The Bias of Autoregressive Coefficient Estimators

PAUL SHAMAN and ROBERT A. STINE*

This article presents simple expressions for the bias of estimators of the coefficients of an autoregressive model of arbitrary, but known, finite order. The results include models both with and without a constant term. The effects of overspecification of the model order on the bias are described. The emphasis is on least-squares and Yule–Walker estimators, but the methods extend to other estimators of similar design. Although only the order T^{-1} component of the bias is captured, where T is the series length, this asymptotic approximation is shown to be very accurate for least-squares estimators through some numerical simulations. The simulations examine fourth-order autoregressions chosen to resemble some data series from the literature. The order T^{-1} bias approximations for Yule–Walker estimators need not be accurate, especially if the zeros of the associated polynomial have moduli near 1. Examples are given where the approximation is accurate and where it is useless. The bias expressions are very simple in the case of least squares, being linear combinations of the unknown true coefficients. No interaction among the coefficients occurs. For example, if the data are a time series from a fourth-order autoregressive model with coefficients $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ and no constant term, the order T^{-1} bias of the least-squares coefficient estimator is $(-\alpha_1, 1 - 2\alpha_2 - \alpha_4, \alpha_1 - 4\alpha_3, 1 - 5\alpha_4)/T$. The results differ slightly for a model with a constant term. An easily programmed algorithm for generating these expressions for any finite-order autoregressive model is given, with or without a constant term. The structure of the order T^{-1} bias for Yule–Walker estimators is not so readily represented, but it is easily evaluated for any model. Thus one can quickly incorporate these results into the study of other time-series problems, such as the effects of estimation error on the mean squared prediction error. Direct methods of analysis are employed to obtain the expressions. By analysis of bias in the frequency domain, infinite series representations that obscure the simple form of the bias are avoided. The key results are several lemmas regarding sums of elements within the inverse covariance matrix of p consecutive observations from an autoregression of order p. The bias approximations follow directly from these lemmas. The derivations are straightforward and yield useful insight into the structure of the estimators.

KEY WORDS: Autoregressive process; Least-squares estimator; Yule–Walker estimator.

1. INTRODUCTION

The bias of least-squares and Yule–Walker estimators of coefficients in an autoregression is well-known to depend on the unknown coefficients. For example, in a first-order autoregression, bias tends to shrink the least-squares estimator toward 0. The larger the coefficient becomes, the greater is the bias. In short series with considerable serial correlation the shrinkage can have a pronounced effect on the estimator.

In applications it is often important to consider the effects of the bias of autoregressive coefficient estimators. In practice, sample estimates $\hat{\alpha}$ replace unknown coefficients α in expressions for spectral estimators, prediction error, prediction mean squared error (PMSE), and hypothesis tests, for example. Autoregressive spectral estimators of the form $\hat{f}(\lambda) = (\hat{\sigma}^2/(2\pi))|1 + \sum_j \hat{\alpha}_j e^{i\lambda j}|^{-2}$ have bias that depends in part on that of the vector of coefficients α, and the bias of the estimator $\hat{\alpha}$ can hide peaks in the spectrum (Lysne and Tjøstheim 1987). Similarly, a common estimate of the PMSE for a forecast f periods into the future is $\text{PMSE}(f) = \hat{\sigma}^2(\sum_{j=0}^{f-1} \hat{\omega}_j^2 + \hat{\eta}_f/T)$, where T is the number of observations, the $\hat{\omega}_j$ are coefficients in the moving average process defined by the autoregressive process with coefficients $\hat{\alpha}$, and $\hat{\eta}_f > 0$ compensates for the increased prediction error due to estimation (Fuller and Hasza 1981). Finding the bias of PMSE(f) requires an expression for the bias of $\hat{\alpha}$. As with spectral estimators, the bias of $\hat{\alpha}$ can have a noticeable effect, even to the extent of canceling out the effect of the additional correction term $\hat{\eta}_f$ (Stine 1987). Also, the bias of coefficient estimators is a factor in the error of predicting a future observation with a misspecified model (Kunitomo and Yamamoto 1985). Granger–Sims tests of causality (Granger 1969; Sims 1972) are used to explore causal relationships between different time series. Construction of the tests involves the use of least-squares estimators of autoregressive parameters. The bias of these estimators can adversely affect the test result.

A further consideration that requires attention is the difference in bias between the least-squares and Yule–Walker methods of estimation. The aim of this article is to derive useful expressions for the bias in least-squares and Yule–Walker estimation for autoregressive models of known order.

Much research has been devoted to estimating the bias. Most efforts have considered the $O(1/T)$ term, the first-order term of the bias. Marriott and Pope (1954) found this first-order term for a model with a single autoregressive coefficient. They showed that the bias of the least-squares estimator is $-2/T$ times the true coefficient when the mean is known and increases to $-3/T$ when the mean is estimated. Kendall (1954) also derived the latter result. Bhansali (1981) obtained a general expression for the bias of least-squares estimators in models of arbitrary fitted order with known mean. His result does not require that the order of the fitted model be that of the true model. Although his result reduces to that of Marriott and Pope in the case of a first-order model, its complexity conceals a simple pattern that was revealed in numerical studies of the expression (Stine 1982). In particular, the first-order term of the bias of the least-squares estimator is a linear function of the unknown coefficients. Kunitomo and Ya-

* Paul Shaman is Professor and Robert A. Stine is Assistant Professor, Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104-6302.

Journal of the American Statistical Association
September 1988, Vol. 83, No. 403, Theory and Methods

mamoto (1985) also developed a representation for the bias when the order of the fitted model does not coincide with that of the true model. Further consideration of the bias was given by A. M. Walker in an unpublished manuscript. Walker obtained the formula given by Bhansali when the fitted and true orders coincide. He also considered the Yule–Walker estimator.

Several approaches have been employed in deriving expressions for the bias. Tanaka (1984) obtained a bias approximation as a by-product of an Edgeworth expansion for the distribution of maximum likelihood estimators in autoregressive moving average models. In a more direct approach, Tjøstheim and Paulsen (1983) used a Taylor series expansion for an estimator viewed as a function of correlation estimators to obtain expressions for the first-order bias of both least-squares and Yule–Walker estimators. Similarly, Yamamoto and Kunitomo (1984) obtained the same bias approximations from series expansions formulated for autoregressive models expressed in vector form. In each case, the authors presented some explicit results for models with at most three autoregressive coefficients; however, these methods are hard to extend to higher-order models. For example, Tjøstheim and Paulsen and Yamamoto and Kunitomo exploited the relationship of the coefficients of the process to the roots of the associated characteristic equation. As the order of the model increases, the algebraic complexity encountered limits such approaches to low-order models.

We present explicit expressions for the first-order term of the bias of both least-squares and Yule–Walker estimators. We also include expressions for the effects of estimating the mean with both types of estimators.

Notation and assumptions are given in Section 2. A general expression for the bias of an estimator of the coefficients in an autoregression is given in Section 3. We also describe the differences between least-squares and Yule–Walker procedures and the effect of estimating the mean. In Section 4, we provide three lemmas about the structure of the inverse of the $p \times p$ covariance matrix of an autoregressive process. These lemmas are used in Section 5 to obtain the results for the least-squares estimator. In Section 6, we discuss the Yule–Walker estimator. We close with a discussion of the implications of our results in Section 7, which includes some numerical calculations of the first-order bias approximation and derivation of the additional bias stemming from overspecification of the model order.

2. AUTOREGRESSIVE MODELS: NOTATION AND ASSUMPTIONS

Let $\{y_t\}$ be a discrete-time autoregressive process of known, finite order p,

$$\sum_{j=0}^{p} \alpha_j (y_{t-j} - \mu) = \varepsilon_t, \qquad \alpha_0 \equiv 1,$$

where $\mu = E(y_t)$. Observations from this process are denoted by $y = (y_1, \ldots, y_T)'$, and the vector of p coefficients to be estimated is $\alpha = (\alpha_1, \ldots, \alpha_p)'$. The error terms $\{\varepsilon_t\}$ are independent and identically distributed and have mean 0 and variance σ^2. In addition, we assume that the zeros of the polynomial $A(z) = \sum_{j=0}^{p} \alpha_j z^j$ lie strictly outside the unit circle so that the process is stationary. The covariances of the process are $r_k = E\{(y_t - \mu)(y_{t-k} - \mu)\}$ $(k = 0, \pm 1, \ldots)$. The covariance matrix of $(y_t, \ldots, y_{t-p+1})$ is $R = (R_{ij})$ with $R_{ij} = r_{i-j}$. The covariances and coefficients are related through the well-known equations

$$\sum_{j=0}^{p} \alpha_j r_{k-j} = \delta(k)\,\sigma^2, \qquad k = 0, 1, 2, \ldots, \tag{2.1}$$

where $\delta(k) = 1$ if $k = 0$ and is 0 otherwise. The spectral density is $f(\lambda) = (\sigma^2/(2\pi))|A(e^{-i\lambda})|^{-2}$ $(-\pi \leqq \lambda \leqq \pi)$, and we abbreviate the transfer function of the process as $A(\lambda) = A(e^{-i\lambda})$.

We use $\hat{}$ to denote least-squares estimators and $\tilde{}$ to denote Yule–Walker estimators. Terms including an estimated mean are distinguished by $*$. Thus we write the least-squares estimator for the coefficients α as

$$\hat{\alpha} = -\hat{R}^{-1}\hat{r}, \tag{2.2}$$

where the elements of the estimated covariance matrix $\hat{R}$ are

$$\hat{r}_{ij} = \sum_{t=p+1}^{T} (y_{t-i} - \mu)(y_{t-j} - \mu)/(T - p), \qquad i, j = 1, 2, \ldots, p, \tag{2.3}$$

and $\hat{r} = (\hat{r}_{01}, \ldots, \hat{r}_{0p})'$, when μ is known. When the mean is unknown, the estimator is $\hat{\alpha}^* = -\hat{R}^{*-1}\hat{r}^*$, where the covariances now are estimated by

$$\hat{r}^*_{ij} = \sum_{t=p+1}^{T} (y_{t-i} - \hat{\mu}_i)(y_{t-j} - \hat{\mu}_j)/(T - p), \qquad i, j = 0, 1, \ldots, p,$$

with $\hat{\mu}_i = \sum_{t=p+1}^{T} y_{t-i}/(T - p)$. The Yule–Walker estimator when μ is known is $\tilde{\alpha} = -\tilde{R}^{-1}\tilde{r}$, where the covariance estimators are

$$\tilde{r}_j = \sum_{t=j+1}^{T} (y_t - \mu)(y_{t-j} - \mu)/T, \qquad j = 0, 1, \ldots, p, \tag{2.4}$$

and $\tilde{R} = (\tilde{r}_{|i-j|})$, $\tilde{r} = (\tilde{r}_1, \ldots, \tilde{r}_p)'$. When the mean is unknown, the Yule–Walker estimator is $\tilde{\alpha}^* = -\tilde{R}^{*-1}\tilde{r}^*$, where $\tilde{R}^*$ and $\tilde{r}^*$ are formed from covariance estimators $\tilde{r}^*_j$ that have the same form as $\tilde{r}_j$ with μ replaced by $\sum_{t=1}^{T} y_t/T$.

To ensure the validity of the approximations to the bias used in this article we assume that the errors ε_t have finite moment of order 16 and that

$$E(|\hat{R}^{-1} - R^{-1}|^8) = O(1) \quad \text{as } T \to \infty, \tag{2.5}$$

where $|A|$ is the largest eigenvalue (in absolute value) of A and $R = E\hat{R}$. These assumptions were used by Lewis and Reinsel (1988). See also Bhansali (1981), whose (A3) is stronger than (2.1).

3. PRELIMINARIES

In this section we begin derivation of expressions for the bias of least-squares and Yule–Walker estimators of α. The results agree with those given for the least-squares case by Bhansali (1981, theorem 4.1) and Kunitomo and Yamamoto (1985, theorem 1). Further development of the bias expressions appears in Sections 5 and 6. Definitions of the estimators are given in Section 2. Unless otherwise stated, the order of the autoregressive process is assumed to be correctly specified in the estimation procedure. There are four cases to treat, least-squares and Yule–Walker estimation, each without and with mean correction. In the development we begin with least-squares estimation when the mean is known. The adjustments required for the other three cases are then described. The assumptions specified in Section 2 are required.

To begin we use the notation of least-squares estimation with known mean as a generic notation. Let $\hat{B} = (\hat{R} - R)R^{-1}$. Following Lewis and Reinsel (1988) (see also Bhansali 1981) we may write

$$TE(\hat{\alpha} - \alpha)$$
$$= -TR^{-1}E(\hat{r} - r - \hat{B}r) \quad (3.1)$$
$$+ TR^{-1}E[(\hat{R} - R)R^{-1} \times \{(\hat{R} - R)\alpha + \hat{r} - r\}] \quad (3.2)$$
$$+ O(T^{-1/2}).$$

In the least-squares case with known mean, (3.1) vanishes for all T. In the other three cases, the limit of (3.1) as $T \to \infty$ is not 0. For all four cases the $O(T^{-1/2})$ remainder result is valid.

For least-squares estimation with known mean, as $T \to \infty$ the limit of (3.2) is easily obtained from the limiting expression for the covariances of the scaled sample covariances (2.3),

$$\lim_{T\to\infty} T \operatorname{cov}(\hat{r}_{gh}, \hat{r}_{kl}) = 2\pi \int_{-\pi}^{\pi} (e^{i\lambda(g-h)} + e^{-i\lambda(g-h)})\, e^{i\lambda(l-k)} f^2(\lambda)\, d\lambda + (\kappa_4/\sigma^4) r_{g-h} r_{k-l}, \quad (3.3)$$

where κ_4 is the fourth cumulant of ε_t. [See, e.g., corollary 8.3.1 of Anderson (1971), which treats the covariance estimator $(T/(T - j))\tilde{r}_j$, where $\tilde{r}_j$ is defined at (2.4).] The right side of (3.3) is valid for the covariance estimators used in connection with all four methods—least-squares and Yule–Walker estimation, each with and without mean correction (see Anderson 1971, pp. 468–471). That is, for all of the methods, as $T \to \infty$ the limit of (3.2) is the same expression.

Define $R^{-1} = (R^{ij})$ and consider least-squares with known mean. Then (2.1), (3.2), (3.3), and the comments following (3.2) show that the jth element of $TE(\hat{\alpha} - \alpha)$ has limit

$$\lim_{T\to\infty} TE(\hat{\alpha}_j - \alpha_j) = 2\pi \sum_{g,h,k=1}^{p} R^{jg} R^{hk} \int_{-\pi}^{\pi} (e^{i\lambda(g-h-k)} + e^{-i\lambda(g-h+k)}) \times A(-\lambda) f^2(\lambda)\, d\lambda, \qquad j = 1, \ldots, p. \quad (3.4)$$

That (3.4) does not depend on the fourth-order cumulant κ_4 was noted by Bhansali (1981). Expression (3.4) is equivalent to (4.1) in Bhansali (1981).

Let $e(\lambda) = (1, e^{i\lambda}, \ldots, e^{i\lambda(p-1)})'$. Then (3.4) may be written as

$$\lim_{T\to\infty} TE(\hat{\alpha} - \alpha) = 2\pi R^{-1} \int_{-\pi}^{\pi} \{e(\lambda)e(-\lambda)'R^{-1}e(-\lambda) + e(-\lambda)e(\lambda)'R^{-1}e(-\lambda)\} e^{-i\lambda} A(-\lambda) f^2(\lambda)\, d\lambda. \quad (3.5)$$

Now consider least-squares estimation with unknown mean. Since $\hat{r}_{ij}^* = \hat{r}_{ij} - (\hat{\mu}_i - \mu)(\hat{\mu}_j - \mu)$, we have

$$TE(\hat{r}_{ij}^* - r_{i-j}) = -TE(\hat{\mu}_i - \mu)(\hat{\mu}_j - \mu)$$
$$= -2\pi f(0) + o(1)$$
$$= -\sigma^2(1 + \alpha_1 + \cdots + \alpha_p)^{-2} + o(1)$$

as $T \to \infty$ (e.g., Anderson 1971, theorem 8.3.1). Then the analog of (3.1) may be written as

$$-TR^{-1}E(\hat{r}^* - r - \hat{B}^* r)$$
$$= -TR^{-1}E\{\hat{r}^* - r + (\hat{R}^* - R)\alpha\}$$
$$= \sigma^2(1 + \alpha_1 + \cdots + \alpha_p)^{-2} R^{-1}(c + cc'\alpha) + o(1)$$
$$= \sigma^2(1 + \alpha_1 + \cdots + \alpha_p)^{-1} R^{-1} c + o(1), \quad (3.6)$$

as $T \to \infty$, where $c = (1, \ldots, 1)'$ is $p \times 1$.

In the case of Yule–Walker estimation with known mean, (2.4) has expected value

$$E(\tilde{r}_j) = (1 - j/T) r_j, \qquad j = 0, 1, \ldots, p,$$

and (3.1) becomes

$$-TR^{-1}E(\tilde{r} - r - \tilde{B}r)$$
$$= -TR^{-1}E\{\tilde{r} - r + (\tilde{R} - R)\alpha\}$$
$$= R^{-1} d, \quad (3.7)$$

where d is a $p \times 1$ vector with jth element

$$d_j = \sum_{k=0}^{p} |j - k| r_{j-k} \alpha_k, \qquad j = 1, \ldots, p. \quad (3.8)$$

When the mean is unknown,

$$TE(\tilde{r}_j^* - r_j)$$
$$= -jr_j - 2\pi f(0) + o(1)$$
$$= -jr_j - \sigma^2(1 + \alpha_1 + \cdots + \alpha_p)^{-2} + o(1)$$

as $T \to \infty$ (e.g., Anderson 1971, theorem 8.3.2). Thus the expression corresponding to (3.1) is the sum of the right sides of (3.6) and (3.7).

Additional discussion of the bias for Yule–Walker estimation is in Section 6.

4. THREE LEMMAS

Derivation of the explicit bias results requires evaluation of the two bilinear forms involving R^{-1} that appear in (3.5) and of the row sums of R^{-1} in (3.6). We give these in three lemmas. Proofs are available in Shaman and Stine (1987). For the elements of R^{-1} we use

$$R^{jk} = \sigma^{-2} \sum_{r=1}^{\min(j,k)} (\alpha_{j-r}\alpha_{k-r} - \alpha_{p-j+r}\alpha_{p-k+r}), \qquad j, k = 1, \ldots, p$$

(Parzen 1961, p. 968). The matrix R^{-1} is symmetric about both the main and transverse diagonals; that is, $R^{jk} = R^{kj} = R^{p+1-k,p+1-j}$ $(j, k = 1, \ldots, p)$.

Lemma 1.

$$\sum_{k=1}^{p} R^{jk} = \sigma^{-2} \sum_{h=0}^{p} \alpha_h \sum_{r=0}^{j-1} (\alpha_r - \alpha_{p-r}), \qquad j = 1, \ldots, p.$$

Lemma 2.

$$\begin{aligned}\sigma^2 e(\lambda)' R^{-1} e(-\lambda) &= A(\lambda) \sum_{j=0}^{p} (1 - j)\alpha_j e^{i\lambda j} + A(-\lambda) \sum_{j=0}^{p} (p - 1 - j)\alpha_j e^{-i\lambda j} \\ &= p|A(\lambda)|^2 + iA(\lambda)A'(-\lambda) - iA(-\lambda)A'(\lambda).\end{aligned}$$

Lemma 3. If p is even,

$$\sigma^2 e(-\lambda)' R^{-1} e(-\lambda) = \{A(\lambda) + e^{-i\lambda p} A(-\lambda)\} \sum_{j=0}^{(1/2)p-1} (\alpha_j - \alpha_{p-j}) e^{-i\lambda j} \times (1 + e^{-i\lambda 2} + \cdots + e^{-i\lambda(p-2-2j)}).$$

If p is odd,

$$\begin{aligned}\sigma^2 e(-\lambda)' R^{-1} e(-\lambda) &= A(\lambda) \sum_{j=0}^{(1/2)(p-1)} (\alpha_{j-1} - \alpha_{p-j}) e^{-i\lambda(j-1)} \times (1 + e^{-i\lambda 2} + \cdots + e^{-i\lambda(p-1-2j)}) \\ &\quad + e^{-i\lambda p} A(-\lambda) \sum_{j=0}^{(1/2)(p-1)} (\alpha_j - \alpha_{p+1-j}) e^{-i\lambda(j-1)} \times (1 + e^{-i\lambda 2} + \cdots + e^{-i\lambda(p-1-2j)}),\end{aligned}$$

where $\alpha_{-1} = \alpha_{p+1} = 0$.

5. THE BIAS OF THE LEAST-SQUARES ESTIMATOR

If the mean is known the $O(1/T)$ bias of the least-squares estimator is given by (3.5). By Lemma 2 the second summand on the right side of (3.5) contributes

$$R^{-1} \int_{-\pi}^{\pi} e(-\lambda) e^{-i\lambda} \sum_{j=0}^{p} (1 - j)\alpha_j e^{i\lambda j} f(\lambda)\, d\lambda,$$

which is

$$-\alpha - (0, \alpha_2, 2\alpha_3, \ldots, (p-1)\alpha_p)' = -(\alpha_1, 2\alpha_2, \ldots, p\alpha_p)', \tag{5.1}$$

where we have used (2.1) and the fact that for $j = 1, \ldots, p$, the integral is equal to successive columns of R. The first summand on the right side of (3.5) yields, by Lemma 3,

$$\begin{aligned}&R^{-1} \int_{-\pi}^{\pi} e(\lambda) e^{-i\lambda} \sum_{j=0}^{(1/2)p-1} (\alpha_j - \alpha_{p-j}) e^{-i\lambda j} \times \sum_{k=0}^{(1/2)p-1-j} e^{-i\lambda 2k} f(\lambda)\, d\lambda, \qquad p \text{ even} \\ &R^{-1} \int_{-\pi}^{\pi} e(\lambda) e^{-i\lambda} \sum_{j=0}^{(1/2)(p-1)} (\alpha_{j-1} - \alpha_{p-j}) e^{-i\lambda(j-1)} \times \sum_{k=0}^{(1/2)(p-1)-j} e^{-i\lambda 2k} f(\lambda)\, d\lambda, \qquad p \text{ odd},\end{aligned} \tag{5.2}$$

with $\alpha_{-1} = 0$. These expressions reduce to

$$\begin{aligned}&\sum_{j=0}^{(1/2)p-1} (\alpha_j - \alpha_{p-j}) a_j, \qquad p \text{ even} \\ &\sum_{j=0}^{(1/2)(p-1)} (\alpha_{j-1} - \alpha_{p-j}) b_j, \qquad p \text{ odd},\end{aligned} \tag{5.3}$$

where a_j is $p \times 1$ and has 1s in rows $j + 2, j + 4, \ldots, p - j$ and 0s elsewhere, and b_j is $p \times 1$ and has 1s in rows $j + 1, j + 3, \ldots, p - j$ and 0s elsewhere.

If there is mean correction, (3.6) is a summand of the $O(1/T)$ bias vector. By Lemma 1 the jth component of (3.6) yields

$$\sum_{r=0}^{j-1} (\alpha_r - \alpha_{p-r}), \qquad j = 1, \ldots, p. \tag{5.4}$$

Table 1 displays the least-squares bias vectors for $p = 1, \ldots, 6$, as constructed from (5.1), (5.3), and (5.4). If one adopts the convention of writing the autoregressive model as $y_t - \mu = \sum_{j=1}^{p} \alpha_j (y_{t-j} - \mu) + \varepsilon_t$, then one must change the algebraic sign of each constant term in our bias expressions.

Many authors have given the first-order bias term for $p = 1$. The result for known mean is in Marriott and Pope (1954), White (1961), and Shenton and Johnson (1965); for unknown mean it is in Marriott and Pope (1954), Kendall (1954), and White (1961). For $p = 2$ Tanaka (1984) and Yamamoto and Kunitomo (1984) gave the bias for known mean; Tjøstheim and Paulsen (1983), Tanaka (1984), Yamamoto and Kunitomo (1984), and A. M.

Table 1. Least-Squares Bias Vectors

p	Known mean: $\lim_{T\to\infty} TE(\hat{\alpha} - \alpha)'$	Unknown mean: $\lim_{T\to\infty} TE(\hat{\alpha}^* - \alpha)'$
1	$(-2\alpha_1)$	$(1 - 3\alpha_1)$
2	$(-\alpha_1, 1 - 3\alpha_2)$	$(1 - \alpha_1 - \alpha_2, 2 - 4\alpha_2)$
3	$(-\alpha_1 - \alpha_3, 1 - 3\alpha_2, -4\alpha_3)$	$(1 - \alpha_1 - 2\alpha_3, 2 + \alpha_1 - 4\alpha_2 - \alpha_3, 1 - 5\alpha_3)$
4	$(-\alpha_1, 1 - 2\alpha_2 - \alpha_4, \alpha_1 - 4\alpha_3, 1 - 5\alpha_4)$	$(1 - \alpha_1 - \alpha_4, 2 + \alpha_1 - 2\alpha_2 - \alpha_3 - 2\alpha_4, 1 + 2\alpha_1 - 5\alpha_3 - \alpha_4, 2 - 6\alpha_4)$
5	$(-\alpha_1 - \alpha_5, 1 - 2\alpha_2 - \alpha_4, \alpha_1 - 4\alpha_3 - \alpha_5, 1 - 5\alpha_4, -6\alpha_5)$	$(1 - \alpha_1 - 2\alpha_5, 2 + \alpha_1 - 2\alpha_2 - 2\alpha_4 - \alpha_5,$ $1 + 2\alpha_1 + \alpha_2 - 5\alpha_3 - \alpha_4 - 2\alpha_5, 2 + \alpha_1 - 6\alpha_4 - \alpha_5, 1 - 7\alpha_5)$
6	$(-\alpha_1, 1 - 2\alpha_2 - \alpha_6, \alpha_1 - 3\alpha_3 - \alpha_5, 1 + \alpha_2 - 5\alpha_4 - \alpha_6,$ $\alpha_1 - 6\alpha_5, 1 - 7\alpha_6)$	$(1 - \alpha_1 - \alpha_6, 2 + \alpha_1 - 2\alpha_2 - \alpha_5 - 2\alpha_6,$ $1 + 2\alpha_1 + \alpha_2 - 3\alpha_3 - \alpha_4 - 2\alpha_5 - \alpha_6, 2 + \alpha_1 + 2\alpha_2 - 6\alpha_4 - \alpha_5 - 2\alpha_6,$ $1 + 2\alpha_1 - 7\alpha_5 - \alpha_6, 2 - 8\alpha_6)$

Walker (unpublished manuscript) gave it for unknown mean. For $p = 3$ Yamamoto and Kunitomo (1984) listed the first-order bias for both mean cases, but their expressions for the second component when the mean is known and for all components when the mean is unknown are in error.

6. THE BIAS OF THE YULE–WALKER ESTIMATOR

The development in Section 3 shows that the $O(1/T)$ bias for Yule–Walker estimators is obtained by adding the vector (3.7) to the least-squares results. Thus, when there is no mean adjustment, the $O(1/T)$ bias of the Yule–Walker estimator is the sum of (3.7), (5.1), and (5.3). When there is correction for the mean, (5.4) is a fourth summand.

Thus for $p = 1$ we find

$$\lim_{T\to\infty} TE(\tilde{\alpha}_1 - \alpha_1) = -3\alpha_1,$$

$$\lim_{T\to\infty} TE(\tilde{\alpha}_1^* - \alpha_1) = 1 - 4\alpha_1,$$

and for $p = 2$ we obtain (after some simplification)

$$\lim_{T\to\infty} TE(\tilde{\alpha} - \alpha) = \begin{bmatrix} -2\alpha_1 \\ 1 - 3\alpha_2 \end{bmatrix} - \frac{2\alpha_2(1 + \alpha_2)}{(1 + \alpha_2)^2 - \alpha_1^2} \begin{bmatrix} \alpha_1 \\ 1 + \alpha_2 \end{bmatrix}$$

and

$$\lim_{T\to\infty} TE(\tilde{\alpha}^* - \alpha) = \begin{bmatrix} 1 - 2\alpha_1 - \alpha_2 \\ 2 - 4\alpha_2 \end{bmatrix} - \frac{2\alpha_2(1 + \alpha_2)}{(1 + \alpha_2)^2 - \alpha_1^2} \begin{bmatrix} \alpha_1 \\ 1 + \alpha_2 \end{bmatrix}.$$

Tjøstheim and Paulsen (1983) gave these expressions for $p = 2$ in terms of the zeros of the associated polynomial $A(z)$. They also reported the result for $\tilde{\alpha}_1^*$. Unlike the least-squares case, the $O(1/T)$ bias vector for the Yule–Walker estimator is a rational function of the parameters, and denominator terms are small when the zeros of $A(z)$ are close to 1 in absolute value.

7. DISCUSSION

We have examined the bias in autoregressive estimation for least-squares and Yule–Walker estimation methods. Other estimation techniques may be obtained by changing the specification of the estimator of the covariances r_j ($j = 0, \pm 1, \ldots, \pm p$). The differences between the methods are concerned with treatment of end values in the observed time series.

It is evident from the rational structure of the bias expressions for the Yule–Walker estimator that it tends to have greater bias than the least-squares estimator. This was explored numerically for some second-order models by Tjøstheim and Paulsen (1983). The large bias for the Yule–Walker estimator becomes more pronounced as the zeros of the polynomial $A(z)$ move closer to the circumference of the unit circle. The estimator is stable in the sense that the zeros of the estimator of $A(z)$ are necessarily greater than 1 in magnitude, a feature not shared by the least-squares estimator. One of the prices of stability is greater bias. A factor responsible for the lower bias of the least-squares estimator is the use of summations, such as at (2.3), with the same number of summands for all entries in R and r. Tjøstheim and Paulsen (1983) made use of this observation to define a modified Yule–Walker estimator with less bias than the Yule–Walker estimator.

If the model used for estimation is a misspecification, our results do not hold except in the special case when the true model is autoregressive and its order is overspecified. One simply replaces each of the unneeded high-order autoregressive coefficients by 0 in the bias expressions derived here. In the least-squares case it is possible to give simple expressions for the added contribution to the $O(1/T)$ bias stemming from overspecification. Let m denote the true autoregressive order and $p (> m)$ the fitted order. Define the $p \times 1$ vectors $\gamma_{m,k}$ as

$$\gamma_{m,k} = (0, \ldots, 0, \alpha_m, \alpha_{m-1}, \ldots, \alpha_1, 1, 0, \ldots, 0)', \qquad m + k + 1 \leqq p,$$

where there are $p - m - k - 1$ 0s at the top and k 0s at the bottom. If there is no mean correction, the additional bias contribution from overspecification is

$$\sum_{j=0}^{[(1/2)(p-m-1)]} \gamma_{m,2j}, \qquad p \text{ even},$$

and

$$\sum_{j=0}^{[(1/2)(p-m-2)]} \gamma_{m,2j+1}, \qquad p \text{ odd},$$

where $[x]$ denotes the integer part of x and $\sum_{j=0}^{-1} \gamma_{m,2j+1}$ designates the $p \times 1$ vector with each element equal to 0 (which occurs for p odd and $m = p - 1$). If there is mean correction, one further adds the sum $\sum_{j=0}^{p-m-1} \gamma_{m,j}$, which arises from (5.4).

For Yule–Walker estimation the additional bias arising from overspecification is the least-squares additional bias plus a contribution from (3.7).

The simplicity of the bias expressions derived in this article suggests several areas for further study. One is the effect of the bias on the stationarity of the estimated model. The formulas in Table 1 show that there are cases in which the first-order bias moves zeros of the polynomial associated with the estimated coefficients closer to the unit circle. The growth of the coefficient of α_p in the bias of the least-squares estimator also leads to questions about conditions leading to large bias, especially for large p. If we let z_j ($j = 1, \ldots, p$) denote the zeros of $z^{-p}A(z^{-1})$, which are less than 1 in absolute value, then

$$\alpha_j = (-1)^j \sum_{k_1<k_2<\cdots<k_j} z_{k_1} \cdots z_{k_j}.$$

Thus α_p tends to get smaller with increasing p and moderates the growth of the coefficient in the bias expression. In contrast, coefficients near the center of the vector α are sums of increasing numbers of products and may become large as p increases. Estimators of these often have substantial bias. Large bias relative to the size of the coefficient generally occurs when the coefficient is rather small. For the least-squares estimator of α_p with p even, one can make the relative bias arbitrarily large by simply making α_p approach 0.

Some examples of the bias calculations suggest when these approximations are effective. In Table 2 we give the first-order bias for least-squares and Yule–Walker procedures for each of four fourth-order autoregressive processes with complex roots. These results are for the more realistic case of an unknown mean; results for the case of known mean are very similar. The roots are $\rho e^{\pm i\theta}$ with $\theta = 2\pi/P$, and the apparent period P is 5 and 8 in two series and 24 and 29 in the other two. The magnitudes of the roots are either .3 and .5 or .6 and .8. Thus in none of the four cases is the process particularly close to nonstationarity. Periods 24 and 29 appear in the variable-star data described by Bloomfield (1976) and lead to a spectrum with two adjacent peaks. Values of the coefficients of these processes appear in Table 3.

The first-order bias approximations for the least-squares

Table 2. T Times the First-Order Bias for Least-Squares (LS) $\{TE(\hat{\alpha}^* - \alpha)\}$ and Yule–Walker (YW) $\{TE(\tilde{\alpha}^* - \alpha)\}$ Estimators With Mean Correction in Several Fourth-Order Autoregressive Processes With Two Pairs of Complex Roots, $\rho e^{\pm i\theta}$ ($\theta = 2\pi$/period)

		Bias			
Model	Estimator	α_1	α_2	α_3	α_4
1	LS	1.87	.23	−.26	1.87
	YW	3.97	−2.07	.91	1.53
2	LS	2.27	−2.16	.99	.62
	YW	18.18	−31.16	26.68	−9.56
3	LS	2.53	−1.18	−.97	1.87
	YW	17.40	−26.86	16.34	−2.98
4	LS	3.49	−5.50	1.85	.62
	YW	2,554.89	−6,966.74	6,668.03	−2,242.67

Table 3. Model Parameters for Several Fourth-Order Autoregressive Processes With Two Pairs of Complex Roots, $\rho e^{\pm i\theta}$ ($\theta = 2\pi$/period)

Model	Period	ρ	α_1	α_2	α_3	α_4
1	5, 8	.3, .5	−.89	.47	−.11	.02
2	5, 8	.6, .8	−1.50	1.42	−.64	.23
3	24, 29	.3, .5	−1.56	.91	−.23	.02
4	24, 29	.6, .8	−2.72	2.81	−1.30	.23

estimator appear reasonable, and we have confirmed their accuracy with some exploratory simulations of Gaussian series. The least-squares approximations are generally within about two standard errors of the simulated values, once T is on the order of 100–200 observations. On the other hand, the Yule–Walker approximation is clearly not reliable. With periods 24 and 29 and magnitudes .6 and .8, $\alpha_1 = -2.72$; the first-order bias term is about 26 for $\tilde{\alpha}_1^*$ when the series length is 100. By comparison, a simulated estimate of the bias of $\tilde{\alpha}_1^*$ is 1.145 (standard error, .003). The size of the first-order bias term stems from using biased covariance estimators in the Yule–Walker procedure. Since the first five covariances for this process (with $\sigma^2 = 1$) are 371, 362, 338, 302, and 259, the contribution to the bias from (3.7) dominates the approximation. A more accurate bias approximation would seem to require using the guaranteed stationarity of the Yule–Walker estimator. Our Taylor expansion does not utilize this constraint and can err considerably when covariances are large relative to σ^2.

[*Received July 1987. Revised March 1988.*]

REFERENCES

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

Bloomfield, P. (1976), *Fourier Analysis of Time Series: An Introduction,* New York: John Wiley.

Fuller, W. A., and Hasza, D. P. (1981), "Properties of Predictors for Autoregressive Time Series," *Journal of the American Statistical Association,* 76, 155–161.

Granger, C. W. J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica,* 37, 424–438.

Kendall, M. G. (1954), "Note on Bias in the Estimation of Autocorrelation," *Biometrika,* 41, 403–404.

Kunitomo, N., and Yamamoto, T. (1985), "Properties of Predictors in Misspecified Autoregressive Time Series Models," *Journal of the American Statistical Association,* 80, 941–950.

Lewis, R. A., and Reinsel, G. C. (1988), "Prediction Error of Multivariate Time Series With Mis-specified Models," *Journal of Time Series Analysis,* 9, 43–57.

Lysne, D., and Tjøstheim, D. (1987), "Loss of Spectral Peaks in Autoregressive Spectral Estimation," *Biometrika,* 74, 200–206.

Marriott, F. H. C., and Pope, J. A. (1954), "Bias in the Estimation of Autocorrelations," *Biometrika,* 41, 390–402.

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

Shaman, P., and Stine, R. A. (1987), "The Bias of Coefficient Estimators for Autoregressive Models," technical report, University of Pennsylvania, Dept. of Statistics.

Shenton, L. R., and Johnson, W. L. (1965), "Moments of a Serial Correlation Coefficient," *Journal of the Royal Statistical Society,* Ser. B, 27, 308–320.

Sims, C. A. (1972), "Money, Income, and Causality," *American Economic Review,* 62, 540–552.

Stine, R. A. (1982), "Prediction Intervals for Time Series," unpublished Ph.D. dissertation, Princeton University, Dept. of Statistics.

——— (1987), "Estimating Properties of Autoregressive Forecasts," *Journal of the American Statistical Association,* 82, 1072–1078.

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.

Tjøstheim, D., and Paulsen, J. (1983), "Bias of Some Commonly Used Time Series Estimates," *Biometrika,* 70, 389–399; Corrigendum (1984), 71, 656.

White, J. S. (1961), "Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient," *Biometrika,* 48, 85–94.

Yamamoto, T., and Kunitomo, N. (1984), "Asymptotic Bias of the Least Squares Estimator for Multivariate Autoregressive Models," *Annals of the Institute of Statistical Mathematics,* 36, 419–430.