doi:10.1111/j.1467-9892.2006.00481.x

PARTIAL AUTOCORRELATION PARAMETERIZATION FOR SUBSET AUTOREGRESSION

By A. I. McLeod and Y. Zhang

The University of Western Ontario and Acadia University

First version received November 2005

Abstract. A new version of the partial autocorrelation plot and a new family of subset autoregressive models are introduced. A comprehensive approach to model identification, estimation and diagnostic checking is developed for these models. These models are better suited to efficient model building of high-order autoregressions with long time series. Several illustrative examples are given.

Keywords. AR model identification and diagnostic checks; forecasting; long time series; monthly sunspot series; partial autocorrelation plot; seasonal or periodic time series.

1. INTRODUCTION

The AR(p) model with mean μ may be written, $\phi(B)(z_t - \mu) = a_t$, where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$. B is the backshift operator on t and a_t, $t = 1, \ldots, n$ are normal and independently distributed with mean zero and variance σ_a^2. The admissible region for stationary-causal autoregressive processes is defined by the region for which all roots of $\phi(B) = 0$ lie outside the unit circle (Brockwell and Davis, 1991). The usual subset autoregressive model is obtained by constraining some of the ϕ-parameters to zero. In this case, we may write, $\phi(B) = 1 - \phi_{i_1} B^{i_1} - \cdots - \phi_{i_m} B^{i_m}$, where $i_1 < \cdots < i_m$. This model will be denoted by $\mathrm{AR}_\phi(i_1, \ldots, i_m)$. Such subset autoregressive models are often used for modelling seasonal or periodic time series as well as for obtaining a more parsimonious representation of autoregressive models. McClave (1975) presented an algorithm using the Yule–Walker estimators which may be used to find the best model according to some criteria such as the Akaike information criterion (AIC) or Bayes information criterion (BIC). However, as pointed out by Haggan and Oyetunji (1984), the algorithm given by McClave (1975) only finds the single best solution although in practice it is often desirable to examine a range of plausible models. The algorithm developed by Haggan and Oyetunji (1984), utilizing least squares, is not as computationally efficient as that of McClave (1975) but it is easy to find the best set of models. One drawback of this approach is that it is based on least squares. Least squares estimates may be preferable to Yule–Walker estimates because of their lower bias but least squares estimates may be inadmissible.

0143-9782/06/04 599–612 JOURNAL OF TIME SERIES ANALYSIS Vol. 27, No. 4

Although an admissible model may not be needed for short-term forecasting, it is required for spectral estimation or data simulation in engineering design (Hipel and McLeod, 1994, Sect. 9.7.3). Zhang and Terrell (1997) have suggested a new criterion, the projection modulus, which is computationally more efficient but their method is based on Yule–Walker estimates that are known to be less accurate than some alternatives (Tjøstheim and Paulsen, 1983; Percival and Walden, 1993, p. 414 and p. 453; Zhang and McLeod, 2006). Bayesian methods of variable selection in regression were introduced by George and McCulloch (1993) and Bayesian methods for subset autoregression have been developed by Chen (1999) and Unnikrishnan (2004). The approach developed in this article is computationally more efficient than previous techniques, based on maximum likelihood and well suited to fitting long time series and high dimensional subset autoregressive models as is illustrated in Section 3.3.

We now introduce the new subset autoregression models. Consider the Durbin–Levinson recursion for the AR(p)

$$\phi_{j,k+1} = \phi_{j,k} - \phi_{k+1,k+1}\phi_{k+1-j,k}, \quad j = 1, \ldots, k, \tag{1}$$

where $k = 1, \ldots, p$ and $\zeta_i = \phi_{i,i}$. This recursion can be used to define a transformation,

$$\mathcal{B} : (\zeta_1, \ldots, \zeta_p) \to (\phi_1, \ldots, \phi_p), \tag{2}$$

that is one-to-one, continuous and differentiable inside the admissible region (Barndorff-Nielsen and Schou, 1973). Both $\mathcal{B}$ and its inverse $\mathcal{B}^{-1}$ are easily computed (Monahan, 1984). To extend this transformation to the subset autoregressive case, we simply constrain some of the ζ-parameters to zero. In general, this subset AR model may be denoted by $\mathrm{AR}_\zeta(i_1, \ldots, i_m)$ where the underlying parameters are $\zeta_{i_1}, \ldots, \zeta_{i_m}$. The $\mathrm{AR}_\phi(i_1, \ldots, i_m)$ and $\mathrm{AR}_\zeta(i_1, \ldots, i_m)$ are similar but distinct models. For example, in the $\mathrm{AR}_\phi(1, 3)$, $\zeta_1 = \phi_1/(1 - \phi_1\phi_3 - \phi_3^2), \zeta_2 = \phi_1\phi_3/(1 - \phi_3^2)$ and $\zeta_3 = \phi_3$, whereas in the $\mathrm{AR}_\zeta(1, 3)$, $\phi_1 = \zeta_1$, $\phi_2 = -\zeta_1\zeta_3$ and $\phi_3 = \zeta_3$.

In general, the $\mathrm{AR}_\zeta(i_1, \ldots, i_m)$ model has only m parameters, not including μ and σ_a^2, but it specifies $p = i_m$ values for the parameters in ϕ-space, $\phi_1(\zeta_1, \ldots, \zeta_{i_m}), \ldots, \phi_p(\zeta_1, \ldots, \zeta_{i_m})$. In ϕ-space, the admissible region is a complex m-dimensional subspace of the original p-dimensional space of $(\phi_1, \ldots, \phi_p)$. In contrast, the admissible region in the ζ-space, $\mathcal{D}_\zeta$, for the $\mathrm{AR}_\zeta(i_1, \ldots, i_m)$ model is simply the m-dimensional cube with boundary surfaces corresponding to ± 1. The transformation $\mathcal{B}$ induces the transformation $\mathcal{B}_{i_1}, \ldots, i_m : (\zeta_{i_1}, \ldots, \zeta_{i_m}) \to (\phi_1, \ldots, \phi_p)$ defined by setting $\zeta_i = 0$ for $i \notin i_1, \ldots, i_m$ in eqn (2). Denote the image of $\mathcal{D}_\zeta$ using the transformation $\mathcal{B}_{i_1}, \ldots, i_m$ by $\mathcal{D}_\phi$. Then $\mathcal{D}_\phi$ is a very complicated subset of the original p-dimensional admissible space of the full AR(p) model. From Barndorff-Nielsen and Schou (1973, Thm 2) it follows that the transformation $\mathcal{B}_{i_1}, \ldots, i_m : \mathcal{D}_\zeta \to \mathcal{D}_\phi$ is one-to-one, onto, continuous and differentiable. Denote the p-functions determined by $\mathcal{B}_{i_1}, \ldots, i_m$ as $\phi_j(\zeta_{i_1}, \ldots, \zeta_{i_m}), j = 1, \ldots, p$. It follows from eqn (1) that each of these

p-functions, $\phi_1(\zeta_{i_1}, \ldots, \zeta_{i_m}), \ldots, \phi_p(\zeta_{i_1}, \ldots, \zeta_{i_m})$, are polynomial functions of $\zeta_{i_1}, \ldots, \zeta_{i_m}$.

2. MODEL FITTING

2.1. *Exact likelihood function*

In many cases the sample mean is an efficient estimator of μ, so we will assume the series $z_1, \ldots, z_n$ after mean correction has mean zero. If necessary, an iterative method may be used to obtain the exact maximum likelihood estimate (MLE) for μ and this is implemented in our *Mathematica* and R packages. Typically this increases the computer time by a factor of about five which is often not unreasonable.

Then the exact log-likelihood function, apart from a constant, for the $\mathrm{AR}_\zeta(\zeta_{i_1}, \ldots, \zeta_{i_m})$ may be written,

$$\mathcal{L}(\zeta, \sigma_a^2) = -\tfrac{1}{2}\log(\det(\Gamma_n)) - \tfrac{1}{2} z' \Gamma_n^{-1} z \tag{3}$$

where $z = (z_1, \ldots, z_n)$ and Γ_n is the $n \times n$ covariance matrix with (i, j)th entry $\gamma_{i-j} = \mathrm{cov}(z_{t-i}, z_{t-j})$. It follows from Box *et al.* (1994, Sect. A7.5), $z' \Gamma_n^{-1} z = \mathcal{S}(\zeta)/\sigma_a^2$, where $\mathcal{S}(\zeta) = \beta' D \beta$, $\beta = (-1, \phi_1(\zeta), \ldots, \phi_p(\zeta))$ and D is the $(p+1) \times (p+1)$ matrix with (i, j)th entry, $D_{i,j} = z_i z_j + \cdots + z_{n-j} z_{n-i}$. Then letting $p = i_m$ and $g_p = \det(\Gamma_n/\sigma_a^2) = \det(\Gamma_p/\sigma_a^2)$, we have from Barndorff-Nielsen and Schou (1973, eqns 5 and 8),

$$g_p = \prod_{i \in \{i_1, \ldots, i_m\}} (1 - \zeta_i^2)^{-i}. \tag{4}$$

The log-likelihood function can now be written,

$$\mathcal{L}(\zeta, \sigma_a^2) = -\frac{n}{2}\log(\sigma_a^2) - \frac{1}{2}\log(g_p) - \frac{1}{2\sigma_a^2}\mathcal{S}(\zeta). \tag{5}$$

Maximizing over σ_a^2 and dropping constant terms, the concentrated log-likelihood is,

$$\mathcal{L}_c(\zeta) = -\frac{n}{2}\log(\hat{\sigma}_a^2) - \frac{1}{2}\log(g_p), \tag{6}$$

where $\hat{\sigma}_a^2 = \mathcal{S}(\zeta)/n$. $\mathcal{L}_c$ can be optimized numerically using a constrained optimization algorithm such as `FindMinimum` in *Mathematica*. The initial evaluation of D requires $O(n)$ flops but this is performed only once and subsequent likelihood evaluations require only $O(1)$ flops.

There are a number of algorithms for ARMA likelihood evaluation and many of these are listed in Box and Luceño (1997, Sect. 12B). Anyone of these algorithms could also be used. However, all of these algorithms require $O(n)$ flops

per likelihood evaluation and so the algorithm presented in this section is much more efficient.

Initial values of the parameters may be obtained using the partial autocorrelations computed by the Burg algorithm. With this approach, it is possible to obtain exact MLE for even quite high-order AR models as illustrated by the monthly sunspot example, Section 3.3, where $m = 70$ coefficients were estimated.

2.2. *Large-sample distribution of the estimates*

For an observed time series $z_1, \ldots, z_n$ generated by an $\mathrm{AR}_\zeta(i_1, \ldots, i_m)$ model, let $\hat{\zeta} = (\hat{\zeta}_{i_1}, \ldots, \hat{\zeta}_{i_m})$ denote the MLE of $\zeta = (\zeta_{i_1}, \ldots, \zeta_{i_m})$. In the full-model case, $m = p$ and $(\zeta_1, \ldots, \zeta_p)$ is a reparameterization of the AR(p) model. However, in the subset case, when $m < p$, the parameters $\phi_1, \ldots, \phi_p$ are constrained and so $\hat{\phi}_1, \ldots, \hat{\phi}_p$ do not have the usual distribution because of these constraints. Theorem 1 is established in Appendix A.

THEOREM 1. *$\hat{\zeta} \xrightarrow{p} \zeta$ and $\sqrt{n}(\hat{\zeta} - \zeta) \xrightarrow[D]{\mathrm{D}} \mathcal{N}(0, \mathcal{I}_\zeta^{-1})$, where $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow[D]{}$ denotes convergence in distribution and $\mathcal{I}_\zeta$ is the large-sample Fisher information matrix per observation of ζ.*

Properties of the information matrix $\mathcal{I}_\zeta$ are discussed in Barndorff-Nielsen and Schou (1973) for the case of the full model, $\mathrm{AR}_\zeta(1, \ldots, p)$, but a general method of computing $\mathcal{I}_\zeta$ is not explicitly given. It is shown in Appendix A that $\mathcal{I}_\zeta = \mathcal{J}_\zeta' \mathcal{I}_\phi \mathcal{J}_\zeta$, where

$$\mathcal{J}_\zeta = \frac{\partial(\phi_1, \ldots, \phi_p)}{\partial(\zeta_{i_1}, \ldots, \zeta_{i_m})} \tag{7}$$

and $\mathcal{I}_\phi$ is the information matrix for $\phi_1, \ldots, \phi_p$ in the unrestricted AR(p) model. Since $\mathcal{I}_\phi = \sigma_a^{-2}\Gamma_p$, $\mathcal{I}_\phi$ may easily be computed using the result of Siddiqui (1958),

$$\mathcal{I}_\phi^{-1} = \left(\sum_{k=1}^{\min(i,j)} \phi_{i-k}\phi_{j-k} - \phi_{p+k-i}\phi_{p+k-j} \right)_{p \times p}, \tag{8}$$

where $\phi_0 = -1$. The Jacobian $\mathcal{J}_\zeta$ is quite complicated. First, consider the full-model case, $\mathrm{AR}_\zeta(1, \ldots, p)$. The required Jacobian may be derived as the product of a sequence of Jacobians of transformations used in the Durbin–Levinson algorithm, eqn (1), to obtain, $\mathcal{J}_\zeta = \mathcal{J}_{p-1} \cdots \mathcal{J}_1$, where

$$\mathcal{J}_k = \frac{\partial(\phi_{k+1,1}, \ldots, \phi_{k+1,p})}{\partial(\phi_{k,1}, \ldots, \phi_{k,p})}, \tag{9}$$

where $\phi_{k,j} = \phi_{k,k}$ for $j \geq k$ and otherwise for $j < k$, $\phi_{k,j}$ is as defined in eqn (1). It may then be shown that

$$\mathcal{J}_k = \begin{pmatrix} J_{p-k} & A_{p-k,k} \\ 0_{k,p-k} & I_k \end{pmatrix}, \tag{10}$$

where J_{p-k} is the $(p-k) \times (p-k)$ matrix with (i, j)th entry, $J_{i,j}$, where

$$J_{i,j} = \begin{cases} 1, & \text{if } \ i = j, \\ -\zeta_{p-k+1}, & \text{if } \ i = p-k+1-j \wedge i \neq j, \\ 1-\zeta_{p-k+1}, & \text{if } \ i = p-k+1-j \wedge i = j, \\ 0, & \text{otherwise}, \end{cases} \tag{11}$$

$A_{p-k,k}$ is the $(p-k) \times k$ matrix whose first column is $(-\phi_{p-k,p-k}, -\phi_{p-k,p-k-1}, \ldots, -\phi_{p-k,1})$ and whose remaining elements are zeros, $0_{p-k,k}$ is the $k \times (p-k)$ matrix with all zero entries, and I_k is the $k \times k$ identity matrix. For example, for the $\mathrm{AR}_\zeta(1, 2, 3, 4)$,

$$\mathcal{J}_\zeta = \begin{pmatrix} 1 & 0 & -\zeta_4 & -\zeta_3 \\ 0 & 1-\zeta_4 & 0 & -\zeta_2 - \zeta_1(1+\zeta_2)\zeta_3 \\ -\zeta_4 & 0 & 1 & -\zeta_1(1+\zeta_2) - \zeta_2\zeta_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1-\zeta_2 & -\zeta_1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\zeta_3 & -\zeta_2 & 0 \\ -\zeta_3 & 1 & -\zeta_1(1+\zeta_2) & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{12}$$

The information matrix of ζ in the subset case, $\mathrm{AR}_\zeta(i_1, \ldots, i_m)$, may be obtained by selecting rows and columns corresponding to $i_1, \ldots, i_m$ from full-model information matrix. Equivalently, the information matrix in the subset case could also be obtained by selecting the columns corresponding to $i_1, \ldots, i_m$ in the Jacobian matrix corresponding to the full-model case to obtain $\mathcal{J}_{(\zeta_{i1}, \ldots, \zeta_{i_m})}$. Then

$$\mathcal{I}_{(\zeta_{i_1}, \ldots, \zeta_{im})} = \mathcal{J}'_{(\zeta_{i_1}, \ldots, \zeta_{im})} \mathcal{I}_\phi \mathcal{J}_{(\zeta_{i_1}, \ldots, \zeta_{im})}. \tag{13}$$

For example, using our *Mathematica* software, for the $\mathrm{AR}_\zeta(1, 12)$,

$$\mathcal{I}_\zeta = \frac{1}{1-\zeta_{12}^2} \begin{pmatrix} (1 - 2\zeta_1^{10}\zeta_{12} + \zeta_{12}^2)/(1-\zeta_1^2) & 0 \\ 0 & 1 \end{pmatrix}. \tag{14}$$

As a check on the formula for $\mathcal{I}_\zeta$, an $\mathrm{AR}_\zeta(1, 2, 3, 4)$ with $\zeta_i = 0.5$, $i = 1, \ldots, 4$ and $n = 500$ was simulated and fit 1000 times. The empirical covariance matrix of the ζ-parameters was found to agree closely with the theoretical covariance matrix $n^{-1}\mathcal{I}_\zeta^{-1}$. This experiment was repeated using the $\mathrm{AR}_\zeta(1, 4)$ model.

As mentioned in Section 2.1, the Burg algorithm can be used to generate initial estimates of the parameters ζ_i, $i = i_1, \ldots, i_m$ for the exact MLE algorithm. We prefer to use the exact MLE for our final model estimates since these estimates are known to be second-order efficient (Taniguchi, 1983) and simulation experiments

have shown that the exact MLE estimates usually perform better than alternatives in small samples (Box and Luceño, 1997, Sect. 12B).

2.3. *Model identification*

Theorem 1 provides the basis for a useful model identification method for $\mathrm{AR}_\zeta(i_1,\ldots,i_m)$ using the partial autocorrelation function. The partial autocorrelations are estimated for a suitable number of lags $k = 1,\ldots,K$. Typically $K < n/4$. For $K \le 20$, we recommend the MLE using the algorithm in Section 2.1 and for $K > 20$, the Burg estimates may be used. Based on the fitted AR(K) model, the estimated standard errors, $\mathrm{EstSd}(\hat{\zeta}_k)$, of $\hat{\zeta}_k$ are obtained. A suitable $\mathrm{AR}_\zeta(i_1,\ldots,i_m)$ model may be selected by examining a plot of $\hat{\zeta}_k \pm 1.96\mathrm{EstSd}(\hat{\zeta}_k)$ vs. k. This modified partial autocorrelation plot is generally more useful than the customary one (Box *et al.*, 1994). The use of this partial autocorrelation plot is illustrated in Sections 3.1 and 3.2.

Another method of model selection can be based on the BIC defined by $\mathrm{BIC} = -2\mathcal{L}_c + m \quad \log(n)$, where n is the length of the time series and m is the number of parameters estimated. From eqn (6), the log-likelihood of the $\mathrm{AR}_\zeta(i_1,\ldots,i_m)$ may be approximated by $\mathcal{L}_c \approx -(n/2)\log(\hat{\sigma}_a^2)$. Since $\hat{\sigma}_a^2 \approx c_0(1-\hat{\phi}_{i_1,i_1}^2)\cdots(1-\hat{\phi}_{i_m,i_m}^2)$, where c_0 is the sample variance. Hence, we obtain the approximation,

$$\mathrm{BIC}_\zeta = \mathrm{BIC}(i_1,\ldots,i_m) = n\,\log\left(\prod_{k\in\{i_1,\ldots,i_m\}}(1-\hat{\phi}_{k,k}^2)\right) + m\,\log(n). \tag{15}$$

The following algorithm can be used to find the minimum BIC_ζ model:

Step 1. Select L as the maximum order for the autoregression. Select M, $M \le L$, as the maximum number of parameters allowed. The partial autocorrelation plot can be used to select L large enough so that all partial autocorrelations larger than L are assumed zero. Also, from the partial autocorrelations, we can obtain an approximate idea of how many partial autocorrelation parameters might be needed.

Step 2. Sort $(|\hat{\phi}_{1,1}|,\ldots,|\hat{\phi}_{L,L}|)$ in descending order to obtain $(|\hat{\phi}_{i_1,i_1}|,\ldots,|\hat{\phi}_{i_L,i_L}|)$.

Step 3. Compute the $\mathrm{BIC}(i_1,\ldots,i_m)$ for $m = 1,\ldots,M$ and select the minimum BIC_ζ model. It is usually desirable to also consider models which are close to the minimum since sometimes these models may perform better for forecasting on a validation sample or perhaps give better performance on a model diagnostic check. So in this last step, we may select the k best models.

This polynomial time algorithm is suitable for use with long time series and with large L and M. Also, other criteria such as the AIC (Akaike, 1974), AIC_C (Hurvich and Tsai, 1989) or that of Hannan and Quinn (1979) could also be used in this algorithm.

2.4. *Residual autocorrelation diagnostics*

Let $\zeta = (\zeta_{i_1}, \ldots, \zeta_{i_m})$ denote the true parameter values in an $\mathrm{AR}_\zeta(i_1, \ldots, i_m)$ model and let $\dot{\zeta} = (\dot{\zeta}_{i_1}, \ldots, \dot{\zeta}_{i_m})$ denote any value in the admissible parameter space. Then the residuals, $\dot{a}_t, t = p+1, \ldots, n$, corresponding to the parameter $\dot{\zeta}$ and data $z_1, \ldots, z_n$ from a mean-corrected time series are defined by $\dot{a}_t = z_t - \dot{\phi}_1 z_{t-1} - \cdots - \dot{\phi}_p z_{t-p}$ where $t = p+1, \ldots, n$, $\dot{\phi}_i = \phi_i(\dot{\zeta})$ and $p = i_m$. The residuals corresponding to the initial values, $t = 1, \ldots, p$, may be obtained using the backforecasting method of Box *et al.* (1994, Ch. 5) or for asymptotic computations they can simply be set to zero. For lag $k \geq 0$, the residual autocorrelations are defined by $\dot{r}_k = \dot{c}_k/\dot{c}_0$, where $\dot{c}_k = n^{-1}\Sigma_{t=k+1}^{n} \dot{a}_{t-k}\dot{a}_t$ for all $k \geq 0$. When $\dot{\zeta} = \hat{\zeta}$, the residuals and residual autocorrelations will be denoted by $\hat{a}_t$ and $\hat{r}_k$ respectively. For any $L > 1$, let $\dot{r} = (\dot{r}_1, \ldots, \dot{r}_L)$ and similarly for $\hat{r}$ and r.

THEOREM 2. $\sqrt{n}\hat{r} \xrightarrow{D} \mathcal{N}(0, \mathcal{V}_r)$ *where*

$$\mathcal{V}_r = I_m - \mathcal{X}\mathcal{J}_\zeta \mathcal{I}_\zeta^{-1} \mathcal{J}_\zeta' \mathcal{X}' \tag{16}$$

where $\mathcal{X}$ is the $L \times p$ matrix with (i, j)th entry ψ_{i-j}, where the ψ_k are determined as the coefficients of B^k in the expansion $1/\phi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots$ and $\mathcal{J}_\zeta$ and $\mathcal{I}_\zeta$ are as defined in Section 2.3 *for the $AR_\zeta(i_1, \ldots, i_m)$ model.*

This theorem is proved in Appendix B.

Since $\mathcal{J}_\zeta'\mathcal{X}'\mathcal{X}\mathcal{J}_\zeta \approx \mathcal{J}_\zeta'\mathcal{I}_\phi\mathcal{J}_\zeta$ for L large enough and since $\mathcal{I}_\zeta = \mathcal{J}_\zeta'\mathcal{I}_\phi\mathcal{J}_\zeta$, it follows that $\mathcal{V}_r$ is approximately idempotent with rank $L - m$ for L large enough. This justifies the use of the modified portmanteau diagnostic test of Ljung and Box (1978) $Q_L = n(n+2)\Sigma_{k=1}^{L} \hat{r}_k^2/(n-k)$. Under the null hypothesis that the model is adequate, Q_L, is approximately chi-squared distributed on $L - m$ degrees of freedom.

It is also useful to plot the residual autocorrelations and show their $(1-\alpha)\%$ simultaneous confidence interval. As pointed out by Hosking and Ravishanker (1993), a simultaneous confidence interval may be obtained by applying the Bonferonni inequality. The estimated standard deviation of $\hat{r}_k$ is $\mathrm{EstSd}(\hat{r}_k) = v_{i,i}/\sqrt{n}$, where $v_{i,i}$ is the (i, i)th element of the covariance matrix $\hat{\mathcal{V}}_r$ obtained by replacing parameter ζ in eqn (16) by its estimate $\hat{\zeta}$. Then, using the approximation with the Bonferonni inequality, it may be shown that a $(1-\alpha)\%$ simultaneous confidence interval for $\hat{r}_1, \ldots, \hat{r}_L$ is given by $\Phi^{-1}(1 - \alpha/(2m))\mathrm{EstSd}(\hat{r}_k)$, where $\Phi^{-1}(\cdot)$ denotes the inverse cumulative distribution function of the standard normal distribution. This diagnostic plot is illustrated in Section 3.1.

3. ILLUSTRATIVE EXAMPLES

3.1. *Chemical process time series*

Cleveland (1971) identified an $\mathrm{AR}_\phi(1, 2, 7)$ and Unnikrishnan (2004) identified an $\mathrm{AR}_\phi(1, 3, 7)$ model for Series A (Box *et al.*, 1994). Either directly from the partial

autocorrelation plot in Figure 1 or using the BIC_ζ algorithm in Section 2.3 with $L = 20$ and $M = 10$, an $\mathrm{AR}_\zeta(1, 2, 6, 7)$ subset is selected. The top five models with this algorithm are shown in Table I. Figure 2 shows the residual autocorrelation plots for the fitted AR_ϕ and AR_ζ models. The respective maximized log-likelihoods were $\mathcal{L}_c = 232.96$ and 231.69. Thus, a slightly better fit is achieved by the AR_ϕ model in this case, but since the difference is small, it may be concluded that both models fit about equally well.

3.2. *Forecasting experiment*

McLeod and Hipel (1995) fitted an $\mathrm{AR}_\phi(1, 9)$ to the treering series identified as NINEMILE in their article. This series consists of $n = 771$ consecutive annual treering width measurements on Douglas fir at Nine Mile Canyon, UT, for the years 1194–1964. For our forecasting experiment, the first 671 values were used as the training data and the last 100 as the test data. The partial autocorrelation plot of the training series is shown in Figure 3. This plot suggests $L = 15$ and $M = 5$ in the algorithm in Section 2.3 will suffice. The three best BIC_ζ models were $\mathrm{AR}_\zeta(1)$, $\mathrm{AR}_\zeta(1, 9)$ and $\mathrm{AR}_\zeta(1, 2, 9)$. After fitting with exact maximum likelihood, the one-step forecast errors were computed for the test data. The $\mathrm{AR}_\phi(1, 9)$ model was also fit to the training series and the one-step forecast errors over the next 100

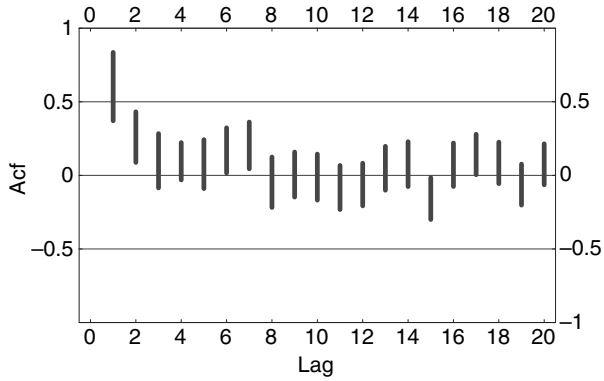

FIGURE 1. Partial autocorrelation plot of Series A. Maximum likelihood estimates used.

TABLE I

TOP FIVE MODELS FOR SERIES A USING THE BIC_ζ ALGORITHM

Model	BIC_ζ
$\mathrm{AR}_\zeta(1, 2, 6, 7)$	−95.91
$\mathrm{AR}_\zeta(1, 2, 6, 7, 15)$	−95.59
$\mathrm{AR}_\zeta(1, 2, 7)$	−95.44
$\mathrm{AR}_\zeta(1, 2, 6, 7, 15, 17)$	−94.32
$\mathrm{AR}_\zeta(1, 2)$	−92.39

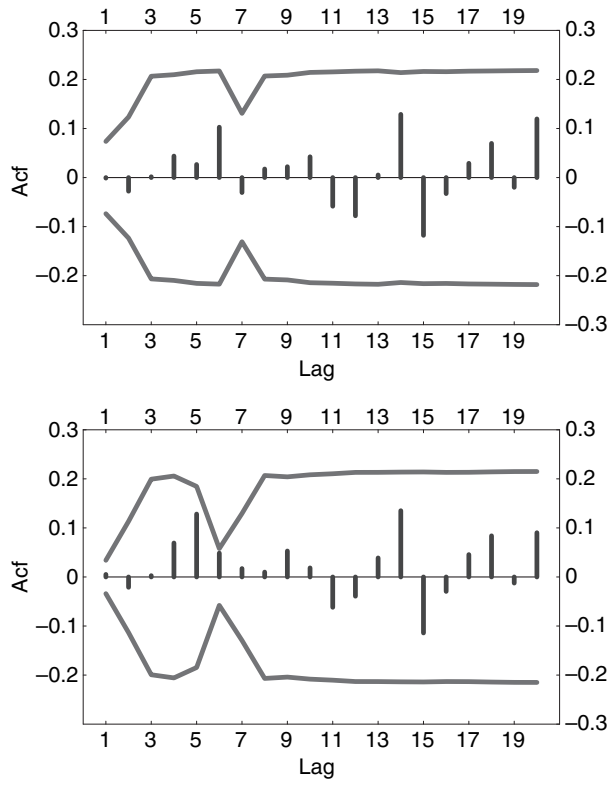

FIGURE 2. Residual autocorrelation plots for the $AR_{\phi}(1, 2, 7)$, upper panel, and the $AR_{\zeta}(1, 2, 6, 7)$, lower panel.

values were computed. Table II compares the fits achieved as well as the root mean square error on the test data. From Table II, with further statistical tests, it was concluded that there is no significant difference in forecast performance.

3.3. *Monthly sunspot series*

The monthly sunspot numbers, 1749–1983 (Andrews and Herzberg, 1985), are comprised of $n = 2820$ consecutive values. Computing the coefficient of skewness for the transformed data, z_t^{λ}, with $\lambda = 1, 2/3, 1/2, 1/3$, we obtained $g_1 = 1.10, 0.48, 0.09, -0.45$ respectively. It is seen that a square-root transformation will improve the normality assumption. For the square-root transformed series, subset AR_{ζ} models were determined using the AIC_{ζ} and BIC_{ζ} algorithms with $L = 300$ and $M = 100$. These algorithms produced subset models with $m = 70$ and $m = 20$ autoregressive coefficients. Maximum likelihood estimation of these two models required only about 30 and 3 minutes, respectively, on a 3 GHz PC using our *Mathematica* software. The best non-subset AR(p) models for the

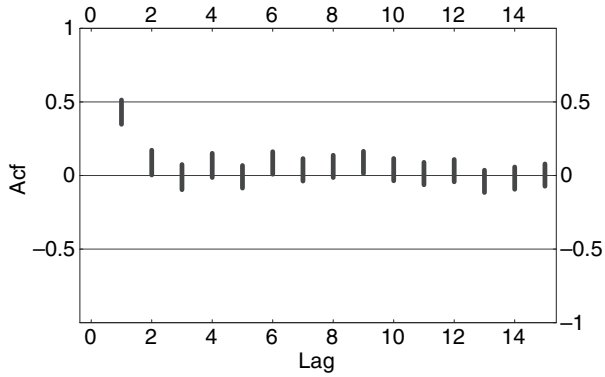

FIGURE 3. Partial autocorrelation plot of the training portion of the Ninemile treering series. Maximum likelihood estimates used.

TABLE II

COMPARISON OF MODELS FIT TO THE TRAINING PORTION OF THE NINEMILE DATASET AND THE ROOT MEAN SQUARE ERROR, RMSE, OF THE ONE-STEP-AHEAD FORECASTS ON THE TEST PORTION

Model	$\mathcal{L}_c$	BIC	$\hat{\sigma}_a$	RMSE
$AR_\phi(1, 9)$	−2462.90	4938.80	39.31	42.34
$AR_\phi(1, 2, 9)$	−2460.65	4947.34	39.16	41.82
$AR_\zeta(1)$	−2467.90	4942.40	39.56	43.29
$AR_\zeta(1, 9)$	−2465.30	4943.60	39.40	42.80
$AR_\zeta(1, 2, 9)$	−2463.00	4945.40	39.27	42.29

square-root monthly sunspots using the AIC and BIC are compared with the subset models in Table III. The AR_ζ fitted with the BIC_ζ algorithm has fewer parameters than each of these non-subset models and it performs better on both the AIC and BIC criteria. Residual autocorrelation diagnostic checks did not indicate any model inadequacy in any of the fitted models.

TABLE III

COMPARISON OF MODELS FITTED TO THE SQUARE-ROOT OF THE MONTHLY SUNSPOT SERIES

Model	m	$\mathcal{L}_c$	AIC	BIC
AR_ζ(AIC)	70	−148.2	436.4	852.6
AR_ζ(BIC)	20	−236.5	513.	631.9
AR(AIC)	27	−241.1	536.3	696.8
AR(BIC)	21	−252.5	547.	671.8

The AR_ζ model fitted using the AIC is denoted by AR_ζ(AIC) and similarly for AR_ζ(BIC). The best non-subset AR models fitted using the AIC and BIC are denoted by AR(AIC) and AR(BIC). The number of coefficients, m, is shown. The values of the AIC and BIC shown are obtained using the exact log-likelihood.

4. CONCLUDING REMARKS

The subset AR model in this article can be extended to the more general case of regression with autocorrelated error. In this case, we would first fit the regression and then fit the subset AR model to the residuals. The Trench algorithm (McLeod and Zinovi, 2006) could then be used to iteratively obtain the joint MLE for all parameters.

The methods presented in this article can be extended to subset MA models. In this case, the ζ-parameters are inverse partial autocorrelations (Hipel and McLeod, 1994, Sect. 5.3.7). Bhansali (1983) showed that the distribution of the inverse partial autocorrelations is equivalent to that of the partial autocorrelations, so the model selection using a modified inverse partial autocorrelation plot or AIC/BIC criterion may be implemented. Similarly, the distribution of the residual autocorrelations is essentially equivalent to that given in Theorem 2.

As discussed by Monahan (1984) and Marriott and Smith (1992), the transformation used in eqn (2) may be extended to reparameterize ARMA models. Hence, the subset AR model may be generalized in this way to the subset ARMA case.

Software written in *Mathematica* is available from the authors' web page for reproducing the examples given in this article as well as for more general usage. An R package implementation is also available.

APPENDIX A: PROOF OF THEOREM 1

Let $\dot{\zeta}$, $\hat{\zeta}$ and ζ denote, respectively, a vector of parameters in the admissible region, the MLE and the true parameter values, and similarly for other functions of these quantities such as the likelihood and residuals. As in McLeod (1978), without loss of generality, we may assume that $\mu = 0$ and $\sigma = 1$ are known. Ignoring terms which are $O_p(1)$, the log-likelihood function of $\dot{\zeta}$ may be written,

$$\dot{\mathcal{L}} = -\tfrac{1}{2}\sum \dot{a}_t^2, \tag{A.1}$$

where $\dot{a}_t = z_t - \dot{\phi}_1 z_{t-1} - \cdots - \dot{\phi}_p z_{t-p}$. Note that for all i and j,

$$\frac{\partial \dot{a}_t}{\partial \dot{\phi}_i} = -z_{t-i} \tag{A.2}$$

and

$$\frac{\partial \dot{a}_t}{\partial \dot{\phi}_i \dot{\phi}_j} = 0. \tag{A.3}$$

It follows that

$$\frac{1}{n}\frac{\partial \mathcal{L}}{\partial \zeta} = \frac{1}{n}\sum a_t \mathcal{J}_\zeta'(z_{t-1}, \ldots, z_{t-p})', \tag{A.4}$$

where $(z_{t-1}, \ldots, z_{t-p})'$ denotes the transpose of the p-dimensional row vector and

$$\mathcal{J}_\zeta = \frac{\partial(\phi_1, \ldots, \phi_p)}{\partial(\zeta_{i_1}, \ldots, \zeta_{i_m})}. \tag{A.5}$$

Since $\Sigma z_{t-j} a_t / n \xrightarrow{p} 0$ when $j > 0$, it follows that from eqn (A.4),

$$n^{-1} \frac{\partial \mathcal{L}}{\partial \zeta} \xrightarrow{p} 0. \tag{A.6}$$

Similarly, neglecting terms which are $O_p(1/\sqrt{n})$,

$$\begin{aligned} -\frac{1}{n}\frac{\partial^2 \mathcal{L}}{\partial \zeta \partial \zeta'} &= \frac{1}{n}\mathcal{J}'_\zeta\Big(\sum z_{t-i} z_{t-j}\Big)\mathcal{J}_\zeta \\ &\xrightarrow{p} \mathcal{J}'_\zeta \mathcal{I}_\phi \mathcal{J}_\zeta \\ &= \mathcal{I}_\zeta, \end{aligned} \tag{A.7}$$

where

$$\mathcal{I}_\phi = (\mathrm{cov}(z_{t-i}, z_{t-j}))_{p\times p}. \tag{A.8}$$

Since the transformation is one-to-one, continuous and differentiable, it follows that $\mathcal{I}_\zeta$ must be positive definite since $\mathcal{I}_\phi$ is positive definite. Expanding $n^{-1}\partial \dot{\mathcal{L}}/\partial \dot{\zeta}$ about $\dot{\zeta} = \hat{\zeta}$ and evaluating at $\dot{\zeta} = \zeta$ and noting that third and higher-order terms are zero,

$$0 = \frac{1}{n}\frac{\partial \mathcal{L}}{\partial \zeta} + (\hat{\zeta} - \zeta)\frac{1}{n}\frac{\partial^2 \mathcal{L}}{\partial \zeta \partial \zeta'}. \tag{A.9}$$

It follows from eqns (A.6), (A.7) and (A.9) that $\hat{\zeta} \xrightarrow{p} \zeta$. Since

$$-\frac{1}{n}\frac{\partial^2 \mathcal{L}}{\partial \zeta \partial \zeta'} = \mathcal{I}_\zeta + O_p\left[\frac{1}{\sqrt{n}}\right], \tag{A.10}$$

it follows that

$$\frac{1}{\sqrt{n}}\frac{\partial \mathcal{L}}{\partial \zeta} = \sqrt{n}(\hat{\zeta} - \zeta)\mathcal{I}_\zeta + o_p(1). \tag{A.11}$$

Since

$$\frac{1}{\sqrt{n}}\frac{\partial \mathcal{L}}{\partial \phi} \xrightarrow{D} N(0, \mathcal{I}_\phi). \tag{A.12}$$

and

$$\frac{\partial \mathcal{L}}{\partial \zeta} = \frac{\partial \mathcal{L}'}{\partial \phi}\,\mathcal{J}_\zeta \tag{A.13}$$

it follows that

$$\sqrt{n}(\hat{\zeta} - \zeta) \xrightarrow{D} N(0, \mathcal{I}_\zeta^{-1}). \tag{A.14}$$

APPENDIX B: PROOF OF THEOREM 2

Without loss of generality and ignoring terms which are $O_p(1)$, we may write the log-likelihood function as $\mathcal{L}(\dot{\zeta}) = -\frac{1}{2}\Sigma \dot{a}_t^2$. Then $r_k = c_k + O_p(1/n)$ and,

$$\zeta - \hat{\zeta} = \mathcal{I}_\zeta^{-1} s_c + O_p(1/n), \tag{B.1}$$

where

$$\begin{aligned} s_c &= \frac{\partial \mathcal{L}}{\partial \zeta} \\ &= \mathcal{J}_\zeta \frac{\partial \mathcal{L}}{\partial \phi} \\ &= \mathcal{J}_\zeta (\sum a_t z_{t-i}). \end{aligned} \tag{B.2}$$

It follows that, neglecting a term which is $O(1)$,

$$n \operatorname{cov}(\hat{\zeta}, r) = -\mathcal{I}_\zeta^{-1} \mathcal{J}_\zeta' \mathcal{X}'. \tag{B.3}$$

Expanding $\dot{r}$ about $\dot{\zeta} = \zeta$ and evaluating at $\dot{\zeta} = \hat{\zeta}$,

$$\hat{r} = r + \mathcal{X} \mathcal{J}_\zeta (\hat{\zeta} - \zeta). \tag{B.4}$$

From eqn (B.4), it follows that $\sqrt{n} r$ is asymptotically normal with mean zero and the given covariance matrix.

This theorem could also be derived using the result of Ahn (1988) on multivariate autoregressions with structured parameterizations.

ACKNOWLEDGEMENT

Support from the NSERC Discovery Grant program is acknowledged with thanks. Valuable suggestions were made by the referee.

NOTE

Corresponding Author: A. I. McLeod, Department of Statistical and Actuarial Sciences, The University of Western Ontario, London, ON, Canada N6A 5B7. E-mail: aimcleod@uwo.ca

REFERENCES

Ahn, S. K. (1988) Distribution for residual autocovariate in multivariate autoregressive models with structured parameterization. *Biometrika* 75, 590–3.

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* AC-19, 716–23.

Andrews, D. F. and Herzberg, A. M. (1985) *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. New York: Springer-Verlag.

Barndorff-Nielsen, O. and Schou G. (1973) On the parametrization of autoregressive models by partial autocorrelations. *Journal of Multivariate Analysis* 3, 408–19.

Bhansali, R. J. (1983) The inverse partial autocorrelation function of a time series and its applications. *Journal of Multivariate Analysis* 13, 310–27.
Box, G. E. P. and Luceño, A. (1997) *Statistical Control by Monitoring and Feedback Adjustment*. New York: Wiley.
Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994) *Time Series Analysis: Forecasting and Control* (3rd edn). San Francisco: Holden-Day.
Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods*, 2nd edn. New York: Springer-Verlag.
Chen, C. W. S. (1999) Subset selection of autoregressive time series models. *Journal of Forecasting* 18, 505–16.
Cleveland, W. S. (1971) The inverse autocorrelations of a time series and their applications. *Technometrics* 14, 277–98.
George, E. I. and McCulloch, R. E. (1993) Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 88, 881–9.
Haggan, V. and Oyetunji, O. B. (1984) On the selection of subset autoregressive time series models. *Journal of Time Series Analysis* 5, 103–13.
Hannan, E. J. and Quinn, B. G. (1979) The determination of the order of an autoregression. *Journal of the Royal Statistical Society* B 41, 190–5.
Hipel, K. W. and McLeod, A. I. (1994) *Time Series Modelling of Water Resources and Environmental Systems*. Amsterdam: Elsevier.
Hosking, J. R. M. and Ravishanker, N. (1993) Approximate simultaneous significance intervals for residual autocorrelations of autoregressive-moving average time series models. *Journal of Time Series Analysis* 14, 19–26.
Hurvich, M. C. and Tsai, C. (1989) Regression and time series model selection in small samples. *Biometrika* 76, 297–307.
Ljung, G. M. and Box, G. E. P. (1978) On a measure of lack of fit in time series models. *Biometrika* 65, 297–303.
Marriott, J. M. and Smith, A. F. M. (1992) Reparameterization aspects of numerical Bayesian methodology for autoregressive-moving average models. *Journal of Time Series Analysis* 13, 327–43.
McClave, J. (1975) Subset autoregression. *Technometrics* 17, 213–20.
McLeod, A. I. (1978) On the distribution of residual autocorrelations in Box–Jenkins models. *Journal of the Royal Statistical Society* B 40, 396–402.
McLeod, A. I. and Hipel, K. W. (1995) Exploratory spectral analysis of hydrological time series. *Journal of Stochastic Hydrology and Hydraulics* 9, 171–205.
McLeod, A. I. and Zinovi, K. (2006) Trench algorithm for time series analysis. *Journal of Statistical Software*. Accepted.
Monahan, J. F. (1984) A note on enforcing stationarity in autoregressive-moving average models. *Biometrika* 71, 403–4.
Percival, D. B. and Walden, A. T. (1993) *Spectral Analysis For Physical Applications*, Cambridge: Cambridge University Press.
Siddiqui, M. M. (1958) On the inversion of the sample covariance matrix in a stationary autoregressive process. *Annals of Mathematical Statistics* 29, 585–8.
Taniguchi, M. (1983) On the second order asymptotic efficiency of estimators of gaussian ARMA processes. *The Annals of Statistics* 11, 157–69.
Tjøstheim, D. and Paulsen, J. (1983) Bias of some commonly-used time series estimates. *Biometrika* 70, 389–99. Correction *Biometrika* 71, p. 656.
Unnikrishnan, N. K. (2004) Bayesian subset model selection for time series. *Journal of Time Series Analysis* 25, 671–90.
Zhang, Y. and McLeod, A. I. (2006) Computer algebra derivation of the bias of Burg estimators. *Journal of Time Series Analysis* 27, 157–65.
Zhang, X. and Terrell, R. D. (1997) Prediction modulus: a new direction for selecting subset autoregressive models. *Journal of Time Series Analysis* 18, 195–212.