTION COVARIANCE MATRIX

It is convenient now to reverse the order of the rows and columns of G to correspond with the sequence of variables $\hat{x}_n(k)$, $k = 1,2 \ldots K$, whose covariance matrix it estimates. This also allows a kind of open-ended approach, so that G can be analysed for increasing order K. The rearrangement corresponds to using, in place of (2.12),

$$G = \Gamma_{FF} - \Psi_F \Psi_F' \sigma_a^2 .$$

We shall write G_p for the upper left $(p + 1) \times (p + 1)$ submatrix of G. The procedure we suggest is to factorize

$$G = LDL' \qquad (7.1)$$

where L is lower triangular with unit diagonal, and D is diagonal. In practice we have used a Choleski factorization from which L may be obtained by simple scaling of the columns, and D by squaring the diagonal. It is valuable to use a method which stops

(without catastrophic failure) when a diagonal element of D is found to be effectively zero (using numerical rather than statistical criteria). For recent use of Choleski factorization in similar contexts see Newton and Pagano (1983) and Hawkins and Eplett (1982).

The motivation is that the diagonal elements $d_0, d_1 \ldots$ of D constitute the sequence of regression error variances $\sigma_f^2 = \sigma_e^2 - \sigma_\alpha^2$, corresponding to the regression (5.2) for increasing $p = q = 0,1,2 \ldots$. Furthermore $L^{-1} = T$ say, is again lower triangular and its successive rows contain the estimates of the AR parameters in (5.2), i.e.

$$t_p = (-\phi_p, -\phi_{p-1} \ldots -\phi_1, 1) \qquad (7.2)$$

for $p = 0,1,2 \ldots$. The row t_p minimises the quadratic form $t_p G_p t_p'$ w.r.t. variations in $\phi_1 \ldots \phi_p$ and yields d_p as the minimum. That d_p is a decreasing sequence is seen from the fact that the special structure of G implies

$$\tilde{t}_{p+1} G_{p+1} \tilde{t}_{p+1}' \leq t_p G_p t_p' = d_p$$

where $\tilde{t}_{p+1} = (0, t_p)$. The minimising vector t_{p+1} must then obviously yield $d_{p+1} \leq d_p$.

Also note that the matrix $H = T\Psi_p$ is lower triangular with rows consisting of the estimates of the MA parameters in (5.2),

$$h_p = (-\theta_p, -\theta_{p-1} \ldots -\theta_1, 1) . \qquad (7.3)$$

Another interpretation of the procedure may be deduced from (2.11). This is that $\phi_1 \ldots \phi_p$ result from the regression of the first column of Ψ_p as defined in (2.9) upon the remaining columns, taking $K = p + 1$. If we were to calculate

$$H(B) = \phi(b)\psi(B) = \sum_{k=0}^{\infty} h_k B^k$$

this would correspond to choosing $\phi(B)$ so that $\sum_{p+1}^{\infty} h_k^2$ was minimised (giving d_p), with $\theta(B)$ given by truncation of $H(B)$ at B^p. It is clear therefore that a value of $d_p = 0$ indicates not only a singularity in the estimated matrix G_p, but also an exact representation of $\psi(B) = 1/\pi(B) = \theta(B)/\phi(B)$.

The calculations needed to construct G are relatively modest, and for moderate values of p there seems little need to exploit the recursion advocated by HR. However, it is well worth examining what they propose. Our calculations are exactly equivalent to specifying the joint covariance structure of the bivariate process (x_t, a_t) as follows

$$\left.\begin{aligned} c_{xx}(k) &= c_k \\ c_{aa}(k) &= \begin{cases} \sigma_a^2, & k = 0 \\ 0, & k \neq 0 \end{cases} \\ c_{ax}(k) &= \begin{cases} \psi_k \sigma_a^2, & k > 0 \\ 0, & k < 0 \end{cases} \end{aligned}\right\} \qquad (7.4)$$

Using these, we see that the bivariate AR(p) fit of the regression (5.2) in the case $p = q$, may be written

$$\begin{pmatrix} (1 - \phi_1 B - \ldots \phi_p B^p) & (\theta_1 B + \ldots \theta_p B^p) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_t \\ a_t \end{pmatrix} = \begin{pmatrix} e_t \\ a_t \end{pmatrix} \qquad (7.5)$$

The insertion of exact zeros in parts of (7.4) leads to the exact 0 and 1 in (7.5), which may help to reduce computation if the recursion of Whittle (1963) is used.

A useful result derives from the fact that the matrix operator in (7.5) will satisfy a bivariate stationarity condition. Because of its simple structure, this implies that $\phi(B)$, as specified by any of the rows (7.2), satisfies the univariate stationarity condition. There is no similar invertibility constraint upon $\theta(B)$ except if $d_p = 0$, whence it follows since $\theta(B) = \phi(b)\psi(B)$ exactly. If a value of p is chosen with d_p small then $\theta(B)$ is close to $\phi(B)\psi(B)$, so that $\theta(B)$ is likely to be invertible. Otherwise, applying one of the many devices for mapping $\theta(B)$ into an invertible form can be expected to result in only minor modification of its parameters. In practice we have always found the selected values of the ARMA parameters to satisfy stationarity and invertibility constraints. This is a major advantage when they are used as initial values in maximum likelihood estimation.

8. STATISTICAL PROCEDURES

We recommend the following procedures. The justification of the statistical properties which we claim is not included here, because we only have proofs in outline form. They are however in accord with the usual ideas of regression.

The selection of the order M used in the AR(M) estimate of G must be considered, but it does not appear to be very critical. It is important that it be sufficiently large that the autoregressive estimates of π_k have relatively small bias, but not so large that the sequence $\hat{\pi}_k$ is unnecessarily extended beyond the point where π_k is effectively zero. If the AIC or similar criterion is used we suggest that the value so chosen for M be doubled. The choice of K should be at most $M/2$. If the MA(M') estimate of G is used, the equivalent choice of M' is such that

$$\sum_{1}^{M'} w_k^2 = M$$

e.g. for the Parzen lag window, $M' = 3.71M$.

We suggest that various statistics be plotted. Because their range of variation is large in some cases, we have had to truncate some of these on the vertical scale in the following figures, in order to reveal the detail of the smaller values.

We first plot the diagonal elements of G because these estimate the prediction variances $v_k = \mathrm{Var}(\hat{x}_n(k))$ for increasing lead time $k = 1,2,\ldots$. If these die rapidly to zero there may be rather little information upon which model selection can be based, unless the series length n is large.

Next, plot the sequence $d_0, d_1 \ldots$ which estimate the residual variance $\sigma_f^2 = \sigma_e^2 - \sigma_a^2$ in the regression. These will usually become effectively zero at index $k = M/2$ provided M is moderately large, because then we should expect $\phi(B)/\theta(B)$ with $2p = M$ parameters to provide a very close match with $\pi(B)$ which has order M. On occasions, d_k has become zero at a lower index, and through bad numerical conditioning this may lead to increasing values. The sequence must be truncated at this point. This is no great problem since it merely means that an ARMA (p,p) model has been found which is virtually coincident with the AR(M) model. In practice we scale the sequence d_k by

taking in its place, Nd_k/σ_a^2, and we shall henceforth mean this scaled value when we refer to d_k.

We also plot the sequence of reductions in the residual variance, $\delta_k = d_{k-1} - d_k$, and a sequence of (bivariate) partial correlations R_k defined from the regression by

$$R_k^2 = \delta_k/(d_{k-1} + N) .$$

Both δ_k and R_k measure the improvement in the regression (5.2) on introducing the pair of variables x_{t-k}, a_{t-k}.

Assuming a true ARMA (p_0,p_0) model for x_t, the approximate properties for large N upon which we base order selection, are that for $k > p_0$, the δ_k form a sequence of IID random variables having chi-squared distributions with 2 d.f., i.e. exponential with expectation 2. For $k \leq p_0$, the expectation of δ_k will be inflated.

We take NR_k^2 to have approximately the same distribution as δ_k for $k > p_0$. The plot of R_k may therefore be scanned and treated in a very similar manner to that recommended by BJ for examining partial autocorrelation functions. It is always positive however, with median of $1.18/\sqrt{N}$ and upper 90% point $2.15/\sqrt{N}$. The 'cut-off point' indicates the true order p_0. Alternatively, the upper 90% point of 4.6 may be used for δ_k.

The appearance of the residual variance plot d_k may be deduced to be a random walk for $i > p_0$, with downward steps δ_k, reaching 0 at $k = M/2$. It may be possible to base tests for the model order on detecting the point at which this plot falls within the expected range of such a random walk.

9. APPLICATIONS

We did not carry out an extensive simulation study. We did however simulate several of the models used by HR and use one of them for illustration. This was the model

$$(1 + .64B + .7B^2)x_t = (1 + .8B)a_t$$

with $\sigma_a^2 = 1$. We took $N = 50$ and set $M = 10$. Figure 1 shows the series and Figure 2 the four plots derived in the analysis. The order $p = 2$ is clearly indicated, with some suggestion that $p = 3$ might be considered. For $p = 2$ the preliminary estimates gave a

model in good agreement with that simulated:

$$(1 + .62B + .80B^2)x_t = (1 + .61B - .11B^2)a_t .$$

We next considered a series which we had previously modelled, and were interested to discover to what extent the method yielded agreement with our previous conclusions. The series was the monthly wholesale price index for Italy from 1970 to 1976. The difference of the logarithms were used, and are shown in Figure 3. The previous analysis appeared in Piccolo and Tunnicliffe Wilson (1981). There are 83 values and we chose $M = 20$. Figure 4 shows the resulting plots. The order $p = 2$ is indicated, with $\delta_2 = 6.754$, exceeding the upper 95% point of χ^2_2. The value of $\delta_3 = 3.97$ is not very large, but stands clear of the remaining part of this plot. The series shows some evidence of outlier behaviour which tends to deflate sample autocorrelations and similar statistics, so we might be persuaded to consider $p = 3$ as a serious possibility. The preliminary estimates for this choice gave the model

$$(1 - 1.63B + 1.24B^2 - .45B^3)x_t = (1 - .59B + .43B^2 - .04B^3)a_t .$$

This agreed very well with our previous modelling experience which after some trial and error, had lead us to an ARMA (3,2) model with very similar parameters. The spectrum of this model had a moderate peak corresponding to a period of 6.6 months.

Finally we considered the sunspot series,and for model selection took the first 230 of the set of 256 values from 1700 to 1955 due to Waldmeier (1961). We applied a square-root transformation giving the series shown in Figure 5. We tried analyses with both $M = 30$ and $M = 40$. Figure 6 shows the resulting plots. In this example G_p was singular for $p = 7$ which is strongly indicated as the order to be chosen, since $\delta_7 = 30.9$ far exceeds the 99% significance level of χ^2_2. Preliminary estimates gave the model

$$(1 - 1.73B + 1.05B^2 - .02B^3 - .02B^4 - .48B^5 + .72B^6 - .41B^7)x_t$$
$$= (1 - .59B + .02B^2 + .04B^3 + .33B^4 - .50B^5 + .25B^6 + .01B^7)a_t$$

This is a stationary and invertible model which was used as the starting point in exact likelihood estimation using the GENSTAT package. Convergence effectively occurred

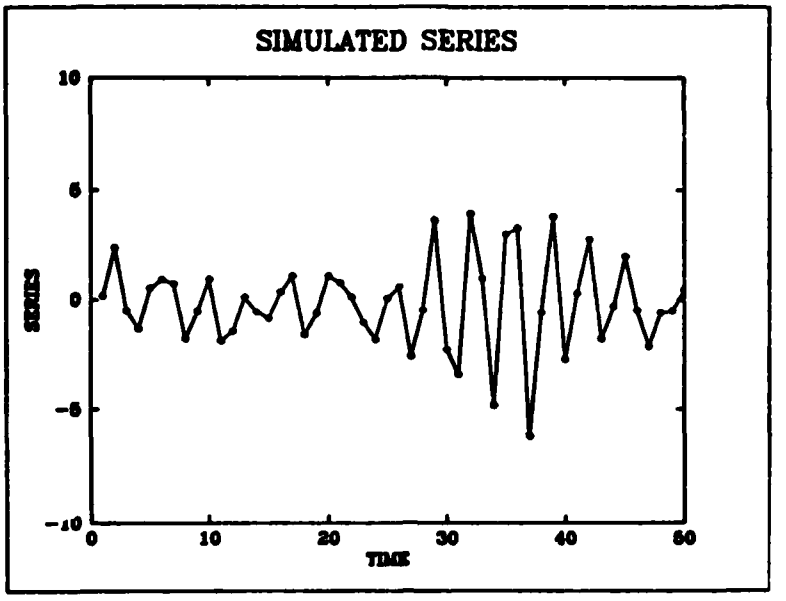


Figure 1.

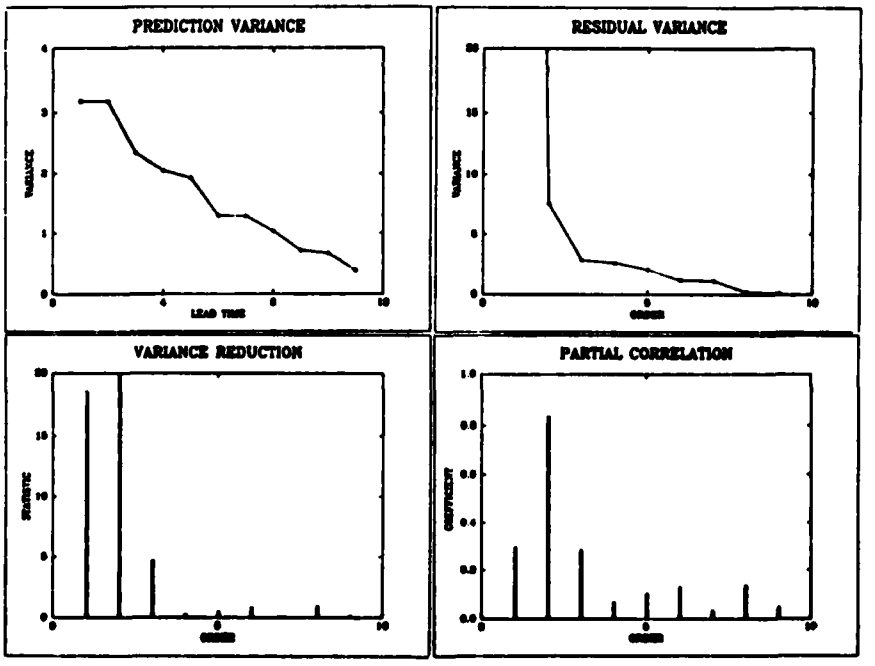


Figure 2. Plots for the simulated series.

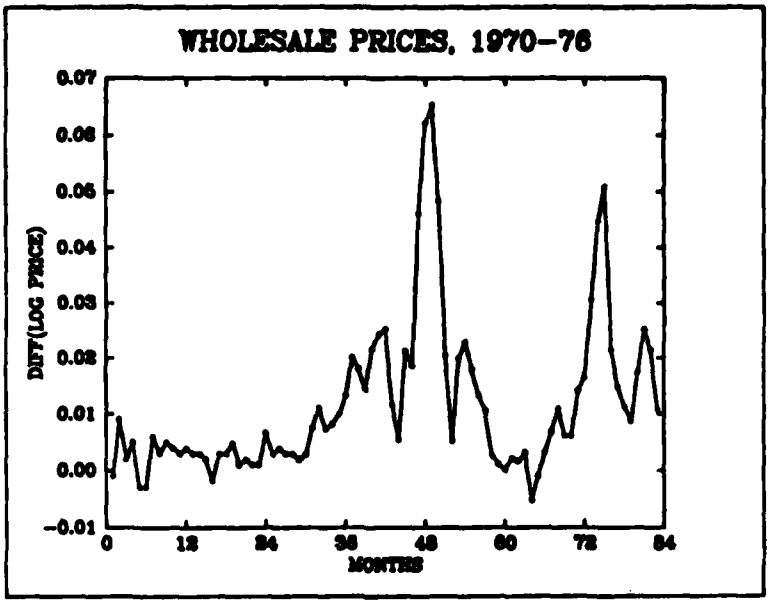


Figure 3.

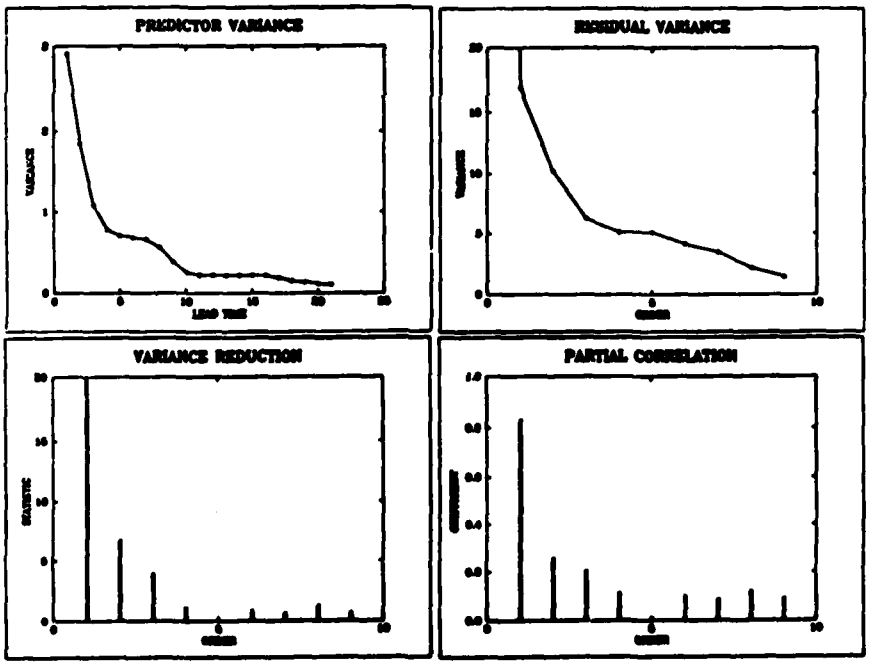


Figure 4. Plots for the wholesale prices series.

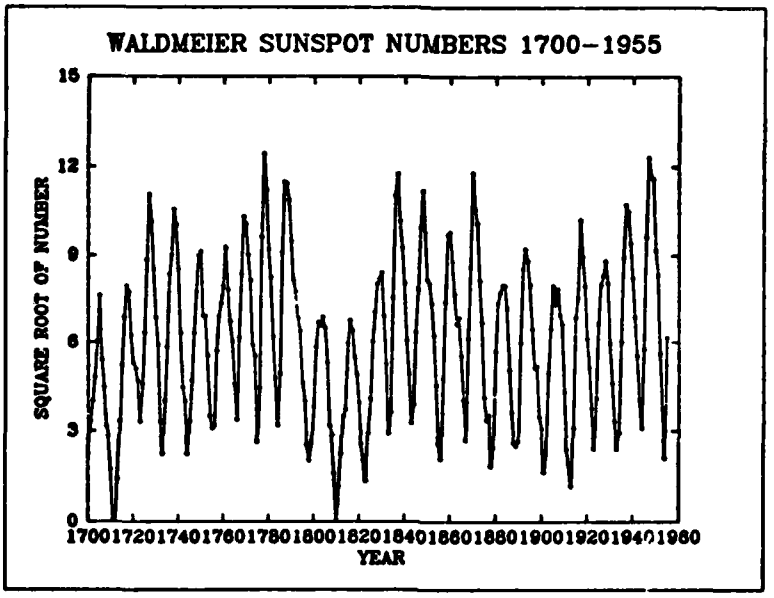


Figure 5.

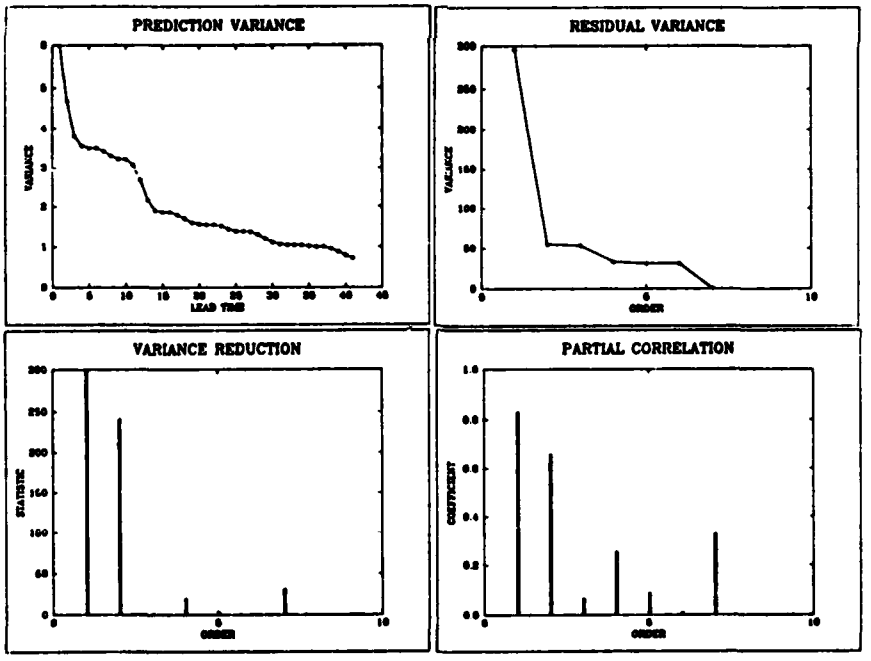


Figure 6. Plots for the sunspot numbers series.

after 15 iterations to the model:

$$(1 - 2.00B - 1.39B^2 + .12B^3 - .25B^4 - .92B^5 + 1.41B^6 - .66B^7)x_t$$
$$= (1 - .80B + .00B^2 + .35B^3 + .52B^4 - .99B^5 + .28B^6 + .25B^7)a_t$$

Examination of the parameter standard errors indicated possible truncation to ARMA (7,5). The estimated residual variance was 0.999, and the residual autocorrelation pattern was extremely white in appearance. The spectrum of the fitted model is shown in Figure 7. The square root of the spectrum is plotted, so as to reveal the smaller peaks, one of which appears to be a harmonic of the very sharp main peak.

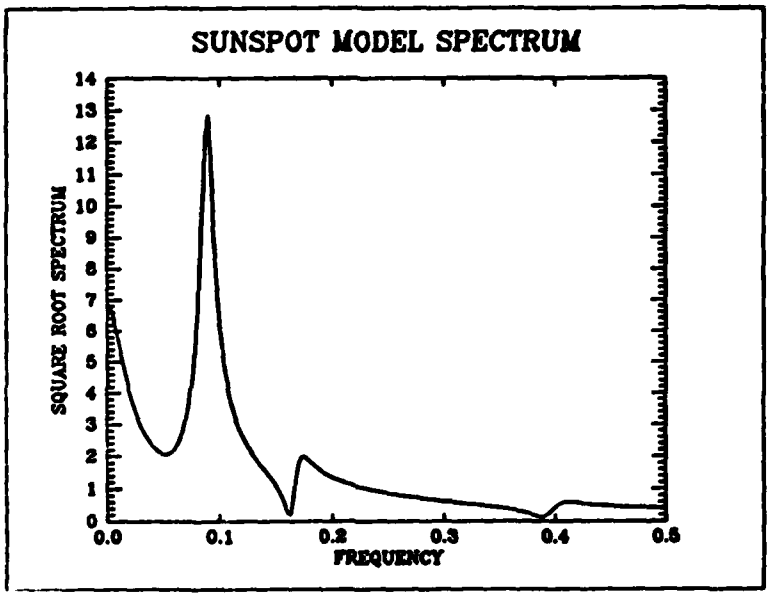


Figure 7.

There was some concern that the use of sample autocorrelations might be inadequate in the analysis of such a highly autocorrelated series. The investigation was therefore repeated, with the AR(40) model for the π coefficients being fitted by exact likelihood, and corresponding theoretical acf values being used in the construction of G.

The resulting plots were however very similar.

Furthermore, this data set was used to investigate the spectral factorization or MA(M') approach to estimating G, as described in section 6. The sample acf was weighted using a Parzen window with a cut-off at lag M' = 110, corresponding to a value of M = 30. Again the results were extremely close to those previously obtained. The rather unexpected result of the singularity of G_7 when very large values of M were used, suggests that perhaps the series is somewhat more deterministic than we had supposed.

Based on these and other examples our conclusion is that the method succeeds in leading us to ARMA models which well approximate the structure of series.

10. AN INVERSE PROCEDURE

In section 7 we showed that the method could be viewed as estimating the AR parameters in the regression

$$\psi_j = \phi_1\psi_{j-1} + \ldots + \phi_p\psi_{j-p} \quad \text{for} \quad j > p$$

using the estimates of ψ_j. This prompts us to consider the inverse problem of estimating the MA parameters in the regression

$$\pi_j = \theta_1\pi_{j-1} + \ldots + \theta_p\pi_{j-p} \quad \text{for} \quad j > p \ ,$$

using the estimates of π_j. This of course ties in very closely with the proposal of Durbin (1959) for estimating parameters in the MA(q) model, though his method starts the regression from j = 1. The possibility of starting the regression from j = p is, we believe, suggested by Fuller (1976) p. 359 in the case of an ARMA(p,q) model.

To implement this idea we merely treat the π weights from the AR(M) model as the ψ weights of the <u>inverse</u> model

$$xi_t = \frac{\phi(B)}{\theta(B)}\, a_t \ .$$

The corresponding inverse autocovariances are simple estimated as

$$ci_k = \sigma_a^2 \sum_{i=0}^{M-k} \pi_i\pi_{i+k}$$

and the G matrix of the inverse model derived. It may however be directly calculated (with rows and columns in the order used for Choleski factorization) as

$$Gi_{jk} = \sigma_a^2 \sum_{m=0}^{M-j} \pi_{m+j}\pi_{m+k} \quad \text{for} \quad 1 \leq k \leq j \ .$$

We applied this idea to the series of annual variations in day length from 1820 to 1970. We used the first differences of the data which are shown in Figure 8, and chose $M = 40$. The plots shown in Figure 9 clearly indicate $p = 2$, and provide the preliminary model for the original series (after interchanging the estimates of θ and ϕ) of

$$(1 - .31B - .05B^2)\nabla x_t = (1 + .55B + .55B^2)a_t \ .$$

Note now that it is the MA operator $\theta(B)$ which has its invertibility ensured. This model is very close to that previously selected by conventional ARMA modelling:

$$(1 - .39B)\nabla x_t = (1 + .55B + .62B^2)a_t \ .$$

We should mention that the method when applied directly to this series, as for the examples in section 9, had indicated a different result, that $p = 3$, with preliminary model:

$$(1 - .95B - .06B^2 + .2B^3)\nabla x_t = (1 - .10B - .00B^2 - .40B^3)a_t \ .$$

After estimation of this ARMA (3,3) model, it was found to be similar to the ARMA (2,2) model, but with a nearly-cancelling extra linear factor on each side.

In this case, we know that ∇x_t has a spectrum which has a near zero around frequency $2\pi/3$, and very little power above that, due to prior data smoothing. This leads to a strong MA operator in the model, which the inverse method seems better at detecting.

11. THE CASE OF UNEQUAL ORDERS.

The foregoing section prompts consideration of this case. We propose that our method may be adapted for $p \neq q$ as follows. In the case that $p - q = r > 0$, the equation (2.2) for the forecast function involves r values $x_n, x_{n-1} \cdots x_{n+1-r}$ in the set P of past values. The whole analysis can then be carried out by allowing the future F and past P to be no longer a simple partition, but to include the above variables as an overlap. This is handled with great simplicity by prefixing rows and columns of Γ_{FF} to G, to give G_{+r} say. For example if $r = 1$,

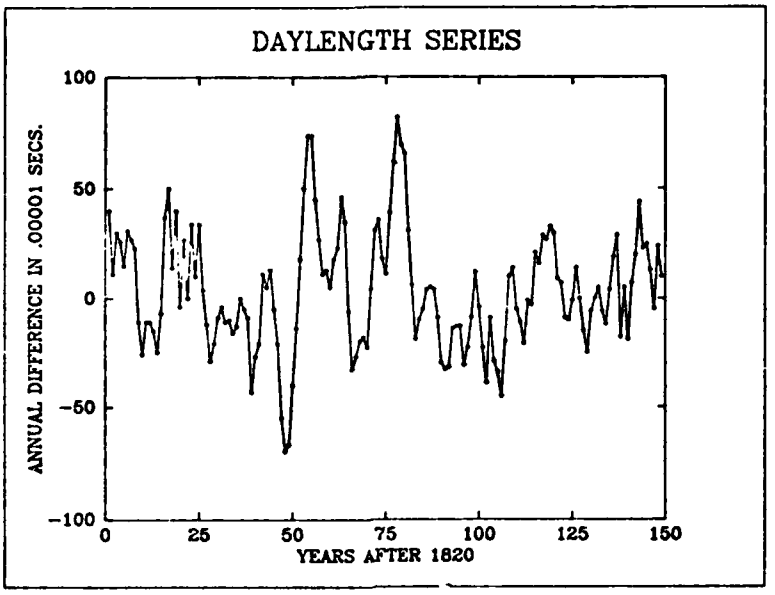


Figure 8.

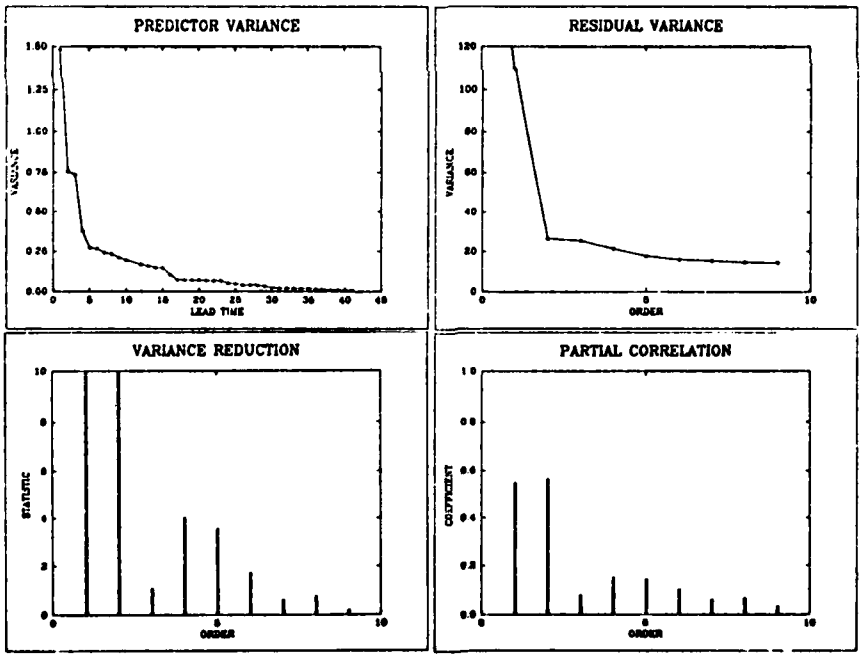


Figure 9. Plots for the daylength series.

$$G_{+1} = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 \cdots \\ \gamma_1 & & \\ \gamma_2 & & G \\ \vdots & & \end{pmatrix}$$

Further identical rows and columns are added for $r = 2,3 \ldots$. Akaike (1974) recommends that F and P overlap by x_n, so that his canonical correlation approach is oriented towards ARMA $(p,p-1)$ models.

Proceeding with the Choleski factorization of G_{+r} now yields $d_0, d_1 \ldots d_{r-1}$ as estimates of the prediction error in AR(k) models for $k = 0 \ldots r-1$. We suggest that the plots previously recommended be now started from $p = r$, i.e. for $q = 0,1,2 \ldots$.

In the case $q - p = r > 0$, we take a gap of length r between the future and the past, and this is implemented merely by removing the first r rows and columns of G, to give G_{-r} say.

By this means the diagonals of an array for all $p > 0$ and $q > 0$ may be built up out of the statistics $d(p,q)$, and used in various ways to select p and q. We are rather reluctant to recommend the use of such a proliferation of statistics, except possibly in the application of an automatic method such as that in HR. It is also of interest to note that for any given p,q the estimates $\hat{\phi}$ may be viewed as being obtained by regressing the first column of Ψ_p in (2.9) upon the remaining columns, where the number of columns $K = p + 1$, but now the partition is obtained by removing the first $q + 1$ rows of Ψ. For $q = 0$ the estimates of the AR(p) model are simply the Yule-Walker estimates which are known to be asymptotically equivalent to the MLEs.

It may also be shown that for any p,q the parameters can be calculated using the bivariate autoregression of HR with the entries in (7.4) slightly modified. For $p - q = r > 0$, the entry ψ_k is replaced by ψ_{k-r} which is taken as 0 for $k < r$. For $q - p = r > 0$, ψ_k is replaced by ψ_{k+r} and c_k replaced by:

$$c_k - \Big(\sum_{i=0}^{r-1} \psi_i \psi_{i+k}\Big)\sigma_a^2 .$$

This reassures us that for all p,q the estimate of $\phi(B)$ is stationary.

There is of course a corresponding array which may be calculated using the inverse autocovariances. The estimates of $\theta(B)$ associated with this array will similarly always be invertible for any p,q. For $p = 0$ they are the Durbin estimates which are proved by Bhansali (1980) to be asymptotically efficient provided M is allowed to grow as a specified function of N.

12. EFFICIENCY FOR MIXED MODELS.

The calculation of fully efficient estimates of ϕ,θ in the case $p > 0$ and $q > 0$ is considered by HR. They use the preliminary estimates in a manner described also in Fuller (1976), to set up a regression for parameter corrections. After applying these corrections to the preliminary estimates, corrected estimates are obtained which have the same asymptotic distribution as the MLEs.

It would be attractive if a direct regression were available which supplied good approximations to the MLEs. This seems unattainable, but we describe a procedure which gives some encouragement. We claim that the regression (5.4) provides such estimates, which agree with the MLE to within $o_p(n^{1/2})$, provided the filter $L(B)$ is appropriately chosen. One such choice is to use

$$x_t^* = 1/\theta(B)x_t, \quad a_t^* = 1/\theta(B)a_t \tag{12.1}$$

and another is

$$x_t^* = 1/\phi(B)a_t, \quad a_t^* = 1/\phi(B)xi_t \tag{12.2}$$

where $xi_t = \pi(B)a_t$ is the estimated inverse series, and $\pi(B)$ the operator estimated in (5.1). Thus (12.1) corresponds to $L(B) = 1/\theta(B)$ and (12.2) to $L(B) = \pi(B)/\phi(B)$. The first choice, in the case of $q = 0$, i.e. $\theta(B) = 1$ gives the Yule-Walker estimates of ϕ, the second choice in the case $p = 0$, $\phi(B) = 1$ gives the Durbin estimates of θ. In any other case some approximation to $\theta(B)$ or $\phi(B)$ must be used in order to get reasonable efficiency. One cannot always rely on getting good preliminary estimates of these by using the main method of this paper. We have for example simulated a sample of size 100 from the signal-plus-noise model

$$(1 - .98B)x_t = (1 - .83B)a_t .$$

The method completely failed to identify the model, although the sample acf hints at the model by showing a tendency to positive values at low lags. There is just sufficient information in the data to obtain well defined MLEs $\hat{\phi} = .84(SE\ .26)$ and $\hat{\theta} = .76(SE\ .31)$, though reasonably good preliminary estimates are necessary to find these.

It would appear to us that part of the skill in identifying ARMA models from the sample acf, lies in visual filtering of the acf in order to pick out the patterns. This seems to be equivalent to selecting the filters in (12.1) or (12.2), to amplify the signal component of the data. Often the signal is at a low frequency and then one may be able to obtain much improved results from the regression (5.4) by a very rough guess at a low frequency selection filter to use in place of $1/\phi(B)$ in (12.2).

We hope to investigate automatic heuristic methods of selecting such filters, e.g. by scanning the series spectrum to locate the frequency bands of signal components. The orders in the regression (5.4) may then be determined by criteria such as HR use, or by other methods of selecting variables in regression. Unfortunately the attractive stationarity property of $\phi(B)$ is no longer certain, but in general, improved estimates will result. Refinement of these by a few iterations of ML estimation will still be recommended.

On a final point, we mention that order identification and preliminary estimation methods have been successfully applied in the multivariate time series context. Cooper and Wood (1982) use canonical cross-correlation analysis, and Tiao and Tsay (1983) have extended their regression methods. We expect that the method proposed in this paper will readily generalise, and it may be worthwhile exploiting the multivariate recursion of Whittle not only to obtain multivariate innovations, but also in an extension of HR's use of this algorithm.

REFERENCES

Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis, New York, Wiley.

Akaike, H. (1974). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. Ann. Inst. Statist. Math. 26, 363-387.

Akaike, H. (1976). Canonical correlation analysis of time series and the use of an information criterion, in: Systems Identification: Advances and Case Studies, eds. R. K. Mehra and D. G. Lainiotis, 27-96, New York, Academic Press.

Bêguin, J. M., Gourieroux, C. S. and Monfort, A. (1980). Identification of a mixed autoregressive-moving average process: the corner method, in: Time Series, ed. O.A. Anderson, 423-436, Amsterdam, North-Holland.

Bhansali, R. J. (1974). Asymptotic properties of the Wiener-Kolmogorov predictor. I. J. Roy. Stat. Soc. B 36, 61-73.

Bhansali, R. J. (1977). Asymptotic properties of the Wiener-Kolmogorov predictor. I. J. Roy. Stat. Soc. B 39, 66-72.

Bhansali, R. J. (1980). Autoregressive and window estimates of the inverse correlation function. Biometrika 67, 3, 551-566.

Box, G.E.P. and Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control, Revised edition 1976, S. Francisco, Holden-Day.

Cooper, D. M. and Wood, E. F. (1982). Identifying Multivariate Time Series Models. Journ. Time Series Analys. 3, 3, 153-164.

Cybenko, G. (1980). The Numerical Stability of the Levinson-Durbin Algorithm for Toeplitz Systems of Equations. SIAM J. Scient. Statist. Comput. 1, 3, 303-319.

Durbin, J. (1959). Efficient estimation of parameters in moving average models. Biometrika 28, 233-244.

Durbin, J. (1960). The fitting of time series models. Rev. Int. Inst. Stat. 28, 233-244.

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

Glasbey, C. A. (1982). A generalization of partial autocorrelations useful in identifying ARMA models. Technometrics 24, 223-228.

Gray, H. L., Kelley, G. D. and McIntire, D. D. (1978). A new approach to ARMA modeling. Communications in Statistics B, 7, 1-77.

Hannan, E. J. and Rissanen, J. (1982). Recursive estimation of mixed autoregressive moving average order. Biometrika 69, 81-94.

Hawkins, D. M. and Eplett, W. J. R. (1982). The Cholesky Factorization of the Inverse Correlation or Covariance Matrix in Multiple Regression. Technometrics 24, 3, 191-198.

Laurie, D. P. (1982). Cramér-Wold Factorization. Algorithm AS175. Applied Statistics 31, 1, 86-93.

Mazzali, A. (1982). Gli spazi di previsione dei modelli ARMA. Rivista Statistica Applicata 15, 4, 295-308.

Newton, J. H. and Pagano, M. (1983). The finite memory prediction of covariance stationary time series. SIAM J. Scient. Statist. Comput. 4, 2, 330-339.

Piccolo, D. and Tunnicliffe-Wilson, G. (1981). Time series models of Wholesale Retail Price relationships in Italy. Economic Notes 10, 2, 35-36.

Tiao, G. C. and Tsay, R. S. (1981). Identification of Nonstationary and Stationary ARMA Models. Techn. Rep. 647, Department of Statistics, Univ. of Wisconsin-Madison.

Tiao, G. C. and Tsay, R. S. (1983). Multiple Time Series Modeling and Extended Sample Cross-Correlations. J. Business and Economic Statistics 1, 1, 43-56.

Tsay, R. S. and Tiao, G. C. (1983a). Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA models. J. Amer. Statist. Assoc. (to appear)

Tsay, R. S. and Tiao, G. C. (1983b). Use of Canonical Analysis in Time Series Model Identification. Techn. Rep. 283, Department of Statistics, Carnegie-Mellon Univ., Pittsburgh, PA.

Young, P. C. (1974). Recursive approaches to time series analysis. Bull. Inst. Math. Appl. 10, 209-224.

Young, P. C., Jakeman, A. and McMurtrie, R. (1980). An Instrumental Variable Method for Model Order Identification. Automatica 16, 281-294.

Waldmeier, M. (1961). The sunspot activity in the years 1610-1960, Zurich, Schultess.

Whittle, P. (1963). Prediction and Regulation by Linear Least Squares Methods, London, Engl. Univ. Press.

Woodward, W. A. and Gray, H. L. (1981). On the relationship between the S array and the Box-Jenkins method of ARMA model identification. Journ. Amer. Statist. Assoc. 76, 579-587.

GTW/DP/ed

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ DISTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 2592 — 2. GOVT ACCESSION NO.: AD-A136426	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): A UNIFIED APPROACH TO ARMA MODEL IDENTIFICATION AND PRELIMINARY ESTIMATION	5. TYPE OF REPORT & PERIOD COVERED: Summary Report - no specific reporting period
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): G. Tunnicliffe Wilson and D. Piccolo	8. CONTRACT OR GRANT NUMBER(s): DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709	12. REPORT DATE: November 1983
	13. NUMBER OF PAGES: 30
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

ARMA model identification; preliminary estimation; Yule-Walker equations; Durbin-Levinson algorithm; prediction spaces; Choleski factorization

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper reviews several different methods for identifying the orders of autoregressive-moving average models for time series data. The case is made that these have a common basis, and that a unified approach may be found in the analysis of a matrix G, defined to be the covariance matrix of forecast values.

(cont.)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT (cont.)

The estimation of this matrix is considered, emphasis being placed on the use of high order autoregression to approximate the predictor coefficients. Statistical procedures are proposed for analysing G, and identifying the model orders.

A simulation example and three sets of real data are used to illustrate the procedure, which appears to be very useful as a tool for order identification and preliminary model estimation.


FILME