GENERALIZED LEVINSON-DURBIN SEQUENCES,

BINOMIAL COEFFICIENTS AND AUTOREGRESSIVE ESTIMATION

BY PAUL SHAMAN

University of Pennsylvania

For $\{ y_t \}$ a discrete time, second-order stationary process, the Levinson-Durbin recursion is used to determine the coefficients α_{jk}, j=1, … , k, of the best linear predictor of y_{k+1}, $\hat{y}_{k+1} = -\alpha_{1k} y_k - \cdots - \alpha_{kk} y_1$, best in the sense of minimizing the mean square error. The coefficients α_{jk} determine a Levinson-Durbin sequence. A generalized Levinson-Durbin sequence, a special case of a sequence of generalized binomial coefficients, is studied. Binomial coefficients form a generalized Levinson-Durbin sequence, and all generalized Levinson-Durbin sequences are shown to obey some summation formulas which generalize summations satisfied by binomial coefficients. The summation formulas are expressed in terms of the partial correlation sequence. Levinson-Durbin sequences arise in the construction of autoregressive model coefficient estimates for the Yule-Walker, tapered Yule-Walker, Burg and Kay estimators. The least squares autoregressive estimator, though, does not give rise to a Levinson-Durbin sequence. However, least squares fixed point processes, which yield least squares estimates of the coefficients unbiased to order $1/T$, where T is the sample length, can be combined to construct Levinson-Durbin sequences. By contrast, analogous Yule-Walker fixed point processes do not combine to construct Levinson-Durbin sequences. The least squares fixed point processes are studied when the mean of the process is a polynomial time trend that is estimated by least squares. For each degree of polynomial trend, the fixed point processes form a sequence of projections from an infinite order fixed point process. The correlation functions and spectral densities of these infinite order fixed point processes are derived for polynomial trends of degree up to 5.

AMS 2000 subject classifications. Primary 62M10; secondary 05A10, 11B65.

Key words and phrases. Bias, binomial coefficients, Burg estimator, fixed-point process, generalized binomial coefficients, generalized Levinson-Durbin sequence, Kay estimator, least squares estimator, Levinson-Durbin recursion, Levinson-Durbin sequence, minimum phase filter, partial correlations, Yule-Walker estimator.

1. Introduction. The Levinson-Durbin recursion has long been a fixture in time series analysis. It is commonly viewed in two contexts. One is that of prediction for a discrete time, second-order stationary process $\{ y_t \}$. Given $y_1, \dots , y_k$, for any $k \geq 1$, the recursion determines the coefficients α_{jk}, j=1, … , k, of the best linear predictor of y_{k+1},

(1) $$\hat{y}_{k+1} = -\alpha_{1k} y_k - \cdots - \alpha_{kk} y_1,$$

best in the sense of minimizing the mean square error. (Minus signs are used here because of the sign convention employed in this paper for the coefficients of autoregressive processes.) The recursion begins with specification of α_{11}, and at the nth stage one obtains $\alpha_{1n}, \dots , \alpha_{nn}$. The mean square error of the predictor is also specified at each step. Levinson (1947) devised the recursion to give a simple procedure for construction of the best linear predictor. His paper was reprinted as Appendix B to Norbert Wiener's monograph on time series (1949). Wiener's work had originally been issued in February 1942 as a classified governmental report. For some details of this history see Kailath (1974).

The second context for the recursion is that of estimation of the coefficients of an autoregressive model of finite order, given data $y_1, \dots , y_T$. The sample Yule-Walker equations are often used to construct an estimator of the coefficients. Bartlett [(1955), pages 264-265] and Durbin (1960) both developed the recursion as a simple method of solving the sample Yule-Walker equations, which are linear in the coefficients.

The recursion determines an infinite sequence α_{jn}, $j = 1, \dots, n$, $n = 1, 2, \dots$. In the first context the values $-\alpha_{nn}$, $n = 1, 2, \dots$, are the partial correlations of the process $\{ y_t \}$, known as the reflection coefficients in the engineering literature. In the second setting the $-\alpha_{nn}$'s form an estimate of the partial correlation function. In both contexts the partial correlations $-\alpha_{nn}$, $n = 1, 2, \ldots$, determine all of the values α_{jn}.

In this paper we study properties of the sequence produced by the Levinson-Durbin recursion, and we further generalize such a sequence.

DEFINITION 1. α_{jn}, j=1, … , n, n = 1, 2, … is a *Levinson-Durbin sequence* if the coefficients, all real-valued, satisfy

(2) $$\alpha_{j,n} = \alpha_{j,n\text{-}1} + \alpha_{nn}\,\alpha_{n\text{-}j\,,\,n\text{-}1}, \qquad j = 1, \ldots, n-1, \quad n = 2, 3, \ldots,$$

and

(3) $$|\alpha_{nn}| < 1, \qquad n = 1, 2, \ldots .$$

If (2) holds and the α_{nn}'s are not subject to (3), the α_{jn}'s are said to form a *generalized Levinson-Durbin sequence*.

The Levinson-Durbin recursion is given by (2) together with the $-\alpha_{nn}$'s defined to be the partial correlations of the process to be predicted, or the sample partial correlations of the time series realization to which an autoregressive model is to be fit. A generalized Levinson-Durbin sequence allows arbitrary specification of the α_{nn}'s. It is a special case of a sequence of generalized binomial coefficients; see, e.g., Ollerton and Shannon (2002). If (3) holds, then the sequence of $-\alpha_{nn}$'s forms the partial correlation function for some second-order stationary process. From (2), all of the α_{jn} values are determined by just the α_{nn}'s.

Binomial coefficients form a generalized Levinson-Durbin sequence. Levinson-Durbin sequences arise, e.g., from (i) Yule-Walker and tapered Yule-Walker estimation of the coefficients of an autoregressive process, (ii) fixed point models arising in least squares estimation of the autoregressive process coefficients, (iii) estimation of the autoregressive process coefficients by Burg's method (1975) and (iv) estimation of the autoregressive process coefficients by Kay's method (1983).

Binomial coefficients. If $\alpha_{nn} = 1$ for each n, then (2) generates the binomial coefficients. Taking into account the symmetric structure of binomial coefficients, we see that (2) is simply an expression of Pascal's triangle if $\alpha_{nn} = 1$ for each n. The binomial coefficients also arise as the limit of a sequence of fixed point models determined by least squares estimation of autoregressive process coefficients, as noted by Stine and Shaman (1989). This will be discussed in Section 3.

Yule-Walker estimation. The Yule-Walker estimator of the coefficients of an autoregressive process of known finite order p is determined by a Levinson-Durbin recursion which defines a sequence for which $\alpha_{jn} = \alpha_{jp}$, $j = 1, \ldots, p$, and $\alpha_{jn} = 0$, $j = p+1, \ldots, n$, for all $n > p$. The values $-\alpha_{nn}$, $n = 1, \ldots, p$, are the Yule-Walker sample partial correlations. For sample length T the order $1/T$ bias of the Yule-Walker estimator has been discussed by Tjøstheim and Paulsen (1983) and Shaman and Stine (1988). For each value of p numerical calculations show that there is a unique autoregressive process of order p (unique up to scale) for which the order $1/T$ bias of the Yule-Walker estimator is 0. This process is called a fixed point process because it is given by the fixed point of a contraction mapping. This result may be extended to the case where a polynomial trend in time is estimated by least squares and the Yule-Walker estimator is subsequently calculated from the trend residuals. The Yule-Walker fixed point processes differ according to the autoregressive order p and the degree of the estimated polynomial trend, and they are determined numerically by iterating the contraction mappings. Although the Yule-Walker estimator itself yields a Levinson-Durbin sequence, it is interesting that the Yule-Walker fixed point processes for a given degree of estimated polynomial trend do not combine to form a Levinson-Durbin sequence. These comments also hold for the tapered Yule-Walker estimator, with the proviso that the fixed point processes depend upon the specific data taper chosen. Dahlhaus (198?) and Zhang (1992) consider the tapered Yule-Walker estimator.

Least squares estimation. The least squares estimator of the coefficients of an autoregressive process of known finite order p does not correspond to a Levinson-Durbin sequence. The order $1/T$ bias of the least squares estimator of the coefficients of an autoregressive process of known finite order p has been derived by Tjøstheim and Paulsen (1983), Shaman and Stine (1988) and Pham (1993). The bias expression is linear in the autoregressive parameters and defines a contraction mapping. A fixed point process which is unique up to scale and for which the least squares estimator is unbiased to order $1/T$ can be derived analytically for each autoregressive order p and degree of estimated polynomial trend in time. Moreover, for each degree of estimated polynomial trend, the fixed point processes form a sequence of projections from an infinite order fixed point process. In contrast to the Yule-Walker situation, the least squares estimator does not yield a Levinson-Durbin sequence, but the least squares fixed point processes for a given degree of estimated polynomial trend do combine to form a Levinson-Durbin sequence.

The Burg and Kay estimators. The Burg and Kay estimators both generate Levinson-Durbin sequences. Burg's algorithm determines the α_{nn}'s by minimizing a sequence of sums of squares of forward and backward one-step prediction errors. The remaining α_{jn} values are then determined from (2). For a description of the Burg estimator see, e.g., Brockwell and Davis [(2002), pages 147-148]. Kay's estimator (1983) of the autoregressive coefficients is a recursive maximum likelihood procedure. The parameter α_{nn} is estimated at the nth stage by maximizing a partial Gaussian likelihood and then (2) is applied to determine $\alpha_{1n}, \ldots, \alpha_{n-1,n}$.

This paper is organized as follows. In Section 2 some properties of generalized Levinson-Durbin sequences are presented. These properties generalize relations satisfied by binomial coefficients. It is also noted that the Levinson-Durbin sequences define minimum phase filters. Least squares estimation bias and least squares fixed point processes are described in Section 3. Section 4 is devoted to Yule-Walker estimation bias and fixed point processes. Concluding discussion appears in Section 5, and proofs are in Section 6.

2. Properties of generalized Levinson-Durbin sequences.

2.1. *Cholesky factorization*. Let $\{ y_t \}$ be a discrete time, second-order stationary process with positive definite covariance sequence $\gamma(j)$, j=0, ±1, ±2, … . Let $\boldsymbol{\Gamma}_n$ denote the covariance matrix of $(y_1, \dots, y_n)'$ and $\boldsymbol{\gamma}_n = (\gamma(1), \dots, \gamma(n))'$, and let $\boldsymbol{\alpha}_n = (\alpha_{1n}, \dots, \alpha_{nn})'$ be the coefficients specifying the best linear predictor of y_{n+1}, as indicated in (1), given knowledge of $y_1, \dots, y_n$. Then $\boldsymbol{\Gamma}_n\boldsymbol{\alpha}_n = -\boldsymbol{\gamma}_n$, and the Levinson-Durbin recursion is given by (2) and

$$\alpha_{nn} = -\frac{\gamma(n) + \alpha_{1,\,n-1}\gamma(n-1) + \cdots + \alpha_{n-1,\,n-1}\gamma(1)}{\gamma(0) + \alpha_{1,\,n-1}\gamma(1) + \cdots + \alpha_{n-1,\,n-1}\gamma(n-1)}.$$

It is well-known that these α_{nn}'s satisfy (3) and that the α_{jn}'s are used to form the lower matrix in the Cholesky factorization of the inverse of $\boldsymbol{\Gamma}_n$.

2.2. *Minimum phase.* If $\{\alpha_{jn}\}$ is a generalized Levinson-Durbin sequence, that is, it satisfies (2), define the polynomials

$$A_n(z) = \sum_{j=0}^{n} \alpha_{jn} z^{n-j}, \qquad n = 1, 2, \dots, \tag{4}$$

where $\alpha_{0n} = 1$. From (2) it follows that

$$A_n(z) = z A_{n-1}(z) + \alpha_{nn} z^{n-1} A_{n-1}(z^{-1}). \tag{5}$$

Proposition 2.1. *If* $\{\alpha_{jn}\}$ *is a Levinson-Durbin sequence, that is, it satisfies both* (2) and (3), *then the zeros of the polynomials* $A_n(z)$ *lie strictly inside the unit circle* $|z| = 1$.

This result is well-known. The proof uses induction on n. It follows from (5) and application of Rouché's theorem, and is identical to the proof of Theorem 5 in Stine and Shaman (1989). The result states that $A_n(z)$ determines a minimum phase filter for each n.

In the context of parametrization of an autoregressive process of order p, Barndorff-Nielsen and Schou (1973) consider the set of all coefficient vectors $(\alpha_{1p}, \ldots, \alpha_{pp})$ for which the zeros of $A_p(z)$ lie strictly inside $|z| = 1$. They show that the mapping which transforms such $(\alpha_{1p}, \ldots, \alpha_{pp})$ to the partial correlations $(\alpha_{11}, \ldots, \alpha_{pp})$ is one-to-one and onto $(-1, 1)^p$. Moreover, the mapping and its inverse are both continuously differentiable. Thus, while it is difficult to specify criteria for the zeros of $A_p(z)$ to lie strictly inside $|z| = 1$ in terms of the coefficients $\alpha_{1p}, \ldots, \alpha_{pp}$, it is trivial to do so in terms of the partial correlations. Wise (1956) gives criteria in terms of the coefficient vectors and lists the results explicitly for $p = 1, \ldots, 4$.

2.3. *Relations for generalized Levinson-Durbin sequences.* In this subsection we state some summation formulas satisfied by generalized Levinson-Durbin sequences. They are generalizations of formulas for binomial coefficients. A list of binomial coefficient summations is given in Section 0.15 of Gradshteyn and Ryzhik (1980), e.g.
First we note that the binomial coefficient sequence and the sequences for which $\alpha_{1n} = \cdots = \alpha_{nn}$ for all $n = 2, 3, \ldots$ are the only symmetric generalized Levinson-Durbin sequences.

THEOREM 2.1. *Suppose* $\{\alpha_{jn}\}$ *is a generalized Levinson-Durbin sequence for which* $\alpha_{jn} = \alpha_{n-j,n}$, $j = 1, \ldots, n-1$, *and assume* $\alpha_{nn} \neq 1$, $n = 2, 3, \ldots$. *Then* $\alpha_{1n} = \cdots = \alpha_{nn}$, $n = 2, 3, \ldots$.

Some binomial coefficient summation formulas are simple and widely used. The next theorem gives the generalizations of $\sum_{j=0}^{n} \binom{n}{j} = 2^n$ and $\sum_{j=0}^{n} (-1)^j \binom{n}{j} = 0$.

THEOREM 2.2. *If* $\{\alpha_{jn}\}$ *is a generalized Levinson-Durbin sequence,*

(6) $$\sum_{j=0}^{n} \alpha_{jn} = \prod_{j=1}^{n} (1+\alpha_{jj}), \qquad n=1,2,\ldots,$$

(7) $$\sum_{j=0}^{n} (-1)^j \alpha_{jn} = \prod_{j=1}^{n} (1+(-1)^j \alpha_{jj}), \qquad n=1,2,\ldots,$$

where $\alpha_{0n} = 1$.

REMARK 2.1. Theorem 2.2 indicates that $\sum_{j=0}^{n} \alpha_{jn} > 0$ and $\sum_{j=0}^{n} (-1)^j \alpha_{jn} > 0$ are necessary conditions for $\{\alpha_{jn}\}$ to be a Levinson-Durbin sequence.

Generalizations of the binomial coefficient summations $\sum_{j=1}^{n} j \binom{n}{j} = n 2^{n-1}$ and $\sum_{j=1}^{n} (-1)^{j-1} j \binom{n}{j} = 0$ can also be expressed in terms of α_{jj}, $j = 1, 2, \ldots, n$.

THEOREM 2.3. *If* $\{\alpha_{jn}\}$ *is a generalized Levinson-Durbin sequence,*

$$\sum_{j=1}^{n} j\alpha_{jn} = \sum_{l=1}^{n} \prod_{j=1}^{l-1}(1+\alpha_{jj})\, l\alpha_{ll} \prod_{k=l+1}^{n}(1-\alpha_{kk}), \qquad n=1,2,\ldots, \tag{8}$$

$$\sum_{j=1}^{n}(-1)^{j-1} j\alpha_{jn} = \sum_{l=1}^{n} \prod_{j=1}^{l-1}(1+(-1)^{j}\alpha_{jj})(-1)^{l-1} l\alpha_{ll} \prod_{k=l+1}^{n}(1-(-1)^{k}\alpha_{kk}), \quad n=1,2,\ldots, \tag{9}$$

where $\prod_1^0(\cdot) = \prod_{n+1}^{n}(\cdot) = 1$.

Each of the binomial sums $1 + \binom{n}{2} + \binom{n}{4} + \cdots$ and $\binom{n}{1} + \binom{n}{3} + \cdots$ is equal to 2^{n-1}.

Define the decomposition

$$\prod_{k=1}^{r}(1+x_k) = \prod_{k=1}^{r}{}_{\text{(even)}}(1+x_k) + \prod_{k=1}^{r}{}_{\text{(odd)}}(1+x_k), \tag{10}$$

where $\prod_{\text{(even)}}$ $\left(\prod_{\text{(odd)}}\right)$ is the sum of terms from the left-hand side of (10), each of which is the product of an even (odd) number of x_k's. For example, if $r = 4$, $\prod_{\text{(even)}}$ is $1 + x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4 + x_1x_2x_3x_4$ and $\prod_{\text{(odd)}}$ is $x_1 + x_2 + x_3 + x_4 + x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4$.

THEOREM 2.4. *If* $\{\alpha_{jn}\}$ *is a generalized Levinson-Durbin sequence,*

$$S_{1n} := 1 + \alpha_{2n} + \alpha_{4n} + \cdots = \prod_{j=1}^{\left[\frac{1}{2}n\right]}(1+\alpha_{2j,2j}) \prod_{k=1}^{\left[\frac{1}{2}(n+1)\right]}{}_{\text{(even)}}(1+\alpha_{2k-1,2k-1}), \quad n=1,2,\ldots, \tag{11}$$

(12) $\quad S_{2n} := \alpha_{1n} + \alpha_{3n} + \cdots = \prod_{j=1}^{[\frac{1}{2}n]}(1+\alpha_{2j,2j}) \prod_{k=1}^{[\frac{1}{2}(n+1)]}{}_{\text{(odd)}}(1+\alpha_{2k-1,2k-1}), \quad n = 1, 2, \ldots,$

where $\alpha_{jn} = 0$ *for* $j > n$, $\prod_1^0(\cdot) = 1$ *and* $[x]$ *denotes the integer part of* x.

2.4. *Autoregressive process of finite order.* An order p autoregressive process $\{y_t\}$, AR(p), is defined by

(13)
$$\sum_{j=0}^{p} \alpha_j (y_{t-j} - \mu) = \varepsilon_t, \qquad t = 0, \pm 1, \pm 2, \ldots,$$

where $\mu = \mathrm{E}(y_t)$ and $\alpha_0 = 1$, and $\{\varepsilon_t\}$ is an iid sequence with mean 0 and variance σ^2. In addition, the zeros of $A_p(z)$ defined at (4) are assumed to lie strictly inside $|z| = 1$. Calculation of the best linear predictor of y_{n+1}, given $y_1, \ldots, y_n$, $n = 1, 2, \ldots$, leads to a Levinson-Durbin sequence with $\alpha_{jp} = \alpha_j$, $j = 1, \ldots, p$, and for all $n > p$, $\alpha_{jn} = \alpha_j$, $j = 1, \ldots, p$, and $\alpha_{jn} = 0$, $j = p + 1, \ldots, n$. The partial correlation at lag p is $-\alpha_p$, and all partial correlations at lags greater than p are 0.

The Yule-Walker equations for the AR(p) process (13) are

(14)
$$\sum_{l=0}^{p} \alpha_l \gamma(j - l) = 0, \qquad j = 1, 2, \ldots.$$

3. Least squares estimation. Let $y_1, \ldots, y_T$ be observations from the AR(p) process defined at (13). We deal with a constant mean μ, as specified in (13), and we also allow the mean to be a polynomial time trend, $\mu(t) = \sum_{j=0}^{k-1} \beta_j t^j$, for $t = 1, \ldots, T$. Define covariance estimators by

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

if μ (or $\mu(t)$) is known. If the mean is unknown, we replace μ in (14) by the sample mean $\bar{y}$ or by the least squares estimator $\hat{\mu}(t)$ of the polynomial time trend. Let $\boldsymbol{\alpha}_p$ = $(\alpha_1, \ldots, \alpha_p)'$. The least squares estimator of $\boldsymbol{\alpha}_p$ is $\hat{\alpha}_p^k = -\mathbf{C}_p^{-1}\mathbf{c}_p$, where $\mathbf{C}_p$ is the $p \times p$ matrix with c_{ij} in row i and column j, $i, j = 1, \ldots, p$, and $\mathbf{c}_p$ is the $p \times 1$ vector with c_{0i} in row i, $i = 1, \ldots, p$. Here the superscript k indicates the degree of the estimated polynomial time trend is $k - 1$, and $k = 0$ is used to designate a known mean. As p varies, this estimator does not determine a Levinson-Durbin sequence, and the coefficients of the estimator do not generally determine a minimum phase filter.

3.1. *The bias approximation.* To ensure the validity of the bias approximations used in this paper we assume that the errors ε_t have finite moment of order 16 and that

(16) $$\mathrm{E}(\,\|\mathbf{C}_p^{-1} - \boldsymbol{\Gamma}_p^{-1}\|^k\,) = O(1) \qquad \text{as } T \to \infty \text{ for } k \le 8$$

[see Lewis and Reinsel(1988)], where $\|\,\mathbf{A}\,\|$ is the largest absolute eigenvalue of $\mathbf{A}$ and $\boldsymbol{\Gamma}_p$ is the covariance matrix of $(y_1, \ldots, y_p)'$. Also see Bhansali (1981), whose assumption (A3) is stronger than (16).

Let $\mathbf{e}_j$ be the $(p + 1) \times 1$ vector with 1's in rows $j + 3$, $j + 5$, … , $p + 1 - j$ and 0's elsewhere, and $\mathbf{d}_j$ be the $(p + 1) \times 1$ vector with 1's in rows $j + 2$, $j + 4$, … , $p + 1 - j$ and 0's elsewhere. Define the $(p + 1) \times (p + 1)$ matrices $\mathbf{B}_{1p} = \mathrm{diag}(0, 1, 2, \ldots, p)$;

$\mathbf{B}_{2p} = [-\mathbf{e}_0, -\mathbf{e}_1, \ldots, -\mathbf{e}_{p/2-1}, \mathbf{0}, \mathbf{e}_{p/2-1}, \ldots, \mathbf{e}_1, \mathbf{e}_0]$ if p is even and $\mathbf{B}_{2p} = [-\mathbf{d}_1, -\mathbf{d}_2, \ldots, -\mathbf{d}_{(p-1)/2}, \mathbf{0}, \mathbf{d}_{(p-1)/2}, \ldots, \mathbf{d}_1, \mathbf{d}_0]$ if p is odd; $\mathbf{B}_{3p}$ with (i, j) entry equal to -1 for $j < i \le p + 2 - j$, 1 for $p + 2 - j < i \le j$, and 0 otherwise, $i, j = 1, \ldots, p + 1$; and $\mathbf{B}_p^k = \mathbf{B}_{1p} + \mathbf{B}_{2p} + k\,\mathbf{B}_{3p}$, where k is the number of unknown parameters in the polynomial time trend. For example,

$$\mathbf{B}_7^k = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -k & 1 & 0 & 0 & 0 & 0 & 0 & k+1 \\ -k-1 & -k & 2 & 0 & 0 & 0 & k+1 & k \\ -k & -k-1 & -k & 3 & 0 & k+1 & k & k+1 \\ -k-1 & -k & -k-1 & -k & k+5 & k & k+1 & k \\ -k & -k-1 & -k & 0 & 0 & k+6 & k & k+1 \\ -k-1 & -k & 0 & 0 & 0 & 0 & k+7 & k \\ -k & 0 & 0 & 0 & 0 & 0 & 0 & k+8 \end{bmatrix},$$

$$\mathbf{B}_8^k = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -k & 1 & 0 & 0 & 0 & 0 & 0 & 0 & k \\ -k-1 & -k & 2 & 0 & 0 & 0 & 0 & k & k+1 \\ -k & -k-1 & -k & 3 & 0 & 0 & k & k+1 & k \\ -k-1 & -k & -k-1 & -k & 4 & k & k+1 & k & k+1 \\ -k & -k-1 & -k & -k-1 & 0 & k+6 & k & k+1 & k \\ -k-1 & -k & -k-1 & 0 & 0 & 0 & k+7 & k & k+1 \\ -k & -k-1 & 0 & 0 & 0 & 0 & 0 & k+8 & k \\ -k-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k+9 \end{bmatrix}.$$

The approximate bias of $\hat{\alpha}_p^k$ is a linear function of the α_j's. It is given by

$$\begin{pmatrix} 1 \\ \mathrm{E}\hat{\alpha}_p^k \end{pmatrix} = (\mathbf{I}_{p+1} - \mathbf{B}_p^k / T) \begin{pmatrix} 1 \\ \alpha_p \end{pmatrix} + o(1/T), \tag{17}$$

where $\mathbf{I}_{p+1}$ is the $(p + 1) \times (p + 1)$ identity matrix. For details of this bias derivation see Shaman and Stine (1988), Stine and Shaman (1989) and Pham (1993).

REMARK 3.1. Estimation of the polynomial time trend by least squares contributes the term involving $k\,\mathbf{B}_{3p}$ to the bias. This follows from Theorems 10.32 and 10.34 in Anderson (1971), for example. See also Cheang and Reinsel [(2000), page 1177].

3.2. *Least squares fixed point model coefficients.* For each order p of autoregressive process and each degree $k-1$ of the polynomial time trend, up to scale there is a unique autoregressive model for which the least squares estimator is unbiased to terms of order $1/T$. This is stated in the following theorem, which was given in Stine and Shaman (1989) for $k = 0, 1$.

THEOREM 3.1. *If* $T > (p + k + 1)/2$, *the expectation mapping* $\mathbf{I}_{p+1} - \mathbf{B}_p^k / T$ *is a contraction with fixed point* $(1,\ \tilde{\alpha}_p^{k\,\prime})'$ *satisfying* $(\mathbf{I}_{p+1} - \mathbf{B}_p^k / T)\begin{pmatrix}1\\ \tilde{\alpha}_p^k\end{pmatrix} = \begin{pmatrix}1\\ \tilde{\alpha}_p^k\end{pmatrix}$,

$k = 0, 1, \ldots,\ p = 1, 2, \ldots$. *The fixed point for which the first coordinate is equal to* 1 *is unique*.

REMARK 3.2. The fixed point vector $\tilde{\alpha}_p^k$ is obtained by solving the equations

$$\mathbf{B}_p^k\begin{pmatrix}1\\ \tilde{\alpha}_p^k\end{pmatrix} = \mathbf{0}_{p+1}, \qquad (18)$$

where $\mathbf{0}_q$ denotes the $q \times 1$ vector of zeros. Define

$$\mathbf{B}_p^k = \begin{pmatrix} 0 & \mathbf{0}'_p \\ \mathbf{b}_p^k & \mathbf{B}_{22p}^k \end{pmatrix}, \tag{19}$$

where $\mathbf{B}_{22p}^k$ is $p \times p$. Then (18) is equivalent to

$$\mathbf{B}_{22p}^k \tilde{\alpha}_p^k = -\mathbf{b}_p^k, \tag{20}$$

where, for $p = 1$, $-\mathbf{b}_p^k$ is the scalar k, and

$$\begin{aligned} -\mathbf{b}_{p+1}^k &= (-\mathbf{b}_p^{k\,\prime},\ k+1)' \quad \text{if } p \text{ is odd,} \\ &= (-\mathbf{b}_p^{k\,\prime},\ k)' \qquad \text{if } p \text{ is even.} \end{aligned} \tag{21}$$

In addition, the matrices $\mathbf{B}_{22p}^k$ are also related for successive values of p. Let $\mathbf{J}_p$ denote the $p \times p$ permutation matrix with 1's along the main skew diagonal and 0's elsewhere. Then

$$\begin{aligned} \mathbf{B}_{22,p+1}^k &= \begin{pmatrix} \mathbf{B}_{22p}^k - (k+1)\mathbf{J}_p & -\mathbf{J}_p \mathbf{b}_p^k \\ \mathbf{0}_p' & p+k+2 \end{pmatrix} \quad \text{if } p \text{ is odd,} \\ &= \begin{pmatrix} \mathbf{B}_{22p}^k - k\mathbf{J}_p & -\mathbf{J}_p \mathbf{b}_p^k \\ \mathbf{0}_p' & p+k+2 \end{pmatrix} \quad \text{if } p \text{ is even.} \end{aligned} \tag{22}$$

The equations (20) are equivalent to the Yule-Walker equations for the fixed point coefficient vector $\tilde{\alpha}_p^k$. Denote by $\tilde{\gamma}_p^k(j)$ the lag j covariance of the AR(p) process with the fixed point coefficient vector, and let $\tilde{\Gamma}_p^k$ be the $p \times p$ covariance matrix for this process and $\tilde{\gamma}_p^k = (\tilde{\gamma}_p^k(1), \ldots, \tilde{\gamma}_p^k(p))'$. The Yule-Walker equations are $\tilde{\Gamma}_p^k \tilde{\alpha}_p^k = -\tilde{\gamma}_p^k$, and if we premultiply them by $\mathbf{B}_{22p}^k \left(\tilde{\Gamma}_p^k\right)^{-1}$ we obtain (20).

For each value of k the coefficients of the least squares fixed point models implied by Theorem 3.1 combine to form a Levinson-Durbin sequence. For $k = 0$ and $k = 1$ these sequences were given in Stine and Shaman (1989). Theorem 3.1 implies that the bias of least squares estimation of the autoregressive coefficients pulls the estimate toward the fixed point coefficient vector.

THEOREM 3.2. *Let* $\tilde{\alpha}_p^k = (\tilde{\alpha}_{1p}^k, \ldots, \tilde{\alpha}_{pp}^k)'$ *be the fixed point defined in Theorem* 3.1. *For each* $k = 0, 1, 2, \ldots$, *the coefficients of the autoregressive models for which least squares estimates are unbiased to terms of order* $1/T$ *combine to form a Levinson-Durbin sequence* $\{\tilde{\alpha}_{jp}^k,\ j = 1, \ldots, p,\ p = 1, 2, \ldots\}$ *with*

$$\begin{aligned} \tilde{\alpha}_{pp}^k &= \frac{k}{p+k+1}, \qquad && p \text{ odd}, \\ &= \frac{k+1}{p+k+1}, \qquad && p \text{ even}. \end{aligned}$$

REMARK 3.3. As $k \to \infty$, $\tilde{\alpha}_{jp}^k$ converges to the binomial coefficient $\binom{p}{j}$.

Table 1 displays the coefficient vectors for the least squares fixed point models for p = 1, …, 6. Each of the coefficients is a ratio of polynomials in k, the number of parameters in the polynomial time trend. As p increases, some of the polynomials in k in the numerator of $\tilde{\alpha}_{jp}^k$ become complicated, and there is no evident completely general pattern.

Table 1

p	$\tilde{\alpha}_p^{k}{}'$
1	$\left(\frac{k}{k+2}\right)$
2	$\left(\frac{2k}{k+3}, \frac{k+1}{k+3}\right)$
3	$\left(\frac{3k}{k+4}, \frac{3k^2+5k+4}{(k+3)(k+4)}, \frac{k}{k+4}\right)$
4	$\left(\frac{4k}{k+5}, \frac{2(3k^2+5k+4)}{(k+4)(k+5)}, \frac{4k(k+2)}{(k+4)(k+5)}, \frac{k+1}{k+5}\right)$
5	$\left(\frac{5k}{k+6}, \frac{2(5k^2+7k+6)}{(k+5)(k+6)}, \frac{2k(5k^2+21k+28)}{(k+4)(k+5)(k+6)}, \frac{5k^2+7k+6}{(k+5)(k+6)}, \frac{k}{k+6}\right)$
6	$\left(\frac{6k}{k+7}, \frac{3(5k^2+7k+6)}{(k+6)(k+7)}, \frac{4k(5k^2+21k+28)}{(k+5)(k+6)(k+7)}, \frac{3(k+3)(5k^2+7k+6)}{(k+5)(k+6)(k+7)}, \frac{6k(k+2)}{(k+6)(k+7)}, \frac{k+1}{k+7}\right)$

The binomial coefficients $\binom{p}{j}$ for a fixed p satisfy $j\binom{p}{j} = (p+1-j)\binom{p}{p+1-j}$, $j = 1,\ldots, p$. According to the following lemma, the least squares fixed point coefficients satisfy this same condition for p odd, but for p even there is a more complicated relation among the coefficients. The lemma will be proved and will be used in the proof of Theorem 3.2 in Section 6.

LEMMA 3.1. *If p is odd,*

(23) $$j\tilde{\alpha}_{jp}^{k} - (p+1-j)\,\tilde{\alpha}_{p+1-j,p}^{k} = 0, \quad j = 1,\ldots, p.$$

If p is even,

$$(24)\qquad j\widetilde{\alpha}_{jp}^{k} - (p+2-j)\,\widetilde{\alpha}_{p+1-j,p}^{k} + \sum_{i=0}^{j-1}(-1)^{j-1-i}\,\widetilde{\alpha}_{ip}^{k} + \sum_{i=0}^{j-2}(-1)^{j-2-i}\,\widetilde{\alpha}_{p-i,p}^{k} = 0, \qquad j = 1,\ldots,p,$$

where $\widetilde{\alpha}_{0p}^{k} = 1$ *and* $\sum_{0}^{-1}(\cdot) = 0.$

REMARK 3.4. From Theorems 2.2 and 3.2 we may calculate, for $k = 0, 1, \ldots,$

$$\sum_{j=0}^{p}\widetilde{\alpha}_{jp}^{k} = 2^{p}\prod_{j=1}^{\left[\frac{1}{2}(p+1)\right]}\frac{k+j}{p+k+2-j}$$

and

$$\sum_{j=0}^{p}(-1)^{j}\,\widetilde{\alpha}_{jp}^{k} = \left[\tfrac{1}{2}(p+1)\right]!\,2^{p}\prod_{j=1}^{\left[\frac{1}{2}(p+1)\right]}\frac{1}{p+k+2-j}.$$

Further, from Theorems 2.3 and 3.2 we have the sums, for $k = 0, 1, \ldots,$

$$\sum_{j=1}^{p} j\,\widetilde{\alpha}_{jp}^{k} = p2^{p-1}\frac{\left(k-1+\left[\frac{3}{2}p\right]/\,p\right)\prod_{j=1}^{\left[\frac{1}{2}(p-1)\right]}(k+j+1)}{\prod_{j=1}^{\left[\frac{1}{2}(p+1)\right]}(p+k+2-j)}$$

and

$$\sum_{j=1}^{p}(-1)^{j-1}\,j\,\widetilde{\alpha}_{jp}^{k} = -\frac{p\left(\frac{1}{2}p\right)!\,2^{p-2}}{\prod_{j=1}^{\frac{1}{2}p}(p+k+2-j)}, \qquad p \text{ even},$$

$$= \frac{\left(\frac{1}{2}(p+1)\right)!\,(k-\frac{1}{2}(p-1))\,2^{p-1}}{\prod_{j=1}^{\frac{1}{2}(p+1)}(p+k+2-j)}, \qquad p \text{ odd}.$$

REMARK 3.5. Let $\widetilde{A}_p^k(z)$ be defined as in (4) for the least squares fixed point model of order p and with degree $k-1$ for the polynomial time trend. Numerical calculations show that the zeros of $\widetilde{A}_p^k(z)$ occur in complex pairs except for a single real zero which is negative for p odd. For $k=0$ their arguments are distributed approximately evenly spaced around the circle. The zeros increase in modulus as p increases, and they tend toward $z=-1$ as k increases. The zeros for (p, k) = (4, 0), (4, 1), (20, 0) and (20, 1) are pictured in Figure 1 of Stine and Shaman (1989).

3.3. *Least squares fixed point model correlation functions and spectra.* As we have noted, for each k the coefficients of the least squares fixed point models combine to form a Levinson-Durbin sequence. The coefficients $-\widetilde{\alpha}_{pp}^k$, $p = 1, 2, \ldots,$ thus form the partial correlation sequence of an infinite order autoregression. The fixed point model with coefficients $\widetilde{\alpha}_{jp}^k$, $j = 1, \ldots, p$, is a least-squares approximation to this infinite order autoregression. The variance of this autoregression is

$$\widetilde{\gamma}_\infty^k(0) = \sigma^2 \prod_{j=1}^{\infty}\left(1-\left(\widetilde{\alpha}_{jj}^k\right)^2\right)^{-1}, \qquad k = 0,1,\ldots. \tag{25}$$

See, e.g., Barndorff-Nielsen and Schou (1973).

The next theorem and its corollary give the covariance functions and spectral densities for these infinite order autoregressions for $k = 0, \ldots, 5$.

THEOREM 3.3. *For each* $k = 0, 1, \ldots$ *and each* $p = 1,2, \ldots$, *the fixed point model with coefficients* $\tilde{\alpha}_{lp}^{k}$, $l = 1, \ldots, p$, *is a least-squares approximation to an infinite order autoregression. The variance of this autoregression is given by* (25), *and for* $k = 0, \ldots, 5$ *the correlation functions are*

$$\tilde{\rho}_{\infty}^{0}(j) = 0, \qquad j \text{ odd},$$
$$= -\frac{1}{j^2 - 1}, \qquad j \text{ even},$$

$$\tilde{\rho}_{\infty}^{1}(j) = \frac{1}{j^2 - 4}, \qquad j \text{ odd},$$
$$= -\frac{1}{j^2 - 1}, \qquad j \text{ even},$$

$$\tilde{\rho}_{\infty}^{2}(j) = \frac{3}{2}\frac{1}{j^2 - 4}, \qquad j \text{ odd},$$
$$= -\frac{3}{2}\frac{j^2 - 6}{(j^2 - 1)(j^2 - 9)}, \qquad j \text{ even},$$

$$\tilde{\rho}_{\infty}^{3}(j) = \frac{2j^2 - 29}{(j^2 - 4)(j^2 - 16)}, \qquad j \text{ odd},$$
$$= -\frac{2j^2 - 9}{(j^2 - 1)(j^2 - 9)}, \qquad j \text{ even},$$

$$\tilde{\rho}_{\infty}^{4}(j) = \frac{5}{2}\frac{j^2 - 13}{(j^2 - 4)(j^2 - 16)}, \qquad j \text{ odd},$$
$$= -\frac{5}{2}\frac{(j^4 - 28j^2 + 90)}{(j^2 - 1)(j^2 - 9)(j^2 - 25)}, \qquad j \text{ even},$$

$$\tilde{\rho}_{\infty}^{5}(j) = \frac{3}{2}\frac{2j^4 - 95j^2 + 843}{(j^2 - 4)(j^2 - 16)(j^2 - 36)}, \qquad j \text{ odd},$$
$$= -\frac{3}{2}\frac{2j^4 - 53j^2 + 150}{(j^2 - 1)(j^2 - 9)(j^2 - 25)}, \qquad j \text{ even}.$$

REMARK 3.6. The correlation function $\tilde{\rho}_p^k(j)$ for the fixed point model with coefficients $\tilde{\alpha}_{lp}^k$, $l = 1, \ldots, p$, is given for $j = 1, \ldots, p$ by $\tilde{\rho}_\infty^k(j)$, and then for $j = p + 1,\ p + 2, \ldots$ by use of the Yule Walker equations (14) written for the correlation function.

COROLLARY 3.1. *The spectral densities of the infinite order autoregressions cited in Theorem* 3.3 *are, for* $k = 0, \ldots, 5$,

$$\tilde{f}_\infty^0(\omega) = \tilde{\gamma}_\infty^0(0)\,\tfrac{1}{4}\left|\sin\omega\right|, \quad -\pi \le \omega \le \pi,$$

$$\tilde{f}_\infty^1(\omega) = \tilde{\gamma}_\infty^1(0)\,\tfrac{1}{8}\left|2\sin\omega - \sin 2\omega\right|, \quad -\pi \le \omega \le \pi,$$

$$\tilde{f}_\infty^2(\omega) = \tilde{\gamma}_\infty^2(0)\,\tfrac{3}{64}\left|5\sin\omega - 4\sin 2\omega + \sin 3\omega\right|, \quad -\pi \le \omega \le \pi,$$

$$\tilde{f}_\infty^3(\omega) = \tilde{\gamma}_\infty^3(0)\,\tfrac{1}{64}\left|14\sin\omega - 14\sin 2\omega + 6\sin 3\omega - \sin 4\omega\right|, \quad -\pi \le \omega \le \pi,$$

$$\tilde{f}_\infty^4(\omega) = \tilde{\gamma}_\infty^4(0)\,\tfrac{5}{1024}\left|42\sin\omega - 48\sin 2\omega + 27\sin 3\omega - 8\sin 4\omega + \sin 5\omega\right|, \quad -\pi \le \omega \le \pi,$$

$$\tilde{f}_\infty^5(\omega) = \tilde{\gamma}_\infty^5(0)\,\tfrac{3}{2048}\left|132\sin\omega - 165\sin 2\omega + 110\sin 3\omega - 44\sin 4\omega + 10\sin 5\omega - \sin 6\omega\right|,$$

$$-\pi \le \omega \le \pi.$$

Figure 1 displays the spectra listed in Corollary 3.1 for $k = 0$, 1 and 2 . As k increases the spectral density concentrates its mass at frequencies $\pm\pi$. Figure 2 shows the fixed point spectra for $k = 2$ and $p = 2$, 10, 30 and infinity. As p increases for a fixed value of k, total spectral mass increases.

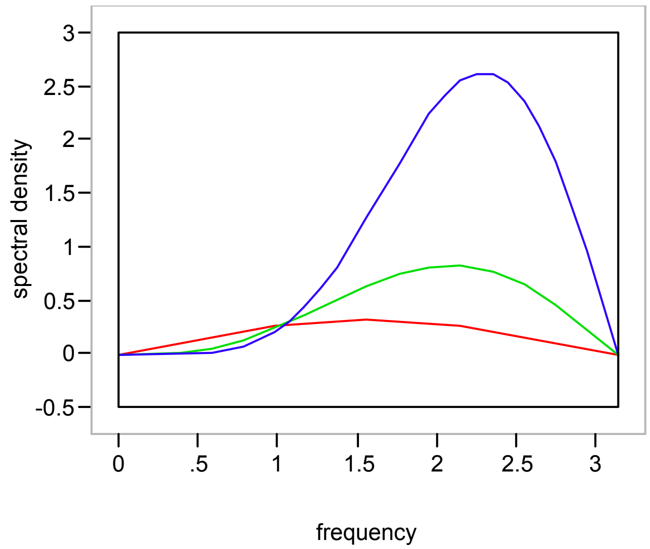

Fig. 1. Spectra $\widetilde{f}_{\infty}^{k}(\omega)$ for $k = 0$ (bottom), 1 (middle) and 2 (top).

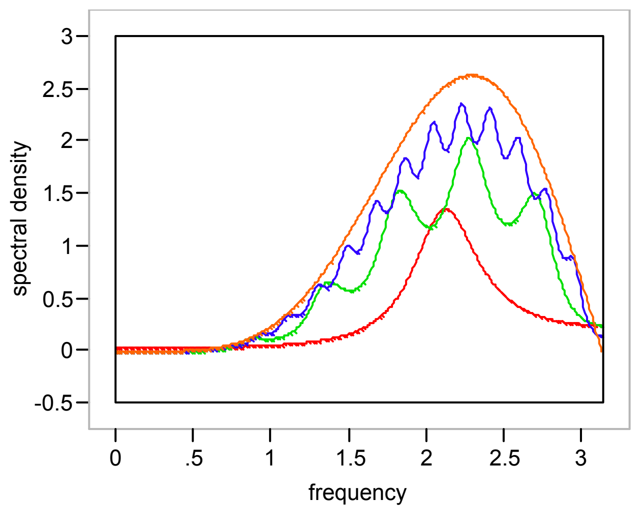

Fig. 2. Fixed point spectra $\widetilde{f}_{p}^{2}(\omega)$ for p =2, 10, 30 and infinity (from bottom to top).

4. Yule-Walker estimation. Although it can have substantial bias [see, e.g., Tjøstheim and Paulsen(1983)], the Yule-Walker estimator is commonly used by practitioners, the main reasons apparently being that it is easy to compute and its coefficients determine a minimum phase filter. The estimator is obtained from the biased covariance estimators

(26) $$g_j = \frac{1}{T}\sum_{t=1}^{T-j}(y_t - \mu)(y_{t+j} - \mu), \qquad j = 0, 1, \ldots, p,$$

for μ (or $\mu(t)$) known. If the mean is unknown, we replace μ in (26) by the sample mean $\bar{y}$ or by the least squares estimator $\hat{\mu}(t)$ of the trend. The Yule-Walker estimator of the parameter vector $\boldsymbol{\alpha}_p$ is the solution of the sample analog of the equation (14) and is given by $\hat{\alpha}_p^{kYW} = -\mathbf{G}_p^{-1}\mathbf{g}_p$, where $\mathbf{G}_p$ is the $p \times p$ Toeplitz matrix with $g_{|i-j|}$ in row i and column j, $i, j = 1, \ldots, p$, and $\mathbf{g}_p$ is the $p \times 1$ vector with g_i in row i, $i = 1, \ldots, p$. As with the notation for the least squares estimator, the superscript k indicates that the degree of the estimated polynomial time trend is $k - 1$, and $k = 0$ designates a known mean.

The expected value of $\hat{\alpha}_p^{kYW}$ to terms of order $1/T$ is

(27) $$\begin{pmatrix} 1 \\ \mathrm{E}\hat{\alpha}_p^{kYW} \end{pmatrix} = (\mathrm{I} - \mathbf{B}_p^k / T)\begin{pmatrix} 1 \\ \alpha_p \end{pmatrix} + \begin{pmatrix} 0 \\ \mathbf{G}_p^{-1}\mathbf{d}_p / T \end{pmatrix} + o(1/T),$$

where $\mathbf{d}_p$ is the $p \times 1$ vector with elements

(28) $$d_{jp} = \sum_{l=0}^{p} \left| j - l \right| \gamma(j - l)\, \alpha_l, \qquad j = 1, \ldots, p$$

[see Tjøstheim and Paulsen(1983) and Shaman and Stine (1988)]. That is, the order $1/T$ bias expression for the Yule-Walker estimator is that of the least squares estimator, plus an additional term. This added term arises from bias introduced in (26) from division by T instead of $T - j$. Unlike the result for the least squares estimator, the order $1/T$ bias expression for the Yule-Walker estimator is not a linear function of the α_j's.

REMARK 4.1. Numerical calculations show that the bias mapping of the Yule-Walker estimator is a contraction, thus for each p and k yielding a unique fixed point model with coefficients $\tilde{\alpha}_{jp}^{kYW}$, $j = 1, \ldots, p$, for which the order $1/T$ terms in (27) are zero. However, for each value of k these fixed point model coefficients do not combine as p varies to form a Levinson-Durbin sequence.

REMARK 4.2. Simulation shows that the order $1/T$ bias expression in (27) does not accurately estimate the bias of the Yule-Walker estimator throughout the region where the AR(p) coefficients vary and define a minimum phase filter. The discrepancy between the actual bias and that implied by (27) can be substantial. However, (27) is accurate in the vicinity of the fixed point model coefficients. Numerical properties of the bias of the Yule-Walker estimator are currently under study and will be reported in a subsequent paper.

Table 2 displays the coefficient vectors for Yule-Walker fixed point models for $k = 1,2$ and $p = 1, \ldots, 6$. For comparison the corresponding least squares fixed point coefficient vectors are included in the table.

Table 2

Fixed point parameter vectors, $k = 1$

p	Least squares	Yule-Walker
1	(0.3333)	(0.25)
2	(0.5, 0.5)	(0.2668, 0.3287)
3	(0.6, 0.6, 0.2)	(0.2474, 0.3396, 0.1132)
4	(0.6667, 0.8, 0.4, 0.3333)	(0.2386, 0.3474, 0.1350, 0.1888)
5	(0.7143, 0.8571, 0.5143, 0.4286, 0.1429)	(0.2205, 0.3450, 0.1386, 0.1976, 0.0717)
6	(0.75, 0.9643, 0.6429, 0.6429, 0.3214, 0.25)	(0.2141, 0.3341, 0.1383, 0.2147, 0.0886, 0.1317)

Fixed point parameter vectors, $k = 2$

p	Least squares	Yule-Walker
1	(0.5)	(0.4)
2	(0.8, 0.6)	(0.4408, 0.4158)
3	(1.0, 0.8667, 0.3333)	(0.4092, 0.4471, 0.1941)
4	(1.1429, 1.2381, 0.7619, 0.4286)	(0.3953, 0.4618, 0.2402, 0.2465)
5	(1.25, 1.4286, 1.0714, 0.7143, 0.25)	(0.3630, 0.4558, 0.2484, 0.2710, 0.1248)
6	(1.3333, 1.6667, 1.4286, 1.1905, 0.6667, 0.3333)	(0.3524, 0.4405, 0.2485, 0.2987, 0.1604, 0.1738)

Let $\widetilde{A}_p^{kYW}(z)$ be defined as in (4) for the Yule-Walker fixed point model of order p and for degree $k-1$ for the polynomial time trend. Figure 3 shows the zeros of $\widetilde{A}_p^{kYW}(z)$ for $p = 10$ and $k = 1, \ldots, 4$. The zeros of $\widetilde{A}_p^{k}(z)$ for the corresponding least squares fixed point models are shown for comparison. The Yule-Walker fixed point zeros increase in modulus as p increases, and they tend toward $z = -1$ as k increases. As Figure 3 illustrates, for given p and k the Yule-Walker fixed point zeros have smaller modulus and are farther from $z = -1$ than the least squares fixed point zeros.

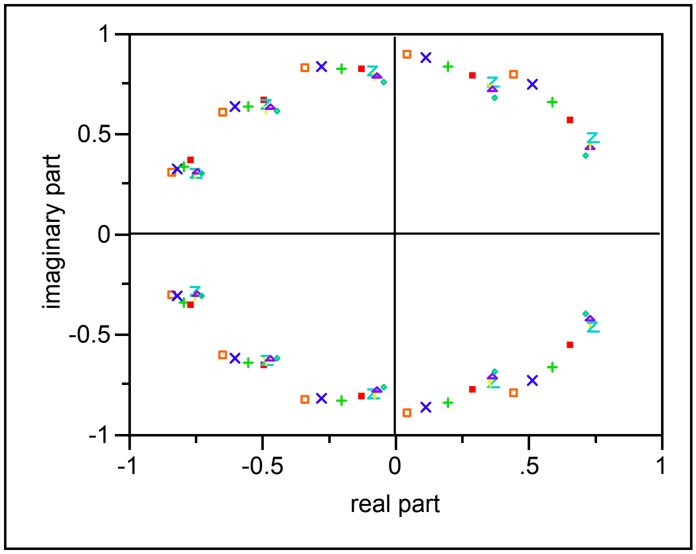

Fig. 3. Zeros of the fixed point polynomial $\widetilde{A}_p^k(z)$ for $p = 10$ for least squares estimation for $k = 1$ (▪), 2 (+), 3 (×) and 4 (▫), and of the fixed point polynomial $\widetilde{A}_p^{kYW}(z)$ for $p = 10$ for Yule-Walker estimation for $k = 1$ (◊), 2 (Δ), 3 (ʏ) and 4 (z).

5. Discussion. This paper has developed and connected two themes. One is that of Levinson-Durbin and generalized Levinson-Durbin sequences. These sequences are shown to have some properties which generalize those of binomial coefficients. The second theme considers estimation of the coefficients of an autoregressive model. Two common estimators are given by the least squares and Yule-Walker procedures. The Yule-Walker procedure generates a Levinson-Durbin sequence, but the least squares procedure does not. However, this contrast between the estimators is reversed for their corresponding fixed point autoregressive processes, which yield estimates unbiased to order $1/T$, where T is the length of the observed time series. Least squares fixed point processes do combine to form a Levinson-Durbin sequence, but Yule-Walker fixed point processes do not do so.

Using the Levinson-Durbin sequence framework and other calculations, we have explored some of the structure of the least squares fixed point autoregressive processes. The results have implications for understanding of the bias and other properties of least squares estimators. Further study is needed to describe these issues more fully.

It is well-known that the Yule-Walker estimator can have substantial bias. The fixed point Yule-Walker autoregressive processes are essentially free of Yule-Walker estimation bias, and so are processes close to these fixed point structures. However, in some parts of the parameter space the estimation bias can be so great as to render the Yule-Walker estimates very misleading. This is especially an issue if one is using the Yule-Walker estimation scheme to construct an estimate of the partial autocorrelation function to aid in selecting values of p and q in fitting an ARMA(p, q) model to data. Tjøstheim and Paulsen (1983) construct a modified estimator, with less bias than the Yule-Walker estimator. It arises from representing an AR(p) model as a vector AR(1) model. Also, the bias of the Yule-Walker estimator may be reduced substantially by applying tapering.

6. Proofs. We begin with the symmetry question addressed by Theorem 2.1.

PROOF OF THEOREM 2.1. From (2), for $n = 2,3, \ldots,$ $\alpha_{j,n+1} = \alpha_{jn} + \alpha_{n+1,n+1}\alpha_{n+1-j,n}$ and $\alpha_{n+1-j,n+1} = \alpha_{n+1-j,n} + \alpha_{n+1,n+1}\alpha_{jn}$, $j = 1,\ldots,n$. Then, if symmetry holds,

$$(1-\alpha_{n+1,n+1})\alpha_{jn} = (1-\alpha_{n+1,n+1})\alpha_{n+1-j,n}, \qquad j = 1,\ldots,n,$$

and if $\alpha_{n+1,n+1} \neq 1$, $\alpha_{jn} = \alpha_{n+1-j,n}$, $j = 1,\ldots,n$. It follows that $\alpha_{1n} = \cdots = \alpha_{nn}$, $n = 2,3,\ldots.$

□

The remaining theorems in Section 2 are proved by induction.

PROOF OF THEOREM 2.2. We give the proof of (7); verification of (6) is similar. Certainly (7) is true for $n = 1$. Then write

$$\sum_{j=0}^{n+1}(-1)^j\alpha_{j,n+1} = 1+\sum_{j=1}^{n}(-1)^j\alpha_{j,n+1} + (-1)^{n+1}\alpha_{n+1,n+1}$$

$$= 1+\sum_{j=1}^{n}(-1)^j\alpha_{jn} + \alpha_{n+1,n+1}\sum_{j=1}^{n}(-1)^j\alpha_{n+1-j,n} + (-1)^{n+1}\alpha_{n+1,n+1}$$

$$= \sum_{j=0}^{n}(-1)^j\alpha_{jn}(1+(-1)^{n+1}\alpha_{n+1,n+1}),$$

where the second step follows from (2). The proof is completed by applying the induction hypothesis. □

PROOF OF THEOREM 2.3. Consider (8), which clearly holds for $n = 1$. As an alternative to (4) define $\mathcal{A}_n(z) = \sum_{j=0}^{n}\alpha_{jn}z^j = z^n A_n(z^{-1})$. Then (5) implies

(29) $$\mathsf{A}_n(z) = \mathsf{A}_{n-1}(z) + \alpha_{nn} z^n \mathsf{A}_{n-1}(z^{-1}).$$

The left-hand side of (8) is equal to $d\mathsf{A}_n(z)/dz\big|_{z=1}$. Differentiating both sides of (29) for $n + 1$ and applying (6), we have

$$\sum_{j=1}^{n+1} j\alpha_{j,n+1} = (1-\alpha_{n+1,n+1})\sum_{j=1}^{n} j\alpha_{jn} + (n+1)\alpha_{n+1,n+1}\prod_{j=1}^{n}(1+\alpha_{jj}).$$

Then (8) follows by substituting the induction hypothesis for $\sum_{j=1}^{n} j\alpha_{jn}$ on the right-hand side and rewriting the resulting expression as a sum ranging from 1 to $n + 1$. The proof of (9) is the same except that one sets z equal to -1 after differentiating and uses (7) instead of (6). □

Proof of Theorem 2.4. We apply induction. It suffices to prove (11), because (12) follows immediately from (11) and (6). First note that (11) is valid for $n = 1$. Consider the case n odd. Using (2), we can write

$$\begin{aligned} S_{1,n+1} &= 1 + \alpha_{2,n+1} + \alpha_{4,n+1} + \cdots + \alpha_{n+1,n+1} \\ &= 1 + \alpha_{2n} + \alpha_{n+1,n+1}\alpha_{n-1,n} + \alpha_{4n} + \alpha_{n+1,n+1}\alpha_{n-3,n} + \cdots + \alpha_{n-1,n} + \alpha_{n+1,n+1}\alpha_{2n} + \alpha_{n+1,n+1} \\ &= S_{1n} + \alpha_{n+1,n+1} S_{1n}, \end{aligned}$$

which by the induction hypothesis reduces to (11) for $n + 1$. If n is even, we have, similarly to the above for n odd,

$$\begin{aligned} S_{1,n+1} &= 1 + \alpha_{2,n+1} + \alpha_{4,n+1} + \cdots + \alpha_{n,n+1} \\ &= S_{1n} + \alpha_{n+1,n+1} S_{2n}, \end{aligned}$$

which reduces to (11) for $n + 1$. □

PROOF OF THEOREM 3.1. The proof is straightforward. It is easy to show by direct calculation that the eigenvalues of the matrix $\mathbf{I} - \mathbf{B}_p^k / T$ are its diagonal elements, and these are all less than 1 in magnitude if $T > (p + k + 1)/2$. □

PROOF OF LEMMA 3.1. The proofs of (23) and (24) are identical. To obtain the result for j, subtract row $p + 1 - j$ from row j in the system of equations (20). Of course, only the subtractions for $j = 1,\ldots, [p/2]$ are needed. □

Stine and Shaman (1989) give the proof of Theorem 3.2 for $k = 0$ (see their Theorem 2) and $k = 1$ (see their Theorem 3 and Lemma 5). The proof of Theorem 3.2 given here is for general k and is more simple.

PROOF OF THEOREM 3.2. To find the least squares fixed point vector $\tilde{\alpha}_p^k$ we need to solve the system of equations (20). The equation from the last row of (20) [see also (21) and (22)] shows immediately that $\tilde{\alpha}_{pp}^k$ is $k/(p + k + 1)$ for p odd and $(k + 1)/(p + k + 1)$ for p even. The remaining values $\tilde{\alpha}_{jp}^k$ can then be determined from (20) by solving in the order $j = 1$, $p - 1, 2$, $p - 2, \ldots$. Next we need to verify that for each value of k the fixed point solutions as p varies combine to form a Levinson-Durbin sequence.

For each k the coefficients $\tilde{\alpha}^k_{pp}$ as p varies are less than 1 in magnitude, and thus verification that the fixed point coefficients form a Levinson-Durbin sequence requires establishing for all p that

$$\tilde{\alpha}^k_{p+1} = \begin{pmatrix} \tilde{\alpha}^k_p \\ 0 \end{pmatrix} + \tilde{\alpha}^k_{p+1,p+1} \begin{pmatrix} \mathbf{J}_p \tilde{\alpha}^k_p \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_p + \tilde{\alpha}^k_{p+1,p+1}\mathbf{J}_p & \mathbf{0}_p \\ \mathbf{0}_p' & 1 \end{pmatrix} \begin{pmatrix} \tilde{\alpha}^k_p \\ \tilde{\alpha}^k_{p+1,p+1} \end{pmatrix}. \tag{30}$$

To do so, it suffices to prove that (30) satisfies (20) for $p + 1$.

(i) Consider p odd. Using (22) and (30), we have

$$\begin{aligned} &\mathbf{B}^k_{22,p+1}\tilde{\alpha}^k_{p+1} \\ &= \begin{pmatrix} \mathbf{B}^k_{22p} - (k+1)\mathbf{J}_p + \tilde{\alpha}^k_{p+1,p+1}\mathbf{B}^k_{22p}\mathbf{J}_p - (k+1)\tilde{\alpha}^k_{p+1,p+1}\mathbf{I}_p & -\mathbf{J}_p\mathbf{b}^k_p \\ \mathbf{0}_p' & p+k+2 \end{pmatrix} \begin{pmatrix} \tilde{\alpha}^k_p \\ \tilde{\alpha}^k_{p+1,p+1} \end{pmatrix}. \end{aligned} \tag{31}$$

The last row of (31) is simply equal to k on the right-hand side, as required. The following lemma provides workable expressions for $\mathbf{B}^k_{22p}\mathbf{J}_p$ for p odd and even.

LEMMA 6.1. *If p is odd,*

$$\mathbf{B}^k_{22p}\mathbf{J}_p = \mathrm{diag}(k+2, k+3, \ldots, k+p+1)(\mathbf{I}_p + \mathbf{J}_p) - \mathbf{B}^k_{22p}. \tag{32}$$

If p is even,

$$\begin{aligned} \mathbf{B}^k_{22p}\mathbf{J}_p &= \mathrm{diag}(k+2, k+3, \ldots, k+p+1)(\mathbf{I}_p + \mathbf{J}_p) - \mathbf{B}^k_{22p} \\ &\quad + \sum_{j=1}^{p-1} (-1)^j \mathbf{U}^j_p (\mathbf{I}_p + \mathbf{J}_p), \end{aligned} \tag{33}$$

where $\mathbf{U}_p$ is the $p \times p$ matrix with 1's along the first superdiagonal and 0's elsewhere.

PROOF. First consider p odd. Inspecting the structure of $\mathbf{B}^k_{22p}$, we may write

$$\mathbf{B}_{22p}^{k} = \operatorname{diag}(k+2, k+3, \ldots, k+p+1) + \sum_{j=1}^{p-1} g_j^k \mathbf{U}_p^j (\mathbf{I}_p - \mathbf{J}_p),$$

where g_j^k is equal to k for j odd and $k+1$ for j even. Then (32) follows if we multiply on the right by $\mathbf{J}_p$. If p is even, the proof is similar. Write

$$\mathbf{B}_{22p}^{k} = \operatorname{diag}(k+2, k+3, \ldots, k+p+1) + \sum_{j=1}^{p-1} \left(g_j^k \mathbf{U}_p^j - g_{j+1}^k \mathbf{U}_p^j \mathbf{J}_p \right)$$

and multiply on the right by $\mathbf{J}_p$. □

By multiplication the first p rows of (31) reduce to

$$\begin{aligned} &-\mathbf{b}_p^k - (k+1)\mathbf{J}_p \tilde{\alpha}_p^k + \tilde{\alpha}_{p+1,p+1}^k \operatorname{diag}(k+2, k+3, \ldots, k+p+1)(\tilde{\alpha}_p^k + \mathbf{J}_p \tilde{\alpha}_p^k) \\ &\qquad - (k+1)\tilde{\alpha}_{p+1,p+1}^k \tilde{\alpha}_p^k, \end{aligned}$$

where we have used (32) and the fact that $\mathbf{J}_p \mathbf{b}_p^k = \mathbf{b}_p^k$ for p odd. This simplifies to

$$-\mathbf{b}_p^k + \tilde{\alpha}_{p+1,p+1}^k \operatorname{diag}(j\tilde{\alpha}_{jp}^k - (p-j+1)\tilde{\alpha}_{p+1-j}^k,\ j = 1, \ldots, p),$$

which is $-\mathbf{b}_p^k$ by Lemma 3.1. Combining this with the last row of (31), we complete the proof of Theorem 3.2 for p odd.

(ii) If p is even the proof follows similarly from (21), (22), (24) and (33). □

Proof of Theorem 3.3. The correlation functions for $k = 0$ and 1 are given in Shaman and Stine (1989). We sketch derivation of the results. The Yule-Walker equations (14) give, for fixed k and p,

$$\tilde{\rho}_p^k(j) = -\sum_{l=1}^{p} \tilde{\alpha}_{lp}^k \tilde{\rho}_p^k(j-l), \qquad j = 1, 2, \ldots.$$

For fixed k and p use the solution for $\tilde{\alpha}_{lp}^k$, $l = 1,\ldots, p$, to verify

(34) $$\tilde{\rho}_\infty^k(j) = -\sum_{l=1}^{p} \tilde{\alpha}_{lp}^k \tilde{\rho}_\infty^k(j-l), \qquad j = 1, \ldots, p.$$

For fixed k (34) holds for $p = 1, 2, \ldots$. □

Corollary 3.1 is proved by transforming the indicated spectral densities to obtain the correlation functions in Theorem 3.3.

REFERENCES

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

Barndorff-Nielsen, O. and Schou, G. (1973). On the parametrization of autogregressive models by partial autocorrelations. *J. Multivariate Anal.* **3** 408-419.

Bartlett, M. S. (1955). *An Introduction to Stochastic Processes*. Cambridge University Press.

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. (2002). *Introduction to Time Series and Forecasting*, 2nd ed. Springer, New York.

Burg, J. P. (1975). Maximum entropy spectral analysis. Ph. D. dissertation, Stanford Univ.

Cheang, W.-K. and Reinsel, G. C. (2000). Bias reduction of autoregressive estimates in time series regression model through restricted maximum likelihood. *J. Amer. Statist. Assoc.* **95** 1173-1184.

Dahlhaus, R. (1984). Parameter estimation of stationary processes with spectra containing strong peaks. In *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 50-86. Springer, New York.

Durbin, J. (1960). The fitting of time-series models. *Rev. Inst. Internat. Statist.* **28** 233-244.

Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Tables of Integrals, Series, and Products*, corrected and revised ed. Academic Press, New York.

Kailath, T. (1974). A view of three decades of linear filtering theory. *IEEE Trans.*

Information Theory **20** 146-181.

Kay, S. M. (1983). Recursive maximum likelihood estimation of autoregressive processes. *IEEE Trans. on Acoustics, Speech and Signal Processing* **31** 56-65.

Levinson, N. (1947). The Wiener RMS (root-mean-square) error criterion in filter design and prediction. *J. Math. Phys. Mass. Inst. Tech.* **25** 261-278.

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.

Ollerton, R. L. and Shannon, A. G. (2002). Generalized binomial coefficients underlying generalized Fibonacci sequences. Preprint.

Pham, D. T. (1993). On the asymptotic expansions for the bias and covariance matrix of autoregressive estimators. In *Developments in Time Series Analysis* (T. Subba Rao, ed.) 80-100. Chapman and Hall, London.

Shaman, P. and Stine, R. A. (1988). The bias of autoregressive coefficient estimators. *J. Amer. Statist. Assoc.* **83** 842-848.

Stine, R. A. and Shaman, P. (1989). A fixed point characterization for bias of auto-regressive estimators. *Ann. Statist.* **17** 1275-1284.

Tjøstheim, D. and Paulsen, J. (1983). Bias of some commonly-used time series estimates. *Biometrika* **70** 389-399.

Whittle, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika* **50** 129-134.

Wiener, N. (1949). *Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications.* MIT Press, Cambridge, MA.

Wise, J. (1956). Stationarity conditions for stochastic processes of the autoregressive and moving-average type. *Biometrika* **43** 215-219.

Zhang, H.-C. (1992). Reduction of the asymptotic bias of autoregressive and spectral estimators by tapering. *J. Time Ser. Anal.* **13** 451-469.