# The Asymptotic Properties of Burg Estimators

by

Günter Hainz

Universität Heidelberg

Address:

Günter Hainz

Institut für Angewandte Mathematik

Universität Heidelberg

Im Neuenheimer Feld 294

D-69120 Heidelberg

Germany

Abbreviated title: Properties of Burg Estimators

**Abstract**

There are estimators for multivariate autoregressive models which are regarded as multivariate versions of Burg's univariate estimator. For two of these multivariate Burg estimators the asymptotic equivalence with the Yule-Walker estimator is established in this paper, so central limit theorems for the Yule-Walker estimator extend to these estimators. Furthermore, the asymptotic bias of the univariate Burg estimator to terms of $n^{-1}$ is shown to be the same as the bias of the least-squares estimator; $n$ is the number of observations. The main results are true even for mis-specified models.

**AMS 1991 subject classifications.** Primary 62 M 10.

**Keywords and phrases.** Burg estimator; asymptotic bias; central limit theorem.

# 1 Preliminaries

The most popular estimators for autoregressive models seem to be the Yule-Walker, the least squares, and the Burg estimators. The Burg estimator was introduced by Burg (1968) for univariate time series, and it was generalized to multivariate models by Morf et al. (1978), Strand (1977) and others (cf. Jones, 1978).

Central limit theorems are well known for the Yule-Walker and the least squares estimators, as is the asymptotic bias of them in the univariate case (Shaman and Stine,1988). Nicholls and Pope (1988) also calculated the asymptotic bias of the multivariate least squares estimator. Kay and Makhoul (1983) showed the asymptotic equivalence of the univariate Burg estimator and the Yule-Walker estimator, but neither the asymptotic distribution of the multivariate Burg estimator nor the bias seem to be known. Simulations indicate that the bias of the univariate Burg estimator is about as large as the bias of the least squares estimator (Lysne and Tjøstheim,1987), which tends to be smaller than the bias of the YW estimator, especially if the process has roots near the unit circle (Shaman and Stine,1988). But unlike the least squares estimator, the Burg estimators are stable (or: causal), which is a property often asked for.

In this paper the asymptotic properties of the Burg estimators are investigated further: After having defined the estimators, the asymptotic equivalence of the multivariate Burg estimator and the Yule-Walker estimator is established in Section 2; the equivalence holds even for mis-specified models. In Section 3 the asymptotic bias of the univariate Burg estimator is shown to be the same as the bias of the least squares estimator.

We assume that $\{X_t\}$, $t \in \mathbf{Z}$, is a $d$-variate stationary ergodic stochastic process of full rank with values in $\mathbf{R}^d$; let the mean of its components be zero and the variance be finite. The process can be forecasted by the linear predictor $\hat{X}_t^{(f)} := \sum_{j=1}^{p} A_j^{(p)} X_{t-j}$ or, in reversed time, by $\hat{X}_t^{(b)} := \sum_{j=1}^{p} B_j^{(p)} X_{t+j}$, where $A_j^{(p)}$ and $B_j^{(p)}$ are the $d \times d$ matrices which minimize the traces of $S_p^{(f)} := E(X_t - \hat{X}_t^{(f)})(X_t - \hat{X}_t^{(f)})^\top$ and $S_p^{(b)} := E(X_t - \hat{X}_t^{(b)})(X_t - \hat{X}_t^{(b)})^\top$.

It is well known that this leads to the Yule-Walker equations

$$\begin{pmatrix} \mathbf{1} & -A_1^{(p)} & \dots & -A_p^{(p)} \\ -B_p^{(p)} & \dots & -B_1^{(p)} & \mathbf{1} \end{pmatrix} \cdot R_{(p+1)} = \begin{pmatrix} S_p^{(f)} & 0 & \dots & 0 \\ 0 & \dots & 0 & S_p^{(b)} \end{pmatrix}, \tag{1.1}$$

where

$$R_{(p+1)} := \begin{pmatrix} R_0 & R_1 & \dots & R_p \\ R_1^\top & R_0 & \dots & R_{p-1} \\ \vdots & & \ddots & \vdots \\ R_p^\top & R_{p-1}^\top & \dots & R_0 \end{pmatrix} \tag{1.2}$$

is a regular matrix of autocovariances $R_i := EX_{t+i}X_t^\top$; $\mathbf{0}$ and $\mathbf{1}$ are the $d \times d$ zero and identity matrices. If $d = 1$, then $A_j^{(p)} = B_j^{(p)}$ and $S_p^{(f)} = S_p^{(b)}$ hold, and we define $\phi_j^{(p)} := A_j^{(p)}$ and $S_p := S_p^{(f)}$. The autocovariances are estimated by $\hat{R}_i := \frac{1}{n}\sum_{i=1}^{n-i} X_{t+i}X_t^\top$, $\hat{R}_{-i} := \hat{R}_i^\top$ $(i \geq 0)$, where $X_1, \dots, X_n$ are the available data. If the mean of the process is unknown, we subtract the arithmetic mean $\frac{1}{n}\sum_{i=1}^{n} X_i$ from the data.

After replacing $R_i$ by $\hat{R}_i$ in (1.2) we get the Yule-Walker (YW) estimator $\hat{A}_1^{(p)}, \dots, \hat{A}_p^{(p)}$, $\hat{B}_1^{(p)}, \dots, \hat{B}_p^{(p)}$,$\hat{S}_p^{(f)}$, $\hat{S}_p^{(b)}$ as solution of (1.1); in the univariate case we get $\hat{\phi}_1^{(p)}, \dots, \hat{\phi}_p^{(p)}$, $\hat{S}_p$. $\hat{R}_{(p+1)}$ denotes the corresponding estimator of $R_{(p+1)}$.

The YW estimator can be calculated recursively using the (multivariate) Levinson-Durbin algorithm (Whittle,1963):

$$\hat{S}_0^{(b)} := \hat{S}_0^{(f)} := \hat{R}_0,$$

recursion for $k \geq 1$:

$$
\begin{aligned}
\hat{A}_k^{(k)} &:= (\hat{R}_k - \sum_{j=1}^{k-1} \hat{A}_j^{(k-1)} \hat{R}_{k-j}) \hat{S}_{k-1}^{(b)-1}, \\
\hat{B}_k^{(k)} &:= (\hat{R}_k^\top - \sum_{j=1}^{k-1} \hat{B}_j^{(k-1)} \hat{R}_{k-j}^\top) \hat{S}_{k-1}^{(f)-1}, \\
\hat{S}_k^{(f)} &:= (\mathbf{1} - \hat{A}_k^{(k)} \hat{B}_k^{(k)}) \hat{S}_{k-1}^{(f)}, \\
\hat{S}_k^{(b)} &:= (\mathbf{1} - \hat{B}_k^{(k)} \hat{A}_k^{(k)}) \hat{S}_{k-1}^{(b)}, \\
\hat{A}_j^{(k)} &:= \hat{A}_j^{(k-1)} - \hat{A}_k^{(k)} \hat{B}_{k-j}^{(k-1)} \qquad (j = 1 \dots k-1), \\
\hat{B}_j^{(k)} &:= \hat{B}_j^{(k-1)} - \hat{B}_k^{(k)} \hat{A}_{k-j}^{(k-1)} \qquad (j = 1 \dots k-1).
\end{aligned}
\tag{1.3}
$$

For $d = 1$, Burg (1968) used the same recursive algorithm but estimated $\phi_k^{(k)}$ by

$$
\tilde{\phi}_k^{(k)} := \frac{\sum_{t=k+1}^n \tilde{e}_t^{(k-1)} \tilde{\eta}_{t-k}^{(k-1)}}{\frac{1}{2} \sum_{t=k+1}^n (\tilde{e}_t^{(k-1)2} + \tilde{\eta}_{t-k}^{(k-1)2})},
\tag{1.4}
$$

where $\tilde{e}_t^{(k-1)}$ and $\tilde{\eta}_t^{(k-1)}$ are the estimated forward and backward prediction errors

$$
\tilde{e}_t^{(k-1)} = X_t - \sum_{j=1}^{k-1} \tilde{\phi}_j^{(k-1)} X_{t-j} \quad \text{and} \quad \tilde{\eta}_t^{(k-1)} = X_t - \sum_{j=1}^{k-1} \tilde{\phi}_j^{(k-1)} X_{t+j},
\tag{1.5}
$$

which can be calculated recursively by $\tilde{e}_t^{(k)} = \tilde{e}_t^{(k-1)} - \tilde{\phi}_k^{(k)} \tilde{\eta}_{t-k}^{(k-1)}$ and $\tilde{\eta}_{t-k}^{(k)} = \tilde{\eta}_{t-k}^{(k-1)} - \tilde{\phi}_k^{(k)} \tilde{e}_t^{(k-1)}$. For the multivariate case several versions of this method were proposed (see Jones, 1978), the most popular seeming to be the one described by Morf et al. (1978), which uses the Levinson-Durbin algorithm with some alterations:

$$
\begin{aligned}
\tilde{\rho}_k &:= (\sum_{t=k+1}^n \tilde{e}_t^{(k-1)} \tilde{e}_t^{(k-1)\top})^{-1/2} (\sum_{t=k+1}^n \tilde{e}_t^{(k-1)} \tilde{\eta}_{t-k}^{(k-1)\top}) (\sum_{t=k+1}^n \tilde{\eta}_{t-k}^{(k-1)} \tilde{\eta}_{t-k}^{(k-1)\top})^{-\top/2}, \\
\tilde{A}_k^{(k)} &:= \tilde{S}_{k-1}^{(f)1/2} \cdot \tilde{\rho}_k \cdot \tilde{S}_{k-1}^{(b)-1/2}, \\
\tilde{B}_k^{(k)} &:= \tilde{S}_{k-1}^{(b)1/2} \cdot \tilde{\rho}_k^\top \cdot \tilde{S}_{k-1}^{(f)-1/2}, \\
\tilde{e}_t^{(k)} &:= \tilde{e}_t^{(k-1)} - \tilde{A}_k^{(k)} \cdot \tilde{\eta}_{t-k}^{(k-1)}, \\
\tilde{\eta}_{t-k}^{(k)} &:= \tilde{\eta}_{t-k}^{(k-1)} - \tilde{B}_k^{(k)} \cdot \tilde{e}_t^{(k-1)},
\end{aligned}
\tag{1.6}
$$

where $A^{1/2}$ is the lower triangular matrix with positive diagonal elements defined by the Cholesky decomposition of the symmetric, positive definite matrix $A = A^{1/2} \cdot A^{\top/2}$; further

$A^{\top/2} := (A^{1/2})^\top$, $A^{-1/2} := (A^{1/2})^{-1}$, $A^{-\top/2} := (A^{-1/2})^\top$. The matrices $\tilde{S}_k^{(f)}$, $\tilde{S}_k^{(b)}$, $\tilde{A}_j^{(k)}$, $\tilde{B}_j^{(k)}$ $(j = 1 \ldots k-1)$ are defined as in the Levinson-Durbin algorithm.

A different version was suggested by Strand (1977): Here $\tilde{A}_k^{(k)}$ is the solution of

$$\begin{aligned}\tilde{A}_k^{(k)}\Big(\frac{1}{n}\sum_{t=k+1}^{n}\tilde{\eta}_{t-k}^{(k-1)}\tilde{\eta}_{t-k}^{(k-1)\top}\Big)\tilde{S}_{k-1}^{(b)-1} &+ \Big(\frac{1}{n}\sum_{t=k+1}^{n}\tilde{e}_t^{(k-1)}\tilde{e}_t^{(k-1)\top}\Big)\tilde{S}_{k-1}^{(f)-1}\tilde{A}_k^{(k)} = \\ &= \frac{2}{n}\sum_{t=k+1}^{n}\tilde{e}_t^{(k-1)}\tilde{\eta}_{t-k}^{(k-1)\top}\tilde{S}_{k-1}^{(b)-1} \quad ,\end{aligned} \tag{1.7}$$

and $\tilde{B}_k^{(k)} := (\tilde{S}_{k-1}^{(f)-1}\tilde{A}_k^{(k)}\tilde{S}_{k-1}^{(b)})^\top$.

Unlike the other method, the algorithm proposed by Strand (1977) reduces to the univariate Burg estimator (1.4) for $d = 1$, but it is more expensive to calculate. Both Burg estimators are known to be stable and are calculable recursively, which is particularly useful if a model selection procedure has to be performed simultaneously; the same is true for the YW estimator.

## 2 Asymptotic Distribution of Multivariate Burg Estimators

We want to find central limit theorems for both multivariate versions of the Burg estimator by showing the asymptotic equivalence with the Yule-Walker estimator. In the following $\tilde{A}^{(p)}$ always denotes one of the Burg estimators of section 1.

**Theorem 1** *Under the assumptions on the process mentioned in section 1,*

$$\tilde{A}^{(p)} = \hat{A}^{(p)} + O_p(1/n) \quad , \quad \tilde{B}^{(p)} = \hat{B}^{(p)} + O_p(1/n),$$
$$\tilde{S}_p^{(f)} = \hat{S}_p^{(f)} + O_p(1/n) \quad , \quad \tilde{S}_p^{(b)} = \hat{S}_p^{(b)} + O_p(1/n)$$

*hold, where* $\tilde{A}^{(p)} = (\tilde{A}_1^{(p)}, \ldots, \tilde{A}_p^{(p)})$ *etc.*

**Proof.**

First the lemma is proved by induction for the method of Morf et al., then for Strand's method.

$$\begin{aligned}\tilde{A}_1^{(1)} &= \tilde{S}_0^{(f)1/2}\Big(\frac{1}{n}\sum_{t=2}^{n}X_tX_t^\top\Big)^{-1/2}\Big(\frac{1}{n}\sum_{t=2}^{n}X_tX_{t-1}^\top\Big)\Big(\frac{1}{n}\sum_{t=2}^{n}X_{t-1}X_{t-1}^\top\Big)^{-\top/2}\tilde{S}_0^{(b)-1/2} = \\ &= \tilde{S}_0^{(f)1/2}(\tilde{S}_0^{(f)} + O_p(1/n))^{-1/2}\hat{R}_1(\tilde{S}_0^{(b)} + O_p(1/n))^{-\top/2}\tilde{S}_0^{(b)-1/2} = \\ &= \hat{R}_1\hat{R}_0^{-1} + O_p(1/n) = \hat{A}_1^{(1)} + O_p(1/n),\end{aligned}$$

and similarly $\tilde{B}_1^{(1)} = \hat{B}_1^{(1)} + O_p(1/n)$, from which

$$\tilde{S}_1^{(f)} = (\mathbf{1} - \hat{A}_1^{(1)}\hat{B}_1^{(1)} + O_p(1/n))(\hat{S}_0^{(f)} + O_p(1/n)) = \hat{S}_1^{(f)} + O_p(1/n)$$

and $\tilde{S}_1^{(b)} = \hat{S}_1^{(b)} + O_p(1/n)$ follow.

Now we assume that the statement of the lemma has already been proved for $1, \ldots, k-1 \leq p-1$, i.e.

$$\begin{aligned} \tilde{A}^{(k-1)} = \hat{A}^{(k-1)} + O_p(1/n) \quad &, \quad \tilde{B}^{(k-1)} = \hat{B}^{(k-1)} + O_p(1/n), \\ \tilde{S}_{k-1}^{(f)} = \hat{S}_{k-1}^{(f)} + O_p(1/n) \quad &, \quad \tilde{S}_{k-1}^{(b)} = \hat{S}_{k-1}^{(b)} + O_p(1/n). \end{aligned} \tag{2.1}$$

Also let the following assertions be shown, which are obviously true for $k = 2$:

$$\begin{aligned} \tilde{S}_{k-2}^{(f)} &= \frac{1}{n}\sum_{t=k}^{n} \tilde{e}_t^{(k-2)}\tilde{e}_t^{(k-2)\top} + O_p(1/n), \\ \tilde{S}_{k-2}^{(b)} &= \frac{1}{n}\sum_{t=k}^{n} \tilde{\eta}_{t-k+1}^{(k-2)}\tilde{\eta}_{t-k+1}^{(k-2)\top} + O_p(1/n), \end{aligned} \tag{2.2}$$

$$\hat{A}_{k-1}^{(k-1)} \cdot \hat{S}_{k-2}^{(b)} = \frac{1}{n}\sum_{t=k}^{n} \tilde{e}_t^{(k-2)}\tilde{\eta}_{t-k+1}^{(k-2)\top} + O_p(1/n), \tag{2.3}$$

$$\hat{B}_{k-1}^{(k-1)} \cdot \hat{S}_{k-2}^{(f)} = \frac{1}{n}\sum_{t=k}^{n} \tilde{\eta}_{t-k+1}^{(k-2)}\tilde{e}_t^{(k-2)\top} + O_p(1/n). \tag{2.4}$$

Next we prove (2.2)–(2.4) for $k+1$ instead of $k$:

Proof of (2.2):

$$\begin{aligned} \frac{1}{n}\sum_{t=k+1}^{n} \tilde{e}_t^{(k-1)}\tilde{e}_t^{(k-1)\top} &= \frac{1}{n}\sum_{t=k+1}^{n} (\tilde{e}_t^{(k-2)} - \tilde{A}_{k-1}^{(k-1)}\tilde{\eta}_{t-k+1}^{(k-2)})(\tilde{e}_t^{(k-2)} - \tilde{A}_{k-1}^{(k-1)}\tilde{\eta}_{t-k+1}^{(k-2)})^\top = \\ &= \tilde{S}_{k-2}^{(f)} - \tilde{A}_{k-1}^{(k-1)}\hat{B}_{k-1}^{(k-1)}\hat{S}_{k-2}^{(f)} - \\ &\quad - \hat{A}_{k-1}^{(k-1)}\hat{S}_{k-2}^{(b)}\tilde{A}_{k-1}^{(k-1)\top} + \tilde{A}_{k-1}^{(k-1)}\tilde{S}_{k-2}^{(b)}\tilde{A}_{k-1}^{(k-1)\top} + O_p(1/n) = \\ &= \tilde{S}_{k-2}^{(f)} - \tilde{A}_{k-1}^{(k-1)}\tilde{B}_{k-1}^{(k-1)}\tilde{S}_{k-2}^{(f)} + O_p(1/n) = \tilde{S}_{k-1}^{(f)} + O_p(1/n), \end{aligned}$$

using (2.1)–(2.4). Similarly $\tilde{S}_{k-1}^{(b)} = \frac{1}{n}\sum_{t=k+1}^{n} \tilde{\eta}_{t-k}^{(k-1)}\tilde{\eta}_{t-k}^{(k-1)\top} + O_p(1/n)$ can be shown.

Proof of (2.3):

$$\begin{aligned}
\frac{1}{n}\sum_{t=k+1}^{n}\tilde{e}_t^{(k-1)}\tilde{\eta}_{t-k}^{(k-1)\top} &= \frac{1}{n}\sum_{t=k+1}^{n}(X_t-\sum_{j=1}^{k-1}\tilde{A}_j^{(k-1)}X_{t-j})(X_{t-k}-\sum_{j=1}^{k-1}\tilde{B}_j^{(k-1)}X_{t-k+j})^\top = \\
&= \hat{R}_k-\sum_{j=1}^{k-1}\hat{R}_{k-j}\hat{B}_j^{(k-1)\top}-\sum_{j=1}^{k-1}\hat{A}_j^{(k-1)}\hat{R}_{k-j}+ \\
&\quad +\sum_{j,i=1}^{k-1}\hat{A}_j^{(k-1)}\hat{R}_{k-j-i}\hat{B}_i^{(k-1)\top}+O_p(1/n)= \\
&= \hat{A}_k^{(k)}\hat{S}_{k-1}^{(b)}+O_p(1/n),
\end{aligned}$$

using (2.1), (1.1), and the definition of $\hat{A}_k^{(k)}$ in the Levinson-Durbin algorithm (1.3).

Proof of (2.4): Because $\hat{A}_k^{(k)}\hat{S}_{k-1}^{(b)}=(\hat{B}_k^{(k)}\hat{S}_{k-1}^{(f)})^\top$ holds and (2.3) has been shown,

$$\begin{aligned}
\frac{1}{n}\sum_{t=k+1}^{n}\tilde{\eta}_{t-k}^{(k-1)}\tilde{e}_t^{(k-1)\top} &= (\frac{1}{n}\sum_{t=k+1}^{n}\tilde{e}_t^{(k-1)}\tilde{\eta}_{t-k}^{(k-1)\top})^\top=(\hat{A}_k^{(k)}\cdot\hat{S}_{k-1}^{(b)}+O_p(1/n))^\top= \\
&= \hat{B}_k^{(k)}\hat{S}_{k-1}^{(f)}+O_p(1/n).
\end{aligned}$$

From these considerations follows (2.1) for $k+1$:

$$\begin{aligned}
\tilde{A}_k^{(k)} &= \tilde{S}_{k-1}^{(f)1/2}(\tilde{S}_{k-1}^{(f)}+O_p(1/n))^{-1/2}(\hat{A}_k^{(k)}\hat{S}_{k-1}^{(b)}+O_p(1/n))(\tilde{S}_{k-1}^{(b)}+O_p(1/n))^{-\top/2}\tilde{S}_{k-1}^{(b)-1/2}= \\
&= \hat{A}_k^{(k)}\hat{S}_{k-1}^{(b)}\tilde{S}_{k-1}^{(b)-1}+O_p(1/n)=\hat{A}_k^{(k)}+O_p(1/n).
\end{aligned}$$

The proof for $\tilde{B}_k^{(k)},\tilde{S}_k^{(f)},\tilde{S}_k^{(b)},\tilde{A}_j^{(k)},\tilde{B}_j^{(k)}\quad(j=1\ldots k-1)$ is obvious.

The lemma can be proved the same way for the estimator of Strand (1977), because

$$\tilde{A}_k^{(k)}=\hat{A}_k^{(k)}+O_p(1/n)$$

follows from (1.7), using (2.2) and (2.3):

$$\tilde{A}_k^{(k)}(\tilde{S}_{k-1}^{(b)}+O_p(1/n))+(\tilde{S}_{k-1}^{(f)}+O_p(1/n))\tilde{S}_{k-1}^{(f)-1}\tilde{A}_k^{(k)}\tilde{S}_{k-1}^{(b)} \quad = \quad 2\cdot(\hat{A}_k^{(k)}\cdot\hat{S}_{k-1}^{(b)}+O_p(1/n)).$$

□

It should be noted that the equivalence still holds if the mean of the process has to be estimated by the arithmetic mean of the data.

Now central limit theorems for the Yule-Walker estimator can be extended to the Burg estimators under the very general assumptions of section 1.

**Theorem 2** *Let $\{X_t\}$, $t \in \mathbf{Z}$, be a d-variate stable AR(p)-process with coefficients $A_1^{(p)}, \ldots, A_p^{(p)}$ and with independent and identically distributed innovations; the innovations have zero mean vector and a regular covariance matrix $\Sigma$. Then*

$$\sqrt{n}(vec\tilde{A}^{(p)} - vecA^{(p)}) \implies \mathcal{N}(0, R_{(p)}^{-1} \otimes \Sigma),$$
$$\tilde{S}_p^{(f)} \longrightarrow \Sigma \text{ in probability}$$

*hold. Here $vecA^{(p)} = vec(A_1^{(p)}, \ldots, A_p^{(p)})$ is the vector which is created by stacking the columns of $A^{(p)}$, and $\otimes$ denotes the Kronecker product.*

This theorem is proved in Hannan (1970), ch. VI.2, Th. 1., for the YW estimator.

The same way central limit theorems for mis-specified models, as stated e.g. in Lewis and Reinsel (1988), (3.5), can be extended to the Burg estimators,too.

## 3 Bias of the Univariate Burg Estimator

In the following we investigate the asymptotic bias of the univariate Burg estimator $\tilde{\phi}^{(p)} = (\tilde{\phi}_1^{(p)}, \ldots, \tilde{\phi}_p^{(p)})^\top$ to terms of order $n^{-1}$; it will be shown that it is as large as the asymptotic bias of the least squares estimator $\check{\phi}^{(p)}$, which was calculated by Shaman and Stine (1988) for a true model.

They used the assumptions $EX_t^{16} < \infty$ and $E(|\hat{R}_{(p)}^{-1} - R_{(p)}^{-1}|^8) = O(1)$, where $|A|$ is the absolute value of the largest eigenvalue of the matrix $A$.

In addition to that we assume for all $r \in \mathbf{N}$(for simplicity)

$$EX_t^r < \infty, \qquad E\hat{N}_p^{-r} = O(1), \qquad E\tilde{N}_p^{-r} = O(1),$$

where $\tilde{N}_p := \frac{1}{2n}\sum_{t=p+1}^{n}(\tilde{e}_t^{(p-1)2} + \tilde{\eta}_{t-p}^{(p-1)2})$ and $\hat{N}_p := \hat{R}_0 - \sum_{j=1}^{p-1}\hat{\phi}_j^{(p-1)}\hat{R}_j$; these assumptions guarantee the uniform integrability of the terms appearing in the proof.

With the definitions $\tilde{Z}_p := \frac{1}{n}\sum_{t=p+1}^{n}\tilde{e}_t^{(p-1)}\tilde{\eta}_{t-p}^{(p-1)}$ and $\hat{Z}_p := \hat{R}_p - \sum_{j=1}^{p-1}\hat{\phi}_j^{(p-1)}\hat{R}_{p-j}$, $\hat{\phi}_p^{(p)} = \hat{Z}_p/\hat{N}_p$ and $\tilde{\phi}_p^{(p)} = \tilde{Z}_p/\tilde{N}_p$ hold.

As the Burg estimator is defined recursively and no other useful representation of it is known, we compare it with the YW estimator, which can be calculated recursively by the Levinson-Durbin algorithm, too. For the difference between YW and least squares estimators

$$\lim_{n\to\infty} n \cdot E(\hat{\phi}^{(p)} - \check{\phi}^{(p)}) = R_{(p)}^{-1} \cdot d_{(p)}$$

holds (see Shaman and Stine,1988, (3.7)), where $\hat{\phi}^{(p)}$ is the YW estimator, $\check{\phi}^{(p)}$ is the least squares estimator and $d_{(p)} := (d_{(p),1}, \ldots, d_{(p),p})^\top$, with

$$d_{(p),j} := \sum_{k=0}^{p} |j-k| \, R_{j-k} \Phi_k^{(p)}$$

and

$$\Phi_{-1}^{(p)} := 0, \quad \Phi_0^{(p)} := -1, \quad \Phi_j^{(p)} := \phi_j^{(p)} \qquad (j = 1 \ldots p).$$

We will prove that

$$\lim_{n\to\infty} n \cdot E(\tilde{\phi}^{(p)} - \hat{\phi}^{(p)}) = -R_{(p)}^{-1} \cdot d_{(p)}, \tag{3.1}$$

so that for the bias of $\tilde{\phi}^{(p)}$

$$\lim_{n\to\infty} n \cdot E(\tilde{\phi}^{(p)} - \phi^{(p)}) = \lim_{n\to\infty} n \cdot E(\check{\phi}^{(p)} - \phi^{(p)})$$

holds.

**Theorem 3** *Under the assumptions mentioned above and in section 1, the bias of the univariate Burg estimator is equal to the bias of the least squares estimator,*

$$\lim_{n\to\infty} n \cdot E(\tilde{\phi}^{(p)} - \phi^{(p)}) = \lim_{n\to\infty} n \cdot E(\check{\phi}^{(p)} - \phi^{(p)}).$$

**Proof.**

The proof follows by induction; if $EX_t$ is estimated, $X_t$ has to be substituted by $X_t - \frac{1}{n}\sum_{j=1}^{n} X_j$ everywhere.

$$\tilde{\phi}_1^{(1)} = \frac{\frac{2}{n}\sum_{t=2}^{n} X_t X_{t-1}}{\frac{1}{n}\sum_{t=2}^{n}(X_t^2 + X_{t-1}^2)} = \frac{\hat{R}_1}{\hat{R}_0 - \frac{1}{2n}(X_1^2 + X_n^2)} = \hat{\phi}_1^{(1)}(1 + \frac{1}{2n\hat{R}_0}(X_1^2 + X_n^2) + \hat{h}_n^{(1)}).$$

The proof of the uniform integrability of the terms appearing here and in the following is not difficult and therefore omitted.

As $\hat{h}_n^{(1)} = O_p(n^{-2})$ and $E(|n \cdot \hat{h}_n^{(1)} \cdot \hat{\phi}_1^{(1)}|) \to 0$, the assertion of the theorem holds for $p = 1$:

$$n \cdot E(\tilde{\phi}_1^{(1)} - \hat{\phi}_1^{(1)}) = E(\frac{\hat{\phi}_1^{(1)}(X_1^2 + X_n^2)}{2 \cdot \hat{R}_0} + n \cdot \hat{h}_n^{(1)}\hat{\phi}_1^{(1)}) \longrightarrow \phi_1^{(1)} = -R_{(1)}^{-1} \cdot d_{(1)}.$$

Let now $nE(\tilde{\phi}^{(i)} - \hat{\phi}^{(i)}) \to -R_{(i)}^{-1} \cdot d_{(i)}$ be shown for $1 \le i \le p-1$. Using a Taylor expansion

$$\tilde{\phi}_p^{(p)} = \frac{\hat{Z}_p + (\tilde{Z}_p - \hat{Z}_p)}{\hat{N}_p + (\tilde{N}_p - \hat{N}_p)} = \frac{1}{\hat{N}_p}\{\hat{Z}_p + (\tilde{Z}_p - \hat{Z}_p)\}\{1 - \frac{\tilde{N}_p - \hat{N}_p}{\hat{N}_p} + \hat{h}_n^{(p)}\},$$

where $\hat{h}_n^{(p)}$ is the remainder term of order $O_p(n^{-2})$, we find from (1.4) and (1.5) that

$$\tilde{\phi}_p^{(p)} =$$

$$= \frac{\frac{2}{n}\sum_{t=p+1}^{n}(X_tX_{t-p} - X_{t-p}\sum_{j=1}^{p-1}X_{t-j}\tilde{\phi}_j^{(p-1)} - X_t\sum_{j=1}^{p-1}X_{t-p+j}\tilde{\phi}_j^{(p-1)} + \ldots}{\frac{1}{n}\sum_{t=p+1}^{n}(X_t^2 - 2X_t\sum_{j=1}^{p-1}X_{t-j}\tilde{\phi}_j^{(p-1)} + \sum_{j,i=1}^{p-1}\tilde{\phi}_j^{(p-1)}\tilde{\phi}_i^{(p-1)}X_{t-j}X_{t-i} + \ldots}\cdots$$

$$\cdots\frac{\ldots + \sum_{j,i=1}^{p-1}\tilde{\phi}_j^{(p-1)}\tilde{\phi}_i^{(p-1)}X_{t-j}X_{t-p+i})}{\ldots + X_{t-p}^2 - 2X_{t-p}\sum_{j=1}^{p-1}X_{t-p+j}\tilde{\phi}_j^{(p-1)} + \sum_{j,i=1}^{p-1}\tilde{\phi}_j^{(p-1)}\tilde{\phi}_i^{(p-1)}X_{t-p+j}X_{t-p+i})} =$$

$$= \frac{1}{\hat{N}_p}\Big\{\hat{Z}_p - 2\cdot\sum_{j=1}^{p-1}(\tilde{\phi}_j^{(p-1)} - \hat{\phi}_j^{(p-1)})\hat{R}_{p-j} + \frac{1}{n}\sum_{j=1}^{p-1}(\tilde{\phi}_j^{(p-1)} - \hat{\phi}_j^{(p-1)})(2j \text{ terms of form } X_tX_{t+p-j}) +$$

$$+ \frac{1}{n}\sum_{j=1}^{p-1}\hat{\phi}_j^{(p-1)}(2j \text{ terms of form } X_tX_{t+p-j}) +$$

$$+ \sum_{j,i=1}^{p-1}(\tilde{\phi}_j^{(p-1)} - \hat{\phi}_j^{(p-1)})(\tilde{\phi}_i^{(p-1)} - \hat{\phi}_i^{(p-1)})\hat{R}_{p-j-i} + 2\sum_{j,i=1}^{p-1}(\tilde{\phi}_j^{(p-1)} - \hat{\phi}_j^{(p-1)})\hat{\phi}_i^{(p-1)}\hat{R}_{p-j-i} -$$

$$- \frac{1}{n}\sum_{j,i=1}^{p-1}\hat{\phi}_{p-j}^{(p-1)}\hat{\phi}_i^{(p-1)}(p - |j-i| \text{ terms of form } X_tX_{t+|j-i|}) -$$

$$- \frac{2}{n}\sum_{j,i=1}^{p-1}(\tilde{\phi}_{p-j}^{(p-1)} - \hat{\phi}_{p-j}^{(p-1)})\hat{\phi}_i^{(p-1)}(p - |j-i| \text{ terms of form } X_tX_{t+|j-i|}) -$$

$$- \frac{1}{n}\sum_{j,i=1}^{p-1}(\tilde{\phi}_{p-j}^{(p-1)} - \hat{\phi}_{p-j}^{(p-1)})(\tilde{\phi}_i^{(p-1)} - \hat{\phi}_i^{(p-1)})(p - |j-i| \text{ terms of form } X_tX_{t+|j-i|})\Big\} \times$$

$$\times\Big\{1 + \frac{1}{2n\hat{N}_p}[(2p \text{ terms of form } X_t^2) + 4n\sum_{j=1}^{p-1}(\tilde{\phi}_j^{(p-1)} - \hat{\phi}_j^{(p-1)})\hat{R}_j -$$

$$- 2\sum_{j=1}^{p-1}(\tilde{\phi}_j^{(p-1)} - \hat{\phi}_j^{(p-1)})(2p - 2j \text{ terms of form } X_tX_{t+j}) -$$

$$- 2\sum_{j=1}^{p-1}\hat{\phi}_j^{(p-1)}(2p - 2j \text{ terms of form } X_tX_{t+j}) -$$

$$- 2n\sum_{j,i=1}^{p-1}(\tilde{\phi}_j^{(p-1)} - \hat{\phi}_j^{(p-1)})(\tilde{\phi}_i^{(p-1)} - \hat{\phi}_i^{(p-1)})\hat{R}_{j-i} - 4n\sum_{j,i=1}^{p-1}(\tilde{\phi}_j^{(p-1)} - \hat{\phi}_j^{(p-1)})\hat{\phi}_i^{(p-1)}\hat{R}_{j-i} +$$

$$+ 2\sum_{j,i=1}^{p-1}\hat{\phi}_j^{(p-1)}\hat{\phi}_i^{(p-1)}(p - |j-i| \text{ terms of form } X_tX_{t+|j-i|}) +$$

$$+ 2\sum_{j,i=1}^{p-1}(\tilde{\phi}_j^{(p-1)} - \hat{\phi}_j^{(p-1)})(\tilde{\phi}_i^{(p-1)} - \hat{\phi}_i^{(p-1)})(p - |j-i| \text{ terms of form } X_tX_{t+|j-i|}) +$$

$$+ 4\sum_{j,i=1}^{p-1}(\tilde{\phi}_j^{(p-1)} - \hat{\phi}_j^{(p-1)})\hat{\phi}_i^{(p-1)}(p - |j-i| \text{ terms of form } X_tX_{t+|j-i|})] + \hat{h}_n^{(p)}\Big\}.$$

Here "j terms of form $X_tX_{t+k}$ " etc. means a sum of such terms, e.g. $X_1X_{k+1} + \ldots + X_jX_{k+j}$, where only data near the boundaries of the data set appear. After multiplying the terms in brackets and using the Yule-Walker equations one finds that

$$
\begin{aligned}
& n \cdot E(\tilde{\phi}_p^{(p)} - \hat{\phi}_p^{(p)}) \longrightarrow \\
& \quad \longrightarrow \frac{2}{N_p} \sum_{j=1}^{p-1} j \phi_j^{(p-1)} R_{p-j} - \frac{1}{N_p} \sum_{j,i=1}^{p-1} \phi_{p-j}^{(p-1)} \phi_i^{(p-1)} (p - |j-i|) R_{j-i} + \\
& \quad\quad + \frac{\phi_p^{(p)}}{N_p} (p \cdot R_0 - 2 \sum_{j=1}^{p-1} (p-j) \phi_j^{(p-1)} R_j + \sum_{j,i=1}^{p-1} \phi_j^{(p-1)} \phi_i^{(p-1)} (p - |j-i|) R_{j-i}),
\end{aligned}
$$

because the expectations of most of the resulting terms are $o(1/n)$.

The limit can be rewritten as

$$
\begin{aligned}
n \cdot E(\tilde{\phi}_p^{(p)} - \hat{\phi}_p^{(p)}) & \longrightarrow & \frac{1}{N_p} \sum_{j=1}^{p-1} (j \phi_j^{(p-1)} R_{p-j} - j \phi_p^{(p)} \phi_{p-j}^{(p-1)} R_{p-j}) + \\
& & + \frac{1}{N_p} \sum_{j=1}^{p-1} \phi_{p-j}^{(p-1)} d_{(p),j} + \frac{\phi_p^{(p)}}{N_p} p R_0 = \\
& = & \frac{1}{N_p} (\sum_{j=1}^{p-1} j \phi_j^{(p)} R_{p-j} + \sum_{j=1}^{p-1} \phi_{p-j}^{(p-1)} d_{(p),j} + \phi_p^{(p)} p R_0) =: g_p,
\end{aligned}
$$

and for the other components of $\tilde{\phi}^{(p)} - \hat{\phi}^{(p)}$

$$
\begin{aligned}
n \cdot E(\tilde{\phi}_j^{(p)} - \hat{\phi}_j^{(p)}) & = & n \cdot E(\tilde{\phi}_j^{(p-1)} - \hat{\phi}_j^{(p-1)}) + n \cdot E(\hat{\phi}_{p-j}^{(p-1)} (\hat{\phi}_p^{(p)} - \tilde{\phi}_p^{(p)}) + \tilde{\phi}_p^{(p)} (\hat{\phi}_{p-j}^{(p-1)} - \tilde{\phi}_{p-j}^{(p-1)})) \\
& & (j = 1 \ldots p-1)
\end{aligned}
$$

hold because of (1.3).

Introducing $\phi_{\downarrow}^{(p-1)} := (\phi_{p-1}^{(p-1)}, \ldots, \phi_1^{(p-1)})^\top$ etc. we find by induction that

$$
\begin{aligned}
& n \cdot E((\tilde{\phi}_1^{(p)}, \ldots, \tilde{\phi}_{p-1}^{(p)})^\top - (\hat{\phi}_1^{(p)}, \ldots, \hat{\phi}_{p-1}^{(p)})^\top) = \\
& \quad = \quad n \cdot E(\tilde{\phi}^{(p-1)} - \hat{\phi}^{(p-1)}) - n \cdot E(\tilde{\phi}_p^{(p)} - \hat{\phi}_p^{(p)}) \hat{\phi}_{\downarrow}^{(p-1)} - n \cdot E \tilde{\phi}_p^{(p)} (\tilde{\phi}_{\downarrow}^{(p-1)} - \hat{\phi}_{\downarrow}^{(p-1)}) \longrightarrow \\
& \quad\quad \longrightarrow -R_{(p-1)}^{-1} d_{(p-1)} - g_p \phi_{\downarrow}^{(p-1)} + \phi_p^{(p)} (R_{(p-1)}^{-1} d_{(p-1)})_{\downarrow}.
\end{aligned}
$$

As $(R_{(p-1)}^{-1} d_{(p-1)})_{\downarrow} = R_{(p-1)}^{-1} d_{(p-1)\downarrow}$ holds, we have to show:

$$
\begin{pmatrix} -R_{(p-1)}^{-1} d_{(p-1)} - g_p \phi_{\downarrow}^{(p-1)} + \phi_p^{(p)} R_{(p-1)}^{-1} d_{(p-1)\downarrow} \\ g_p \end{pmatrix} = - \begin{pmatrix} R_0 & \ldots & R_{p-1} \\ \vdots & \ddots & \vdots \\ R_{p-1} & \ldots & R_0 \end{pmatrix}^{-1} \cdot \begin{pmatrix} d_{(p),1} \\ \vdots \\ d_{(p),p} \end{pmatrix}.
$$

After multiplication with $R_{(p)}$, using $N_p = R_0 - \sum_{k=1}^{p-1} \phi_k^{(p-1)} R_k$, $\sum_{k=1}^{p-1} \phi_{p-k}^{(p-1)} R_{p-i-k} = R_i \quad (i = 1 \ldots p-1)$, and $\phi_\downarrow^{(p-1)} = R_{(p-1)}^{-1} \cdot (R_{p-1}, \ldots, R_1)^\top$, we see that

$$\begin{pmatrix} R_0 & \ldots & R_{p-1} \\ \vdots & \ddots & \vdots \\ R_{p-1} & \ldots & R_0 \end{pmatrix} \cdot \begin{pmatrix} -R_{(p-1)}^{-1} d_{(p-1)} - g_p \phi_\downarrow^{(p-1)} + \phi_p^{(p)} R_{(p-1)}^{-1} d_{(p-1)\downarrow} \\ g_p \end{pmatrix} =$$

$$= \begin{pmatrix} -d_{(p-1)} - g_p R_{(p-1)} \phi_\downarrow^{(p-1)} + \phi_p^{(p)} d_{(p-1)\downarrow} + \begin{pmatrix} R_{p-1} \\ \vdots \\ R_1 \end{pmatrix} g_p \\ (R_{p-1}, \ldots, R_1) \cdot (-R_{(p-1)}^{-1} d_{(p-1)} - g_p \phi_\downarrow^{(p-1)} + \phi_p^{(p)} R_{(p-1)}^{-1} d_{(p-1)\downarrow}) + R_0 g_p \end{pmatrix} =$$

$$= \begin{pmatrix} -\sum_{k=0}^{p-1} |1-k| R_{1-k} \Phi_k^{(p-1)} + \phi_p^{(p)} \sum_{k=0}^{p-1} |p-1-k| R_{p-1-k} \Phi_k^{(p-1)} - g_p (\sum_{k=1}^{p-1} \phi_{p-k}^{(p-1)} R_{k-1} - R_{p-1}) \\ \vdots \\ -\sum_{k=0}^{p-1} |p-1-k| R_{p-1-k} \Phi_k^{(p-1)} + \phi_p^{(p)} \sum_{k=0}^{p-1} |1-k| R_{1-k} \Phi_k^{(p-1)} - g_p (\sum_{k=1}^{p-1} \phi_{p-k}^{(p-1)} R_{p-k-1} - R_1) \\ -\phi_\downarrow^{(p-1)\top} d_{(p-1)} + \phi_p^{(p)} \phi_\downarrow^{(p-1)\top} d_{(p-1)\downarrow} - g_p (\sum_{k=1}^{p-1} R_k \phi_k^{(p-1)} - R_0) \end{pmatrix} =$$

$$= \begin{pmatrix} -\sum_{k=0}^{p-1} |1-k| R_{1-k} \Phi_k^{(p-1)} + \phi_p^{(p)} \sum_{k=0}^{p-1} |p-1-k| R_{p-1-k} \Phi_k^{(p-1)} \\ \vdots \\ -\sum_{k=0}^{p-1} |p-1-k| R_{p-1-k} \Phi_k^{(p-1)} + \phi_p^{(p)} \sum_{k=0}^{p-1} |1-k| R_{1-k} \Phi_k^{(p-1)} \\ -\phi_\downarrow^{(p-1)\top} d_{(p-1)} + \phi_p^{(p)} \phi_\downarrow^{(p-1)\top} d_{(p-1)\downarrow} + g_p N_p \end{pmatrix} \stackrel{!}{=} -d_{(p)};$$

the last identity still has to be proved.

It holds for all rows:

1. rows $i = 1, \ldots, p-1$:

$$\begin{aligned} & -(\sum_{k=0}^{p-1} |i-k| R_{i-k} \Phi_k^{(p-1)} - \phi_p^{(p)} \sum_{k=0}^{p-1} |p-i-k| R_{p-i-k} \Phi_k^{(p-1)}) = \\ = & \ -(\sum_{k=1}^{p-1} |i-k| R_{i-k} (\Phi_k^{(p-1)} - \phi_p^{(p)} \Phi_{p-k}^{(p-1)}) - i R_i + \phi_p^{(p)} (p-i) R_{p-i}) = \\ = & \ -\sum_{k=0}^{p} |i-k| R_{i-k} \Phi_k^{(p)} = -d_{(p),i} \ , \end{aligned}$$

2. last row:

$$\begin{aligned}
&-\phi_{\downarrow}^{(p-1)\top} d_{(p-1)} + \phi_p^{(p)} \phi_{\downarrow}^{(p-1)\top} d_{(p-1)\downarrow} + g_p N_p = \\
&= \quad -\phi_{\downarrow}^{(p-1)\top} (d_{(p),1}, \ldots, d_{(p),p-1})^\top + g_p N_p = \\
&= \quad -\sum_{k=1}^{p-1} \phi_{p-k}^{(p-1)} d_{(p),k} + \sum_{k=1}^{p-1} k\phi_k^{(p)} R_{p-k} + \sum_{k=1}^{p-1} \phi_{p-k}^{(p-1)} d_{(p),k} + \phi_p^{(p)} p R_0 = \\
&= \quad \sum_{k=1}^{p} k\phi_k^{(p)} R_{p-k} = \sum_{k=1}^{p} k\phi_k^{(p)} R_{p-k} - p(\sum_{k=1}^{p} R_{p-k}\phi_k^{(p)} - R_p) = \\
&= \quad -\sum_{k=0}^{p} (p-k) R_{p-k} \Phi_k^{(p)} = -d_{(p),p} \; ,
\end{aligned}$$

so (3.1) and the statement of the theorem follow. □

The same theorem can be shown in a very similar way for the Burg estimator used by Morf et al. (1978) in the case $d = 1$.

As a consequence we get the bias of the univariate Burg estimator for a true model from the result of Shaman and Stine (1988):

**Theorem 4** *Let $\{X_t\}, t \in \mathsf{Z}$, be a univariate stable AR(p)-process, $X_t = \sum_{j=1}^{p} \phi_j^{(p)} X_{t-j} + \epsilon_t$, where $\epsilon_t$ are independent and identically distributed innovations, $E\epsilon_t = 0$, $0 < E\epsilon_t^2 < \infty$, and let the assumptions of Theorem 3 hold. Then*

$$n \cdot E(\tilde{\phi}^{(p)} - \phi^{(p)}) \longrightarrow \begin{cases} v + \sum_{j=0}^{p/2-1} (\Phi_j^{(p)} - \Phi_{p-j}^{(p)}) \cdot a_j, & \text{if } p \text{ is even,} \\ \\ v + \sum_{j=0}^{1/2\cdot(p-1)} (\Phi_{j-1}^{(p)} - \Phi_{p-j}^{(p)}) \cdot b_j, & \text{if } p \text{ is odd,} \end{cases}$$

*if the mean of $X_t$ is known. If $EX_t$ has to be estimated by $\frac{1}{n}\sum_{t=1}^{n} X_t$, the additional bias term*

$$c = (c_1, \ldots, c_p)^\top, \qquad c_j := \sum_{i=0}^{j-1} (\Phi_i^{(p)} - \Phi_{p-i}^{(p)}),$$

*appears.*

Here $a_j$, $b_j$ and $v$ are defined by

$$a_j := (a_{j,1}, \ldots, , a_{j,p})^\top, \qquad a_{j,i} := \begin{cases} 1 & \text{if } i = j+2, j+4, \ldots, p-j, \\ 0 & \text{else,} \end{cases}$$

$$b_j := (b_{j,1}, \ldots, b_{j,p})^\top, \qquad b_{j,i} := \begin{cases} 1 & \text{if } i = j+1, j+3, \ldots, p-j, \\ 0 & \text{else,} \end{cases}$$

$$v := -(\phi_1^{(p)}, 2 \cdot \phi_2^{(p)}, \ldots, p \cdot \phi_p^{(p)})^\top.$$

We still have to investigate the asymptotic bias of $\tilde{S}_p = \prod_{j=1}^p (1 - \tilde{\phi}_j^{(j)2}) \hat{R}_0$:

**Theorem 5** *Under the assumptions of Theorem 4, the bias of $\tilde{S}_p$ is given by*

$$nE(\tilde{S}_p - S_p) \quad \longrightarrow \quad -pS_p$$

*for known $EX_t$, and*

$$nE(\tilde{S}_p - S_p) \quad \longrightarrow \quad -(p+1)S_p$$

*for estimated mean.*

Proof: Shaman (1983) showed that the YW estimator $\hat{S}_p = \hat{N}_{p+1}$ of $S_p$ has the asymptotic bias

$$nE(\hat{S}_p - S_p) \longrightarrow -pS_p + 2\sum_{j=1}^p j\phi_j^{(p)} R_j - \sum_{j,i=1}^p |j-i| \phi_j^{(p)} \phi_i^{(p)} R_{j-i}, \qquad (3.2)$$

if $EX_t$ is known. From the proof of Theorem 3 follows

$$\begin{aligned} nE(\tilde{N}_{p+1} - \hat{N}_{p+1}) \longrightarrow -(p+1)R_0 + 2\sum_{j=1}^p \phi_j^{(p)} (p+1-j) R_j - \\ - \sum_{j,i=1}^p \phi_j^{(p)} \phi_i^{(p)} (p+1-|j-i|) R_{j-i} = \\ = \quad -(p+1)S_p - 2\sum_{j=1}^p j\phi_j^{(p)} R_j + \sum_{j,i=1}^p |j-i| \phi_j^{(p)} \phi_i^{(p)} R_{j-i}, \end{aligned}$$

so the bias of $\tilde{N}_{p+1}$ is

$$nE(\tilde{N}_{p+1} - S_p) \longrightarrow -(p+1)S_p - pS_p = -(2p+1)S_p.$$

It is easy to see that

$$\tilde{N}_{p+1} - \tilde{S}_p = -\sum_{j=0}^p \left(\frac{1}{2n}(\tilde{e}_{j+1}^{(j)2} + \tilde{\eta}_{n-j}^{(j)2}) \prod_{k=j+1}^p (1 - \tilde{\phi}_k^{(k)2})\right),$$

and consequently

$$nE(\tilde{N}_{p+1} - \tilde{S}_p) \quad \longrightarrow \quad -\sum_{j=0}^p S_j \prod_{k=j+1}^p (1 - \phi_k^{(k)2}) = -\sum_{j=0}^p S_p = -(p+1)S_p$$

holds, which proves the assertion.

For unknown mean, the theorem can be shown the same way using

$$nE(\hat{S}_p - S_p) \longrightarrow -(p+1)S_p + 2\sum_{j=1}^{p} j\phi_j^{(p)} R_j - \sum_{j,i=1}^{p} |j-i|\phi_j^{(p)}\phi_i^{(p)} R_{j-i}$$

instead of (3.2) (cf. Zhang, Th. 4.2, (2.10) and Th. 3.1). □

The asymptotic bias of the YW estimator can become very large if the process has roots near the complex unit circle, but it can be reduced by (variable) tapering (Zhang,1992). For the Burg estimator no improvement can be achieved in this way, because the untapered Burg estimator has the same asymptotic bias as the tapered YW estimator, and $nE(\tilde{\phi}_p^{(p)} - \hat{\phi}_p^{(p)}) \to 0$ if both estimators are tapered.

## 4 Conclusion

It is known from simulations that the Burg estimator has a smaller bias than the YW estimator if a root of the process is near the unit circle, but a proof has been missing. We have shown that the univariate Burg estimator has the same asymptotic bias as the least squares estimator, which is usually smaller than the bias of the YW estimator. This makes the reduction of the bias possible. Also the Burg estimator is recursively computable and stable, in contrast to the least squares estimator, so the Burg estimator has the major advantages of both of these well known estimators. We have also shown that the multivariate Burg estimators have the same asymptotic distribution as the multivariate YW estimator, and some simulations (Strand,1977, Morf et al.,1978) seem to indicate that their bias is smaller than that of the YW estimator, but no analytic results on this are known yet.

**Acknowledgement:** I would like to thank Prof. Dr. R. Dahlhaus, who supervised my Diplomarbeit, the major results of which are presented here, and Dr. D. Janas for their support.

## References

**Burg,J. P. (1968)** A New Analysis Technique for Time Series Data. *NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics.*

**Hannan,E. J. (1970)** Multiple Time Series. New York:Wiley.

**Jones,R. H. (1978)** Multivariate Autoregression Estimation Using Residuals. In: *Applied Time Series Analysis.* Proceedings of the First Applied Time Series Symposium, Tulsa, Okla., 1976 (ed. D.F.Findley). New York: Academic Press, pp. 139-162.

**Kay, S. and Makhoul,J. (1983)** Statistics of the estimated Reflection Coefficients of an Autoregressive Process. *IEEE Trans. Acoust., Speech, Signal Processing 31*, 1447-1455.

**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.

**Lysne,D. and Tjøstheim,D. (1987)** Loss of spectral peaks in autoregressive spectral estimation. *Biometrika 74*, 200-206.

**Morf,M. , Vieira,A. , Lee,D. T. L. and Kailath,T. (1978)** Recursive Multichannel Maximum Entropy Spectral Estimation. *IEEE Trans. Geosci. Electron. 16*, 85-94.

**Nicholls, D. F. and Pope,A. L. (1988)** Bias in the Estimation of Multivariate Autoregressions. *Austral. J. Statist. 30A*, Spec. Issue, 296-309.

**Shaman,P. (1983)** Properties of Estimates of the Mean Square Error of Prediction in Autoregressive Models. In: *Studies in Econometrics, Time Series, and Multivariate Statistics* (eds. S. Karlin, T. Amemiya, L. A. Goodman), New York: Academic Press, pp. 331-343.

— **and Stine,R. A. (1988)** The Bias of Autoregressive Coefficient Estimators. *J.Amer.Statist.Assoc. 83*, 842-848.

**Strand, O. N. (1977)** Multichannel Complex Maximum Entropy (Autoregressive) Spectral Analysis. *IEEE Trans. Automat. Control 22*, 634-640.

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

**Zhang,H.-C. (1992)** Reduction of the Asymptotic Bias of Autoregressive and Spectral Estimators by Tapering. *J. Time Ser. Anal. 13*, 451-469.

Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg
Germany