The Annals of Statistics
1989, Vol. 17, No. 3, 1275–1284

A FIXED POINT CHARACTERIZATION FOR BIAS OF AUTOREGRESSIVE ESTIMATORS

BY ROBERT A. STINE AND PAUL SHAMAN

University of Pennsylvania

Least squares estimators of the coefficients of an autoregression of known, finite order are biased to order $1/T$, where T is the sample length, unless the observed time series is generated by a unique model for that order. The coefficients of this special model are the fixed point of a linear mapping defined by the bias of the least squares estimator. Separate results are given for models with known mean and unknown mean. The "fixed point models" for different orders of autoregression are least squares approximations to an infinite-order autoregression which is unique but for arbitrary scaling. Explicit expressions are given for the coefficients of the fixed point models at each order. The autocorrelation function and spectral density of the underlying infinite-order process are also presented. Numerical calculations suggest similar properties hold for Yule–Walker estimators. Implications for bootstrapping autoregressive models are discussed.

1. Introduction. The effect of bias on least squares estimators in autoregressions has often been studied, but seldom fully understood. The bias expressions of Bhansali (1981) and Tjøstheim and Paulsen (1983) are sufficiently complex that one cannot tell, for example, if bias is moving the estimated model closer to nonstationarity. Shaman and Stine (1988) recently showed that the bias of least squares estimators for models of known, finite order is a linear function of the unknown model coefficients, to order $1/T$. We employ a matrix representation of the bias to develop our results, which are illustrated by the following example. If α_1, α_2 and α_3 are the coefficients of an autoregressive model of order 3 with known mean, then the bias to order $1/T$ of the least squares estimator is $(-\alpha_1 - \alpha_3, 1 - 3\alpha_2, -4\alpha_3)/T$. Only the autoregressive model with coefficients $(0, \frac{1}{3}, 0)$ has no bias to this order of approximation. This simple example generalizes to autoregressive models of any known, finite order. In fact, the finite-order models for which least squares estimation is unbiased are projections of an infinite-order process that is unique up to scale.

The coefficient vector of each of these projections is the fixed point of the linear mapping defined by the bias approximation. If the time series used in estimation is not generated by this unique model, then the bias of least squares pulls the estimator closer to its coefficients. The fixed point coefficients may define a model whose roots are closer to the region of nonstationarity. Thus the tendency for bias to shrink the least squares coefficient estimator toward 0 in a first-order model with known mean does not extend to higher-order models.

The following section presents notation, estimators and the linear bias mapping. Section 3 summarizes the results, with the proofs in Section 4. The

Received March 1988; revised October 1988.
AMS 1980 *subject classification*. Primary 62M10.
Key words and phrases. Autoregressive process, least squares estimator, bias, fixed point, contraction, Durbin–Levinson recursion, Yule–Walker estimator.

discussion of Section 5 describes the effect on Yule–Walker estimators and implications for bootstrap methods.

2. The bias of least squares estimators. Let $\{y_t\}$ be a discrete time autoregression of known, finite order p,

$$\sum_{j=0}^{p} \alpha_{jp}(y_{t-j} - \mu) = \varepsilon_t, \qquad t = \cdots, -2, -1, 0, 1, 2, \ldots, \tag{2.1}$$

where $\mu = E(y_t)$ and $\alpha_{0p} = 1$. Observations from this process are denoted $y = (y_1, \ldots, y_T)'$ and the vector of coefficients is $\alpha_p = (\alpha_{0p}, \alpha_{1p}, \ldots, \alpha_{pp})'$. The error terms $\{\varepsilon_t\}$ are iid with mean 0 and variance σ^2. The zeros of $\sum_{j=0}^{p} \alpha_{jp} z^{p-j}$ lie strictly inside the unit circle so that the process $\{y_t\}$ is stationary.

The least squares estimator is the solution of an approximation to the Yule–Walker equations. When μ is known, the covariance estimators are

$$c_{ij} = \sum_{t=p+1}^{T} (y_{t-i} - \mu)(y_{t-j} - \mu)/(T-p), \qquad i, j = 0, 1, \ldots, p.$$

Let C_p be the $p \times p$ matrix with elements $\{c_{ij},\ i, j = 1, 2, \ldots, p\}$ and define $c_p = (c_{01}, c_{02}, \ldots, c_{0p})'$. The least squares estimator of α_p is then $\hat{\alpha}_p = (1, -c_p' C_p^{-1})'$. When μ is not known, the covariances are estimated by

$$c_{ij}^* = \sum_{t=p+1}^{T} (y_{t-i} - \bar{y}_i)(y_{t-j} - \bar{y}_j)/(T-p), \qquad i, j = 0, 1, \ldots, p,$$

where $\bar{y}_i = \sum_{t=p+1}^{T} y_{t-i}/(T-p)$ and C_p^* and c_p^* are defined analogously. The least squares estimator is thus $\hat{\alpha}_p^* = (1, -c_p^{*\prime} C_p^{*-1})'$.

An additional assumption is needed to ensure the validity of the approximations to the bias used in this paper. We assume [see Lewis and Reinsel (1988)] that the errors $\{\varepsilon_t\}$ have finite moment of order 16 and that

$$E\left(\|C_p^{-1} - \Gamma_p^{-1}\|^k\right) = O(1) \quad \text{as } T \to \infty \text{ for } k \le 8, \tag{2.2}$$

where $\|A\|$ is the matrix norm given by the largest absolute eigenvalue of A and $\Gamma_p = E(C_p)$ with elements $\gamma_{ij} = \gamma_{|i-j|} = \mathrm{Cov}(y_t, y_{t+|i-j|})$, $i, j = 1, \ldots, p$. See also Bhansali (1981), whose assumption (A3) is stronger than (2.2).

The approximate bias of $\hat{\alpha}_p$ has a simple, linear form. If μ is known, $E(\hat{\alpha}_p) = (I - B_p/T)\alpha_p + o(1/T)$, where I is the $(p+1) \times (p+1)$ identity matrix and the $(p+1) \times (p+1)$ matrix $B_p = B_{1p} + B_{2p}$. The matrix $B_{1p} = \mathrm{diag}(0, 1, 2, \ldots, p)$. The columns of B_{2p} are arrangements of vectors e_j or d_j where e_j is $(p+1) \times 1$ with 1's in rows $j+3, j+5, \ldots, p+1-j$ and 0's elsewhere and d_j is $(p+1) \times 1$ with 1's in rows $j+2, j+4, \ldots, p+1-j$ and 0's elsewhere. When p is even, $B_{2p} = [-e_0, -e_1, \ldots, -e_{p/2-1}, 0, e_{p/2-1}, \ldots, e_1, e_0]$, and when p is odd, $B_{2p} = [-d_1, -d_2, \ldots, -d_{(p-1)/2}, 0, d_{(p-1)/2}, \ldots,$

$d_1, d_0]$. For example,

$$B_{10} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 & 4 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & -1 & 0 & -1 & 0 & 5 & 0 & 1 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 & -1 & 0 & 7 & 0 & 1 & 0 & 1 \\ 0 & -1 & 0 & -1 & 0 & 0 & 0 & 8 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 9 & 0 & 1 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 10 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 11 \end{bmatrix}.$$

Estimation of μ requires adding a term to the bias. If the $O(1/T)$ bias of $\hat{\alpha}_p^*$ is $-B_p^*\alpha_p/T$, then $B_p^* = B_p + B_{3p}$, where the (i, j) element of B_{3p} is -1 for $j < i \le p - j$, 1 for $p - j < i \le j$ and 0 elsewhere. See Shaman and Stine (1988), Section 5 for details.

3. A fixed point characterization for the bias.

The first theorem shows for each order of autoregression that only one model exists for which least squares is unbiased to terms of order $1/T$.

THEOREM 1. *If $T > (p + 1)/2$, the expectation mapping $I - B_p/T$ is a contraction with a unique fixed point $\tilde{\alpha}_p$ satisfying $(I - B_p/T)\tilde{\alpha}_p = \tilde{\alpha}_p$ with first coordinate $\tilde{\alpha}_{0p} = 1$. Similarly, if $T > (p + 2)/2$, $I - B_p^*/T$ is a contraction with a unique fixed point $\tilde{\alpha}_p^*$ having first coordinate $\tilde{\alpha}_{0p}^* = 1$.*

The next two theorems give formulas for the coefficients of these models.

THEOREM 2. *The autoregressive model of even order p with known mean for which least squares is unbiased to order $1/T$ has coefficients*

$$\tilde{\alpha}_{2k,p} = \prod_{j=1}^{k} \frac{(p + 2 - 2j)/(2j)}{(p + 3 - 2j)/(2j - 1)}, \qquad k = 1, \dots, p/2, \tag{3.1}$$

and 0 *otherwise. If the order p is odd, the coefficients are*

$$\tilde{\alpha}_{2k,p} = \prod_{j=1}^{k} \frac{(p + 1 - 2j)/(2j)}{(p + 2 - 2j)/(2j - 1)}, \qquad k = 1, \dots, (p - 1)/2, \tag{3.2}$$

and 0 *otherwise.*

THEOREM 3. *The autoregressive model with unknown mean for which least squares is unbiased to order $1/T$ has coefficients*

$$\tilde{\alpha}_{kp}^* = \prod_{j=1}^{k} \frac{(p + 2 - j - \delta_j)/(j + \delta_j)}{(p + 3 - j)/(j + 1)}, \qquad k = 1, \dots, p, \tag{3.3}$$

when p is even, where $\delta_j = 1$ *if j is odd and is* 0 *otherwise. If p is odd, the coefficients are*

$$\tilde{\alpha}^*_{kp} = \prod_{j=1}^{k} \frac{(p+1-j)/j}{(p+2-j+\delta_j)/(j+1-\delta_j)}, \qquad k = 1, \ldots, p. \tag{3.4}$$

The nonzero elements of $\tilde{\alpha}_p$ are monotonically decreasing; those of $\tilde{\alpha}^*_p$ are not. As p increases, some elements of $\tilde{\alpha}^*_p$ become greater than 1.

Our next pair of theorems shows that the autoregressive processes defined by the fixed point coefficients are stationary. Define the polynomials

$$\tilde{\mathscr{A}}_p(z) = \sum_{j=0}^{p} \tilde{\alpha}_{jp} z^{p-j} = \sum_{k=0}^{[p/2]} \tilde{\alpha}_{2k,p} z^{p-2k}, \tag{3.5}$$

$$\tilde{\mathscr{A}}^*_p(z) = \sum_{j=0}^{p} \tilde{\alpha}^*_{jp} z^{p-j}, \tag{3.6}$$

where $[x]$ denotes the greatest integer less than or equal to x.

THEOREM 4. *The zeros* z_j *of* $\tilde{\mathscr{A}}_p(z)$ *satisfy* $|z_j| < 1$, $j = 1, \ldots, p$.

THEOREM 5. *The zeros* z_j *of* $\tilde{\mathscr{A}}^*_p(z)$ *satisfy* $|z_j| < 1$, $j = 1, \ldots, p$.

Figure 1 shows the location of the fixed points for orders $p = 4$ and 20. Estimation of the mean shifts the zeros toward -1 and the zeros increase in norm as the order increases. Theorem 1 implies for any coefficient vector α that $\lim_{j \to \infty}(I - B_p/T)^j \alpha = \tilde{\alpha}_p$. To interpret this convergence, begin with a time series generated by a pth order autoregressive model with coefficients α so that

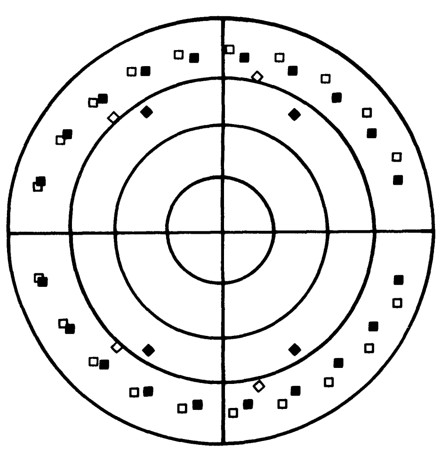

FIG. 1. *Zeros of the fixed point models for order* 4 *with mean known* (◆) *and unknown* (◇) *and for order* 20 *with mean known* (■) *and unknown* (□).

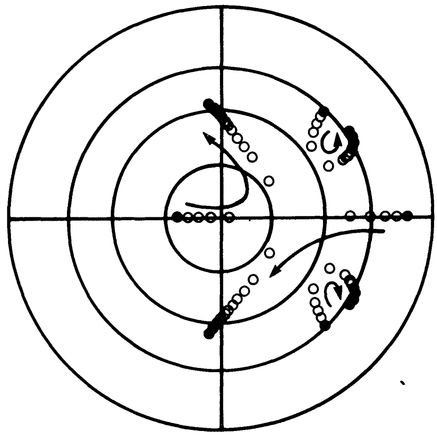

FIG. 2. *Zeros defined by iterating the bias mapping* $I - B_4/50$ *twenty times with* $AR(4)$ *models having known mean. The zeros defined by the initial model are the filled circles at* -0.2, 0.9 *and* $0.5 \pm i0.5$ *and the arrows indicate the movement of the zeros.*

$E\hat{\alpha}_p = (I - B_p/T)\alpha + o(1/T)$. Let $(I - B_p/T)\alpha$ become the vector of coefficients of a second generating model. The expected value of the least squares estimator applied to data from this second model is $(I - B_p/T)^2\alpha_p + o(1/T)$. Continuing in this fashion, bias eventually pulls the coefficients to the fixed point $\tilde{\alpha}_p$. As Figure 2 shows, the movement of the zeros associated with this sequence of coefficients toward those of the fixed point is rather chaotic and depends upon the location of the zeros of the underlying model. Notice that bias favors models with complex zeros. In this example, two real zeros merge and form a complex pair at considerable distance from each other.

Our final theorem describes how these fixed point models are related for different orders of autoregression. A single infinite-order autoregression generates all of the fixed point models.

THEOREM 6. *The fixed point models defined by* $\tilde{\alpha}_p$, $p = 1, 2, \ldots$, *are least squares approximations to an infinite-order autoregression with correlation function*

$$\tilde{\rho}_j = \begin{cases} 0, & j \text{ odd}, \\ -1/(j^2 - 1), & j \text{ even}. \end{cases} \tag{3.7}$$

The models defined by $\tilde{\alpha}_p^*$ *are the least squares approximations to an infinite-order autoregression with correlation function*

$$\tilde{\rho}_j^* = \begin{cases} 1/(j^2 - 4), & j \text{ odd}, \\ -1/(j^2 - 1), & j \text{ even}. \end{cases} \tag{3.8}$$

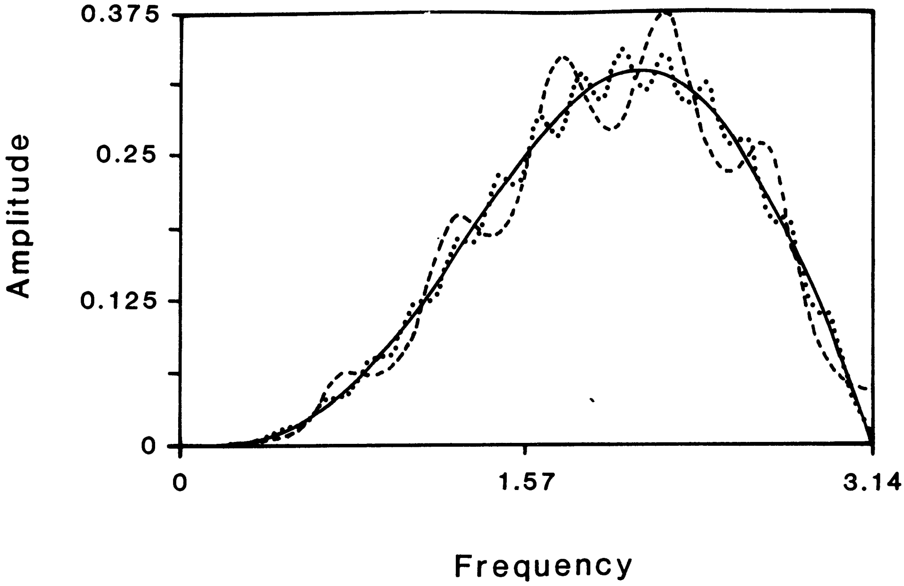

FIG. 3. *Spectral densities of fixed point models with mean unknown and orders* 10 (---), 30 (· · ·) *and infinite* (—).

Since Theorem 6 only specifies the correlation function, the underlying process is unique up to choice of scale. The following corollary gives the spectral density of the underlying process.

COROLLARY 1. *The spectral density of the process defined by* (3.7) *is* ($\gamma_0 = 1$)

$$\tilde{g}(\omega) = |\sin \omega|/4, \qquad -\pi \leq \omega \leq \pi. \tag{3.9}$$

The spectral density defined by the mean-unknown case (3.8) *is* ($\gamma_0 = 1$)

$$\tilde{g}^*(\omega) = |\sin \omega - (\sin 2\omega)/2|/4, \qquad -\pi \leq \omega \leq \pi. \tag{3.10}$$

Notice that (3.9) is not differentiable at 0. This corollary is easily verified by computing the Fourier coefficients. Figure 3 shows $\tilde{g}^*$ and the spectral densities of the finite-order approximations defined by $\tilde{\alpha}_p^*$ for $p = 10$ and 30. Plots of $\tilde{g}$ and its approximations are similar, though symmetric about $\pi/2$.

4. Proofs. Proofs of the first two lemmas rely upon the triangular structure of the matrices B_p and B_p^* that define the $O(1/T)$ component of the bias.

LEMMA 1. *The matrices B_p and B_p^* are similar to lower triangular matrices.*

LEMMA 2. *The eigenvalues of B_p and B_p^* are their diagonal elements.*

PROOF OF THEOREM 1. If $T > (p+1)/2$, $I - B_p/T$ has a single eigenvalue of 1, with the remaining less than 1 in absolute value. The normalization $\alpha_{0p} \equiv 1$ yields a unique solution which is the fixed point coefficient vector. A similar argument applies if the mean is unknown. □

The next two lemmas simplify the proof of Theorem 2.

LEMMA 3. *For p odd, $\tilde{\alpha}_{jp} = \tilde{\alpha}_{j,p-1}$, $j = 0, 1, \dots, p-1$, and $\tilde{\alpha}_{pp} = 0$.*

LEMMA 4. *For j odd, $\tilde{\alpha}_{jp} = 0$.*

PROOF OF THEOREM 2. Lemmas 3 and 4 imply that it is sufficient to consider even-indexed coefficients with p even. Considering, in order, $\tilde{\alpha}_{pp}, \tilde{\alpha}_{2p}, \tilde{\alpha}_{p-2,p}, \tilde{\alpha}_{4p}, \dots,$ we obtain $\tilde{\alpha}_{p-2k,p} = \tilde{\alpha}_{2k,p}(2k+1)/(p+1-2k)$, $k = 0, 1, \dots, [(p-2)/4]$, and $\tilde{\alpha}_{2k,p} = \tilde{\alpha}_{p+2-2k,p}(p-2k)/2k$, $k = 1, 2, \dots, [p/4]$, and (3.1) and (3.2) follow. □

PROOF OF THEOREM 3. Again use an alternating recursion. If p is even, $\tilde{\alpha}^*_{jp} = \tilde{\alpha}^*_{p+1-j,p}(p+2-j-\delta_j)/(j+\delta_j)$, $j = 1, 2, \dots, p/2$, and $\tilde{\alpha}^*_{p-j,p} = \tilde{\alpha}^*_{jp}(j+2)/(p+2-j)$, $j = 0, 1, \dots, p/2 - 1$, where $\delta_j = 1$ if j is odd and is 0 otherwise. If p is odd, $\tilde{\alpha}^*_{jp} = \tilde{\alpha}^*_{p+1-j,p}(p+1-j)/j$, $j = 1, 2, \dots, (p-1)/2$, and $\tilde{\alpha}^*_{p-j,p} = \tilde{\alpha}^*_{jp}(j+1+\delta_j)/(p+2-j-\delta_j)$, $j = 0, 1, \dots, (p-1)/2$. □

PROOF OF THEOREM 4. Lemma 3 implies that it is sufficient to consider p even. Use Lemma 4 and define $\mathscr{B}_p(z^2) = \mathscr{A}_p(z)$. Theorem 2 shows that the coefficients of $\mathscr{B}_p$ are strictly decreasing and the Enestrom–Kakeya theorem [Marden (1966), Section 30] implies that its zeros lie strictly inside the unit circle. Since the zeros of $\mathscr{A}_p(z)$ are square roots of the zeros of $\mathscr{B}_p(z)$, they also lie inside the unit circle. □

The coefficients of $\tilde{\alpha}^*_p$ do not satisfy the monotonicity needed to apply the Enestrom–Kakeya theorem and we give an inductive proof for Theorem 5. This proof requires the following lemma which is important in understanding the relationship among the fixed points for different orders of autoregression.

LEMMA 5. *The coefficients of the fixed point with unknown mean satisfy*

$$\tilde{\alpha}^*_{kp} - \tilde{\alpha}^*_{k,p-1} = \tilde{\alpha}^*_{pp}\tilde{\alpha}^*_{p-k,p-1}, \qquad k = 1, \dots, p-1,\ p = 2, 3, \dots. \tag{4.1}$$

PROOF. By Theorem 3, the $\tilde{\alpha}^*_{kp}$ are products $v_{1p} \cdots v_{kp}$, where for j odd, $v_{jp} = (p+1-j)/(p+3-j)$. For j even, $v_{jp} = [(j+1)(p+2-j)]/[j(p1+3-j)]$ if p is even and $v_{jp} = [(j+1)(p+1-j)]/[j(p+2-j)]$ if p is odd. Thus if j is even, $v_{jp}v_{p+2-j,p} = 1$ if p is even and $v_{jp}v_{p+1-j,p} = 1$ if p is odd. We establish (4.1) for p even; the details of the proof for p odd are similar.

Suppose k is even. Then

$$\tilde{\alpha}^*_{kp} - \tilde{\alpha}^*_{k,p-1} = (v_{1p} \cdots v_{k-1,p})(v_{2p} \cdots v_{kp}) - (v_{1,p-1} \cdots v_{k-1,p-1})(v_{2,p-1} \cdots v_{k,p-1})$$

becomes

$$\begin{aligned} &\frac{p+2-k}{p+2} \prod_{j=1}^{k/2} \frac{2j+1}{2j} \frac{p+2-2j}{p+3-2j} \\ &- \frac{p+1-k}{p+1} \prod_{j=1}^{k/2} \frac{2j+1}{2j} \frac{p-2j}{p+1-2j} \qquad (4.2) \\ &= \frac{2}{p+2} \frac{k+1}{p+1} \prod_{j=1}^{k/2-1} \frac{2j+1}{2j} \frac{p-2j}{p+1-2j}, \end{aligned}$$

where $\prod_{j=1}^{0} = 1$. Theorem 3 gives $\tilde{\alpha}^*_{pp} = 2/(p+2)$ and the remaining factors on the last line of (4.2) are (with $v_{2,p-1} \cdots v_{k-2,p-1} = 1$ for $k = 2$)

$$\frac{k+1}{p+1} v_{2,p-1} \cdots v_{k-2,p-1} = (v_{1,p-1} \cdots v_{p-k-1,p-1})(v_{2,p-1} \cdots v_{p-k,p-1}) = \tilde{\alpha}^*_{p-k,p-1}.$$

If k is odd, the left side of (4.1) is

$$(v_{1p} \cdots v_{kp})(v_{2p} \cdots v_{k-1,p}) - (v_{1,p-1} \cdots v_{k,p-1})(v_{2,p-1} \cdots v_{k-1,p-1}) = \frac{2}{p+2} \frac{k}{p+1} \prod_{j=1}^{(k-1)/2} \frac{2j+1}{2j} \frac{p-2j}{p+1-2j} = \tilde{\alpha}^*_{pp} \tilde{\alpha}^*_{p-k,p-1}. \qquad \square$$

PROOF OF THEOREM 5. We use induction on the order p of the autoregression. For $p = 1$, the zero is $-\frac{1}{3}$. Assume that the zeros of $\tilde{\mathscr{A}}^*_{p-1}(z)$ lie inside the unit circle. Lemma 5 implies $|\tilde{\mathscr{A}}^*_p(z) - z\tilde{\mathscr{A}}^*_{p-1}(z)| < |z\tilde{\mathscr{A}}^*_{p-1}(z)| = |\tilde{\mathscr{A}}^*_{p-1}(z)|$ for $|z| = 1$ and, by Rouché's theorem, the zeros of $\tilde{\mathscr{A}}^*_p(z)$ also lie inside the unit circle. $\square$

Equation (4.1) is the well-known Durbin–Levinson recursion [see, e.g., Brockwell and Davis (1987)]. Given a correlation function, this recursion produces the coefficients of the autoregressive models of orders $1, 2, \ldots,$ which minimize the expected squared error of one-step-ahead prediction. Thus, each fixed point model is the sequential approximation to some infinite-order model and we need only find the correlation function of the latter.

PROOF OF THEOREM 6. We sketch the argument for the case of known mean. The Yule–Walker equations are

$$\tilde{\rho}_p = -\sum_{k=1}^{p} \tilde{\alpha}_{kp} \tilde{\rho}_{p-k}, \qquad p = 1, 2, \ldots.$$

If p is odd, Lemma 4 implies the first half of (3.7) since the Yule–Walker equations pair odd-indexed coefficients, which are 0, with even-indexed correlations. For p even, one verifies the second half of (3.7) by summing $-\sum_{k=1}^{p}\tilde{\alpha}_{kp}\tilde{\rho}_{p-k}$ in reverse order. The partial sums are

$$-\sum_{k=(p/2)-j}^{p/2} \tilde{\alpha}_{2k,p}\tilde{\rho}_{p-2k} = -\frac{1\cdot 3\cdots(2j-1)(p-2)(p-4)\cdots(p-2j)}{2\cdot 4\cdots 2j(p+1)(p-1)\cdots(p-2j+1)},$$

$$j = 1,2,\ldots,(p/2)-1.$$

For $j = (p/2) - 1$, the factors cancel and this reduces to $-1/(p^2 - 1)$. □

5. Discussion and extensions. That least squares tends to "shrink" roots toward the origin is a part of the folklore of time series analysis. This description shows that the bias does not necessarily have such an effect. Our results show that the effect of bias, on average, is to move $\hat{\alpha}_p$ toward the fixed point. The zeros of the fixed point polynomials are inside the unit circle, but move close to the boundary as the order of autoregression increases.

Though one seldom has occasion to iterate the estimation process, bootstrapping does involve two or more such iterations. To bootstrap an autoregression, one initially obtains $\hat{\alpha}_p$ from the observed realization of $\{y_t\}$. One then uses this estimate in place of the true coefficients α_p to generate bootstrap replications of the observations. For the bootstrap series, $\hat{\alpha}_p$ becomes the "true" coefficient vector [see, e.g., Efron and Tibshirani (1986)]. Thus, when one computes estimates of α_p from the bootstrap series, the expected value of these is approximately $(I - B_p/T)^2\alpha_p$, not $(I - B_p/T)\alpha_p$. The bootstrap estimate of bias is approximately $-(B_p/T)(I - B_p/T)\alpha_p$, not $-B_p\alpha_p/T$.

Our results may be extended to polynomial trends. Suppose that the mean μ in (2.1) is replaced by the trend in time $\mu(t) = \sum_{j=0}^{k-1}\beta_j t^j$, where we use $k = 0$ to denote a known mean. If $\hat{\mu}(t)$ is the least squares estimator of $\mu(t)$, let $\tilde{\alpha}_p(k)$ be the fixed point vector associated with least squares estimation of an autoregression of order p in the residuals $y_t - \hat{\mu}(t)$. Then $B_p(k)\tilde{\alpha}_p(k) = 0$, where $B_p(k) = B_p + kB_{3p}$ for $k = 0, 1, 2, \ldots$. Thus, $\tilde{\alpha}_p = \tilde{\alpha}_p(0)$ and $\tilde{\alpha}_p^* = \tilde{\alpha}_p(1)$. As k increases, $\tilde{\alpha}_p(k)$ moves toward the vector with elements $\binom{p}{j}$, $j = 0, 1, \ldots, p$, and the zeros of the associated polynomial move toward -1. The asymmetric location of the zeros of $\tilde{\mathscr{A}}_p^*(z)$ in Figure 1 illustrates this effect. The limiting behavior of $\tilde{\alpha}_p(k)$ follows from arguments similar to those used above and Theorems 10.3.2 and 10.3.4 of Anderson (1971). Related results appear in Pantula and Fuller (1985).

The bias of the Yule–Walker estimator has nonlinear dependence on α_p. Numerical calculations based on the bias approximation in Shaman and Stine (1988) suggest that a unique fixed point again exists. The zeros associated with Yule–Walker estimation are also symmetric about the complex axis when μ is known, but lie closer to the origin than those defined by least squares. One should note that the $O(1/T)$ bias approximation is not particularly accurate for Yule–Walker estimators. However, Zhang (1988) has shown that an appropriately tapered Yule–Walker estimator has $O(1/T)$ bias and variance equal to those of the least squares estimator.

REFERENCES

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

BHANSALI, R. J. (1981). Effects of not knowing the order of an autoregressive process on the mean squared error of prediction—I. *J. Amer. Statist. Assoc.* **76** 588–597.

BROCKWELL, P. J. and DAVIS, R. A. (1987). *Time Series: Theory and Methods*. Springer, New York.

EFRON, B. and TIBSHIRANI, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with discussion). *Statist. Sci.* **1** 54–77.

LEWIS, R. A. and REINSEL, G. C. (1988). Prediction error of multivariate time series with mis-specified models. *J. Time Ser. Anal.* **9** 43–57.

MARDEN, M. (1966). *The Geometry of Polynomials*. Amer. Math. Soc., Providence, R.I.

PANTULA, S. G. and FULLER, W. A. (1985). Mean estimation bias in least squares estimation of autoregressive processes. *J. Econometrics* **27** 99–121.

SHAMAN, P and STINE, R. A. (1988). The bias of autoregressive coefficient estimators. *J. Amer. Statist. Assoc.* **83** 842–848.

TJØSTHEIM, D. and PAULSEN, J. (1983). Bias of some commonly-used time series estimates. *Biometrika* **70** 389–399. [Corrigendum: **71** 656 (1984).]

ZHANG, H.-C. (1988). Reduction of the asymptotic bias of autoregressive parameter estimators by tapering. Technical Report, Dept. Statistics, Univ. Pennsylvania.

DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104-6302