Bias Reduction of Autoregressive Estimates in Time Series Regression Model Through Restricted Maximum Likelihood

Wai-Kwong CHEANG and Gregory C. REINSEL

This article considers maximum likelihood (ML) and restricted maximum likelihood (REML) estimation of time series regression models with autoregressive AR(p) noise. Approximate biases of the ML and REML estimators of the AR parameters, based on their approximate representations, are derived. In addition, a bias result for the ML estimator (MLE) of the error variance is established. Numerical results are presented to illustrate the biases of the MLE and REML estimator for the AR parameters, and simulation results are provided to assess the adequacy of our approximations. The impact of bias of the AR estimates on testing of linear trend in a regression trend model is also investigated. For a time series of short or moderate sample length, the REML estimator is generally much less biased than the MLE. Consequently, the REML approach leads to more accurate inferences for the regression parameters.

KEY WORDS: Autoregression; Bias; Maximum likelihood estimator; Restricted maximum likelihood estimator; Time series regression model.

1. INTRODUCTION

For a linear regression model involving time series data with stationary autoregressive moving average (ARMA) error terms, the maximum likelihood estimator (MLE) of the ARMA model parameters can be substantially biased in its finite-sample distribution. This bias may have a significant impact on inferences of the regression parameters based on generalized least squares (GLS) estimation, particularly when the sample size is not large and the AR or the MA polynomial has roots near the unit circle. One standard "corrective" approach for reducing the bias is to use a bias-corrected MLE, obtained by subtracting a bias estimate from the MLE. We consider use of the restricted maximum likelihood (REML) estimation procedure to be a "preventive" approach to bias reduction for the ARMA parameters.

Although REML has been rather popular and widely used in the estimation of variance components (see, e.g., Harville 1977; Searle, Casella, and McCulloch 1992) in mixed-effects linear models, its use in time series regression is less developed. The use of this procedure for ARMA models has been described by Cooper and Thompson (1977) and Tunnicliffe Wilson (1989), among others. Through a simulation study by Cooper and Thompson (1977) of an MA(1) model with unknown mean, as well as a study by McGilchrist (1989) for polynomial regression with ARMA errors, REML estimates were demonstrated to be less biased than ML estimates. In this article we consider and compare ML and REML estimation of a time series regression model with autoregressive AR(p) noise. Accurate approximate representations of the MLE and REML estimator are developed in terms of the (unconditional) least squares (LS) estimator. The approximate biases of the MLE and REML estimator of the AR(p) parameters, based on their approximate representations, are derived. In addition, a bias result for the MLE of the error variance is established. Numerical results are presented to illustrate the biases, and simulation results are provided to assess the adequacy of our approximations. To illustrate the impact of bias on inferences for regression parameters, simulation results for testing of linear trend are also presented.

Wai-Kwong Cheang is Assistant Professor, Division of Mathematics, Nanyang Technological University, Singapore 259756. Gregory C. Reinsel is Professor, Department of Statistics, University of Wisconsin, Madison, WI 53706 (E-mail: *reinsel@stat.wisc.edu*). The authors thank the referee, associate editor, and editor for useful comments that led to improvements in the presentation of the article.

2. TIME SERIES REGRESSION MODEL WITH AR(p) NOISE

Suppose that the time series $\{Y_t\}$ of interest follows a linear regression model of the form

$$Y_t = \mathbf{x}_t'\boldsymbol{\beta} + N_t, \qquad t = 1, \ldots, T, \tag{1}$$

where $\mathbf{x}_t = (x_{t1}, \ldots, x_{tr})'$ is a r-dimensional vector of deterministic or stochastic regressors and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)'$ is the vector of unknown regression parameters to be estimated. The noise series $\{N_t\}$ is assumed to be a stationary process following an AR(p) model,

$$N_t = \phi_1 N_{t-1} + \phi_2 N_{t-2} + \cdots + \phi_p N_{t-p} + \varepsilon_t, \tag{2}$$

where $\{\varepsilon_t\}$ is a white noise process with mean 0 and variance σ_ε^2. The AR(p) model may be written as $\phi(B)N_t = \varepsilon_t$, where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$. For $\{N_t\}$ to be stationary, all roots of $\phi(B) = 0$ must be greater than 1 in absolute value. The autocovariance function $\gamma(l) = \operatorname{cov}(N_t, N_{t+l})$ of $\{N_t\}$ satisfies the difference equation

$$\gamma(l) = \phi_1\gamma(l-1) + \phi_2\gamma(l-2) + \cdots + \phi_p\gamma(l-p), \qquad l \ge 1, \tag{3}$$

with $\gamma(0) = \sigma_\varepsilon^2/\Delta$, where $\Delta = 1 - \phi_1\rho_1 - \phi_2\rho_2 - \cdots - \phi_p\rho_p$ and $\rho_l = \gamma(l)/\gamma(0)$.

Given a sample of T observations, let $\mathbf{Y} = (Y_1, \ldots, Y_T)'$ and $\mathbf{N} = (N_1, \ldots, N_T)'$ be the $T \times 1$ data and noise vectors,

Journal of the American Statistical Association
December 2000, Vol. 95, No. 452, Theory and Methods

and let $\boldsymbol{\varepsilon} = (\varepsilon_1^*, \ldots, \varepsilon_p^*, \varepsilon_{p+1}, \ldots, \varepsilon_T)'$. Define the $T \times r$ matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]'$, and assume that $\mathbf{X}$ is of full rank r and satisfies the Grenander conditions (see, e.g., Anderson 1971, p. 572). Briefly, with $d_{ij}(h) = \sum_{t=1}^{T-h} x_{i,t+h} x_{jt}, h = 0, 1, \ldots$, the Grenander conditions assume that $d_{ii}(0) \to \infty, x_{i,T+1}^2/d_{ii}(0) \to 0$, and $\lim d_{ij}(h)/\sqrt{d_{ii}(0)d_{jj}(0)} \equiv r_{ij}(h)$ exists as $T \to \infty$, for $i, j = 1, \ldots, r$, and $h = 0, \pm 1, \pm 2, \ldots$, and we assume the matrix of the limits of elements $\{r_{ij}(0)\}$ is positive definite (nonsingular). Then the models (1) and (2) may be expressed in matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}, \qquad \mathbf{P}'\mathbf{N} = \boldsymbol{\varepsilon}, \tag{4}$$

where the $T \times T$ transformation matrix $\mathbf{P}'$ is lower triangular with its first p diagonal elements equal to $\Delta^{1/2}, (\Delta/\Delta_1)^{1/2}, \ldots, (\Delta/\Delta_{p-1})^{1/2}$, its remaining diagonal elements equal to 1, elements in the (i, j)th position equal to $-\phi_{i-j,i-1}^*$ for $j = 1, \ldots, i-1$ and $i = 2, \ldots, p$, equal to $-\phi_{i-j}$ for $j = i-p, \ldots, i-1$ when $i > p$, and 0 elsewhere. In the first p rows of $\mathbf{P}'$, the elements $\phi_{ik}^* = \phi_{ik}(\Delta/\Delta_k)^{1/2}, i = 1, \ldots, k$, where $\phi_{1k}, \ldots, \phi_{kk}$, for $k = 1, \ldots, p-1$, are solutions for coefficients $\phi_1, \ldots, \phi_k$ in the system of the first k Yule–Walker equations in (3) with p set equal to k, and $\Delta_k = 1 - \phi_{1k}\rho_1 - \phi_{2k}\rho_2 - \cdots - \phi_{kk}\rho_k$. For example, $\phi_{11} = \gamma(1)/\gamma(0) = \rho_1$ and $\Delta_1 = 1 - \rho_1^2$. Because $\operatorname{cov}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}$, it follows that the covariance matrix of $\mathbf{N}$ is $\operatorname{cov}(\mathbf{N}) = \operatorname{cov}(\mathbf{P}'^{-1}\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{P}'^{-1}\mathbf{P}^{-1} = \sigma_\varepsilon^2 \mathbf{V}$, where $\mathbf{V}^{-1} = \mathbf{P}\mathbf{P}'$.

3. MAXIMUM LIKELIHOOD ESTIMATION

Under the regression model (4) and the normality assumption $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I})$, the log-likelihood function is

$$l(\boldsymbol{\beta}, \boldsymbol{\phi}, \sigma_\varepsilon^2) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma_\varepsilon^2) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2\sigma_\varepsilon^2} S(\boldsymbol{\beta}, \boldsymbol{\phi}),$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ and $S(\boldsymbol{\beta}, \boldsymbol{\phi}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ is the sum of squares function. Note that $|\mathbf{V}| = |\mathbf{V}_p|$, where $\sigma_\varepsilon^2 \mathbf{V}_p$ is the covariance matrix of p consecutive values from the AR(p). For example, in the AR(2) model, $p = 2$, and so $|\mathbf{V}| = |\mathbf{V}_2| = (1 - \rho_1^2)/\Delta^2$.

To derive the likelihood equations, we note that $S(\boldsymbol{\beta}, \boldsymbol{\phi})$ can be expressed as a quadratic function of the parameter $\boldsymbol{\phi}$ (see, e.g., Box, Jenkins, and Reinsel 1994, p. 298). We then have the following two equivalent expressions for $S(\boldsymbol{\beta}, \boldsymbol{\phi})$:

$$S(\boldsymbol{\beta}, \boldsymbol{\phi}) = \mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta}$$

and

$$\begin{aligned} S(\boldsymbol{\beta}, \boldsymbol{\phi}) &= \sum_{t=p+1}^{T} \varepsilon_t^2 + \varepsilon_1^{*2} + \cdots + \varepsilon_p^{*2} \\ &= \mathbf{N}'\mathbf{P}\mathbf{P}'\mathbf{N} = C_{00} - 2\boldsymbol{\phi}'\mathbf{c}_p + \boldsymbol{\phi}'\mathbf{C}_p\boldsymbol{\phi}, \end{aligned}$$

where $C_{00} = \sum_{t=1}^{T} N_t^2, \mathbf{c}_p = (C_{10}, C_{20}, \ldots, C_{p0})', \mathbf{C}_p$ is a $p \times p$ symmetric matrix with (i, j)th element C_{ij}, and the elements C_{ij} are "symmetric" sums of squares and lagged cross-products of the N_t ($N_t = Y_t - \mathbf{x}_t'\boldsymbol{\beta}$), given by $C_{ij} = \sum_{t=i+1}^{T-j} N_t N_{t-i+j} = C_{ji}$, with $T - i - j$ terms in the sum. The (i, j)th element of $\mathbf{C}_p$ has expected value $E(C_{ij}) = (T - i - j)\gamma(i - j), i, j = 1, \ldots, p$.

From the foregoing expression, the first partial derivatives of $S(\boldsymbol{\beta}, \boldsymbol{\phi})$ are

$$\frac{\partial S}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} + 2\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta}$$

and

$$\frac{\partial S}{\partial \boldsymbol{\phi}} = 2(\mathbf{C}_p\boldsymbol{\phi} - \mathbf{c}_p),$$

with $E[\partial S/\partial \boldsymbol{\beta}] = \mathbf{0}$ and $E[\partial S/\partial \boldsymbol{\phi}] = 2\{E(\mathbf{C}_p)\boldsymbol{\phi} - E(\mathbf{c}_p)\}$. Notice that the ith element of $E(\mathbf{C}_p)\boldsymbol{\phi} - E(\mathbf{c}_p), \sum_{j=1}^{p} E(C_{ij})\phi_j - E(C_{i0})$, is equal to

$$\sum_{j=1}^{p} (T - i - j)\gamma(i - j)\phi_j - (T - i)\gamma(i) = -\sum_{j=1}^{p} j\gamma(i - j)\phi_j,$$

using the Yule–Walker equations (3). The likelihood equations are

$$\frac{\partial l}{\partial \boldsymbol{\phi}} = -\frac{1}{2} \frac{\partial}{\partial \boldsymbol{\phi}} \log |\mathbf{V}| - \frac{1}{\sigma_\varepsilon^2} (\mathbf{C}_p\boldsymbol{\phi} - \mathbf{c}_p) = \mathbf{0}, \tag{5}$$

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma_\varepsilon^2} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} - \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta}) = \mathbf{0}, \tag{6}$$

and

$$\frac{\partial l}{\partial \sigma_\varepsilon^2} = -\frac{T}{2\sigma_\varepsilon^2} + \frac{S}{2\sigma_\varepsilon^4} = 0. \tag{7}$$

Taking expectations of the likelihood equations in (5) and using the fact that $E(\partial l/\partial \boldsymbol{\phi}) = \mathbf{0}$, it follows that $-(\sigma_\varepsilon^2/2)(\partial/\partial\boldsymbol{\phi}) \log |\mathbf{V}| = E(\mathbf{C}_p)\boldsymbol{\phi} - E(\mathbf{c}_p)$. Therefore, in (5) we obtain

$$-\frac{\sigma_\varepsilon^2}{2} \frac{\partial}{\partial \phi_i} \log |\mathbf{V}| = -\sum_{j=1}^{p} j\gamma(i - j)\phi_j, \quad i = 1, \ldots, p. \tag{8}$$

From (6) and (7), the MLEs $(\hat{\boldsymbol{\beta}}_{\mathrm{M}}, \hat{\boldsymbol{\phi}}_{\mathrm{M}}, \hat{\sigma}_\varepsilon^2)$ are given by

$$\hat{\boldsymbol{\beta}}_{\mathrm{M}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{Y}, \quad \hat{\sigma}_\varepsilon^2 = \frac{1}{T} S(\hat{\boldsymbol{\beta}}_{\mathrm{M}}, \hat{\boldsymbol{\phi}}_{\mathrm{M}}), \tag{9}$$

where $\hat{\mathbf{V}} = \mathbf{V}(\hat{\boldsymbol{\phi}}_{\mathrm{M}})$ and $\hat{\boldsymbol{\phi}}_{\mathrm{M}}$ satisfies the likelihood equations (5).

The second partial derivatives of $S(\boldsymbol{\beta}, \boldsymbol{\phi})$ are $\partial^2 S/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = 2\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ and $\partial^2 S/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}' = 2\mathbf{C}_p$, and also note that $E[\partial^2 S/\partial\boldsymbol{\beta}\partial\boldsymbol{\phi}'] = \mathbf{0}$. Hence we have

$$\begin{aligned} E\left[\frac{\partial^2 S}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'}\right] &= 2E(\mathbf{C}_p) \\ &= 2\{(T - i - j)\gamma(i - j)\} \approx 2(T - 2)\sigma_\varepsilon^2 \mathbf{V}_p. \end{aligned}$$

3.1 Approximate Maximum Likelihood Estimator of ϕ

We now use the method suggested by Box et al. (1994, p. 300) for estimation of an AR(p) model to obtain an approximation to the MLE of $\boldsymbol{\phi}$. Based on this approximate MLE, we derive an approximate representation for the REML estimator in Section 4, and also obtain approximations for bias of the estimators. The basic factors involved in this approach to both estimation and bias approximation are that $S(\boldsymbol{\beta}, \boldsymbol{\phi})$ can be written as a quadratic function of $\boldsymbol{\phi}$ using the C_{ij}, so that $\partial S/\partial \boldsymbol{\phi}$ is a linear form in $\boldsymbol{\phi}$, and that $\partial \log|\mathbf{V}|/\partial \phi_i$ in (8) is replaced by an "approximation" (estimate) that is a linear form in $\boldsymbol{\phi}$. Consequently, an accurate approximation to the likelihood equations in (5) can be obtained that are linear in $\boldsymbol{\phi}$, involving the C_{ij}.

Let $\hat{N}_t = Y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}$ be the "residuals" from the regression model (1), where $\hat{\boldsymbol{\beta}}$ is an estimate of $\boldsymbol{\beta}$ of the form of (9) based on preliminary ML estimation of $\boldsymbol{\phi}$. First, the unconditional LS estimator of $\boldsymbol{\phi}$, obtained by minimizing $S(\boldsymbol{\beta}, \boldsymbol{\phi})$ with respect to $\boldsymbol{\phi}$, is $\hat{\boldsymbol{\phi}}_{\mathrm{L}} = \hat{\mathbf{C}}_p^{-1}\hat{\mathbf{c}}_p$, where now the $\hat{C}_{ij} \equiv C_{ij}(\hat{\mathbf{N}})$ are based on the $\hat{N}_t$. From (8), we have

$$
\begin{aligned}
-\frac{\sigma_\varepsilon^2}{2}\frac{\partial}{\partial \phi_i}\log|\mathbf{V}|
&= -\phi_1\gamma(i-1) - 2\phi_2\gamma(i-2) - \cdots - p\phi_p\gamma(i-p) \\
&\approx -\frac{1}{T-i-1}\,C_{i1}\phi_1 - \frac{2}{T-i-2}\,C_{i2}\phi_2 - \cdots \\
&\quad -\frac{p}{T-i-p}\,C_{ip}\phi_p, \qquad i = 1, \ldots, p,
\end{aligned}
$$

on using $C_{ij}^* \equiv C_{ij}/(T-i-j)$ as an "estimate" of $\gamma(i-j)$. Substituting these approximations into (5), evaluating at $\hat{\mathbf{N}}$, the likelihood equations become linear in $\boldsymbol{\phi}$,

$$
\left(\sum_{j=1}^{p}\hat{C}_{ij}\phi_j - \hat{C}_{i0}\right) + \sum_{j=1}^{p}\frac{j}{T-i-j}\,\hat{C}_{ij}\phi_j = 0, \qquad i = 1, \ldots, p. \tag{10}
$$

Equivalently, we have $\hat{\mathbf{C}}_p^*\boldsymbol{\phi} = \hat{\mathbf{c}}_p^*$, where $\hat{\mathbf{C}}_p^*$ and $\hat{\mathbf{c}}_p^*$ are defined similarly to $\hat{\mathbf{C}}_p$ and $\hat{\mathbf{c}}_p$, with elements $\hat{C}_{ij}^* = \hat{C}_{ij}/(T-i-j)$. An approximate ML estimate $\hat{\boldsymbol{\phi}}_{\mathrm{M}}$ is then obtained as

$$
\hat{\boldsymbol{\phi}}_{\mathrm{M}} = \hat{\mathbf{C}}_p^{*-1}\hat{\mathbf{c}}_p^*. \tag{11}
$$

For example, for AR(1) noise ($p = 1$), we obtain $\hat{\phi}_{\mathrm{L}} = \hat{C}_{10}/\hat{C}_{11}$ and $\hat{\phi}_{\mathrm{M}} = \hat{C}_{10}^*/\hat{C}_{11}^* = \{1 - [1/(T-1)]\}\hat{\phi}_{\mathrm{L}}$, where $\hat{C}_{10}^* = \sum_{t=2}^{T}\hat{N}_t\hat{N}_{t-1}/(T-1)$ and $\hat{C}_{11}^* = \sum_{t=2}^{T-1}\hat{N}_t^2/(T-2)$.

3.2 Bias of the Maximum Likelihood Estimator

We consider the estimator $\hat{\boldsymbol{\phi}}_{\mathrm{M}} = \hat{\mathbf{C}}_p^{*-1}\hat{\mathbf{c}}_p^*$, and set $\boldsymbol{\Gamma}_p^* = E(\hat{\mathbf{C}}_p^*)$ and $\boldsymbol{\gamma}_p^* = E(\hat{\mathbf{c}}_p^*)$. Then

$$
\begin{aligned}
\hat{\boldsymbol{\phi}}_{\mathrm{M}} - \boldsymbol{\phi} &= \hat{\mathbf{C}}_p^{*-1}[\hat{\mathbf{c}}_p^* - \hat{\mathbf{C}}_p^*\boldsymbol{\phi}] \\
&= \hat{\mathbf{C}}_p^{*-1}[(\hat{\mathbf{c}}_p^* - \boldsymbol{\gamma}_p^*) - (\hat{\mathbf{C}}_p^* - \boldsymbol{\Gamma}_p^*)\boldsymbol{\phi} + (\boldsymbol{\gamma}_p^* - \boldsymbol{\Gamma}_p^*\boldsymbol{\phi})].
\end{aligned}
$$

We also use the approximation that $\hat{\mathbf{C}}_p^{*-1} - \boldsymbol{\Gamma}_p^{*-1} = -\hat{\mathbf{C}}_p^{*-1}(\hat{\mathbf{C}}_p^* - \boldsymbol{\Gamma}_p^*)\boldsymbol{\Gamma}_p^{*-1} \approx -\boldsymbol{\Gamma}_p^{*-1}(\hat{\mathbf{C}}_p^* - \boldsymbol{\Gamma}_p^*)\boldsymbol{\Gamma}_p^{*-1}$. Therefore, we have the expansion for the MLE as

$$
\begin{aligned}
\hat{\boldsymbol{\phi}}_{\mathrm{M}} - \boldsymbol{\phi} &\approx [\boldsymbol{\Gamma}_p^{*-1} - \boldsymbol{\Gamma}_p^{*-1}(\hat{\mathbf{C}}_p^* - \boldsymbol{\Gamma}_p^*)\boldsymbol{\Gamma}_p^{*-1}] \\
&\quad \times [(\hat{\mathbf{c}}_p^* - \boldsymbol{\gamma}_p^*) - (\hat{\mathbf{C}}_p^* - \boldsymbol{\Gamma}_p^*)\boldsymbol{\phi} + (\boldsymbol{\gamma}_p^* - \boldsymbol{\Gamma}_p^*\boldsymbol{\phi})].
\end{aligned} \tag{12}
$$

So a representation for the bias will be obtained from

$$
\begin{aligned}
&E(\hat{\boldsymbol{\phi}}_{\mathrm{M}} - \boldsymbol{\phi}) \approx \boldsymbol{\Gamma}_p^{*-1}(\boldsymbol{\gamma}_p^* - \boldsymbol{\Gamma}_p^*\boldsymbol{\phi}) \\
&\quad - \boldsymbol{\Gamma}_p^{*-1}E[(\hat{\mathbf{C}}_p^* - \boldsymbol{\Gamma}_p^*)\boldsymbol{\Gamma}_p^{*-1}[(\hat{\mathbf{c}}_p^* - \boldsymbol{\gamma}_p^*) - (\hat{\mathbf{C}}_p^* - \boldsymbol{\Gamma}_p^*)\boldsymbol{\phi}]].
\end{aligned} \tag{13}
$$

We focus on the first term in the bias representation (13), by obtaining more specific forms for $\boldsymbol{\Gamma}_p^*$ and $\boldsymbol{\gamma}_p^*$. Consider the "residual" vector from GLS regression, $\tilde{\mathbf{N}} = \mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}_{\mathrm{G}} = \mathbf{G}\mathbf{Y}$, where $\tilde{\boldsymbol{\beta}}_{\mathrm{G}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$ is the GLS estimator of $\boldsymbol{\beta}$ and $\mathbf{G} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$. Define $\mathbf{L}_j = (\mathbf{L}^j + \mathbf{L}'^j)/2$, where $\mathbf{L}$ is a $T \times T$ lag matrix with 1's on the first subdiagonal and 0's elsewhere. In addition, define $\mathbf{M}_{11} = \mathbf{I} - \mathbf{e}_1\mathbf{e}_1' - \mathbf{e}_T\mathbf{e}_T'$ and $\mathbf{M}_{ii} = \mathbf{M}_{i-1,i-1} - \mathbf{e}_i\mathbf{e}_i' - \mathbf{e}_{T-i+1}\mathbf{e}_{T-i+1}'$, for $i = 2, \ldots, p$, where $\mathbf{e}_i$ is a $T \times 1$ unit vector with 1 in the ith position and 0's elsewhere, and $2\mathbf{M}_{12} = 2\mathbf{L}_1 - \mathbf{e}_1\mathbf{e}_2' - \mathbf{e}_2\mathbf{e}_1' - \mathbf{e}_{T-1}\mathbf{e}_T' - \mathbf{e}_T\mathbf{e}_{T-1}' = 2\mathbf{M}_{21}$, with other matrices $\mathbf{M}_{ij}, i, j = 1, \ldots, p, i \neq j$, being defined similarly. That is, $\mathbf{M}_{ii}$ has 1's on the main diagonal except for the first and last i elements and 0's elsewhere, and for $i < j, 2\mathbf{M}_{ij}$ has $T - i - j$ 1's on the $(j-i)$th diagonals above and below the main diagonal, excluding the first and last i elements on these diagonals, and 0's elsewhere. Then the quadratic forms $\hat{C}_{ij} \equiv C_{ij}(\tilde{\mathbf{N}})$ may be written as

$$
\hat{C}_{i0} = \mathbf{Y}'\mathbf{G}'\mathbf{L}_i\mathbf{G}\mathbf{Y}, \qquad i = 1, \ldots, p,
$$

and

$$
\hat{C}_{ij} = \mathbf{Y}'\mathbf{G}'\mathbf{M}_{ij}\mathbf{G}\mathbf{Y}, \qquad i, j = 1, \ldots, p.
$$

Using a standard first-moment result for quadratic forms, say $E(\hat{C}_{ij}) = \sigma_\varepsilon^2\,\mathrm{tr}(\mathbf{M}_{ij}\mathbf{G}\mathbf{V}\mathbf{G}') = \sigma_\varepsilon^2[\mathrm{tr}(\mathbf{M}_{ij}\mathbf{V}) - \mathrm{tr}((\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_{ij}\mathbf{X})]$, we obtain

$$
E(\hat{C}_{i0}) = (T-i)\gamma(i) - \sigma_\varepsilon^2 b_i, \qquad i = 1, \ldots, p,
$$

and

$$
E(\hat{C}_{ij}) = (T-i-j)\gamma(i-j) - \sigma_\varepsilon^2 a_{ij}, \qquad i, j = 1, \ldots, p,
$$

where $b_i = \mathrm{tr}[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{L}_i\mathbf{X}], i = 1, \ldots, p$, and $a_{ij} = a_{ji} = \mathrm{tr}[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_{ij}\mathbf{X}], i, j = 1, \ldots, p$. Therefore, we find that

$$
\boldsymbol{\Gamma}_p^* = E(\hat{\mathbf{C}}_p^*) = \boldsymbol{\Gamma}_p - \frac{\sigma_\varepsilon^2}{T}\,\mathbf{A}^* \approx \boldsymbol{\Gamma}_p - \frac{\sigma_\varepsilon^2}{T}\,\mathbf{A} \tag{14}
$$

and

$$
\boldsymbol{\gamma}_p^* = E(\hat{\mathbf{c}}_p^*) = \boldsymbol{\gamma}_p - \frac{\sigma_\varepsilon^2}{T}\,\mathbf{b}^* \approx \boldsymbol{\gamma}_p - \frac{\sigma_\varepsilon^2}{T}\,\mathbf{b}, \tag{15}
$$

where $\boldsymbol{\Gamma}_p = \sigma_\varepsilon^2\mathbf{V}_p, \boldsymbol{\gamma}_p = [\gamma(1), \ldots, \gamma(p)]' = \boldsymbol{\Gamma}_p\boldsymbol{\phi}, \mathbf{A} = \{a_{ij}\}, \mathbf{b} = (b_1, \ldots, b_p)', \mathbf{A}^* \approx \mathbf{A}$ has (i,j)th element equal to $a_{ij}^* = [T/(T-i-j)]a_{ij}$, and $\mathbf{b}^* \approx \mathbf{b}$ has ith element equal to $b_i^* = [T/(T-i)]b_i$.

Define $\boldsymbol{\tau}(\boldsymbol{\phi}) = [\tau_1(\boldsymbol{\phi}), \ldots, \tau_p(\boldsymbol{\phi})]'$, where

$$\tau_j(\boldsymbol{\phi}) = \frac{1}{2}\operatorname{tr}\left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\frac{\partial}{\partial\phi_j}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\right], \quad j = 1, \ldots, p. \tag{16}$$

From the form $\mathbf{V}^{-1} = \mathbf{PP}'$ and $S(\boldsymbol{\beta}, \boldsymbol{\phi}) = \mathbf{N}'\mathbf{V}^{-1}\mathbf{N} = C_{00} - 2\boldsymbol{\phi}'\mathbf{c}_p + \boldsymbol{\phi}'\mathbf{C}_p\boldsymbol{\phi}$, it follows that

$$\mathbf{V}^{-1} = \mathbf{I} + \sum_{i=1}^{p}\sum_{j=1}^{p}\phi_i\phi_j\mathbf{M}_{ij} - 2\sum_{i=1}^{p}\phi_i\mathbf{L}_i,$$

so that $\partial\mathbf{V}^{-1}/\partial\phi_j = 2(\sum_{i=1}^{p}\phi_i\mathbf{M}_{ij} - \mathbf{L}_j), j = 1, \ldots, p$. Then the derivatives in (16) are obtained as

$$\frac{\partial}{\partial\phi_j}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) = 2\left(\sum_{i=1}^{p}\phi_i\mathbf{X}'\mathbf{M}_{ij}\mathbf{X} - \mathbf{X}'\mathbf{L}_j\mathbf{X}\right), \quad j = 1, \ldots, p.$$

Therefore, we see that $\boldsymbol{\tau}(\boldsymbol{\phi}) = \mathbf{A}\boldsymbol{\phi} - \mathbf{b}$, and we also define $\boldsymbol{\tau}^*(\boldsymbol{\phi}) = \mathbf{A}^*\boldsymbol{\phi} - \mathbf{b}^* \approx \boldsymbol{\tau}(\boldsymbol{\phi})$. (An alternate derivation for the form of $\boldsymbol{\tau}(\boldsymbol{\phi})$ related to REML estimation is noted in Section 4.) Hence the first term in the bias expression (13) is

$$\begin{aligned}
&\boldsymbol{\Gamma}_p^{*-1}(\boldsymbol{\gamma}_p^* - \boldsymbol{\Gamma}_p^*\boldsymbol{\phi}) \\
&= \left[\boldsymbol{\Gamma}_p - \frac{\sigma_\varepsilon^2}{T}\mathbf{A}^*\right]^{-1}\left[\boldsymbol{\gamma}_p - \frac{\sigma_\varepsilon^2}{T}\mathbf{b}^* - \left(\boldsymbol{\Gamma}_p - \frac{\sigma_\varepsilon^2}{T}\mathbf{A}^*\right)\boldsymbol{\phi}\right] \\
&= \left[\boldsymbol{\Gamma}_p - \frac{\sigma_\varepsilon^2}{T}\mathbf{A}^*\right]^{-1}\frac{\sigma_\varepsilon^2}{T}(\mathbf{A}^*\boldsymbol{\phi} - \mathbf{b}^*) \\
&= \frac{1}{T}\left[\mathbf{V}_p - \frac{1}{T}\mathbf{A}^*\right]^{-1}\boldsymbol{\tau}^*(\boldsymbol{\phi}) \\
&\approx \frac{1}{T-2}\left[\mathbf{V}_p - \frac{1}{T-2}\mathbf{A}\right]^{-1}\boldsymbol{\tau}(\boldsymbol{\phi}).
\end{aligned} \tag{17}$$

As a further approximation, we could simplify to order $1/T$ terms as $[1/(T-2)]\mathbf{V}_p^{-1}\boldsymbol{\tau}(\boldsymbol{\phi})$.

The second and third terms in the bias expression (13) are obtainable from the covariances between $\hat{\mathbf{C}}_p^*$ and $\hat{\mathbf{c}}_p^*$ and variances of elements of $\hat{\mathbf{C}}_p^*$. Under the Grenander conditions for $\mathbf{X}$, these quantities are "dominated" (to order $1/T$) by the values obtained if $\mathbf{N} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$ were being used to form the elements in the matrices $\hat{\mathbf{C}}_p^*$ and $\hat{\mathbf{c}}_p^*$ instead of $\tilde{\mathbf{N}} = \mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}_{\mathrm{G}}$. For example, the Grenander conditions imply that quantities such as $\mathbf{D}_T^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{D}_T^{-1}$ and $\mathbf{D}_T^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}\mathbf{D}_T^{-1}$ are $O(1)$ as $T \to \infty$, where $\mathbf{D}_T = \operatorname{diag}\{\sqrt{d_{11}(0)}, \ldots, \sqrt{d_{rr}(0)}\}$. We denote the corresponding matrices by $\mathbf{C}_p^*$ and $\mathbf{c}_p^*$. In fact, these "dominating" terms can be determined from previous results of Shaman and Stine (1988) for bias of conditional LS estimators in an AR(p) model with no regression component, or by application of the method of Cordeiro and Klein (1994) on bias of the MLE to an AR(p) model (see the Appendix). Using an approach related to that of Cordeiro and Klein (1994), we find that

$$\begin{aligned}
&\boldsymbol{\Gamma}_p^{*-1}E[(\hat{\mathbf{C}}_p^* - \boldsymbol{\Gamma}_p^*)\boldsymbol{\Gamma}_p^{*-1}[(\hat{\mathbf{c}}_p^* - \boldsymbol{\gamma}_p^*) - (\hat{\mathbf{C}}_p^* - \boldsymbol{\Gamma}_p^*)\boldsymbol{\phi}]] \\
&\approx \boldsymbol{\Gamma}_p^{-1}E[(\mathbf{C}_p^* - \boldsymbol{\Gamma}_p)\boldsymbol{\Gamma}_p^{-1}[(\mathbf{c}_p^* - \boldsymbol{\gamma}_p) - (\mathbf{C}_p^* - \boldsymbol{\Gamma}_p)\boldsymbol{\phi}]] \\
&\approx -\frac{1}{T-2}\sum_{k=1}^{p}\frac{\partial\mathbf{V}_p^{-1}}{\partial\phi_k}\mathbf{e}_k^{(p)},
\end{aligned}$$

where $\mathbf{e}_k^{(p)}$ is a $p \times 1$ unit vector with 1 in the kth position and 0's elsewhere. So, combining with (17), from (13) we obtain

$$\begin{aligned}
E(\hat{\boldsymbol{\phi}}_{\mathrm{M}} - \boldsymbol{\phi}) &= \frac{1}{T-2}\left[\mathbf{V}_p - \frac{1}{T-2}\mathbf{A}\right]^{-1}\boldsymbol{\tau}(\boldsymbol{\phi}) \\
&\quad + \frac{1}{T-2}\sum_{k=1}^{p}\frac{\partial\mathbf{V}_p^{-1}}{\partial\phi_k}\mathbf{e}_k^{(p)} + O\left(\frac{1}{T^2}\right) \qquad (18) \\
&= \frac{1}{T-2}\mathbf{V}_p^{-1}\boldsymbol{\tau}(\boldsymbol{\phi}) + \frac{1}{T-2}\sum_{k=1}^{p}\frac{\partial\mathbf{V}_p^{-1}}{\partial\phi_k}\mathbf{e}_k^{(p)} \\
&\quad + O\left(\frac{1}{T^2}\right). \qquad (18')
\end{aligned}$$

The validity of (18′) can be justified under suitable conditions essentially from the work developed by Bhansali (1981), Tanaka (1984), and Shaman and Stine (1988) for bias approximations in AR models, but without the regression component. The validity can be established under the existence of sufficient moments of the ε_t and the assumption of the Grenander conditions on $\mathbf{X}$. Similar to the results shown in these related works, these conditions ensure the convergence of $TE(\hat{\boldsymbol{\phi}}_{\mathrm{M}} - \boldsymbol{\phi})$ to $\mathbf{V}_p^{-1}\bar{\boldsymbol{\tau}}(\boldsymbol{\phi}) + \boldsymbol{\nu}_p(\boldsymbol{\phi})$, where $\bar{\boldsymbol{\tau}}(\boldsymbol{\phi}) = \lim_{T\to\infty}\boldsymbol{\tau}(\boldsymbol{\phi})$ exists and $\boldsymbol{\nu}_p(\boldsymbol{\phi}) = \sum_{k=1}^{p}(\partial\mathbf{V}_p^{-1}/\partial\phi_k)\mathbf{e}_k^{(p)}$.

Following Parzen (1961) and Galbraith and Galbraith (1974), the (i,j)th element of $\mathbf{V}_p^{-1}$ is

$$v^{ij} = \sum_{k=1}^{\min(i,j)}(\phi_{i-k}\phi_{j-k} - \phi_{p-i+k}\phi_{p-j+k}), \quad i, j = 1, \ldots, p,$$

where $\phi_0 = -1$. The elements v^{ij} of $\mathbf{V}_p^{-1}$ can be directly differentiated to obtain the terms required in (18). As developed in the Appendix, we thus obtain the explicit expression for $\sum_{k=1}^{p}(\partial\mathbf{V}_p^{-1}/\partial\phi_k)\mathbf{e}_k^{(p)}$ given by (A.5) of the Appendix.

For example, for AR(1) noise, we have $\mathbf{V}_1 = (1-\phi^2)^{-1}$, and from (18), we obtain

$$\begin{aligned}
E(\hat{\phi}_{\mathrm{M}} - \phi) &\approx \frac{1}{T-2}\left[\frac{1}{1-\phi^2} - \frac{A}{T-2}\right]^{-1}\tau(\phi) - \frac{2\phi}{T-2} \\
&= \frac{(1-\phi^2)\tau(\phi)}{T-2-(1-\phi^2)A} - \frac{2\phi}{T-2},
\end{aligned}$$

with $\mathbf{V}^{-1} = \mathbf{I} + \phi^2\mathbf{M}_{11} - 2\phi\mathbf{L}_1, \tau(\phi) = A\phi - b, A = a_{11} = \operatorname{tr}[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_{11}\mathbf{X}]$, and $b = b_1 = \operatorname{tr}[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{L}_1\mathbf{X}]$. Therefore, in the more common situation where $\phi > 0$, the magnitude of ML bias will be increased by the presence of a regression component with

$\tau(\phi) < 0$, which typically might hold when regressors have smooth trend-like behavior. Next, consider the example of an AR(2) model ($p = 2$). By inverting $\mathbf{V}_2$ directly, we obtain

$$\mathbf{V}_2^{-1} = \frac{\Delta}{1-\rho_1^2}\begin{bmatrix} 1 & -\rho_1 \\ -\rho_1 & 1 \end{bmatrix} = \begin{bmatrix} 1-\phi_2^2 & -\phi_1(1+\phi_2) \\ -\phi_1(1+\phi_2) & 1-\phi_2^2 \end{bmatrix},$$

where we have used the relation $\Delta/(1-\rho_1^2) = 1-\phi_2^2$ for an AR(2). Therefore,

$$\sum_{k=1}^{2} \frac{\partial \mathbf{V}_2^{-1}}{\partial \phi_k}\mathbf{e}_k^{(2)} = \begin{bmatrix} 0 \\ -(1+\phi_2) \end{bmatrix} + \begin{bmatrix} -\phi_1 \\ -2\phi_2 \end{bmatrix} = -\begin{bmatrix} \phi_1 \\ 1+3\phi_2 \end{bmatrix}.$$

Thus the bias of the MLE of $\boldsymbol{\phi}$, to order $1/T$ terms using (18′), is

$$\begin{aligned} E(\hat{\boldsymbol{\phi}}_{\mathrm{M}} - \boldsymbol{\phi}) &= -\frac{1}{T-2} \\ &\times \begin{bmatrix} \phi_1 - (1-\phi_2^2)\tau_1(\boldsymbol{\phi}) + \phi_1(1+\phi_2)\tau_2(\boldsymbol{\phi}) \\ 1+3\phi_2 - (1-\phi_2^2)\tau_2(\boldsymbol{\phi}) + \phi_1(1+\phi_2)\tau_1(\boldsymbol{\phi}) \end{bmatrix} \\ &+ O\left(\frac{1}{T^2}\right). \end{aligned} \tag{19}$$

In addition, consider the special case of polynomial regression (of degree $r-1$) with AR(p) noise. It then follows (e.g., Anderson 1971, sec. 10.2.3) that we have the approximation $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \approx (1/g(0;\boldsymbol{\phi}))\mathbf{X}'\mathbf{X}$, where $g(0;\boldsymbol{\phi}) = |\phi(1)|^{-2} = (1-\sum_{j=1}^{p}\phi_j)^{-2}$, which is $2\pi/\sigma_\varepsilon^2$ times the spectral density function $f(\lambda) = (\sigma_\varepsilon^2/2\pi)g(\lambda;\boldsymbol{\phi}) = (\sigma_\varepsilon^2/2\pi)|\phi(e^{-i\lambda})|^{-2}$ of the AR(p) process evaluated at $\lambda = 0$. Using this approximation, we have $\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| \approx r\log[g(0;\boldsymbol{\phi})^{-1}] + \log|\mathbf{X}'\mathbf{X}|$. Thus for $j = 1,\ldots,p$, $\tau_j(\boldsymbol{\phi}) = \frac{1}{2}(\partial/\partial\phi_j)\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| \approx -r(1-\sum_{k=1}^{p}\phi_k)^{-1}$, so that $\boldsymbol{\tau}(\boldsymbol{\phi}) \approx -r(1-\sum_{k=1}^{p}\phi_k)^{-1}\mathbf{1}_p$. From lemma 1 of Shaman and Stine (1988), we have

$$\sum_{j=1}^{p} v^{ij} = \left(1-\sum_{j=1}^{p}\phi_j\right)\sum_{j=0}^{i-1}(\phi_{p-j}-\phi_j), \qquad i = 1,\ldots,p,$$

with $\phi_0 = -1$. The kth element of $\mathbf{V}_p^{-1}\boldsymbol{\tau}(\boldsymbol{\phi}) \approx -r(1-\phi_1-\cdots-\phi_p)^{-1}\mathbf{V}_p^{-1}\mathbf{1}_p$ is then given approximately by $-r\sum_{j=0}^{k-1}(\phi_{p-j}-\phi_j)$. Therefore, for polynomial regressors, the ML bias for ϕ_k given by (18′) may be further approximated by

$$E(\hat{\phi}_{k\mathrm{M}} - \phi_k) = \frac{1}{T-2}\,\mathbf{e}_k^{(p)\prime}\sum_{j=1}^{p}\frac{\partial \mathbf{V}_p^{-1}}{\partial \phi_j}\,\mathbf{e}_j^{(p)} - \frac{r}{T-2}\sum_{j=0}^{k-1}(\phi_{p-j}-\phi_j) + O\left(\frac{1}{T^2}\right). \tag{20}$$

For example, in the AR(1) noise model ($p = 1$), we have $\mathbf{V}_1^{-1}\boldsymbol{\tau}(\phi) = -r(1-\phi^2)(1-\phi)^{-1} = -r(1+\phi)$, and the ML bias approximation (20) gives $E(\hat{\phi}_{\mathrm{M}}-\phi) \approx -[2\phi/(T-2)] - [r(1+\phi)/(T-2)] = -[2\phi+r(1+\phi)]/(T-2)$. When $p = 2$, we get $\mathbf{V}_2^{-1}\boldsymbol{\tau}(\boldsymbol{\phi}) \approx -r(1-\phi_1-\phi_2)^{-1}\mathbf{V}_2^{-1}\mathbf{1}_2 = -r(1+\phi_2)\mathbf{1}_2$, and the ML bias (19) in this case reduces to

$$E(\hat{\boldsymbol{\phi}}_{\mathrm{M}} - \boldsymbol{\phi}) = -\frac{1}{T-2}\begin{bmatrix} \phi_1 + r(1+\phi_2) \\ 1+3\phi_2+r(1+\phi_2) \end{bmatrix} + O\left(\frac{1}{T^2}\right). \tag{21}$$

When $\mathbf{X} = \mathbf{1}_T$, so that there is only a constant term with $r = 1$, these results are in agreement with ML bias results for the special cases of the AR(1) and AR(2) models as given by Tanaka (1984) and Cordeiro and Klein (1994). The result (20) for the bias of the MLE in the AR(p) model with only a constant term also coincides with the bias of the conditional LS estimator given by Shaman and Stine (1988).

We also mention the relation of approximation (18) to a few other results in the literature. First, as indicated, the bias of the MLE $\hat{\boldsymbol{\phi}}_{\mathrm{M}}$, to order $1/T$ can be simplified as in (18′). It can be shown that this is the approximation obtained by extension of the approach of Cordeiro and Klein (1994) to a regression model with AR(p) noise. In their work, a general formula for the first-order bias of the MLE in an ARMA model without any regression component was derived. In the special case of AR(1) noise, the use of an approximate form of $\boldsymbol{\tau}(\boldsymbol{\phi})$ in (18′) leads to a further bias representation that was presented by Magee (1989). Specifically, for this case Magee (1989) used $\tau(\phi) = (1/2)\mathrm{tr}[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\partial/\partial\phi)(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})] \approx (1/2\phi)[r-(1-\phi^2)\mathrm{tr}[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}]]$ for $\phi \neq 0$, and with this approximation obtained $E(\hat{\phi}_{\mathrm{M}}-\phi) = -(2\phi/T) + [(1-\phi^2)/2\phi T][r-(1-\phi^2)\mathrm{tr}[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}]] + O(1/T^2)$. However, numerical investigations with AR(1) noise indicate that these approximate forms are not sufficiently accurate compared to (18), especially for moderately large values of ϕ (e.g., $\phi \geq .7$).

4. RESTRICTED MAXIMUM LIKELIHOOD ESTIMATION

Using the representation derived by Harville (1974), the restricted log-likelihood function for $\mathbf{Y}$ based on the regression model (4) is

$$l^*(\boldsymbol{\phi},\sigma_\varepsilon^2) = -\frac{T}{2}\log(2\pi) - \frac{T-r}{2}\log(\sigma_\varepsilon^2) - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2\sigma_\varepsilon^2}S(\tilde{\boldsymbol{\beta}}_{\mathrm{G}},\boldsymbol{\phi}) - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|,$$

where $r = \text{rank}(\mathbf{X})$ and $\tilde{\beta}_{\text{G}} \equiv \tilde{\beta}_{\text{G}}(\phi) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$. For σ_ε^2, the restricted likelihood equation is $\partial l^*/\partial\sigma_\varepsilon^2 = -(T-r)/(2\sigma_\varepsilon^2) + S/(2\sigma_\varepsilon^4) = 0$. Comparing to the likelihood equations (5), the restricted likelihood equations for ϕ are

$$\frac{\partial l^*}{\partial\phi} = -\frac{1}{2}\frac{\partial}{\partial\phi}\log|\mathbf{V}| - \frac{1}{\sigma_\varepsilon^2}(\hat{\mathbf{C}}_p\phi - \hat{\mathbf{c}}_p) - \frac{1}{2}\frac{\partial}{\partial\phi}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| = \mathbf{0}, \quad (22)$$

where the $\hat{C}_{ij}$ in $\hat{\mathbf{C}}_p$ and $\hat{\mathbf{c}}_p$ are now defined in terms of the $\tilde{N}_t = Y_t - \mathbf{x}_t'\tilde{\beta}_{\text{G}}, t = 1, \ldots, T$. The REML estimators $(\hat{\beta}_{\text{R}}, \hat{\phi}_{\text{R}}, \tilde{\sigma}_\varepsilon^2)$ are

$$\hat{\beta}_{\text{R}} = (\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{Y}, \qquad \tilde{\sigma}_\varepsilon^2 = \frac{1}{T-r}S(\hat{\beta}_{\text{R}}, \hat{\phi}_{\text{R}}),$$

where $\tilde{\mathbf{V}} = \mathbf{V}(\hat{\phi}_{\text{R}})$, and $\hat{\phi}_{\text{R}}$ satisfies (22). The REML estimator of σ_ε^2 takes into account the loss in degrees of freedom that results from estimating the regression parameters.

For the "correction" term in (22), it is known (e.g., Searle 1982, p. 337) that

$$\frac{\partial}{\partial\phi_j}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| = \text{tr}\left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\frac{\partial}{\partial\phi_j}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\right] \equiv 2\tau_j(\phi),$$

for $j = 1, \ldots, p$. Recall that the form of $\boldsymbol{\tau}(\phi)$ defined by (16) has been established in Section 3.2 as $\boldsymbol{\tau}(\phi) = \mathbf{A}\phi - \mathbf{b}$. Equivalently, note that this form can also be readily deduced from the equivalent representation of $\boldsymbol{\tau}(\phi)$ in the restricted likelihood equations (22)—that is, $\boldsymbol{\tau}(\phi) = \frac{1}{2}(\partial/\partial\phi)\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|$—using the fact that $E(\partial l^*/\partial\phi) = \mathbf{0}$ and results on $E(\hat{\mathbf{C}}_p)$ and $E(\hat{\mathbf{c}}_p)$ from Section 3.2. This also provides an interpretation for the "preventive" nature of the REML approach in bias reduction relative to ML. That is, the bias of the MLE $\hat{\phi}_{\text{M}}$ can be due mainly to the "bias" of the efficient score $\partial l/\partial\phi$ in (5) when β is replaced by its GLS estimator $\tilde{\beta}_{\text{G}}$, with $E[\partial l/\partial\phi|_{(\tilde{\beta}_{\text{G}},\phi)}] = \boldsymbol{\tau}(\phi)$. This bias is corrected for in the REML equation (22), which provides an unbiased estimating equation for ϕ.

4.1 Approximate Restricted Maximum Likelihood Estimator

In Section 3.1 we gave an approximate form $\hat{\phi}_{\text{M}} = \hat{\mathbf{C}}_p^{*-1}\hat{\mathbf{c}}_p^*$ of the MLE. Here we derive an approximate representation of the REML estimator in terms of $\hat{\phi}_{\text{M}}$. From (10) and (22), the restricted likelihood equations can be expressed approximately as

$$\hat{C}_{i0} - \frac{T-i}{T-i-1}\hat{C}_{i1}\phi_1 - \cdots - \frac{T-i}{T-i-p}\hat{C}_{ip}\phi_p - \sigma_\varepsilon^2(a_{i1}\phi_1 + \cdots + a_{ip}\phi_p - b_i) = 0,$$

for $i = 1, \ldots, p$. Equivalently, in matrix form we have $\hat{\mathbf{c}}_p^* - \hat{\mathbf{C}}_p^*\phi - (\sigma_\varepsilon^2/T)(\mathbf{A}^*\phi - \mathbf{b}^*) = \mathbf{0}$, where now $\mathbf{A}^*$ is defined such that the (i,j)th element is equal to $[T/(T-i)]a_{ij}$ and $\mathbf{b}^*$ is defined as previously in Section 3.2. Evaluating $\mathbf{A}^*, \mathbf{b}^*$, and σ_ε^2 at $\hat{\phi}_{\text{M}}$, denoting as $\hat{\mathbf{A}}^*, \hat{\mathbf{b}}^*$, and $\hat{\sigma}_\varepsilon^2 = S(\hat{\beta}, \hat{\phi}_{\text{M}})/(T-r)$, the restricted likelihood equations are then approximated by linear equations in ϕ, $(\hat{\mathbf{C}}_p^* + (\hat{\sigma}_\varepsilon^2/T)\hat{\mathbf{A}}^*)\phi = \hat{\mathbf{c}}_p^* + (\hat{\sigma}_\varepsilon^2/T)\hat{\mathbf{b}}^*$. So an approximate representation for the REML estimator of ϕ is obtained as

$$\hat{\phi}_{\text{R}} = \left(\hat{\mathbf{C}}_p^* + \frac{\hat{\sigma}_\varepsilon^2}{T}\hat{\mathbf{A}}^*\right)^{-1}\left(\hat{\mathbf{c}}_p^* + \frac{\hat{\sigma}_\varepsilon^2}{T}\hat{\mathbf{b}}^*\right) \equiv \hat{\mathbf{D}}_p^{*-1}\hat{\mathbf{d}}_p^*, \quad (23)$$

where $\hat{\mathbf{D}}_p^* = \hat{\mathbf{C}}_p^* + (\hat{\sigma}_\varepsilon^2/T)\hat{\mathbf{A}}^*$ and $\hat{\mathbf{d}}_p^* = \hat{\mathbf{c}}_p^* + (\hat{\sigma}_\varepsilon^2/T)\hat{\mathbf{b}}^*$. We can express $\hat{\phi}_{\text{R}}$ in relation to the approximation $\hat{\phi}_{\text{M}} = \hat{\mathbf{C}}_p^{*-1}\hat{\mathbf{c}}_p^*$ for the MLE as

$$\begin{aligned}\hat{\phi}_{\text{R}} &= \left(\hat{\mathbf{C}}_p^* + \frac{\hat{\sigma}_\varepsilon^2}{T}\hat{\mathbf{A}}^*\right)^{-1}\left(\hat{\mathbf{C}}_p^*\hat{\phi}_{\text{M}} + \frac{\hat{\sigma}_\varepsilon^2}{T}\hat{\mathbf{b}}^*\right)\\ &= \hat{\phi}_{\text{M}} - \left(\hat{\mathbf{C}}_p^* + \frac{\hat{\sigma}_\varepsilon^2}{T}\hat{\mathbf{A}}^*\right)^{-1}\frac{\hat{\sigma}_\varepsilon^2}{T}(\hat{\mathbf{A}}^*\hat{\phi}_{\text{M}} - \hat{\mathbf{b}}^*). \quad (24)\end{aligned}$$

4.2 Bias of the Restricted Maximum Likelihood Estimator

We can use the expression (24) for $\hat{\phi}_{\text{R}}$ in terms of $\hat{\phi}_{\text{M}}$, and the bias expression (18) for $\hat{\phi}_{\text{M}}$, to obtain a bias approximation for the REML estimator $\hat{\phi}_{\text{R}}$. Alternatively, we can consider the bias approximation of $\hat{\phi}_{\text{R}}$ directly from the representation (23), in the same manner as for $\hat{\phi}_{\text{M}}$, by using an expansion similar to (12) and obtaining expectations of $\hat{\mathbf{D}}_p^*$ and $\hat{\mathbf{d}}_p^*$. Note, in fact, from (14)–(15), that $E(\hat{\mathbf{D}}_p^*) \approx E(\hat{\mathbf{C}}_p^*) + (\sigma_\varepsilon^2/T)\mathbf{A}^* \approx \boldsymbol{\Gamma}_p$ and $E(\hat{\mathbf{d}}_p^*) \approx E(\hat{\mathbf{c}}_p^*) + (\sigma_\varepsilon^2/T)\mathbf{b}^* = \boldsymbol{\gamma}_p$ to order $1/T$ terms. From (23) and using the approximation $\hat{\mathbf{D}}_p^{*-1} - \boldsymbol{\Gamma}_p^{-1} = -\hat{\mathbf{D}}_p^{*-1}(\hat{\mathbf{D}}_p^* - \boldsymbol{\Gamma}_p)\boldsymbol{\Gamma}_p^{-1} \approx -\boldsymbol{\Gamma}_p^{-1}(\hat{\mathbf{D}}_p^* - \boldsymbol{\Gamma}_p)\boldsymbol{\Gamma}_p^{-1}$, we have the expansion for the REML estimator as

$$\begin{aligned}\hat{\phi}_{\text{R}} - \phi &= \hat{\mathbf{D}}_p^{*-1}[\hat{\mathbf{d}}_p^* - \hat{\mathbf{D}}_p^*\phi]\\ &\approx [\boldsymbol{\Gamma}_p^{-1} - \boldsymbol{\Gamma}_p^{-1}(\hat{\mathbf{D}}_p^* - \boldsymbol{\Gamma}_p)\boldsymbol{\Gamma}_p^{-1}]\\ &\quad\times[(\hat{\mathbf{d}}_p^* - \boldsymbol{\gamma}_p) - (\hat{\mathbf{D}}_p^* - \boldsymbol{\Gamma}_p)\phi], \quad (25)\end{aligned}$$

noting that $\boldsymbol{\Gamma}_p\phi - \boldsymbol{\gamma}_p = \mathbf{0}$. Hence a bias approximation for the REML estimator will be obtained, similar to (13), from

$$\begin{aligned}&E(\hat{\phi}_{\text{R}} - \phi)\\ &\approx \boldsymbol{\Gamma}_p^{-1}E(\hat{\mathbf{d}}_p^* - \hat{\mathbf{D}}_p^*\phi) - \boldsymbol{\Gamma}_p^{-1}\\ &\quad\times E[(\hat{\mathbf{D}}_p^* - \boldsymbol{\Gamma}_p)\boldsymbol{\Gamma}_p^{-1}[(\hat{\mathbf{d}}_p^* - \boldsymbol{\gamma}_p) - (\hat{\mathbf{D}}_p^* - \boldsymbol{\Gamma}_p)\phi]]\\ &\approx -\boldsymbol{\Gamma}_p^{-1}E[(\mathbf{C}_p^* - \boldsymbol{\Gamma}_p)\boldsymbol{\Gamma}_p^{-1}[(\mathbf{c}_p^* - \boldsymbol{\gamma}_p) - (\mathbf{C}_p^* - \boldsymbol{\Gamma}_p)\phi]]\\ &\approx \frac{1}{T-2}\sum_{k=1}^{p}\frac{\partial\mathbf{V}_p^{-1}}{\partial\phi_k}\mathbf{e}_k^{(p)}. \quad (26)\end{aligned}$$

Essentially, the expectation terms in (26) are "dominated" by variance and covariance terms of $\mathbf{C}_p^*$ and $\mathbf{c}_p^*$ and behave similar to the second and third terms in the expansion (13) for $E(\hat{\phi}_{\text{M}} - \phi)$. Thus the first bias term from $\hat{\phi}_{\text{M}}$ due to the regression component is eliminated for $\hat{\phi}_{\rm R}$. For example, in the AR(1) model the approximate REML bias expression (26) reduces simply to $E(\hat{\phi}_{\rm R}-\phi)\approx -2\phi/(T-2)$. For polynomial regression, comparison with the "simplified" ML bias approximation $E(\hat{\phi}_{\rm M}-\phi)\approx -[2\phi+r(1+\phi)]/(T-2)$ shows that REML biases are smaller in magnitude than ML biases when $-r/(r+4)<\phi<1$.

5. BIAS OF MAXIMUM LIKELIHOOD ESTIMATOR OF ERROR VARIANCE

We also consider the bias of the MLE $\hat{\sigma}_\varepsilon^2 = S(\hat{\boldsymbol\beta}_{\rm M},\hat{\boldsymbol\phi}_{\rm M})/T$ of the error variance $\sigma_\varepsilon^2=\mathrm{var}(\varepsilon_t)$. We consider the asymptotically equivalent version with

$$\begin{aligned} S &= S(\tilde{\boldsymbol\beta}_{\rm G},\hat{\boldsymbol\phi}_{\rm M})\\ &= (\mathbf{Y}-\mathbf{X}\tilde{\boldsymbol\beta}_{\rm G})'\hat{\mathbf{V}}^{-1}(\mathbf{Y}-\mathbf{X}\tilde{\boldsymbol\beta}_{\rm G}) = \mathbf{Y}'\mathbf{G}'\hat{\mathbf{V}}^{-1}\mathbf{G}\mathbf{Y}, \end{aligned}$$

where $\hat{\mathbf{V}}^{-1} = \mathbf{V}(\hat{\boldsymbol\phi}_{\rm M})^{-1}$, and recall that $\mathbf{G}=\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$. To obtain an approximate expected value of S to order $O(1)$ terms, we use the representation

$$\begin{aligned} \hat{\mathbf{V}}^{-1} = \mathbf{V}^{-1} &+ \sum_{i=1}^{p}\frac{\partial\mathbf{V}^{-1}}{\partial\phi_i}(\hat{\phi}_{i\rm M}-\phi_i)\\ &+\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}\frac{\partial^2\mathbf{V}^{-1}}{\partial\phi_i\partial\phi_j}(\hat{\phi}_{i\rm M}-\phi_i)(\hat{\phi}_{j\rm M}-\phi_j), \end{aligned}\tag{27}$$

with $\mathbf{V}^{-1}$ a quadratic function of the ϕ_i as given in Section 3.2. From the form $\partial\mathbf{V}^{-1}/\partial\phi_j = 2(\sum_{i=1}^{p}\phi_i\mathbf{M}_{ij}-\mathbf{L}_j), j=1,\ldots,p$, it follows that $\mathbf{Y}'\mathbf{G}'(\partial\mathbf{V}^{-1}/\partial\phi_j)\mathbf{G}\mathbf{Y}$ is equal to the jth element of $2(\hat{\mathbf{C}}_p\boldsymbol\phi-\hat{\mathbf{c}}_p)$. Therefore,

$$\begin{aligned} \sum_{i=1}^{p}\mathbf{Y}'\mathbf{G}'\frac{\partial\mathbf{V}^{-1}}{\partial\phi_i}\mathbf{G}\mathbf{Y}(\hat{\phi}_{i\rm M}-\phi_i) &= 2(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)'(\hat{\mathbf{C}}_p\boldsymbol\phi-\hat{\mathbf{c}}_p)\\ &= -2(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)'\hat{\mathbf{C}}_p(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)\\ &\quad + 2(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)'(\hat{\mathbf{C}}_p-\tilde{\mathbf{C}}_p)\hat{\boldsymbol\phi}_{\rm M}, \end{aligned}$$

noting from (10) that $\hat{\boldsymbol\phi}_{\rm M}=\tilde{\mathbf{C}}_p^{-1}\hat{\mathbf{c}}_p$, where $\tilde{\mathbf{C}}_p\approx\hat{\mathbf{C}}_p$ has $[(T-i)/(T-i-j)]\hat{C}_{ij}\approx\hat{C}_{ij}$ in the (i,j)th position. In addition, we know that $\mathbf{Y}'\mathbf{G}'(\partial^2\mathbf{V}^{-1}/\partial\phi_i\partial\phi_j)\mathbf{G}\mathbf{Y}=2\hat{C}_{ij}$, so that

$$\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}\mathbf{Y}'\mathbf{G}'\frac{\partial^2\mathbf{V}^{-1}}{\partial\phi_i\partial\phi_j}\mathbf{G}\mathbf{Y}(\hat{\phi}_{i\rm M}-\phi_i)(\hat{\phi}_{j\rm M}-\phi_j) = (\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)'\hat{\mathbf{C}}_p(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi).$$

Hence we see that the residual sum of squares S has the asymptotic representation as

$$\begin{aligned} S &= \mathbf{Y}'\mathbf{G}'\hat{\mathbf{V}}^{-1}\mathbf{G}\mathbf{Y}\\ &= \mathbf{Y}'\mathbf{G}'\mathbf{V}^{-1}\mathbf{G}\mathbf{Y} - 2(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)'\hat{\mathbf{C}}_p(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)\\ &\quad + 2(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)'(\hat{\mathbf{C}}_p-\tilde{\mathbf{C}}_p)\hat{\boldsymbol\phi}_{\rm M}\\ &\quad + (\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)'\hat{\mathbf{C}}_p(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)\\ &= \mathbf{Y}'\mathbf{G}'\mathbf{V}^{-1}\mathbf{G}\mathbf{Y} - (\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)'\hat{\mathbf{C}}_p(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)\\ &\quad + 2(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)'(\hat{\mathbf{C}}_p-\tilde{\mathbf{C}}_p)\hat{\boldsymbol\phi}_{\rm M}\\ &\approx \mathbf{Y}'\mathbf{G}'\mathbf{V}^{-1}\mathbf{G}\mathbf{Y} - (\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)'\hat{\mathbf{C}}_p(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi). \end{aligned}\tag{28}$$

Then, because $E[(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)']\approx[\sigma_\varepsilon^2/(T-2)]\boldsymbol\Gamma_p^{-1}$ and $E(\hat{\mathbf{C}}_p)\approx(T-2)\boldsymbol\Gamma_p-\sigma_\varepsilon^2\mathbf{A}=(T-2)\boldsymbol\Gamma_p+O(1)$, we obtain

$$\begin{aligned} E(S) &\approx E(\mathbf{Y}'\mathbf{G}'\mathbf{V}^{-1}\mathbf{G}\mathbf{Y}) - E[(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)'\hat{\mathbf{C}}_p(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)]\\ &\approx \sigma_\varepsilon^2\mathrm{tr}(\mathbf{V}^{-1}\mathbf{G}\mathbf{V}\mathbf{G}')\\ &\quad - \mathrm{tr}\{E(\hat{\mathbf{C}}_p)E[(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)(\hat{\boldsymbol\phi}_{\rm M}-\boldsymbol\phi)']\}+O(1/T)\\ &= \sigma_\varepsilon^2(T-r)-\sigma_\varepsilon^2 p+O(1/T)\\ &= \sigma_\varepsilon^2(T-r-p)+O(1/T). \end{aligned}$$

Therefore, we find that

$$\begin{aligned} E(\hat{\sigma}_\varepsilon^2) &= \frac{1}{T}E(S)\\ &= \frac{T-r-p}{T}\sigma_\varepsilon^2+O\left(\frac{1}{T^2}\right)\\ &= \sigma_\varepsilon^2-\frac{r+p}{T}\sigma_\varepsilon^2+O\left(\frac{1}{T^2}\right). \end{aligned}\tag{29}$$

This suggests $\bar{\sigma}_\varepsilon^2=\{T/[T-(r+p)]\}\hat{\sigma}_\varepsilon^2\equiv\{1/[T-(r+p)]\}S(\hat{\boldsymbol\beta}_{\rm M},\hat{\boldsymbol\phi}_{\rm M})$ as an approximately unbiased estimator of σ_ε^2. This form is consistent with results stated by Tanaka (1984), Cordeiro and Klein (1994), and Mentz, Morettin, and Toloi (1998) for the asymptotic bias of the MLE of σ_ε^2 for an AR(p) model with unknown mean ($r=1$), $E(\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2)=-[(p+1)/T]\sigma_\varepsilon^2+O(1/T^2)$.

6. NUMERICAL RESULTS FROM THEORY AND SIMULATION

The time series regression model (1) is useful in modeling processes in various fields, including economics and geophysics. Using (18) and (26), we may assess the bias magnitude of the ML and REML estimates of the AR parameters for some typical models. For example, in trend analysis of geophysical variables such as total ozone and ultraviolet radiation, a regression model with seasonal and trend components may be appropriate for monthly seasonal time series data $\{Y_t\}$. Consider a regression trend model with an AR noise process for such seasonal series of the form

$$Y_t=\mu+S_t+\omega X_t+N_t,\qquad t=1,\ldots,T,\tag{30}$$

where μ is a constant level term, $S_t=\sum_{j=1}^{4}[\beta_{1j}\cos(2\pi jt/12)+\beta_{2j}\sin(2\pi jt/12)]$ is a seasonal component comprising sinusoidal terms of fundamental period 12 months and their harmonics, and ωX_t with $X_t=t/12$ represents a linear trend. For the full model, we have $r=10$. When S_t is excluded from the model for nonseasonal data series, we have $r=2$.

6.1 Theoretical Biases of Maximum Likelihood and Restricted Maximum Likelihood Estimators

For an AR(1) noise model, $(1-\phi B)N_t=\varepsilon_t$, Table 1 presents the theoretical first-order bias approximations of MLEs and REML estimators for the trend model (30), for

Table 1. Theoretical First-order Bias Approximations (Minus Sign Omitted) of AR(1) Parameter Estimators $\hat{\phi}_M$ and $\hat{\phi}_R$ Given by (18) and (26) for the Trend Model (30)

	ϕ									
T	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95
	ML model without seasonal component (r = 2)									
60	.041	.049	.056	.063	.070	.078	.087	.096	.111	.125
120	.020	.024	.027	.031	.034	.038	.042	.046	.051	.057
	ML model with seasonal component (r = 10)									
60	.077	.079	.079	.080	.080	.080	.081	.085	.099	.116
120	.035	.035	.036	.036	.036	.036	.037	.039	.046	.053
	REML (independent of r)									
60	.003	.007	.010	.014	.017	.021	.024	.028	.031	.033
120	.002	.003	.005	.007	.008	.010	.012	.014	.015	.016

$.1 \leq \phi \leq .95$ and $T = 60$ and 120. Only values $\phi > 0$ are considered, because these are the most commonly encountered values in practice. For comparison, Table 1 also gives the biases when the seasonal component S_t is excluded from the model. The results indicate that the REML method clearly provides estimates of ϕ that are less biased than the ML estimates. For example, for the trend model (30) with seasonal component ($r = 10$) and a sample size of $T = 60$, the reduction in bias with the use of REML is about .065 on average for $.1 \leq \phi \leq .9$. Note that from $\text{var}(\hat{\phi}_\text{M}) \approx (1-\phi^2)/(T-2)$, the asymptotic standard deviation of the MLE ranges from .13 to .06 for $.1 \leq \phi \leq .9$. Furthermore, the bias of the REML estimate is independent of r to order $1/T$. Although the ML bias due to the seasonal component is not too substantial, including S_t results in a mild increase in magnitude of bias when $0 < \phi \leq .5$ and a slight decrease when $.6 < \phi < 1$.

For the AR(2) noise model, Figure 1 shows contour plots of the theoretical approximate biases, given by (18), of the MLE $\hat{\phi}_\text{M}$ for the trend model (30) with $T = 60$, as well as contour plots of the biases when the seasonal component S_t is excluded from the model. In Figure 1, the triangle gives the boundary for the stationarity region of an AR(2). The quadratic region defined by $\phi_1^2 + 4\phi_2 < 0$ gives complex roots for the characteristic equation $m^2 - \phi_1 m - \phi_2 = 0$.

Figure 1 shows that without the seasonal component ($r = 2$), the approximation (18) for the bias of $\hat{\phi}_{1\text{M}}$ gives nearly linear contours with slope close to $-1/2$, except for ϕ approaching the boundary $\phi_1 + \phi_2 = 1$. This feature is consistent with the linear contours given by the "simplified" bias approximation (21) for the special case of polynomial regression, which yields $-(2+\phi_1+2\phi_2)/(T-2)$ for $\hat{\phi}_{1\text{M}}$. For the bias of $\hat{\phi}_{2\text{M}}$, the approximation (18) results in contours that are mostly parallel to the ϕ_1-axis. Because the "simplified" bias approximation $-(3+5\phi_2)/(T-2)$ for $\hat{\phi}_{2\text{M}}$ does not depend on ϕ_1, it also gives horizontal contours.

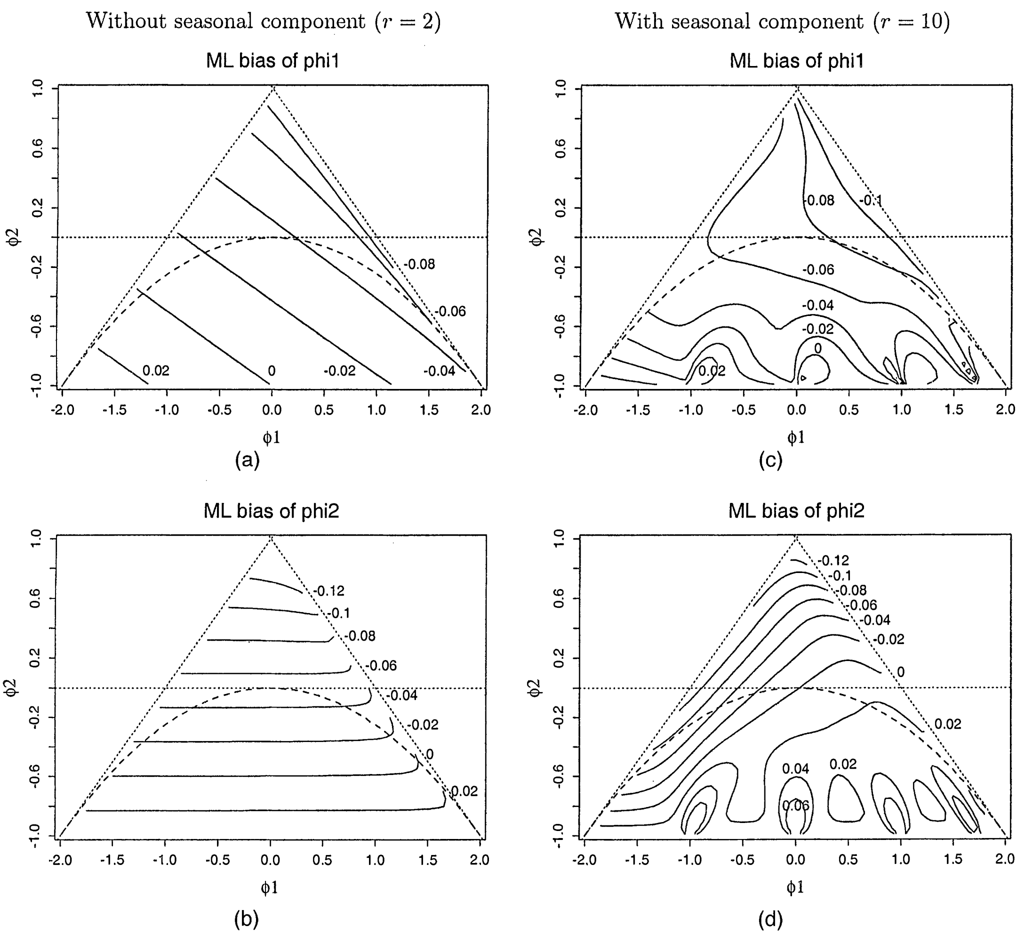

Figure 1. Theoretical Approximate Biases of MLE $\hat{\phi}_M$ of AR(2) Parameters, Given by (18), for the Trend Model (30) and T = 60. (a) Without seasonal component (r = 2), ML bias of phi1; (b) Without seasonal component (r = 2), ML bias of phi2; (c) with seasonal component (r = 10), ML bias of phi1; (d) with seasonal component (r = 10), ML bias of phi2.

Table 2. Empirical Biases (Minus Sign Omitted) of AR(1) Parameter Estimates $\hat{\phi}$ for Trend Model (30), Based on ML and REML (1,000 Replications)

T	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95
ML model without seasonal component (r = 2)										
60	.037	.044	.051	.058	.065	.073	.082	.094	.110	.125
120	.019	.023	.027	.031	.034	.038	.041	.045	.051	.056
ML model with seasonal component (r = 10)										
60	.072	.074	.075	.076	.077	.079	.081	.086	.100	.115
120	.034	.035	.036	.036	.037	.037	.038	.040	.045	.052
REML model without seasonal component (r = 2)										
60	.000	.003	.006	.010	.013	.017	.022	.027	.036	.048
120	.001	.003	.005	.007	.009	.011	.012	.014	.016	.018
REML model with seasonal component (r = 10)										
60	.000	.004	.008	.012	.017	.021	.025	.029	.038	.050
120	.001	.003	.005	.007	.010	.012	.013	.015	.017	.018

When the seasonal component in included ($r = 10$), Figure 1 shows that the linear feature in the contours of the ML bias is "distorted." Local maxima emerge near the nonstationary boundary $\phi_2 = -1$. Notice that these local maxima near $\phi_2 = -1$ occur around values for $\phi_1 \equiv 2\cos(\theta)$ such that the roots of $m^2 - \phi_1 m - \phi_2 = 0$ are complex with θ equal to $\pi j/6, j = 1, \ldots, 4$, the frequencies of the seasonal component. The impact of S_t on the ML bias can be substantial. For example, when $\boldsymbol{\phi} = (.5, .25)'$, the biases of $\hat{\boldsymbol{\phi}}_{\mathrm{M}}$ are $(-.061, -.074)'$ and $(-.098, -.007)'$ for $r = 2$ and $r = 10$. In contrast, without or with the seasonal component, the REML biases for $\hat{\phi}_{1\mathrm{R}}$ and $\hat{\phi}_{2\mathrm{R}}$ from (26) give $E(\hat{\phi}_{1\mathrm{R}} - \phi_1) \approx -\phi_1/(T-2)$ and $E(\hat{\phi}_{2\mathrm{R}} - \phi_2) \approx -(1 + 3\phi_2)/(T-2)$, so the contours of the REML biases are parallel to the ϕ_2-axis and the ϕ_1-axis. These REML biases are generally smaller in magnitude than the ML biases (e.g., when $\phi_1 > 0$ for ϕ_1 and $\phi_2 > -.5$ for ϕ_2).

6.2 Empirical Biases of Maximum Likelihood and Restricted Maximum Likelihood Estimators

To assess the adequacy of the bias expressions given by (18) and (26), we consider the trend model (30) without and with the seasonal component S_t. Using an AR(1) noise model, for each selected ϕ-value in the interval [.1, .95], a simulation was performed for two sample sizes, $T = 60$ and $T = 120$, based on 1,000 replications with $\sigma_\varepsilon^2 = 1$. Because both the MLE and REML estimator of ϕ are invariant with respect to the true value of $\boldsymbol{\beta}$, without loss of generality, we take the regression coefficients as $\boldsymbol{\beta} = \mathbf{0}$ in generating the simulated data.

Table 2 presents the empirical biases of $\hat{\phi}$, based on exact ML and REML. First, we mention that empirical biases (not shown) of the approximate ML and REML estimates, (11) and (23), are close to those of the exact ML and REML estimates. Next, we observe that the impact of the seasonal component S_t on the empirical biases of the ML estimate is generally in agreement with the theoretical bias results of Table 1. For sample size $T = 60$, the theoretical biases of both the ML and REML estimates are in good agreement with the empirical biases for all values of ϕ, for both seasonal and nonseasonal models.

We also consider simulation results with AR(2) noise model and a moderate sample size $T = 60$. For each set of parameter values $\boldsymbol{\phi} = (.5, .25)', (.5, -.25)'$, and $(1.5, -.8)'$, a simulation was performed using 1,000 replications of AR(2) noise with $\sigma_\varepsilon^2 = 1$. For the trend model (30), Table 3 presents the empirical biases of the MLE and REML estimator $\hat{\boldsymbol{\phi}}_{\mathrm{M}}$ and $\hat{\boldsymbol{\phi}}_{\mathrm{R}}$. For sample size $T = 60$, the theoretical bias results of the ML and REML estimates provide good approximations to the empirical biases, for models with and without a seasonal component. Table 3 also presents the empirical standard deviations and root mean squared errors (MSEs) of the MLE and REML estimator. For the parameter values considered, the ML and REML estimates have approximately the same empirical standard

Table 3. Empirical Biases, Standard Deviations, and RMSEs of $\hat{\boldsymbol{\phi}}_M = (\hat{\phi}_{1M}, \hat{\phi}_{2M})'$ and $\hat{\boldsymbol{\phi}}_R = (\hat{\phi}_{1R}, \hat{\phi}_{2R})'$ for the Trend Model (30) with an AR(2) Noise Process and T = 60 (1,000 Replications). The Theoretical Biases of $\hat{\boldsymbol{\phi}}_M$ and $\hat{\boldsymbol{\phi}}_R$ Are Given By (18) and (26), and (m_1, m_2) or $Re^{\pm i\theta}$ Are Roots of $m^2 - \phi_1 m - \phi_2 = 0$.

			Theoretical bias		Empirical results: bias		Standard deviation		RMSE	
$\boldsymbol{\phi}$	m_1 / m_2	R / $\cos\theta$	$\hat{\boldsymbol{\phi}}_M$	$\hat{\boldsymbol{\phi}}_R$	$\hat{\boldsymbol{\phi}}_M$	$\hat{\boldsymbol{\phi}}_R$	$\hat{\boldsymbol{\phi}}_M$	$\hat{\boldsymbol{\phi}}_R$	$\hat{\boldsymbol{\phi}}_M$	$\hat{\boldsymbol{\phi}}_R$
Model without seasonal component (r = 2)										
.5	−.31		−.061	−.009	−.053	−.004	.143	.143	.153	.143
.25	.81		−.074	−.030	−.078	−.033	.131	.138	.153	.141
.5		.5	−.036	−.009	−.034	−.007	.132	.130	.136	.130
−.25		.5	−.029	−.004	−.030	−.004	.120	.124	.124	.125
1.5		.89	−.034	−.026	−.028	−.020	.083	.080	.088	.083
−.8		.84	.019	.024	.012	.019	.081	.084	.082	.086
Model with seasonal component (r = 10)										
.5	−.31		−.098	−.009	−.089	−.006	.148	.148	.173	.148
.25	.81		−.007	−.030	−.013	−.034	.138	.144	.138	.148
.5		.5	−.070	−.009	−.067	−.008	.143	.139	.158	.140
−.25		.5	.022	−.004	.019	−.004	.138	.135	.139	.135
1.5		.89	−.072	−.026	−.064	−.029	.098	.100	.117	.104
−.8		.84	.053	.024	.043	.025	.101	.098	.110	.101

deviations. This suggests that differences in root MSEs between the estimators will essentially be determined by their biases. For example, because of their smaller bias magnitude, all the REML estimates of ϕ_1 in Table 3 have smaller empirical MSEs than those of the ML estimates.

McGilchrist (1989) performed a simulation study of small sample biases of MLEs and REML estimators in regression models with ARMA errors. The simple trend model (30) without the seasonal component was considered with AR(1) and AR(2) noise, as well as other low-order ARMA noise models, and sample size of $T = 25$. Empirical biases of MLEs and REML estimators were obtained for different parameter value settings based on 100 replications for each case. Even for the small sample size of $T = 25$, our theoretical bias approximations for ML and REML estimation agree quite well with those simulation results. For example, in the AR(1) noise case, for $\phi = .75, .5, 0, -.5, -.75$, the theoretical ML biases from (18) are $-.241, -.181, -.086, .001$, and $.044$ compared to the simulated ML mean biases of $-.219, -.165, -.072, -.010$, and $.058$ from McGilchrist (1989), and the theoretical REML biases from (26) are $-.065, -.043, 0, .043$, and $.065$, compared to simulated REML mean biases of $-.066, -.037, .013, .036$, and $.080$.

6.3 Empirical Levels of Tests for Trend

We now investigate the impact of biases of $\hat{\phi}$ on inferences of the regression parameters. We consider testing the hypothesis H_0: $\omega = \omega_0$ in the regression trend model (30) using the t statistic $t = (\hat{\omega} - \omega_0)/$(estimated standard deviation $(\hat{\omega})$), where $\hat{\omega}$ equals $\hat{\omega}_{\mathrm{M}}$ or $\hat{\omega}_{\mathrm{R}}$. As mentioned earlier, the ML and REML estimates of ϕ do not depend on the true value of $\boldsymbol{\beta}$. Furthermore, under H_0: $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, we have $\tilde{\boldsymbol{\beta}}_{\mathrm{G}} - \boldsymbol{\beta}_0 = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{N}$, so that the distribution of $\tilde{\boldsymbol{\beta}}_{\mathrm{G}} - \boldsymbol{\beta}_0$, and hence that of the corresponding t statistic, does not depend on $\boldsymbol{\beta}_0$ when H_0 is true. Thus, without loss of generality, we may consider $\omega_0 = 0$.

Table 4. Empirical Levels of Nominal 5%-Level t Test of H_0: $\omega = 0$ for Trend Model (30) With AR(1) Noise, Based on ML and REML Estimation (1,000 Replications).

T	ϕ .1	.2	.3	.4	.5	.6	.7	.8	.9	.95
ML model without seasonal component (r = 2)										
60	.059	.065	.069	.073	.079	.095	.111	.150	.226	.333
120	.048	.048	.049	.053	.058	.061	.067	.090	.156	.239
ML model with seasonal component (r = 10)										
60	.072	.079	.085	.091	.103	.109	.116	.142	.206	.304
120	.047	.052	.051	.051	.058	.062	.067	.085	.140	.220
REML model without seasonal component (r = 2)										
60	.052	.052	.053	.059	.066	.070	.082	.101	.136	.183
120	.041	.042	.044	.043	.046	.050	.055	.064	.089	.137
REML model with seasonal component (r = 10)										
60	.052	.051	.054	.066	.069	.078	.082	.105	.140	.185
120	.043	.042	.043	.044	.048	.051	.055	.064	.087	.135

Table 5. Empirical Levels of Nominal 5%-Level t Test of H_0: $\omega = 0$ for Trend Model (30) With AR(2) Noise and $T = 60$, Based on ML and REML Estimation (1,000 Replications)

ϕ_1	.5	.5	1.5
ϕ_2	.25	−.25	−.8
Model without seasonal component (r = 2)			
ML	.171	.071	.073
REML	.101	.058	.062
Model with seasonal component (r = 10)			
ML	.147	.071	.083
REML	.103	.056	.066

In Section 5 we established an approximate bias for the MLE $\hat{\sigma}_\varepsilon^2$ given in (29). Thus in our simulations with AR noise, we used $\hat{\sigma}_{\mathrm{M}}^2 = S(\hat{\boldsymbol{\beta}}_{\mathrm{M}}, \hat{\boldsymbol{\phi}}_{\mathrm{M}})/(T - r - p)$ and $\hat{\sigma}_{\mathrm{R}}^2 = S(\hat{\boldsymbol{\beta}}_{\mathrm{R}}, \hat{\boldsymbol{\phi}}_{\mathrm{R}})/(T - r - p)$ to obtain less biased estimates of σ_ε^2 and to estimate the standard deviations of elements in $\hat{\boldsymbol{\beta}}_{\mathrm{M}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ via $\widehat{\operatorname{cov}}(\hat{\boldsymbol{\beta}}_{\mathrm{M}}) = \hat{\sigma}_{\mathrm{M}}^2(\mathbf{X}'\mathbf{V}(\hat{\boldsymbol{\phi}}_{\mathrm{M}})^{-1}\mathbf{X})^{-1}$, for example.

We now assess the improvement on the empirical levels of the t test of H_0: $\omega = 0$ based on a REML trend estimate. This improvement is expected to result from more accurate estimation of the standard deviation of the trend estimate based on the REML method. Table 4 reports empirical levels of the nominal 5%-level t-statistic test of the hypothesis H_0: $\omega = 0$, for simulations with an AR(1) noise model, and Table 5 reports corresponding empirical levels for selected AR(2) noise models. The rejection region for H_0 is $|t| > t_{T-r-p}^{(.025)}$. For both $T = 60$ and $T = 120$, results in Tables 4 and 5 indicate that the REML method performs better than the ML method in providing empirical levels closer to the nominal 5% level, for all values of AR parameters considered. With 1,000 replications, for empirical rejection proportions, a 95% probability interval around .05 is approximately $.05 \pm 2\{(.95)(.05)/1,000\}^{1/2} = (.036, .064)$. From Table 4, for AR(1) noise, we see that when $T = 120$, the REML method yields empirical levels within the 95% probability interval for values of ϕ as large as .8 for models without and with a seasonal component. By comparison, the ML method provides acceptable empirical levels only for $\phi \leq .6$. The REML method seems to provide substantial improvement over ML for both nonseasonal and seasonal models, especially for the AR(1) noise case when $\phi \geq .8$ and the AR(2) noise case with $\boldsymbol{\phi} = (.5, .25)'$, although for these cases REML still leads to higher rejection levels than the nominal 5%.

7. SUMMARY AND CONCLUDING REMARKS

This article has compared ML and REML estimation of a time series regression model with AR(p) noise. We derived the first-order biases of the MLE and REML estimator of the AR parameters, based on their approximate representations. In addition, a bias result for the corresponding MLE of the error variance σ_ε^2 was established. This leads to the recommendation to use $\bar{\sigma}_\varepsilon^2 = S(\hat{\boldsymbol{\beta}}_{\mathrm{M}}, \hat{\boldsymbol{\phi}}_{\mathrm{M}})/(T - r - p)$ as an approximately unbiased estimator of σ_ε^2. We also investigated the impact of bias of the autocorrelation estimates on testing of linear trend in a regression trend model. For a time series of moderate length, the REML estimator is generally less biased than the MLE. Consequently, the REML approach leads to more accurate inferences for the regression parameters.

APPENDIX: DETAILS ON DERIVATION AND FORM OF ML BIAS

Bias for MLE in AR(p) With Zero Mean

We present the bias results mentioned in Section 3.2 for the MLE $\hat{\phi}_{\mathrm{M}}$ of $\phi = (\phi_1, \ldots, \phi_p)'$ for an AR(p) model with known mean 0, $N_t = \phi_1 N_{t-1} + \cdots + \phi_p N_{t-p} + \varepsilon_t$, where $\mathrm{var}(\varepsilon_t) = \sigma_\varepsilon^2$. An approach related to that of Cordeiro and Klein (1994) is used, and the connection between the form given in Section 3.2 and this approach is also established. Let $\mathbf{C}_p^*$ and $\mathbf{c}_p^*$ be the $p \times p$ matrix and $p \times 1$ vector, with elements $C_{ij}^* = C_{ij}/(T-i-j)$ based on the T "observations" $\{N_t\}$ from the AR(p) process with known mean 0. From Section 3.2, we know that $E(\mathbf{C}_p^*) = \mathbf{\Gamma}_p = \sigma_\varepsilon^2 \mathbf{V}_p$ and $E(\mathbf{c}_p^*) = \gamma_p$. Under the present model, we are interested in determination of the "bias" term,

$$E[\hat{\phi}_{\mathrm{M}} - \phi] \approx -\mathbf{\Gamma}_p^{-1} E[(\mathbf{C}_p^* - \mathbf{\Gamma}_p)\mathbf{\Gamma}_p^{-1}[(\mathbf{c}_p^* - \gamma_p) - (\mathbf{C}_p^* - \mathbf{\Gamma}_p)\phi]]$$
$$\equiv -\mathbf{\Gamma}_p^{-1} E[\mathbf{C}_p^* \mathbf{\Gamma}_p^{-1}(\mathbf{c}_p^* - \mathbf{C}_p^* \phi)], \qquad \text{(A.1)}$$

where we use the fact that $E(\mathbf{c}_p^* - \mathbf{C}_p^*\phi) = \gamma_p - \mathbf{\Gamma}_p \phi = \mathbf{0}$. From the log-likelihood for an AR(p) model as given in Section 3, for large T we have

$$\frac{\partial l}{\partial \phi} = -\frac{1}{2}\frac{\partial}{\partial \phi}\log|\mathbf{V}_p| - \frac{1}{\sigma_\varepsilon^2}(\mathbf{C}_p\phi - \mathbf{c}_p)$$
$$\approx -\frac{T-2}{\sigma_\varepsilon^2}(\mathbf{C}_p^*\phi - \mathbf{c}_p^*),$$

and

$$\frac{\partial^2 l}{\partial \phi \partial \phi'} = -\frac{1}{2}\frac{\partial^2}{\partial \phi \partial \phi'}\log|\mathbf{V}_p| - \frac{1}{\sigma_\varepsilon^2}\mathbf{C}_p$$
$$\approx -\frac{T-2}{\sigma_\varepsilon^2}\mathbf{C}_p^*.$$

The information matrix $I(\phi)$ for ϕ is $I(\phi) = -E[\partial^2 l/\partial\phi\partial\phi'] \approx [(T-2)/\sigma_\varepsilon^2]\mathbf{\Gamma}_p = (T-2)\mathbf{V}_p$.

From the foregoing, we see that the quantity in (A.1) can be interpreted as

$$-\mathbf{\Gamma}_p^{-1}E[\mathbf{C}_p^*\mathbf{\Gamma}_p^{-1}(\mathbf{c}_p^* - \mathbf{C}_p^*\phi)]$$
$$\approx I(\phi)^{-1}E\left[\frac{\partial^2 l}{\partial\phi\partial\phi'}I(\phi)^{-1}\frac{\partial l}{\partial \phi}\right]$$
$$= I(\phi)^{-1}\sum_{j=1}^{p}\mathbf{B}_j I(\phi)^{-1}\mathbf{e}_j^{(p)}, \qquad \text{(A.2)}$$

where $\mathbf{B}_j = E[(\partial l/\partial\phi_j)(\partial^2 l/\partial\phi\partial\phi')], j = 1, \ldots, p$, and $\mathbf{e}_j^{(p)}$ denotes the $p \times 1$ unit vector with 1 in the jth position and 0's elsewhere. Notice that interchanging the order of expectation (integration) and differentiation (under assumed regularity conditions) makes the $\mathbf{B}_j$ satisfy the relation

$$-\frac{\partial}{\partial\phi_j}I(\phi) \equiv \frac{\partial}{\partial\phi_j}E\left[\frac{\partial^2 l}{\partial\phi\partial\phi'}\right]$$
$$= E\left[\frac{\partial}{\partial\phi_j}\left(\frac{\partial^2 l}{\partial\phi\partial\phi'}\right)\right] + E\left[\frac{\partial l}{\partial\phi_j}\frac{\partial^2 l}{\partial\phi\partial\phi'}\right]$$
$$= \mathbf{B}_j + O(1),$$

because $E\{(\partial/\partial\phi_j)[(\partial^2 l/\partial\phi\partial\phi')]\} = -\frac{1}{2}(\partial/\partial\phi_j)[(\partial^2/\partial\phi\partial\phi')\log|\mathbf{V}_p|] = O(1), j = 1, \ldots, p$. Hence we find that $\mathbf{B}_j \approx -(\partial/\partial\phi_j)I(\phi) \approx -(T-2)\partial\mathbf{V}_p/\partial\phi_j$. Noting that $\partial\mathbf{V}_p^{-1}/\partial\phi_j = -\mathbf{V}_p^{-1}(\partial\mathbf{V}_p/\partial\phi_j)\mathbf{V}_p^{-1}$, it follows that the term in (A.2) can be expressed approximately as

$$I(\phi)^{-1}\sum_{j=1}^{p}\mathbf{B}_j I(\phi)^{-1}\mathbf{e}_j^{(p)}$$
$$\approx \frac{1}{T-2}\mathbf{V}_p^{-1}\sum_{j=1}^{p}(-\partial\mathbf{V}_p/\partial\phi_j)(\mathbf{V}_p^{-1})\mathbf{e}_j^{(p)}$$
$$= \frac{1}{T-2}\sum_{j=1}^{p}(\partial\mathbf{V}_p^{-1}/\partial\phi_j)\mathbf{e}_j^{(p)}.$$

That is, the ML bias result for the AR(p) model with mean 0 is obtained as

$$E[\hat{\phi}_{\mathrm{M}} - \phi] \approx \frac{1}{T-2}\sum_{j=1}^{p}(\partial\mathbf{V}_p^{-1}/\partial\phi_j)\mathbf{e}_j^{(p)}. \qquad \text{(A.3)}$$

The representation for the bias of the MLE $\hat{\phi}_{\mathrm{M}}$ of ϕ, from the approach presented by Cordeiro and Klein (1994), is obtained as

$$E[\hat{\phi}_{\mathrm{M}} - \phi] = I(\phi)^{-1}\mathbf{B}\,\mathrm{vec}[I(\phi)^{-1}] + O(1/T^2)$$
$$\equiv I(\phi)^{-1}\sum_{j=1}^{p}\mathbf{B}_j I(\phi)^{-1}\mathbf{e}_j^{(p)} + O(1/T^2),$$

where $\mathbf{B} = [\mathbf{B}_1, \ldots, \mathbf{B}_p]$ and $\mathbf{B}_j = -(\partial/\partial\phi_j)I(\phi) - \frac{1}{2}E[(\partial/\partial\phi_j)(\partial^2 l/\partial\phi\partial\phi')], j = 1, \ldots, p$. Thus the connection between the bias term as represented in (A.1) and the representation based on the approach of Cordeiro and Klein (1994) is exhibited directly for the case of the AR(p) model with known mean.

Form and Derivatives of $\mathbf{V}_p^{-1}$ for AR(p)

From Galbraith and Galbraith (1974) and Ljung and Box (1979), the $p \times p$ matrix $\mathbf{V}_p^{-1}$ can be expressed in the form

$$\mathbf{V}_p^{-1} = \mathbf{P}_*'\mathbf{P}_* - \mathbf{Q}_*'\mathbf{Q}_* \equiv \mathbf{P}_*\mathbf{P}_*' - \mathbf{Q}_*\mathbf{Q}_*',$$

where $\mathbf{P}_*' = \mathbf{I} - \sum_{i=1}^{p-1}\phi_i\mathbf{L}^i$ and $\mathbf{Q}_* = \phi_p\mathbf{I} + \sum_{i=1}^{p-1}\phi_{p-i}\mathbf{L}'^i$ are $p \times p$ matrices and $\mathbf{L}$ is a $p \times p$ lag matrix with 1's on the first subdiagonal and 0's elsewhere. Thus we obtain a form for the matrix of derivatives as

$$\frac{\partial\mathbf{V}_p^{-1}}{\partial\phi_j} = \left(\frac{\partial\mathbf{P}_*'}{\partial\phi_j}\mathbf{P}_* + \mathbf{P}_*'\frac{\partial\mathbf{P}_*}{\partial\phi_j}\right) - \left(\frac{\partial\mathbf{Q}_*'}{\partial\phi_j}\mathbf{Q}_* + \mathbf{Q}_*'\frac{\partial\mathbf{Q}_*}{\partial\phi_j}\right)$$
$$= -(\mathbf{L}^j\mathbf{P}_* + \mathbf{P}_*'\mathbf{L}'^j) - (\mathbf{L}^{p-j}\mathbf{Q}_* + \mathbf{Q}_*'\mathbf{L}'^{p-j}),$$
$$j = 1, \ldots, p,$$

noting the convention $\mathbf{L}^0 = \mathbf{I}_p$ and the fact that $\mathbf{L}^p = \mathbf{0}$. Then, because $\mathbf{L}'^j\mathbf{e}_j^{(p)} = \mathbf{0}$ for $j = 1, \ldots, p$, we obtain the representation

$$\sum_{j=1}^{p}\frac{\partial\mathbf{V}_p^{-1}}{\partial\phi_j}\mathbf{e}_j^{(p)} = -\sum_{j=1}^{p}\mathbf{L}^j\mathbf{P}_*\mathbf{e}_j^{(p)}$$
$$-\sum_{j=1}^{p}\mathbf{L}^{p-j}\mathbf{Q}_*\mathbf{e}_j^{(p)} - \mathbf{Q}_*'\left(\sum_{j=1}^{p}\mathbf{L}'^{p-j}\mathbf{e}_j^{(p)}\right), \qquad \text{(A.4)}$$

where $\sum_{j=1}^{p}\mathbf{L}'^{p-j}\mathbf{e}_j^{(p)} = (0,1,0,1,\ldots,0,1)'$ or $(1,0,1,\ldots,0,1)'$, depending on whether p is even or odd. Note that in (A.4),

$\mathbf{L}^j\mathbf{P}_*\mathbf{e}_j^{(p)}$ is interpretable as the jth column of $\mathbf{P}_*$ shifted down by j elements, with j 0's inserted above, and similarly for $\mathbf{L}^{p-j}\mathbf{Q}_*\mathbf{e}_j^{(p)}$; that is,

$$\mathbf{L}^j\mathbf{P}_*\mathbf{e}_j^{(p)} = -(0,\ldots,0,\phi_{j-1},\phi_{j-2},\ldots,\phi_{j-(p-j)})',$$

and

$$\mathbf{L}^{p-j}\mathbf{Q}_*\mathbf{e}_j^{(p)} = (0,\ldots,0,\phi_{p+1-j},\phi_{p+2-j},\ldots,\phi_p)',$$

with the conventions that $\phi_0 = -1$ and $\phi_j = 0$ for $j < 0$ or $j > p$. Thus, in particular, $\sum_{j=1}^p \mathbf{L}^{p-j}\mathbf{Q}_*\mathbf{e}_j^{(p)} = (\phi_1, 2\phi_2, \ldots, p\phi_p)'$. Therefore, we can readily determine the explicit form of (A.4), for p even or odd, as

$$\sum_{j=1}^p \frac{\partial \mathbf{V}_p^{-1}}{\partial \phi_j}\mathbf{e}_j^{(p)} = \begin{bmatrix} -\phi_1 \\ -1-2\phi_2-\phi_p \\ \phi_1 - 3\phi_3 - \phi_{p-1} \\ -1+\phi_2-4\phi_4-\phi_{p-2}-\phi_p \\ \vdots \\ -1+\phi_2-(p-1)\phi_{p-2}-\phi_p \\ \phi_1 - p\phi_{p-1} \\ -1-(p+1)\phi_p \end{bmatrix}$$

or

$$= \begin{bmatrix} -\phi_1-\phi_p \\ -1-2\phi_2-\phi_{p-1} \\ \phi_1-3\phi_3-\phi_{p-2}-\phi_p \\ -1+\phi_2-4\phi_4-\phi_{p-3}-\phi_{p-1} \\ \vdots \\ \phi_1-(p-1)\phi_{p-2}-\phi_p \\ -1-p\phi_{p-1} \\ -(p+1)\phi_p \end{bmatrix}. \quad \text{(A.5)}$$

[Received June 1999. Revised March 2000.]

REFERENCES

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, New York: Wiley.

Bhansali, R. J. (1981), "Effects of Not Knowing the Order of an Autoregressive Process on the Mean Squared Error of Prediction-I," *Journal of the American Statistical Association*, 76, 588–597.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994), *Time Series Analysis: Forecasting and Control* (3rd ed.), Englewood Cliffs, NJ: Prentice-Hall.

Cooper, D. M., and Thompson, R. (1977), "A Note on the Estimation of the Parameters of the Autoregressive Moving Average Process," *Biometrika*, 64, 625–628.

Cordeiro, G. M., and Klein, R. (1994), "Bias Correction in ARMA Models," *Statistics and Probability Letters*, 19, 169–176.

Galbraith, R. F., and Galbraith, J. I. (1974), "On the Inverses of Some Patterned Matrices Arising in the Theory of Stationary Time Series," *Journal of Applied Probability*, 11, 63–71.

Harville, D. A. (1974), "Bayesian Inference for Variance Components Using Only Error Contrasts," *Biometrika*, 61, 383–385.

——— (1977), "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems," *Journal of the American Statistical Association*, 72, 320–340.

Ljung, G. M., and Box, G. E. P. (1979), "The Likelihood Function of Stationary Autoregressive-Moving Average Models," *Biometrika*, 66, 265–270.

Magee, L. (1989), "Bias Approximations for Covariance Parameter Estimators in the Linear Model With AR(1) Errors," *Communications in Statistics, Part A—Theory and Methods*, 18, 395–422.

McGilchrist, C. A. (1989), "Bias of ML and REML Estimators in Regression Models With ARMA Errors," *Journal of Statistical Computation and Simulation*, 32, 127–136.

Mentz, R. P., Morettin, P. A., and Toloi, C. M. C. (1998), "On Residual Variance Estimation in Autoregressive Models," *Journal of Time Series Analysis*, 19, 187–208.

Parzen, E. (1961), "An Approach to Time Series Analysis," *Annals of Mathematical Statistics*, 32, 951–989.

Searle, S. R. (1982), *Matrix Algebra Useful for Statistics*, New York: Wiley.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, New York: Wiley.

Shaman, P., and Stine, R. A. (1988), "The Bias of Autoregressive Coefficient Estimators," *Journal of the American Statistical Association*, 83, 842–848.

Tanaka, K. (1984), "An Asymptotic Expansion Associated With the Maximum Likelihood Estimators in ARMA Models," *Journal of the Royal Statistical Society*, Ser. B, 46, 58–67.

Tunnicliffe Wilson, G. (1989), "On the Use of Marginal Likelihood in Time Series Model Estimation," *Journal of the Royal Statistical Society*, Ser. B, 51, 15–27.