

Finite-Sample Bias Propagation in Autoregressive Estimation With the Yule–Walker Method

Piet M. T. Broersen

Abstract—The Yule–Walker (YW) method for autoregressive (AR) estimation uses lagged-product (LP) autocorrelation estimates to compute an AR parametric spectral model. The LP estimates only have a small triangular bias in the estimated autocorrelation function and are asymptotically unbiased. However, using them in finite samples with the YW method for AR estimation can give a strong distortion in the weak parts of the power spectral density. The distortion is shown to be influential in an example without strong spectral peaks. The true biased AR model, which is computed by applying the triangular bias to the true autocorrelation function, has an infinite order. A new objective measure is introduced to determine the smallest sample size for which the unbiased asymptotic theory can be considered as a fair approximation.

Index Terms—Autocorrelation bias, autoregressive (AR) model, parameter estimation, spectral estimation, time series analysis, triangular bias.

I. INTRODUCTION

TIME series models are a modern and useful perspective for spectral analysis. An excellent tutorial survey of the many time series algorithms that have been developed has been given in [1]. Yule–Walker (YW) equations describe the relation between the p autoregressive (AR) parameters of an AR(p) process and the first p lags of the autocorrelation function. The YW method of AR estimation solves the parameters from the YW equations by substituting them with the lagged-product (LP) estimates for the autocorrelations [1]. The recursive Levinson–Durbin algorithm is an efficient numerical solution of the YW equations [2]. It computes the reflection coefficients for increasing AR orders, which are defined as the negatives of the partial correlation coefficients [1], [3], [4]. The parameters of AR models of arbitrary orders q can be computed from the first q reflection coefficients, which, in turn, are completely determined by the first q lags of the autocorrelation function.

The finite-sample bias of the AR models depends on the estimation method [5]. The bias of the YW method of AR estimation is given by the general terms for all the AR methods, with an additional contribution caused by the triangular bias of the LP autocorrelation estimate [6]. Explicit expressions for the bias of the first two estimated AR parameters have been given. The bias of the second parameter is a rational function of

the parameters, which can become large if the poles of the AR polynomial are close to 1 in absolute value.

The variance of the AR estimates has also been studied. The asymptotical theory gives identical results for all the AR estimation methods. Well-known AR estimation methods include the YW method, the method of Burg [7], several least-squares solutions, and the maximum-likelihood method [1]. Asymptotic expressions have been derived for the variance of estimated reflection coefficients for orders of up to the true AR order [8]. Explicit expressions were given for AR processes until order 3, and an efficient recursive procedure can compute the variance for higher order AR processes. Empirical finite-sample approximations for the variance of the estimated parameters depend on the method of AR estimation [4]. The empirical variances can be used in a finite-sample order selection criterion that is adapted to the estimation method. It turned out [4] that the finite-sample selection criterion for the YW estimation method was almost identical with the asymptotic Akaike information criterion (AIC) [9]. Therefore, the YW method did not show the peculiar order selection behavior of many other methods. This may be a reason for the popularity of the YW method if asymptotical criteria are used for the selection of the AR model order.

A detailed theoretical analysis derived the expectations of the squared reflection coefficients of YW estimates in small samples of a white noise signal [10]. They are applicable to the variance of the reflection coefficients of AR processes for all orders higher than the true AR order, which is important for a finite-sample order selection criterion [11]. Both the YW and Burg methods recursively estimate the AR reflection coefficients for increasing AR orders; most other methods estimate the AR parameters. AR models are stable or stationary if all poles are within the unit circle. It has been shown that this is guaranteed if all the reflection coefficients are between -1 and $+1$. This is explicitly proven for the YW method [12], which has a triangular bias, and for the Burg method without the additional bias [3]. The AR models of these two methods are guaranteed to be always stable.

The triangular bias turned out to be important if a reflection coefficient of any order lower than the true process order p comes close to $+1$ or -1 . If a true reflection coefficient of order m that is less than the true order is given by $1 - m/N$, where N is the sample size, the bias in the estimated reflection coefficient of order $m + 1$ will be 50% [13]. However, the YW method is still often used in practice. The reassuring property of being asymptotically unbiased, as well as the close agreement between the YW estimation method and the theoretical description of the autocorrelation function of AR models

Manuscript received June 19, 2008; revised November 3, 2008. First published January 20, 2009; current version published April 7, 2009. The Associate Editor coordinating the review process for this paper was Dr. J. Blair.

The author is with the Department of Multi-Scale Physics, Delft University of Technology, 2628 CJ Delft, The Netherlands (e-mail: p.m.t.broersen@tudelft.nl).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIM.2008.2009400

with the YW equations, might be a reason. However, it is a historical misconception that the usual LP estimates for the autocorrelation function are based on the accurate statistical theory [14].

The value of the last parameter of order p never causes this extra YW bias for an AR(p) process [15]. The importance of the bias can be related to the condition number of the Toeplitz autocorrelation matrix, which contains only the autocorrelations until order $p - 1$. However, the condition number is independent of the number of observations N . Furthermore, the triangular bias of the LP estimates diminishes with N and asymptotically vanishes. Therefore, no critical value for the condition number can be given. Processes with a high condition number may yield a large bias in finite samples, but, asymptotically, the bias will disappear.

This paper is an extension of a conference proceeding paper [16], with new examples. The expectation of the bias is computed by applying the triangular autocorrelation bias to the true autocorrelation function and solving the YW equations with such bias. The triangular bias changes the order of the true biased process to infinity for each AR(p) process. An objective measure can be defined for the significance of the bias for a given number of observations [16]. This measure determines the minimum sample size for which the influence of the bias in the YW method is not greater than the inevitable inaccuracy due to the estimation variance of the AR parameters.

II. AR ESTIMATES WITH THE YW METHOD

A discrete-time AR(p) process is a time series with observations x_n that can be written as [17]

$$x_n + a_1x_{n-1} + \dots + a_px_{n-p} = \varepsilon_n \quad (1)$$

where ε_n is a purely random process with zero mean and variance σ_ε^2 . In theory, any stationary stochastic process with a continuous spectral density can be written as an exact AR(∞) process. In practice, finite-order models are often sufficient, because high-order parameters tend to be small. The poles of the AR(p) process are the roots of the AR polynomial $A(z)$ given by

$$A(z) = 1 + a_1z^{-1} + \dots + a_pz^{-p}. \quad (2)$$

Processes are called stable or stationary if the AR poles are strictly within the unit circle. The parametric power spectrum $h(\omega)$ of the AR(p) model for $-\pi < \omega \leq \pi$ is given by

$$h(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{|A(e^{j\omega})|^2} = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{|1 + \sum_{i=1}^p a_i e^{-j\omega i}|^2}. \quad (3)$$

Autocovariance function $r(q)$ and normalized autocorrelation $\rho(q)$ at lag q are defined for a signal with a mean value of zero by

$$\rho(q) = \frac{r(q)}{r(0)} = \frac{E(x_n x_{n+q})}{E(x_n^2)}. \quad (4)$$

All the lags of the infinitely long true autocorrelation function of an AR(p) process are determined by the p true AR

parameters with the YW equations, as given by [1]

$$\begin{aligned} \rho(q) + a_1\rho(q-1) + \dots + a_p\rho(q-p) &= 0, & q \geq 1 \\ \rho(-q) &= \rho(q). \end{aligned} \quad (5)$$

Different efficient numerical methods have been developed to solve the YW equations. Recursive algorithms compute the reflection coefficients for increasing orders. Afterward, the parameters for arbitrary AR orders q can be found from the first q reflection coefficients [3]. Equation (5) can also be used for indexes q greater than p to extrapolate the autocorrelation function. The same relations (5) can be used to compute the parameters from correlations and vice versa. They relate true or estimated AR parameters with true or estimated autocorrelations, respectively.

To obtain a positive semidefinite LP estimator for the autocorrelation function of N observations x_n , the estimator for lag q should be biased and is given by

$$\hat{\rho}_{LP}(q) = \hat{r}_{LP}(q)/\hat{r}_{LP}(0) = \frac{1}{N} \sum_{n=1}^{N-q} x_n x_{n+q} / \frac{1}{N} \sum_{n=1}^N x_n^2. \quad (6)$$

Likewise, the positive semidefinite LP estimator for the autocovariance at lag q uses the divisor N for $N - q$ contributions. This gives the triangular bias $1 - q/N$

$$\begin{aligned} E[\hat{r}_{LP}(q)] &= E\left[\frac{1}{N} \sum_{n=1}^{N-q} x_n x_{n+q}\right] \\ &= \frac{N-q}{N} E[x_n x_{n+q}] = r(q) \left(1 - \frac{|q|}{N}\right). \end{aligned} \quad (7)$$

The bias of the normalized autocorrelation in (6) has some additional terms that arise from the expectation of quotients [14] and the subtraction of the mean value [17].

The theoretical expectation of the influence of the triangular bias on the biased values of the AR(p) model parameters $a_{i,b}$ can be evaluated by substituting the biased expectations of the autocovariances in the YW relations (5), i.e.,

$$\begin{aligned} r(q)[1 - q/N] + a_{1,b}r(q-1)[1 - (q-1)/N] + \dots \\ + a_{p,b}r(q-p)[1 - (q-p)/N] = 0, \quad q = 1, \dots, p. \end{aligned} \quad (8)$$

For negative indexes, symmetry relation $r(q) = r(-q)$ has been used. The YW method of AR estimation has triangular bias influence $a_{i,b}$, because it substitutes LP estimates (6) in (5) to compute the AR parameters \hat{a}_i of an AR(p) model with p equations, i.e.,

$$\begin{aligned} \hat{\rho}_{LP}(q) + \hat{a}_1\hat{\rho}_{LP}(q-1) + \dots \\ + \hat{a}_p\hat{\rho}_{LP}(q-p) = 0, \quad 1 \leq q \leq p. \end{aligned} \quad (9)$$

Extended YW equations for $q > p$ describe the parametric estimator for the continuation of the autocorrelation function.

The usual accuracy measure for time series models is the squared error of prediction PE. Prediction \hat{x}_n is obtained by

using the estimated parameters to predict a fresh realization of x_n , i.e.,

$$\hat{x}_n = -\hat{a}_1 x_{n-1} - \dots - \hat{a}_p x_{n-p}. \quad (10)$$

The PE is defined as the long-term average of $(\hat{x}_n - x_n)^2$. Fresh or new data x_n have been generated with the same process equation (1), but they have not been used to estimate the parameters. Both the true and estimated parameter values are known through Monte Carlo simulation experiments. For these particular situations where the true parameters are known, the PE can be computed without generating new data by using an efficient algorithm based on

$$\text{PE} = \frac{\sigma_\varepsilon^2}{2\pi} \int_{-\pi}^{\pi} \frac{h(\omega)}{\hat{h}(\omega)} d\omega = \frac{\sigma_\varepsilon^2}{2\pi} \int_{-\pi}^{\pi} \left| \frac{\hat{A}(e^{j\omega})}{A(e^{j\omega})} \right|^2 d\omega \quad (11)$$

where $\hat{A}(z)$ is an estimated AR polynomial.

The model error (ME) has been defined for AR models as

$$\text{ME} = \text{ME}(\hat{A}, A) = N \left(\frac{\text{PE} - \sigma_\varepsilon^2}{\sigma_\varepsilon^2} \right) \quad (12)$$

by scaling the PE with σ_ε^2 and multiplying with the sample size. The argument of the ME is left out if no confusion is possible. The expectation of the ME for unbiased efficiently estimated AR models of orders K , which are greater than or equal to true order p , is equal to K . The variance of each estimated parameter has a minimal contribution of 1 to the ME expectation. The Cramér–Rao bound for the ME of the estimates of an AR(p) process is given by p , independent of the sample size.

The condition number κ of Toeplitz correlation matrix R with elements $r(0) - r(p - 1)$ has been proposed to classify the sensitivity of an AR process to the YW bias [15]. It is defined as

$$\kappa = \llbracket R \rrbracket \bullet \llbracket R^{-1} \rrbracket \quad (13)$$

where $\llbracket \cdot \rrbracket$ is some matrix norm, such as the largest singular value. The norm is computed for the expectation of the autocorrelation matrix and is independent of sample size N . Therefore, it cannot be a useful accuracy measure for asymptotically unbiased estimates where the bias depends on N .

An objective measure for the quality of AR models estimated with the YW method for a given sample size N is found by using the expectations (7) of the autocovariances with triangular bias to compute the parameters $a_{i,b}$ of the biased AR(p) model $\hat{A}(z)$ with (8) and, finally, the $\text{ME}(\hat{A}, A)$ with (12). The asymptotical expectation of $\text{ME}(\hat{A}, A)$ is zero, because the bias asymptotically disappears. A critical sample size for which the bias is no longer considered significant is defined to be the N for which $\text{ME}(\hat{A}, A) = p$. This is the sample size where the bias error of the AR(p) model becomes of the same magnitude as the variance error, which is caused by the efficient estimation of the p parameters of the AR(p) model. This limit for $\text{ME}(\hat{A}, A)$ seems to be a little arbitrary. However, equal bias and variance contributions are known from the derivation of order selection criterion AIC [9]. Such is an estimator for the

Kullback–Leibler index, which is a distance measure between probability density functions. Therefore, this choice has a firm statistical background.

III. FINITE-SAMPLE AR THEORY

A number of methods for AR estimation have been described [1]. The method of YW was the first of such methods and is still popular, whereas the method of Burg [7] has good properties, such as guaranteed stability, and is often advised for AR estimation in practice. The two methods use recursive estimates of the reflection coefficients. The asymptotical theory for residual variance s_q^2 and prediction error $\text{PE}(q)$ is the same for all estimation methods. The residual variance is most easily expressed as a function of the reflection coefficients. For an AR model of order 0, the expectation of s_0^2 is given by $r(0)$, which is the variance of process x_n . For any arbitrary order q , the residual variance s_q^2 is given by

$$s_q^2 = s_0^2 \prod_{i=1}^q (1 - k_i^2) \quad (14)$$

where the true or estimated reflection coefficient of order i can be substituted for k_i . If true reflection coefficients are used in (14), their value is zero for orders greater than true order p , and the residual variance gives the final value $s_p^2 = \sigma_\varepsilon^2$. For estimated values of k_i , the residual variance will keep decreasing above p because of the estimation variance of k_i .

An explanation for the strong triangular bias of the YW AR method has been given in [13]. Residual variance s_q^2 appears in the denominator of the Levinson–Durbin recursion to compute the reflection coefficient for order $q + 1$ [1], [3]. Suppose that a true k_q has an unbiased value of $1 - 1/N$. A small bias of magnitude $1/N$ would give it the biased value $1 - 2/N$. Substitution of the unbiased and biased values for k_q in (14) gives the approximations $2/N$ and $4/N$ for $(1 - k_q^2)$, respectively. A small bias of magnitude $1/N$ in a reflection coefficient k_q makes a difference of a factor of 2 in residual variance s_q^2 . The error in s_q^2 propagates to all higher AR orders, because the product of (14) is in the denominator of the Levinson–Durbin recursion for the computation of successive reflection coefficients. The triangular bias propagation is always influential if one of the true reflection coefficients of an order lower than true order p is close to unity, with a distance of less than $1/N$. However, for any value $k_q = 1 - \delta$, with a very small δ , the finite-sample influence of the bias will asymptotically disappear, if sample size N is much greater than $1/\delta$.

If order K is at least equal to true order p , the asymptotical AR expectations for unbiased models are determined by

$$\text{E} [s_K^2] = \sigma_\varepsilon^2 \left(1 - \frac{K}{N} \right), \quad K \geq p \quad (15)$$

$$\text{E} [\text{PE}(K)] = \sigma_\varepsilon^2 \left(1 + \frac{K}{N} \right), \quad K \geq p. \quad (16)$$

This result can be applied in practice with reasonable accuracy if order K is less than $0.1N$. For still higher AR orders, the finite-sample theory gives an improved accuracy [4]. Such

theory is based on approximations of the variance of reflection coefficients estimated from finite samples of a white noise process. The asymptotical variance $1/N$ is replaced by empirical finite-sample variance coefficients. The approximations are given by [4]

$$v_{i,YW} = \frac{N-i}{N(N+2)}$$

$$v_{i,Burg} = \frac{1}{N+1-i} \quad (17)$$

for the YW and Burg estimates, respectively. YW variances are smaller than the asymptotical value $1/N$ because of the triangular bias in the autocorrelation function. In contrast, Burg variances are greater, because each new Burg reflection coefficient is estimated from a shorter filtered signal.

Above the true order, all true reflection coefficients are zero. Their squared expectation is given by their variance. The finite-sample expressions for the expectations of s_K^2 and $PE(K)$ for the true order and higher can be derived as

$$E_{FS} [s_K^2] = \sigma_\varepsilon^2 \prod_{i=1}^K (1 - v_{i,\cdot}), \quad K \geq p \quad (18)$$

$$E_{FS}[PE(K)] = \sigma_\varepsilon^2 \prod_{i=1}^K (1 + v_{i,\cdot}), \quad K \geq p. \quad (19)$$

By substituting YW or Burg variance coefficients into $v_{i,\cdot}$, the finite-sample results become dependent on the estimation method.

In practice, residual variance s_q^2 is known for YW or Burg estimates for all orders q , because it has been minimized to compute the AR parameters. However, $PE(q)$ is not known. A finite-sample approximation uses the known s_q^2 to estimate the prediction accuracy in practice with the finite-sample criterion $FSC(q)$ [4] given by

$$FSC(q) = s_q^2 \frac{\prod_{i=1}^q (1 + v_{i,\cdot})}{\prod_{i=1}^q (1 - v_{i,\cdot})}. \quad (20)$$

This approximation can be computed for N given observations x_n . It has a strong relation with the final prediction error [9] and may be used for all model orders q , independent of the true AR order. The values of s_q^2 and $FSC(q)$ are mainly determined by the true values of the reflection coefficients for model orders less than the true order and by (18) and (19) for higher orders. The expectation of $FSC(q)$ for higher model orders is given by $E_{FS}[PE(q)]$. For finite samples, it markedly depends on the estimation method if q is greater than about $N/10$.

IV. YW METHOD APPLIED TO AR(3) EXAMPLES

To demonstrate the bias propagation, an AR(3) example with three equal reflection coefficients has been studied. The triangular bias of (7) is applied to the true autocorrelation function for $N = 50$, and the biased values for the reflection coefficients are found by substituting those biased correlations in (8). As

TABLE I
TRUE REFLECTION COEFFICIENT k AND EXPECTATIONS FOR THE BIASED REFLECTION COEFFICIENTS FOR AN AR(3) PROCESS GENERATED WITH THREE EQUAL TRUE REFLECTION COEFFICIENTS k . THE ME OF THE BIASED AR(3) MODEL, AS COMPUTED WITH (8), AND CONDITION NUMBER κ ARE GIVEN FOR $N = 50$ AND FOR DIFFERENT VALUES OF THE TRUE k

true	biased expectations			measures	
k	k_1	k_2	k_3	$ME(\hat{A}, A)$	κ
0.2	0.196	0.192	0.187	0.02	1.9
-0.4	-0.392	-0.381	-0.363	0.11	3.7
0.6	0.588	0.564	0.501	1.27	13.2
-0.7	-0.686	-0.647	-0.514	4.36	16.3
0.8	0.784	0.718	0.430	25.8	74.8
-0.9	-0.882	-0.738	-0.063	224.9	150.2
0.9	0.882	0.741	0.072	226.7	364.5
0.98	0.960	0.485	-0.276	6771.7	10775.1

it is hardly illustrative to derive the analytical expressions for this example, numerical values are presented. This example with equal true reflection coefficients has the advantage that the bias for different orders can easily be appreciated. Table I gives true value k , the biased reflection coefficients, and the two measures $ME(\hat{A}, A)$ and κ . The first reflection coefficient has a bias of magnitude $1/N$ for all the true values of reflection coefficient k . However, higher orders q have a bias that is only close to q/N for $k = 0.2$; the strongest bias is found for the highest orders. The rows with $k = 0.9$ and -0.9 show that only the value κ strongly depends on the sign of the reflection coefficients. Finally, true value $k = 0.98 = 1 - 1/N$ already gives a theoretical bias of 50% for the second reflection coefficient k_2 , and the sign of the third reflection coefficient is inverted.

It is remarkable how the bias propagates to higher orders for values of k that are much smaller than $1 - p/N$, which are rather far from the unit circle. Moderate values of condition number κ already lead to strong biases, and low condition numbers do not always guarantee a small bias. The example with $k = 0.8$ demonstrates that the moderate condition number 74.8 can give a significant ME value of 25.8. The demand that $ME(\hat{A}, A)$ has a value of less than 3 is only met for $k < 0.666$ in this AR(3) example, with $\kappa = 21.1$ being the condition number. Sample size N should be greater than 33 500 to obtain the reduced bias influence $ME(\hat{A}, A) < 3$ for $k = 0.9$; the theoretical value for the biased k_3 would then be 0.896. The variance influence will become really small only for a very high N . However, the bias influence is still asymptotically vanishing.

This is also presented in Table II. The values of k_3 (biased) and $ME(\hat{A}, A)$ of the AR(3) model are given for increasing N . Due to the multiplication in the ME with the sample size in (12), the influence of the bias increases for N less than 200. In the example with three equal true reflection coefficients, it proportionally decreases to $1/N$ for N greater than about 25 000.

The model accuracy can be determined for the estimated models of increasing orders. Software for estimating many AR models and automatically selecting the best model order is available [18]. Fig. 1 shows the results for an AR(3) example with equal true reflection coefficients of 0.6, where the bias is not influential. Both the YW and Burg AR estimates are shown.

TABLE II
EXPECTATION OF THE BIASED REFLECTION COEFFICIENT OF ORDER 3
AND THE CORRESPONDING MODEL ERROR $ME(\hat{A}, A)$ OF THE BIAS IN (8)
FOR AN AR(3) PROCESS WITH THREE EQUAL TRUE REFLECTION
COEFFICIENTS OF 0.9 AS A FUNCTION OF SAMPLE SIZE N

N	$k_3(\text{biased})$	$ME(\hat{A}, A)$
25	-0.050	177.8
10^2	0.268	249.2
10^3	0.781	83.2
10^4	0.887	9.92
10^5	0.899	1.01

TABLE III
TRUE PARAMETERS a_i OF AN AR(7) PROCESS, THE REFLECTION
COEFFICIENTS k_i , THE BIASED EXPECTATIONS, AND THE ESTIMATED
AR(7) REFLECTION COEFFICIENTS OF THE BURG METHOD
IN A SINGLE SIMULATION RUN WITH $N = 100$

order i	a_i	$k_i(\text{true})$	$k_i(\text{biased})$	$k_i(\text{estimated})$
1	-1.90	-0.37	-0.366	-0.303
2	3.46	0.85	0.830	0.847
3	-3.68	-0.33	-0.267	-0.365
4	3.59	0.77	0.607	0.804
5	-2.26	-0.32	0.057	-0.359
6	1.21	0.70	0.145	0.672
7	-0.30	-0.30	0.267	-0.409
8	0	0	-0.081	
9	0	0	-0.180	
10	0	0	0.034	

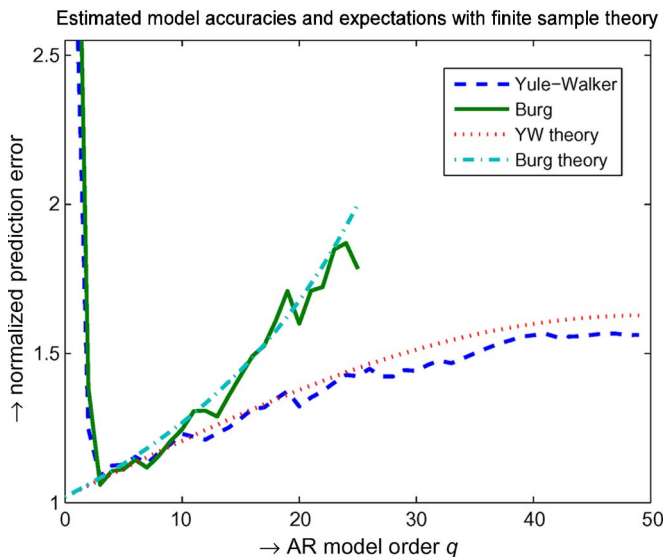


Fig. 1. Estimated accuracy $FSC(q)$ and theoretical accuracy $E_{FS}[PE(q)]$ for AR models estimated from a single realization of $N = 50$ observations of an AR(3) process with three equal true reflection coefficients of 0.6.

The estimated accuracies $FSC(q)$ are computed with (20). Furthermore, the theoretical finite-sample expectations for unbiased models computed with (19) are presented for the two methods. Both methods follow their expectations rather closely in this single realization of 50 observations. The YW lines flatten at higher orders due to the triangular autocorrelation bias of (7) at high correlation lags but have no further negative influence on the accuracy, because all true AR parameters are zero for orders greater than 3. The results for Burg and YW are very close for model orders of less than $N/10$, but differences become noticeable for higher orders.

V. YW METHOD APPLIED TO THE AR(7) EXAMPLE

The YW bias has two important properties: The first is that the bias is only strong after a previous significant reflection coefficient. No bias is found if all reflection coefficients are small. The second is that it cannot be compensated by using higher model orders with additional AR parameters. The influence of the bias is demonstrated with 100 observations of an AR(7) example. The true parameters, reflection coefficients, their expectations with the triangular bias applied, and the estimated AR(7) reflection coefficients of the Burg method are given in Table III. The bias is small for the first two reflection coefficients and increases for higher orders due to the rather high value of k_2 . The bias is very strong for all orders higher

than 4. The Burg estimates from a single simulation run of 100 observations in the final column demonstrate some estimation variance but remain rather close to their expectations for all model orders.

It is shown that the YW bias also has influence on the parameters of orders higher than true order 7. Due to the triangular bias, all reflection coefficients belonging to the true biased autocorrelation function are nonzero. Therefore, the triangularly biased YW model gets theoretical order ∞ , independent of true order p . However, the values may be small for high orders. All true biased reflection coefficients above order 9 are less than 0.1 for $N = 100$ in this example; above order 16, they are less than 0.01.

The expected best AR model order for Burg models is true order 7. The best AR order to be selected with the AIC [9] for estimates with the YW method can theoretically be computed. The reduction of the residual variance is given by (14) and can be used in the AIC. The best order that can be selected for estimated YW models would become 9 in the true AR(7) example. The AR(8) model of only one order higher would be the best order for selection with the AIC if the first and second extra true biased reflection coefficients would be greater and smaller than $1/\sqrt{N}$, respectively. The AR(9) model of two orders higher is the best if the last reflection coefficient is greater than $1/\sqrt{N}$ and the expected sum of squares of the last two biased reflection coefficients is greater than $2/N$. In this example with $N = 100$, their sum of squares is 0.039 and fulfils this requirement.

The theoretical optimal order of the biased model becomes higher than the true order due to the triangular bias. However, the differences are small, and the comparison between the true and true biased models is made here for the AR(7) model. The $ME(\hat{A}, A)$ is 117.4 for this example, with $\hat{A}(z)$ and $A(z)$ being the triangularly biased polynomial of (8) and the true AR(7) polynomial, respectively. Condition number κ is 14 547. The AR(7) example would require 125 000 observations before the condition $ME(\hat{A}, A) = 7$ is met.

The AR(7) example can be characterized by its poles, which are shown in Fig. 2. The true complex conjugated poles have radii of 0.953, 0.940, and 0.936. All the biased complex poles are located much further away from the unit circle, with radii of 0.886, 0.859, and 0.858. The complex estimated Burg poles have radii of 0.960, 0.942, and 0.935, which are much closer to the true radius.

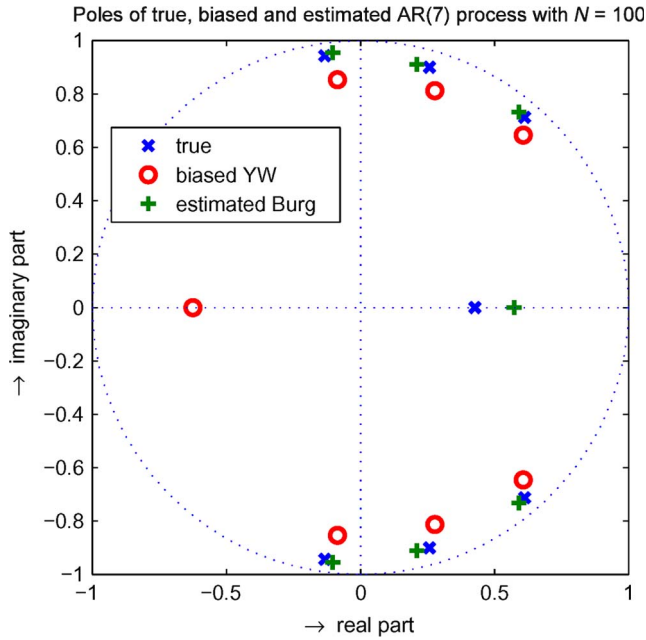


Fig. 2. Poles of the true AR(7) process, the biased AR(7) expectation for the YW method, and the estimated AR(7) model of the Burg method; the triangular AR(7) bias is computed for $N = 100$.

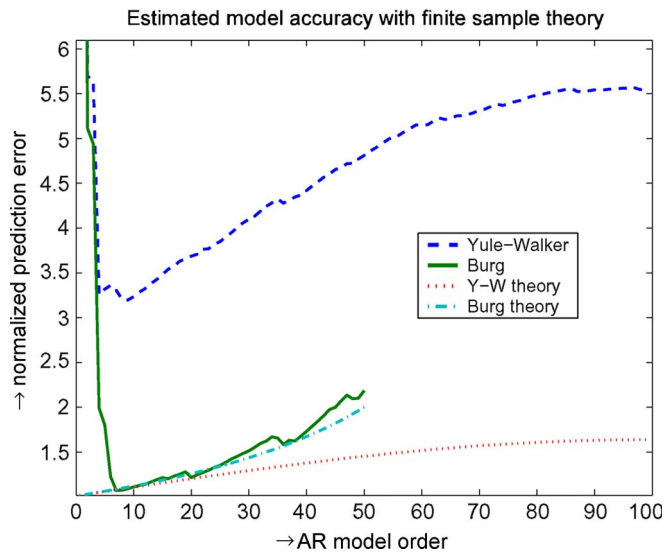


Fig. 3. Estimated accuracy $FSC(q)$ of (20) and theoretical accuracy $E_{FS}[PE(q)]$ of (19) for the AR models estimated from $N = 100$ AR(7) observations as a function of the model order. The strong bias of the YW estimates yields a very large difference with the theoretical finite-sample accuracy that does not take the bias into account.

In addition, the angles of the poles are shifted, which will cause a shift in the frequencies of the spectral peaks. The large displacement of the biased pole on the real axis will give a biased spectrum, which is much too high at high frequencies.

The accuracy of the estimated AR models of different orders is shown in Fig. 3. The estimated model accuracy of the Burg method is close to the theoretical white noise expectation for orders of 7 and higher. Lower order models have a larger ME value, because not all truly nonzero parameters are included. The Burg models of orders higher than 6 have only a very small bias. The influence of the additional triangular bias on

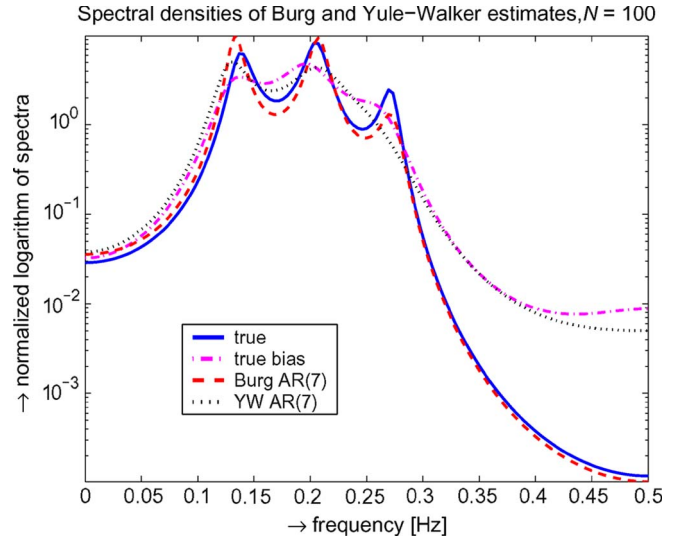


Fig. 4. Estimated and theoretical spectra for the AR models estimated from $N = 100$ AR(7) observations with the Burg and YW methods. The order 7 was selected for both the Burg and YW estimates in this simulation run.

the YW estimates is very great. It is seen that the accuracy may still become somewhat better for orders higher than 7; theoretically, the best order should be 9 for estimates with triangular bias. However, the general shape of the YW model accuracy looks like the theoretical shape of unbiased models, with the same tendency to become flattened at very high orders. Order selection between the YW candidates cannot select an accurate model with a small prediction error, because all YW models are poor in Fig. 3. The bias propagates to all higher order models. AR models estimated with the Burg method are much better.

The estimated spectra of the Burg and YW method are compared with the true spectrum and the expectation of the biased spectrum in Fig. 4 on a logarithmic scale. According to Parseval's theorem, the small triangular bias in the autocorrelation can only cause a small bias in the spectrum, because the sum of the squared differences is equal in both the time and frequency domains. However, the same difference can become very significant in the logarithmic presentation of the power spectral density. The shift of the real pole to the left in Fig. 2 causes the distortion of the YW estimates at higher frequencies. The shift of the complex poles away from the unit circle makes the three spectral peaks less strong for the YW estimates.

The Burg estimates are a close approximation to the true spectrum, whereas the YW estimates approximate the biased true spectrum. The YW estimate of AR order $N - 1$ approximately gives the periodogram as a spectral estimate. Its accuracy can be found in Fig. 3 at the highest model order. The periodogram suffers from the same triangular bias. Hence, it will globally follow the biased spectrum in Fig. 4. However, the variations around the true curve will be much greater. As shown in Fig. 3, YW models of all higher orders have the bias error that was already present in the YW estimate of true order 7.

The average model qualities, i.e., the MEs, of the YW and Burg estimates, both with selected AR orders, have been determined in 1 000 simulation runs, with 100 observations

each. The average ME for YW was 102.3, and the ME value for Burg was 10.7. The quality of the estimated Burg models with the automatically selected model orders of the ARMAsef program [18] is rather close to the Cramér–Rao lower bound 7 for the ME of an AR(7) process. This average includes the uncertainty of order selection; the average quality of the AR(7) Burg models without selection of the order was 8.8.

Not all processes with poles close to the unit circle are very sensitive to the triangular bias of the YW method. As an example, consider an AR(7) process with the first six true parameters zero and a_7 having a value of 0.7. All seven poles would then have equal radii of 0.95. The triangular bias moves them all to 0.94 for $N = 100$ at almost the same angles as the true AR(7) process. The poles of this second example are as close to the unit circle as the poles shown in Fig. 2. However, the influence of the bias is very much smaller now, with $\text{ME}(\hat{A}, A)$ equal to 0.47 and condition number $\kappa = 5.7$.

VI. CONCLUDING REMARK

Biased LP autocorrelation estimates are used in the YW method of AR parameter estimation. The small triangular autocorrelation bias can cause a significant bias of the AR parameters and its logarithmic spectrum in finite samples.

The ME is introduced as an objective measure for quantifying the statistical significance of the triangular bias. It determines the smallest sample size where bias properties can be considered as asymptotic, because the influence of the bias becomes smaller than the inevitable inaccuracy due to the estimation variance.

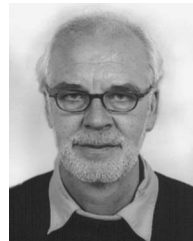
The triangular bias modifies the true order to ∞ for all AR(p) processes estimated with the YW method. The best approximating model order may become higher than true order p and depend on the sample size. Both the locations and the height of spectral peaks may be considerably changed by the YW bias.

The AR model estimated with the Burg method do not have such problems and are preferred.

REFERENCES

- [1] S. M. Kay and S. L. Marple, "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, Nov. 1981.
- [2] J. Durbin, "The fitting of time series models," *Rev. Inst. Int. Stat.*, vol. 28, pp. 233–243, 1960.
- [3] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.

- [4] P. M. T. Broersen, *Automatic Autocorrelation and Spectral Analysis*. London, U.K.: Springer-Verlag, 2006.
- [5] D. Tjøstheim and J. Paulsen, "Bias of some commonly-used time series estimates," *Biometrika*, vol. 7, no. 2, pp. 389–399, 1983.
- [6] P. Shaman and R. A. Stine, "The bias of autoregressive coefficient estimators," *J. Amer. Stat. Assoc.*, vol. 83, no. 403, pp. 842–848, Sep. 1988.
- [7] J. P. Burg, "Maximum entropy spectral analysis," in *Proc. 37th Meeting Soc. Exploration Geophys.*, 1967. Also in *Modern Spectrum Analysis*, D.G. Childers, Ed., New York: IEEE Press, pp. 34–41.
- [8] S. Kay and J. Makhoul, "On the statistics of the estimated reflection coefficients of an autoregressive process," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 6, pp. 1447–1455, Dec. 1983.
- [9] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
- [10] E. Wensink and W. J. Dijkhof, "On finite sample statistics for Yule–Walker estimates," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 509–516, Feb. 2003.
- [11] P. M. T. Broersen and H. E. Wensink, "Autoregressive model order selection by a finite sample estimator for the Kullback–Leibler discrepancy," *IEEE Trans. Signal Process.*, vol. 46, no. 7, pp. 2058–2061, Jul. 1998.
- [12] B. Porat, *Digital Processing of Random Signals: Theory and Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
- [13] J. S. Erkelens and P. M. T. Broersen, "Bias propagation in the autocorrelation method of linear prediction," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 2, pp. 116–119, Mar. 1997.
- [14] P. M. T. Broersen, "Historical misconceptions in autocorrelation estimation," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 4, pp. 1189–1197, Aug. 2007.
- [15] M. J. L. de Hoon, T. H. J. J. van der Hagen, H. Schoonewelle, and H. van Dam, "Why Yule–Walker should not be used for autoregressive modelling," *Ann. Nucl. Energy*, vol. 23, no. 15, pp. 1219–1228, Oct. 1996.
- [16] P. M. T. Broersen, "Finite-sample bias in the Yule–Walker method of autoregressive estimation," in *Proc. IEEE IMTC*, Victoria, BC, Canada, 2008, pp. 342–347.
- [17] M. B. Priestley, *Spectral Analysis and Time Series*. London, U.K.: Academic, 1981.
- [18] P. M. T. Broersen, *ARMASA Matlab toolbox*, 2002. [Online]. Available: <http://www.mathworks.com/matlabcentral/fileexchange/?term=armasa>



Piet M. T. Broersen was born in Zijdewind, The Netherlands, in 1944. He received the M.Sc. degree in applied physics and the Ph.D. degree from Delft University of Technology, Delft, The Netherlands, in 1968 and 1976, respectively.

He is currently with the Department of Multi-Scale Physics, Delft University of Technology. He developed statistical measures to let the measured data speak for themselves in estimated time series models. This provides a practical and accurate solution for the spectral and autocorrelation analyses of stochastic data. Software for the automatic estimation of the spectra and autocorrelation functions of equidistant random data for missing data problems and irregularly sampled observations is available online. His research interests include the automatic and unambiguous identification of the character of stationary random measurement data.