Journal of Multivariate Analysis 101 (2010) 1263–1273

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Generalized Levinson–Durbin sequences, binomial coefficients and autoregressive estimation

Paul Shaman

Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104-6340, United States

ARTICLE INFO

Article history:
Received 9 March 2009
Available online 18 January 2010

AMS subject classifications:
62M10
05A10
11B65

Keywords:
Binomial coefficients
Generalized Levinson–Durbin sequence
Least squares estimator
Levinson–Durbin sequence
Partial correlations
Yule–Walker estimator

ABSTRACT

For a discrete time second-order stationary process, the Levinson–Durbin recursion is used to determine the coefficients of the best linear predictor of the observation at time $k+1$, given k previous observations, best in the sense of minimizing the mean square error. The coefficients determined by the recursion define a Levinson–Durbin sequence. We also define a generalized Levinson–Durbin sequence and note that binomial coefficients form a special case of a generalized Levinson–Durbin sequence. All generalized Levinson–Durbin sequences are shown to obey summation formulas which generalize formulas satisfied by binomial coefficients. Levinson–Durbin sequences arise in the construction of several autoregressive model coefficient estimators. The least squares autoregressive estimator does not give rise to a Levinson–Durbin sequence, but 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 a Levinson–Durbin sequence. By contrast, analogous fixed point processes arising from the Yule–Walker estimator do not combine to construct a Levinson–Durbin sequence, although the Yule–Walker estimator itself does determine a Levinson–Durbin sequence. The least squares and Yule–Walker fixed point processes are further studied when the mean of the process is a polynomial time trend that is estimated by least squares.

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 $\{y_t\}$ with known structure. Given $y_1, \ldots, y_k$, for any $k \geq 1$ the recursion determines the coefficients $\alpha_{j,k}$, $j = 1, \ldots, k$, of the best linear predictor of y_{k+1},

$$\hat{y}_{k+1} = -\alpha_{1,k} y_k - \cdots - \alpha_{k,k} y_1, \tag{1.1}$$

best in the sense of minimizing the mean square error. The recursion begins with specification of $\alpha_{1,1}$, and at the nth stage one obtains $\alpha_{1,n}, \ldots, \alpha_{n,n}$. The mean square error of the predictor is also specified at each step. Levinson [1] devised the recursion to give a simple procedure for construction of the best linear predictor when the structure of the process is known. His paper was reprinted as Appendix B to Wiener's monograph on time series [2]. Wiener's work had originally been issued in February 1942 as a classified government report. For some details of this history see [3]. The second context for the recursion is that of estimation of the coefficients of an autoregressive model of finite order, given data $y_1, \ldots, y_T$. The sample Yule–Walker equations are commonly used to construct an estimator of the coefficients. Bartlett [4, pp. 264–265], Daniels [5, p. 183] and

E-mail address: shaman@wharton.upenn.edu.

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.01.004

Durbin [6] all derived the recursion as a simple method of solving the sample Yule–Walker equations, which are linear in the coefficients.

The recursion determines a double sequence $\alpha_{j,n}$, $j = 1, \ldots, n$, $n = 1, 2, \ldots$. The sequence $\alpha_{n,n}$, $n = 1, 2, \ldots$, determines all of the values $\alpha_{j,n}$.

In this paper we study properties of sequences produced by the Levinson–Durbin recursion, and we further study generalization of such sequences.

Definition 1. $\alpha_{j,n}, j = 1, \ldots, n$, $n = 1, 2, \ldots$, is a *Levinson–Durbin sequence* if the coefficients, all real-valued, satisfy

$$\alpha_{j,n} = \alpha_{j,n-1} + \alpha_{n,n}\alpha_{n-j,n-1}, \quad j = 1, \ldots, n-1, \ n = 2, 3, \ldots, \tag{1.2}$$

and

$$\left|\alpha_{n,n}\right| < 1, \quad n = 1, 2, \ldots. \tag{1.3}$$

If (1.2) holds and the $\alpha_{n,n}$'s are not subject to (1.3), we say that the $\alpha_{j,n}$'s form a *generalized Levinson–Durbin sequence*.

For the prediction problem the recursion (1.2) is used together with (1.3) and the $-\alpha_{n,n}$'s defined to be the partial correlations of the process being predicted. In the context of autoregressive estimation, the Yule–Walker estimator uses (1.2) and (1.3) and defines the $-\alpha_{n,n}$'s to be the sample partial correlations. Other estimators (the Burg and Kay procedures, mentioned below) employ (1.2) and (1.3) and define the $\alpha_{n,n}$'s differently. A generalized Levinson–Durbin sequence allows arbitrary specification of the $\alpha_{n,n}$'s. If (1.3) does hold, the sequence of $-\alpha_{n,n}$'s forms the partial correlation function for some second-order stationary process.

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 and (iv) estimation of the autoregressive process coefficients by Kay's method [7].

If $\alpha_{n,n} = 1$ for each n, then (1.2) generates the binomial coefficients. Taking into account the symmetric structure of binomial coefficients, we see that (1.2) is simply an expression of Pascal's triangle if $\alpha_{n,n} = 1$ 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 in [8]. This will be discussed in Section 3.

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 $\alpha_{j,n} = \alpha_{j,p}$, $j = 1, \ldots, p$, and $\alpha_{j,n} = 0, j = p+1, \ldots, n$, for all $n > p$. The values $-\alpha_{n,n}$, $n = 1, \ldots, p$, are the Yule–Walker sample partial correlations. For sample length T the order $1/T$ bias of the Yule–Walker estimator has been discussed in [9,10]. 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 can be 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. The tapered Yule–Walker estimator is considered in [11,12].

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 in [9,10,13]. 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 both generate Levinson–Durbin sequences. Burg's algorithm determines the $\alpha_{n,n}$'s by minimizing a sequence of sums of squares of forward and backward one-step prediction errors. The remaining $\alpha_{j,n}$ values are then determined from (1.2). For a description of the Burg estimator see, for example, [14, pp. 147–8]. Kay's estimator [7] of the autoregressive coefficients is a recursive maximum likelihood procedure. The parameter $\alpha_{n,n}$ is estimated at the nth stage by maximizing a partial Gaussian likelihood and then (1.2) is applied to determine $\alpha_{1,n}, \ldots, \alpha_{n-1,n}$.

This paper is organized as follows. In Section 2 some properties of generalized Levinson–Durbin sequences are presented. These results 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 $\gamma(j)$, $j = 0, \pm1, \pm2, \ldots$. Let $\mathbf{\Gamma}_n$ denote the covariance matrix of $(y_1, \ldots, y_n)'$ and $\gamma_n = (\gamma(1), \ldots, \gamma(n))'$, and let $\boldsymbol{\alpha}_n = (\alpha_{1,n}, \ldots, \alpha_{n,n})'$ be the vector of coefficients specifying the best linear predictor of y_{n+1}, as indicated in (1.1), given knowledge of $y_1, \ldots, y_n$. Then $\mathbf{\Gamma}_n\boldsymbol{\alpha}_n = -\gamma_n$, and the Levinson–Durbin recursion is given by (1.2) and

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

See, e.g., (7) in [6]. It is well-known that these $\alpha_{n,n}$'s satisfy (1.3) and that the $\alpha_{j,n}$'s are used to form the lower matrix in the Cholesky factorization of the inverse of $\mathbf{\Gamma}_n$.

2.2. Minimum phase

If $\{\alpha_{j,n}\}$ is a generalized Levinson–Durbin sequence, that is, it satisfies (1.2), define the polynomials

$$A_n(z) = \sum_{j=0}^{n} \alpha_{j,n} z^{n-j}, \quad n = 1, 2, \ldots, \tag{2.1}$$

where $\alpha_{0,n} = 1$, $n \geq 1$. From (1.2) it follows that

$$A_n(z) = zA_{n-1}(z) + \alpha_{n,n} z^{n-1} A_{n-1}(z^{-1}). \tag{2.2}$$

Proposition 1. *If $\{\alpha_{j,n}\}$ is a Levinson–Durbin sequence, then the zeros of the polynomials $A_n(z)$ lie strictly inside the unit circle $|z| = 1$.*

This result is well-known. The proof can be given using induction on n. It follows from (2.2) and application of Rouché's theorem and is identical to the proof of Theorem 5 in [8]. 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, [15] considers the set of all coefficient vectors $(\alpha_{1,p}, \ldots, \alpha_{p,p})$ for which the zeros of $A_p(z)$ lie strictly inside $|z| = 1$. The mapping which transforms such $(\alpha_{1,p}, \ldots, \alpha_{p,p})$ to the partial correlations $(\alpha_{1,1}, \ldots, \alpha_{p,p})$ is shown to be 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 $\alpha_{1,p}, \ldots, \alpha_{p,p}$, it is trivial to do so in terms of the partial correlations. Criteria in terms of the coefficient vectors and explicit results for $p = 1, \ldots, 4$ are given in [16].

2.3. Relations for generalized Levinson–Durbin sequences

In this section 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 [17], for example. First we note that if none of the partial correlations equals -1 and if $\alpha_{1,1}$ is not equal to $1/n$ for $n = 2, 3, \ldots$, a symmetric generalized Levinson–Durbin sequence has trivial structure.

Theorem 1. *Suppose $\{\alpha_{j,n}\}$ is a generalized Levinson–Durbin sequence for which $\alpha_{j,n} = \alpha_{n-j,n}$, $j = 1, \ldots, n-1$, $n = 2, 3, \ldots$. If $\alpha_{n,n} \neq 1$, $n = 1, 2, \ldots$, then $\alpha_{j,n} = \alpha_{1,1}/(1 - (n-1)\alpha_{1,1})$, $j = 1, \ldots, n$, $n = 2, 3, \ldots$, for any choice of $\alpha_{1,1}$ not equal to $1/k$, $k = 1, 2, \ldots$.*

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. *If $\{\alpha_{j,n}\}$ is a generalized Levinson–Durbin sequence,*

$$\sum_{j=0}^{n} \alpha_{j,n} = \prod_{j=1}^{n} (1 + \alpha_{j,j}), \quad n = 1, 2, \ldots, \tag{2.3}$$

$$\sum_{j=0}^{n} (-1)^j \alpha_{jn} = \prod_{j=1}^{n} \{1 + (-1)^j \alpha_{jj}\}, \quad n = 1, 2, \ldots, \tag{2.4}$$

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

Formula (2.3) was given by Daniels [5]; see (11.4) and the expression directly above it in his paper.

Remark 1. Theorem 2 indicates that $\sum_{j=0}^{n}\alpha_{j,n} > 0$ and $\sum_{j=0}^{n}(-1)^j\alpha_{j,n} > 0$ are necessary conditions for $\{\alpha_{j,n}\}$ to be a Levinson–Durbin sequence.

Generalizations of the binomial coefficient summations $\sum_{j=1}^{n} j\binom{n}{j} = n2^{n-1}$ and $\sum_{j=1}^{n}(-1)^{j-1}j\binom{n}{j} = 0$ can also be expressed in terms of $\alpha_{j,j}, j = 1, 2, \ldots, n$.

Theorem 3. *If $\{\alpha_{j,n}\}$ is a generalized Levinson–Durbin sequence,*

$$\sum_{j=1}^{n} j\alpha_{j,n} = \sum_{l=1}^{n}\prod_{j=1}^{l-1}(1+\alpha_{j,j})l\alpha_{l,l}\prod_{k=l+1}^{n}(1-\alpha_{k,k}), \quad n = 1, 2, \ldots, \tag{2.5}$$

$$\sum_{j=1}^{n}(-1)^{j-1}j\alpha_{j,n} = \sum_{l=1}^{n}\prod_{j=1}^{l-1}\{1+(-1)^j\alpha_{j,j}\}(-1)^{l-1}l\alpha_{l,l}\prod_{k=l+1}^{n}\{1-(-1)^k\alpha_{k,k}\}, \quad n = 1, 2, \ldots, \tag{2.6}$$

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

$$\prod_{k=1}^{r}(1+x_k) = \prod_{k=1}^{r}{}_{(\text{even})}(1+x_k) + \prod_{k=1}^{r}{}_{(\text{odd})}(1+x_k), \tag{2.7}$$

where $\prod_{(\text{even})}\left(\prod_{(\text{odd})}\right)$ is the sum of terms from the left-hand side of (2.7), 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 4. *If $\{\alpha_{i,n}\}$ is a generalized Levinson–Durbin sequence,*

$$S_{1,n} := 1 + \alpha_{2,n} + \alpha_{4,n} + \cdots = \prod_{j=1}^{\left[\frac{1}{2}n\right]}(1+\alpha_{2j,2j})\prod_{k=1}^{\left[\frac{1}{2}(n+1)\right]}{}_{(\text{even})}(1+\alpha_{2k-1,2k-1}), \quad n = 1, 2, \ldots, \tag{2.8}$$

$$S_{2,n} := \alpha_{1,n} + \alpha_{3,n} + \cdots = \prod_{j=1}^{\left[\frac{1}{2}n\right]}(1+\alpha_{2j,2j})\prod_{k=1}^{\left[\frac{1}{2}(n+1)\right]}{}_{(\text{odd})}(1+\alpha_{2k-1,2k-1}), \quad n = 1, 2, \ldots, \tag{2.9}$$

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

2.4. Autoregressive processes of finite order

An order p autoregressive process $\{y_t\}$, AR(p), is defined by

$$\sum_{j=0}^{p}\alpha_j(y_{t-j} - \mu) = \varepsilon_t, \quad t = 0, \pm1, \pm2, \ldots, \tag{2.10}$$

where $\mu = E(y_t)$, $\alpha_1, \ldots, \alpha_p$ are the autoregressive coefficients, $\alpha_0 = 1$, and $\{\varepsilon_t\}$ is an iid sequence with mean 0 and variance σ^2. In addition, the zeros of $A_p(z)$ defined at (2.1) 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 $\alpha_{j,p} = \alpha_j, j = 1, \ldots, p$, and for all $n > p$, $\alpha_{j,n} = \alpha_j$, $j = 1, \ldots, p$, and $\alpha_{j,n} = 0, j = p+1, \ldots, n$. 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 (2.10) are

$$\sum_{l=0}^{p}\alpha_l\gamma(j-l) = 0, \quad j = 1, 2, \ldots. \tag{2.11}$$

3. Least squares estimation

Let $y_1, \ldots, y_T$ be observations from the AR(p) process defined at (2.10). We deal with a constant mean μ, as specified in (2.10), 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

$$c_{i,j} = \frac{1}{T-p}\sum_{t=p+1}^{T}(y_{t-i} - \mu)(y_{t-j} - \mu), \quad i, j = 0, 1, \ldots, p, \tag{3.1}$$

if μ (or $\mu(t)$) is known. If the mean is unknown, we replace μ in (3.1) 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 $c_{i,j}$ in row i and column j, $i, j = 1, \ldots, p$, and $\mathbf{c}_p$ is the $p \times 1$ vector with $c_{0,i}$ in row i, $i = 1, \ldots, p$. Here the superscript k indicates that 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, because the coefficients of the estimator do not generally satisfy Proposition 1.

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

$$E(\|\mathbf{C}_p^{-1} - \boldsymbol{\Gamma}_p^{-1}\|^k) = O(1) \quad \text{as } T \to \infty \text{ for } k \le 8; \tag{3.2}$$

see [18]. Here $\|\mathbf{A}\|$ is the largest absolute eigenvalue of $\mathbf{A}$ and $\boldsymbol{\Gamma}_p$ is the covariance matrix of $(y_1, \ldots, y_p)'$. Also see [19], whose assumption (A3) is stronger than (3.2).

The details following in this section were given explicitly in [8] for processes with zero mean and constant mean, and are presented here for the general case of a polynomial time trend for reference in the proofs of Theorem 6 and Lemma 2.

Let $\mathbf{e}_j$ be the $(p+1) \times 1$ vector with 1's in rows $j+3, j+5, \ldots, p+1-j$ and 0's elsewhere, $\mathbf{d}_j$ the $(p+1) \times 1$ vector with 1's in rows $j+2, j+4, \ldots, p+1-j$ and 0's elsewhere, and $\mathbf{0}_{p+1}$ the $(p+1) \times 1$ vector of 0's. Define the $(p+1) \times (p+1)$ matrices $\mathbf{B}_{1p} = \text{diag}\,(0, 1, 2, \ldots, p)$; $\mathbf{B}_{2p} = [-\mathbf{e}_0, -\mathbf{e}_1, \ldots, -\mathbf{e}_{p/2-1}, \mathbf{0}_{p+1}, \mathbf{e}_{p/2-1}, \ldots, \mathbf{e}_1, \mathbf{e}_0]$ if p is even and $\mathbf{B}_{2p} = [-\mathbf{d}_1, -\mathbf{d}_2, \ldots, -\mathbf{d}_{(p-1)/2}, \mathbf{0}_{p+1}, \mathbf{d}_{(p-1)/2}, \ldots, \mathbf{d}_1, \mathbf{d}_0]$ 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

$$\begin{pmatrix} 1 \\ \mathrm{E}(\hat{\alpha}_p^k) \end{pmatrix} = (\mathbf{I}_{p+1} - \mathbf{B}_p^k/T) \begin{pmatrix} 1 \\ \alpha_p \end{pmatrix} + o(1/T), \tag{3.3}$$

where $\mathbf{I}_{p+1}$ is the $(p+1) \times (p+1)$ identity matrix and here α_p denotes the $p \times 1$ vector of autoregressive coefficients. For details of this bias derivation see [10,8,13]. 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 [20]. See also [21, p. 1177].

3.2. Least squares fixed point model coefficients

For each order p of the 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 [8] for $k = 0, 1$.

Theorem 5. *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 $(\mathbf{I}_{p+1} - \mathbf{B}_p^k/T)(1, \tilde{\alpha}_p^{k\prime})' = (1, \tilde{\alpha}_p^{k\prime})'$, $k = 0, 1, \ldots, p = 1, 2, \ldots$. The fixed point for which the first coordinate is equal to 1 is unique.*

Remark 2. The fixed point vector $\tilde{\alpha}_p^k$ is obtained by solving the equations

$$\mathbf{B}_p^k \begin{pmatrix} 1 \\ \tilde{\alpha}_p^k \end{pmatrix} = \mathbf{0}_{p+1}. \tag{3.4}$$

Define

$$\mathbf{B}_p^k = \begin{pmatrix} 0 & \mathbf{0}_p' \\ \mathbf{b}_p^k & \mathbf{B}_{22p}^k \end{pmatrix}, \tag{3.5}$$

where $\mathbf{B}_{22p}^k$ is $p \times p$. Then (3.4) is equivalent to

$$\mathbf{B}_{22p}^k \tilde{\alpha}_p^k = -\mathbf{b}_p^k, \tag{3.6}$$

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

$$-\mathbf{b}_{p+1}^k = (-\mathbf{b}_p^{k\prime}, f_p(k))', \tag{3.7}$$

where $f_p(k)$ is $k+1$ if p is odd and is k if p is even. 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

$$\mathbf{B}_{22,p+1}^k = \begin{pmatrix} \mathbf{B}_{22p}^k - f_p(k)\mathbf{J}_p & -\mathbf{J}_p\mathbf{b}_p^k \\ \mathbf{0}_p' & p+k+2 \end{pmatrix}. \tag{3.8}$$

The equations (3.6) 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 (3.6).

For each value of k the coefficients of the least squares fixed point models implied by Theorem 5 combine to form a Levinson–Durbin sequence. For $k = 0$ and $k = 1$ these sequences were given in [8]. Theorem 5 implies that the bias of least squares estimation of the autoregressive coefficients pulls the estimate toward the fixed point coefficient vector.

Parts of the next theorem were given in [8]. The expressions for $\tilde{\alpha}_{p,p}^k$ for $k = 0$ and 1 can be deduced from Theorems 2 and 3, respectively, in [8], and the Levinson–Durbin structure cited in the theorem was proved by a different method for $k = 1$ in Lemma 5 of [8].

Theorem 6. *Let $\tilde{\alpha}_p^k = (\tilde{\alpha}_{1,p}^k, \ldots, \tilde{\alpha}_{p,p}^k)'$ be the fixed point defined in Theorem 5. 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 $\{\tilde{\alpha}_{j,p}^k,\ j = 1, \ldots, p,\ p = 1, 2, \ldots\}$ with*

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

Remark 3. As $k \to \infty$, $\tilde{\alpha}_{j,p}^k$ converges to the binomial coefficient $\binom{p}{j}$. This was noted in [8].

Table 1 displays the coefficient vectors for the least squares fixed point models for $p = 1, \ldots, 6$. 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 $\tilde{\alpha}_{j,p}^k$ become complicated, and there is no evident completely general pattern.

For a fixed p the binomial coefficients 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 6.

Table 1
Least squares fixed point parameter vectors for $p = 1, \ldots, 6$.

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

Lemma 1. *If p is odd,*

$$j\tilde{\alpha}_{j,p}^k - (p+1-j)\tilde{\alpha}_{p+1-j,p}^k = 0, \quad j = 1, \ldots, p. \tag{3.9}$$

If p is even,

$$j\tilde{\alpha}_{j,p}^k - (p+2-j)\tilde{\alpha}_{p+1-j,p}^k + \sum_{i=0}^{j-1}(-1)^{j-1-i}\tilde{\alpha}_{i,p}^k + \sum_{i=0}^{j-2}(-1)^{j-2-i}\tilde{\alpha}_{p-i,p}^k = 0, \quad j = 1, \ldots, p, \tag{3.10}$$

where $\tilde{\alpha}_{0,p}^k = 1$ and $\sum_0^{-1}(\cdot) = 0$.

Remark 4. Let $\tilde{A}_p^k(z)$ be defined as in (2.1) 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 $\tilde{A}_p^k(z)$ 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 [8].

4. Yule–Walker estimation

Although it can have substantial bias [9], 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

$$g_j = \frac{1}{T}\sum_{t=1}^{T-j}(y_t - \mu)(y_{t+j} - \mu), \quad j = 0, 1, \ldots, p, \tag{4.1}$$

for μ (or $\mu(t)$) known. If the mean is unknown, we replace μ in (4.1) 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 analogue of the Eq. (2.11) 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 $\mathbf{g}_p$ denotes 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

$$\begin{pmatrix} 1 \\ \mathrm{E}(\hat{\alpha}_p^{kYW}) \end{pmatrix} = (\mathbf{I}_{p+1} - \mathbf{B}_p^k/T)\begin{pmatrix} 1 \\ \alpha_p \end{pmatrix} + \begin{pmatrix} 0 \\ \mathbf{G}_p^{-1}\mathbf{d}_p/T \end{pmatrix} + o(1/T), \tag{4.2}$$

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

$$d_{j,p} = \sum_{l=0}^{p}|j-l|\,\gamma(j-l)\alpha_l, \quad j = 1, \ldots, p, \tag{4.3}$$

with the α_l's in (4.3) the autoregressive coefficients; see [9,10]. 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 (4.1). 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 autoregressive coefficients.

Table 2
Comparison of least squares and Yule–Walker fixed point parameter vectors.

p	Least squares	Yule–Walker
Fixed point parameter vectors, $k = 1$		
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, $k = 2$		
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)

Remark 5. 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 $\tilde{\alpha}_{j,p}^{kYW}, j = 1, \ldots, p$, for which the order $1/T$ terms in (4.2) 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 6. Simulation shows that the order $1/T$ bias expression in (4.2) 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 (4.2) can be substantial. However, (4.2) 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.

Let $\tilde{A}_p^{kYW}(z)$ be defined as in (2.1) for the Yule–Walker fixed point model of order p and for degree $k-1$ for the polynomial time trend. The zeros of $\tilde{A}_p^{kYW}(z)$ and of $\tilde{A}_p^{k}(z)$ increase in modulus as p increases, and they tend toward $z = -1$ as k increases. For given p and k the Yule–Walker fixed point zeros have smaller modulus and are farther from $z = -1$ than the corresponding least squares fixed point zeros.

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. Namely, 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 [9] study 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.

Whittle [22] extended the Levinson–Durbin recursion to vector stationary processes. His formulation requires specification of two sequences of matrices, one for each direction of time. In [23] a generalized Levinson–Durbin–Whittle sequence is defined from these two sequences of matrices, and the analogues of Theorems 2–4 for Levinson–Durbin–Whittle sequences are presented.

Other extensions of the Levinson–Durbin recursion have been considered by several authors. For example, Bondon [24] gives Levinson–Durbin algorithms for h-step prediction, one allowing n to advance, and another allowing h to advance. Brockwell and Dahlhaus [25] considers subset prediction generally, for which h-step prediction is a special case, and also presents the h-step algorithm with advance of n. Theorem 2 in the present paper can be extended to the h-step algorithm with advance of n, but the summations obtained are slightly more complicated than (2.3) and (2.4). The h-step algorithms are being studied and will be discussed in a subsequent paper.

6. Proofs

We begin with the symmetry question addressed by Theorem 1.

Proof of Theorem 1. From (1.2), for $n = 2, 3, \ldots, \alpha_{j,n+1} = \alpha_{j,n} + \alpha_{n+1,n+1}\alpha_{n+1-j,n}$ and $\alpha_{n+1-j,n+1} = \alpha_{n+1-j,n} + \alpha_{n+1,n+1}\alpha_{j,n}$, $j = 1, \ldots, n$. Then, if symmetry holds,

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

and if $\alpha_{n+1,n+1} \neq 1$, $\alpha_{j,n} = \alpha_{n+1-j,n}$, $j = 1, \ldots, n$. It follows that $\alpha_{1,n} = \cdots = \alpha_{n,n}$, $n = 2, 3, \ldots$, and then (1.2) implies that $\alpha_{j,n} = \alpha_{1,1}/(1 - (n-1)\alpha_{1,1})$, $j = 1, \ldots, n$, $n = 2, 3, \ldots$. □

The remaining theorems in Section 2 are proved by induction.

Proof of Theorem 2. First consider (2.3). It is certainly true for $n = 1$. Then (2.3) follows by writing (2.2) for $n + 1$ with $z = 1$ and applying the induction hypothesis. The proof of (2.4) follows similarly by writing (2.2) for $n+1$ with $z = -1$. □

Proof of Theorem 3. Consider (2.5), which clearly holds for $n = 1$. As an alternative to (2.1) define $B_n(z) = \sum_{j=0}^{n} \alpha_{j,n} z^j = z^n A_n(z^{-1})$. Then (2.2) implies

$$B_n(z) = B_{n-1}(z) + \alpha_{n,n} z^n B_{n-1}(z^{-1}). \tag{6.1}$$

The left-hand side of (2.5) is equal to $\mathrm{d}B_n(z)/\mathrm{d}z|_{z=1}$. Differentiating both sides of (6.1) for $n + 1$ and applying (2.3), we have

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

Then (2.5) follows by substituting the induction hypothesis for $\sum_{j=1}^{n} j\alpha_{j,n}$ on the right-hand side and rewriting the resulting expression as a sum ranging from 1 to $n + 1$. The proof of (2.6) is the same except that one sets z equal to -1 after differentiating and uses (2.4) instead of (2.3). □

Proof of Theorem 4. We apply induction. It suffices to prove (2.8), because (2.9) follows immediately from (2.8) and (2.3). First note that (2.8) is valid for $n = 1$. Consider the case n odd. Using (1.2), we can write

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

which by the induction hypothesis reduces to (2.8) for $n + 1$. If n is even, we have, similarly to the above for n odd,

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

which reduces to (2.8) for $n + 1$. □

Proof of Theorem 5. The proof is straightforward. It is easy to show by direct calculation that the eigenvalues of the matrix $\mathbf{I}_{p+1} - \mathbf{B}_p^k/T$ are its diagonal elements. Thus there is one eigenvalue equal to 1 and all the others are less than 1 in magnitude if $T > (p + k + 1)/2$. The normalization with first coordinate equal to 1 gives the unique fixed point vector. By direct calculation we find that a vector with first coordinate zero must be the zero vector if it defines a fixed point. □

Proof of Lemma 1. The proofs of (3.9) and (3.10) are identical. To obtain the result for j, subtract row $p + 1 - j$ from row j in the system of Eq. (3.6). Of course, only the subtractions for $j = 1, \ldots, [p/2]$ are needed. □

Stine and Shaman [8, Theorems 2,3 and Lemma 5] gives a proof of Theorem 6 for the cases $k = 0$ and $k = 1$. The proof of Theorem 6 given here is for general k and is simpler.

Proof of Theorem 6. To find the least squares fixed point vector $\tilde{\alpha}_p^k$ we need to solve the system of equations (3.6). The equation from the last row of (3.6) [see also (3.7) and (3.8)] shows that $\tilde{\alpha}_{p,p}^k$ is $k/(p+k+1)$ for p odd and $(k+1)/(p+k+1)$ for p even. The remaining values $\tilde{\alpha}_{j,p}^k$ can then be determined from (3.6) 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 $\tilde{\alpha}_{p,p}^k$ 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

$$\tilde{\alpha}_{p+1}^k = \begin{pmatrix} \tilde{\alpha}_p^k \\ 0 \end{pmatrix} + \tilde{\alpha}_{p+1,p+1}^k \begin{pmatrix} \mathbf{J}_p \tilde{\alpha}_p^k \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_p + \tilde{\alpha}_{p+1,p+1}^k \mathbf{J}_p & \mathbf{0}_p \\ \mathbf{0}_p' & 1 \end{pmatrix} \begin{pmatrix} \tilde{\alpha}_p^k \\ \tilde{\alpha}_{p+1,p+1}^k \end{pmatrix}. \tag{6.2}$$

To do so, it suffices to prove that (6.2) satisfies (3.6) for $p+1$.

(i) Consider p odd. Using (3.8) and (6.2), we have

$$\mathbf{B}_{22,p+1}^k \tilde{\alpha}_{p+1}^k = \begin{pmatrix} \mathbf{B}_{22p}^k - (k+1)\mathbf{J}_p + \tilde{\alpha}_{p+1,p+1}^k \mathbf{B}_{22p}^k \mathbf{J}_p - (k+1)\tilde{\alpha}_{p+1,p+1}^k \mathbf{I}_p & -\mathbf{J}_p \mathbf{b}_p^k \\ \mathbf{0}_p' & p+k+2 \end{pmatrix} \begin{pmatrix} \tilde{\alpha}_p^k \\ \tilde{\alpha}_{p+1,p+1}^k \end{pmatrix}. \tag{6.3}$$

The last row of (6.3) is simply equal to $k+1$ on the right-hand side, as required. The following lemma provides workable expressions for $\mathbf{B}_{22p}^k \mathbf{J}_p$.

Lemma 2. *If p is odd,*

$$\mathbf{B}_{22p}^k \mathbf{J}_p = \operatorname{diag}(k+2, k+3, \ldots, k+p+1)(\mathbf{I}_p + \mathbf{J}_p) - \mathbf{B}_{22p}^k. \tag{6.4}$$

If p is even,

$$\mathbf{B}_{22p}^k \mathbf{J}_p = \operatorname{diag}(k+2, k+3, \ldots, k+p+1)(\mathbf{I}_p + \mathbf{J}_p) - \mathbf{B}_{22p}^k + \sum_{j=1}^{p-1} (-1)^j \mathbf{U}_p^j (\mathbf{I}_p + \mathbf{J}_p), \tag{6.5}$$

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

Proof. (i) First consider p odd. Inspecting the structure of $\mathbf{B}_{22p}^k$, we may write

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

where n_j^k is equal to k for j odd and $k+1$ for j even. Then (6.4) follows if we multiply on the right by $\mathbf{I}_p + \mathbf{J}_p$ (note that $\mathbf{J}_p^2 = \mathbf{I}_p$). If p is even, the proof is similar. Write

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

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

By multiplication the first p rows of (6.3) reduce to

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

where we have used (6.4) and the fact that $\mathbf{J}_p \mathbf{b}_p^k = \mathbf{b}_p^k$ for p odd. With $k+1 = (k+p+2)\tilde{\alpha}_{p+1,\,p+1}^k$, this simplifies to

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

which is $-\mathbf{b}_p^k$ by Lemma 1. Combining this with the last row of (6.3), we complete the proof of Theorem 6 for p odd.

(ii) If p is even the proof follows similarly from (3.7), (3.8), (3.10) and (6.5). □

Acknowledgment

The author is grateful to the referees for careful reading and very helpful comments and suggestions.

References

[1] N. Levinson, The Wiener RMS (root-mean-square) error criterion in filter design and prediction, J. Math. Phys. Mass. Inst. Tech. 25 (1947) 261–278.
[2] N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications, MIT Press, Cambridge, MA, 1949.

[3] T. Kailath, A view of three decades of linear filtering theory, IEEE Trans. Inform. Theory 20 (1974) 146–181.
[4] M.S. Bartlett, An Introduction to Stochastic Processes, Cambridge Univ. Press, Cambridge, 1955.
[5] H.E. Daniels, The approximate distribution of serial correlation coefficients, Biometrika 43 (1956) 169–185.
[6] J. Durbin, The fitting of time-series models, Rev. Inst. Internat. Statist. 28 (1960) 233–244.
[7] S.M. Kay, Recursive maximum likelihood estimation of autoregressive processes, IEEE Trans. Acoust. Speech Signal Process. 31 (1983) 56–65.
[8] R.A. Stine, P. Shaman, A fixed point characterization for bias of autoregressive estimators, Ann. Statist. 17 (1989) 1275–1284.
[9] D. Tjøstheim, J. Paulsen, Bias of some commonly-used time series estimates, Biometrika 70 (1983) 389–399.
[10] P. Shaman, R.A. Stine, The bias of autoregressive coefficient estimators, J. Amer. Statist. Assoc. 83 (1988) 842–848.
[11] R. Dahlhaus, Parameter estimation of stationary processes with spectra containing strong peaks, in: J. Franke, W. Härdle, D. Martin (Eds.), Robust and Nonlinear Time Series Analysis, Springer-Verlag, New York, 1984, pp. 50–67.
[12] H.-C. Zhang, Reduction of the asymptotic bias of autoregressive and spectral estimators by tapering, J. Time Ser. Anal. 13 (1992) 451–469.
[13] D.T. Pham, On the asymptotic expansions for the bias and covariance matrix of autoregressive estimators, in: T. Subba Rao (Ed.), Developments in Time Series Analysis, Chapman and Hall, London, 1993, pp. 80–100.
[14] P.J. Brockwell, R.A. Davis, Introduction to Time Series and Forecasting, second ed., Springer-Verlag, New York, 2002.
[15] O. Barndorff-Nielsen, G. Schou, On the parametrization of autoregressive models by partial autocorrelations, J. Multivariate Anal. 3 (1973) 408–419.
[16] J. Wise, Stationarity conditions for stochastic processes of the autoregressive and moving-average type, Biometrika 43 (1956) 215–219.
[17] I.S. Gradshteyn, I.M. Ryzhik, Tables of Integrals, Series, and Products, Corrected and Revised ed., Academic Press, New York, 1980.
[18] R.A. Lewis, G.C. Reinsel, Prediction error of multivariate time series with misspecified models, J. Time Ser. Anal. 9 (1988) 43–57.
[19] R.J. Bhansali, Effects of not knowing the order of an autoregressive process on the mean squared error of prediction. I, J. Amer. Statist. Assoc. 76 (1981) 588–597.
[20] T.W. Anderson, The Statistical Analysis of Time Series, Wiley, New York, 1971.
[21] W.-K. Cheang, G.C. Reinsel, Bias reduction of autoregressive estimates in time series regression model through restricted maximum likelihood, J. Amer. Statist. Assoc. 95 (2000) 1173–1184.
[22] P. Whittle, On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix, Biometrika 50 (1963) 129–134.
[23] P. Shaman, Properties of generalized Levinson–Durbin–Whittle sequences, J. Statist. Plann. Inference 138 (2008) 2808–2814.
[24] P. Bondon, Recursive relations for multistep prediction of a stationary time series, J. Time Ser. Anal. 22 (2001) 399–410.
[25] P.J. Brockwell, R. Dahlhaus, Generalized Levinson–Durbin and Burg algorithms, J. Econometrics 118 (2004) 129–149.