Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

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

Piet M.T. Broersen

Department of Multi Scale Physics
Delft University of Technology, The Netherlands
Tel: +31 15 2786419, e-mail p.m.t.broersen@tudelft.nl

Abstract: Lagged-product autocorrelation estimates have a small triangular bias. However, using them to compute an autoregressive model with the Yule-Walker method can give a strongly distorted spectral model in finite samples. The distortion is shown for examples where the reflection coefficients are not very close to one in absolute value. It will disappear asymptotically. An objective measure is presented to determine the smallest sample size for which the unbiased asymptotic theory is a fair approximation.

1. INTRODUCTION

The Yule-Walker (YW) equations describe the relation between the p autoregressive (AR) parameters and the first p lags of the autocorrelation function. The YW estimation method solves the AR parameters from the YW equations, with the lagged-product estimates substituted for the autocorrelations (Kay and Marple, 1981). The Levinson-Durbin recursion is an efficient solution of the YW equations. For increasing AR orders, it computes reflection coefficients, which are defined as the negatives of the partial correlation coefficients (e.g. Kay and Marple, 1981; Stoica and Moses, 1997; Broersen, 2006). The parameters of an AR model of an arbitrary order q can be computed from the first q reflection coefficients. Those in turn are completely determined by the first q lags of the autocorrelation function.

Shaman and Stine (1988) showed that the bias of the YW method of AR estimation is given by the general terms for AR models, with an additional contribution caused by the triangular bias of the lagged-product autocorrelation estimate. They have given explicit expressions for the first two estimated AR parameters. The bias of the second parameter is a rational function of the parameters, which can become large if poles of the AR polynomial are close to one in absolute value. De Hoon *et al.* (1996) showed that the last parameter of order p never causes this extra bias for an AR(p) process. They related the magnitude of the bias to the condition number of the Toeplitz autocorrelation matrix, which contains only the estimated autocorrelations until order $p - 1$.

Also the variance of AR estimates has been studied. The asymptotical theory gives identical results for all AR estimation methods. Well known AR estimation methods are the YW method, the method of Burg (1967), several least squares solutions and the maximum likelihood method (e.g. Kay and Marple, 1981; Stoica and Moses, 1997; Broersen, 2006). Kay and Makhoul (1983) derived asymptotic expressions for the variance of estimated reflection coefficients, for orders up to the true AR order. Explicit expressions were given for AR processes until order 3 and an efficient recursive procedure can compute that variance for higher order AR processes. Broersen (2006) showed that empirical finite-sample approximations for the variance of estimated parameters depend on the method of AR estimation. He used the empirical variances in a finite-sample order selection criterion that is adapted to the estimation method. It turned out that the finite-sample selection criterion for the YW estimation method was almost identical with the asymptotic AIC criterion of Akaike (1974). A detailed theoretical analysis of Wensink and Dijkhof (2003) derived the expectations of the squared reflection coefficients in small samples of a white noise signal. They determine the variance of the reflection coefficients of general AR processes for all orders higher than the true AR order, which is important for a finite-sample order selection criterion.

The triangular bias turned out to be important if a reflection coefficient of any order lower than the true process order p is close to +1 or –1. If the true reflection coefficient of order m, 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 % (Erkelens and Broersen, 1997). The condition number of de Hoon *et al.* (1996) for the true autocorrelation function is independent of the number of observations N. On the other hand, the triangular bias of the lagged-product estimates diminishes with N and vanishes asymptotically. Therefore, no critical value for that condition number can be given. AR models estimated with the YW method are asymptotically unbiased. Hence, processes with a high condition number may give a large bias in finite samples, but the bias will disappear for growing sample sizes.

This paper presents examples with heavy YW bias, although the poles are not very close to the unit circle or the reflection coefficients are not close to unity. The bias can be computed by applying the triangular autocorrelation bias to the true autocorrelation function, and solving the YW equations with that bias. An objective measure is given that determines the minimum sample size for which the influence of the bias in the Yule-Walker method is not greater than the inaccuracy due to the estimation variance of the AR parameters.

2. AR ESTIMATES WITH THE YULE-WALKER METHOD

A discrete-time AR(p) process is a time series x_n that can be written as (e.g. Kay and Marple, 1981; Priestley, 1981; Stoica and Moses, 1997; Broersen, 2006)

$$x_n + a_1 x_{n-1} + \cdots + a_p x_{n-p} = \varepsilon_n \,, \tag{1}$$

where ε_n is a purely random process with mean zero 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 sufficient because high order parameters tend to be small. The poles of the AR(p) process are the roots of the AR polynomial $A(z)$:

$$A(z) = 1 + a_1 z^{-1} + \cdots + a_p z^{-p} \,. \tag{2}$$

Processes are called stationary if the poles are strictly within the unit circle. The parametric power spectrum $h(\omega)$ of the AR(p) model is for $-\pi < \omega \le \pi$ computed with

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

The normalized spectrum $\varphi(\omega)$ is defined as the spectrum divided by the variance of the signal:

$$\phi(\omega) = h(\omega) / \sigma_x^2 . \tag{4}$$

The autocovariance $r(q)$ and the normalized autocorrelation $\rho(q)$ at lag q are defined for a signal with mean value zero by

$$\rho(q) = \frac{r(q)}{r(0)} = \frac{\mathrm{E}\left(x_n x_{n+q}\right)}{\mathrm{E}\left(x_n^2\right)} \,. \tag{5}$$

All lags of the infinitely long true autocorrelation function of an AR(p) process are determined by p true AR parameters with the *YW equations*, given by (Kay and Marple, 1981):

$$\begin{aligned} &\rho(q) + a_1 \rho(q-1) + \cdots + a_p \rho(q-p) = 0 \,, \quad q \ge 1 \\ &\rho(-q) = \rho(q) \,. \end{aligned} \tag{6}$$

Different efficient numerical methods have been developed to solve the YW equations. Recursive algorithms compute the reflection coefficients for increasing orders. The first m reflection coefficients are determined by the first m lags of the true AR autocorrelation function, for all m. Afterward, parameters for arbitrary AR orders q can be found from the first q reflection coefficients; see Stoica and Moses (1997). Equation (6) can also be used for indexes q greater than p, to extrapolate the autocorrelation function. The same relations (6) can be used to compute parameters from correlations and vice versa. They relate true or estimated AR parameters with true or estimated autocorrelations, respectively. Broersen (2007) showed that efficiently estimated AR parameters serve as an efficient estimator for the autocovariance function.

To obtain a positive semi-definite lagged-product (LP) estimator for the autocorrelation function of N observations x_n, the estimator for lag q should be biased and is given by

$$\hat{\rho}_{LP}(q) = \hat{r}_{LP}(q) / \hat{r}_{LP}(0) = \frac{1}{N} \sum_{n=1}^{N-q} x_n x_{n+q} \Big/ \frac{1}{N} \sum_{n=1}^{N} x_n^2 \,. \tag{7}$$

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

$$\mathrm{E}\left[\hat{r}_{LP}(q)\right] = \mathrm{E}\left[\frac{1}{N} \sum_{n=1}^{N-q} x_n x_{n+q}\right] = \frac{N-q}{N} \mathrm{E}\left[x_n x_{n+q}\right] = r(q)\left(1 - \frac{q}{N}\right). \tag{8}$$

The bias of the normalized autocorrelation in (7) has some additional terms given by Broersen (2007). Those terms arise from the expectation of quotients, but they are not important for the sequel. 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 (6):

$$\begin{aligned} &r(q)\left[1 - q/N\right] + a_{1,b} r(q-1)\left[1 - (q-1)/N\right] + \cdots \\ &\qquad \cdots + a_{p,b} r(q-p)\left[1 - (q-p)/N\right] = 0 \,, \;\; q = 1, \cdots, p \,. \end{aligned} \tag{9}$$

The YW method of AR estimation has this expectation $a_{i,b}$ because it substitutes LP estimates (7) in (6) to compute the AR parameters $\hat{a}_i$ of an AR(p) model with the p equations

$$\hat{\rho}_{LP}(q) + \hat{a}_1 \hat{\rho}_{LP}(q-1) + \cdots + \hat{a}_p \hat{\rho}_{LP}(q-p) = 0 \,, \;\; 1 \le q \le p \,. \tag{10}$$

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, which can be computed by using estimated parameters to predict a fresh realization of x_n:

$$\hat{x}_n = -\hat{a}_1 x_{n-1} - \cdots - \hat{a}_p x_{n-p} \,. \tag{11}$$

The PE is defined as the long-term average of $\left(\hat{x}_n - x_n\right)^2$. 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 the estimated parameters values are known in Monte Carlo simulation experiments. For these particular situations where the true parameters are known, Broersen (2006) showed that the expectation of the PE can be computed without generating new data by using

$$\mathrm{PE} = \frac{\sigma_\varepsilon^2}{2\pi} \int_{-\pi}^{\pi} \frac{h(\omega)}{\hat{h}(\omega)} \, d\omega = \frac{\sigma_\varepsilon^2}{2\pi} \int_{-\pi}^{\pi} \left| \frac{\hat{A}\left(e^{j\omega}\right)}{A\left(e^{j\omega}\right)} \right|^2 d\omega \,, \tag{12}$$

where $\hat{A}(z)$ denotes an estimated AR polynomial. He derived an efficient algorithm to compute the PE of (12). The model error ME for has been defined for AR models by scaling the PE with σ_ε^2 and by multiplying with the sample size as

$$\mathrm{ME} = \mathrm{ME}\left(\hat{A}, A\right) = N\left(\frac{\mathrm{PE} - \sigma_\varepsilon^2}{\sigma_\varepsilon^2}\right). \tag{13}$$

The argument of the ME is left if no confusion is possible. The expectation of the ME for unbiased efficiently estimated AR models of orders K, greater than or equal to the true order p, equals K. The variance of each estimated parameter has a minimal contribution 1 to the ME expectation. The Cramér-Rao bound for the ME of an AR(p) process is given by p.

The condition number κ of the Toeplitz correlation matrix R with the elements $r\,(0)$ until $r\,(p$ - 1) has been used by de Hoon *et al.* (1996) to classify the sensitivity of an AR process to the YW bias. It is defined as

$$\kappa = \left[\!\left[R \right]\!\right] \cdot \left[\!\left[R^{-1} \right]\!\right], \tag{14}$$

where $[\![\ ]\!]$ is some matrix norm, like the largest singular value.

The norm is computed for the expectation of the auto-correlation matrix and is independent of the sample size N.

An objective measure for the quality of AR models estimated with the YW method for a given sample size N is found by calculating the biased expectations of the autocovariances with (8), the parameters $a_{i,b}$ of the biased AR(p) model $\hat{A}(z)$ with (9) and finally the ME($\hat{A},A$) with (13). The asymptotical expectation of this ME($\hat{A},A$) is zero. A critical sample size for which the bias is no longer considered significant is defined to be ME($\hat{A},A$) = p. That is the sample size where the bias error becomes of the same magnitude as the variance error that is caused by efficiently estimating the p parameters of the AR model. This limit for ME($\hat{A},A$) seems to be a little bit arbitrary. However, equal bias and variance contributions are known from the derivation of the order selection criteria of Akaike (1974) as an estimator for the Kullback-Leibler index. Therefore, this choice has a firm statistical background.

3. FINITE-SAMPLE AR THEORY

Kay and Marple (1981) describe many existing methods for AR estimation. The method of YW was the first; the method of Burg (1967) has nice properties and is often used for AR estimation in practice. The asymptotical theory for the residual variance s_q^2 and the prediction error PE(q) do not depend on the estimation method. 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 $r(0)$, the variance of the process x_n. For any arbitrary order q, the residual variance s_q^2 is given by

$$s_q^2 = s_0^2 \prod_{i=1}^{q} \left(1 - k_i^2\right) \tag{15}$$

where k_i is the true or the estimated reflection coefficient of order i. If true reflection coefficients are substituted in (15), their value is zero for orders greater than the true 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 the true order p because of the estimation variance of k_i.

A simple explanation for the triangular bias of the YW AR method has been given by Erkelens and Broersen (1997). The residual variance s_q^2 appears in the denominator of the Levinson-Durbin recursion to compute the reflection coefficient for the order $q + 1$ (Kay and Marple, 1981; Stoica and Moses, 1997). Suppose that a true k_q has the unbiased value 1– 1/N. A small bias of the magnitude 1/N would give it the biased value 1– 2/N. Substitution of the unbiased and of the biased values for k_q in (1- k_q^2) gives the approximations 2/N and 4/N, respectively. A small bias of magnitude 1/N in a reflection coefficient k_q makes a difference of a factor 2 in the residual variance s_q^2. The error in s_q^2 propagates to all higher AR orders because (15) is in the denominator of 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 the true order p is close to unity, with a distance less than 1/N. However, for any value k_q = 1 - δ, with a very small δ, the finite-sample influence of the bias disappears asymptotically if the sample size N is much greater than 1/ δ.

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

$$\mathrm{E}\left[s_K^2 \right] = \sigma_\varepsilon^2 \left(1 - \frac{K}{N} \right),\ K \geq p \tag{16}$$

$$\mathrm{E}\left[\mathrm{PE}(K) \right] = \sigma_\varepsilon^2 \left(1 + \frac{K}{N} \right),\ K \geq p\ . \tag{17}$$

This result can be applied in practice with a reasonable accuracy if K is less then 0.1N. For still higher AR orders, Broersen (2006) showed that finite-sample theory gives an improved accuracy. That theory is based on approximations for 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. Broersen (2006) gives the approximations

$$v_{i,YW} = \frac{N - i}{N(N+2)} \tag{18}$$

$$v_{i,Burg} = \frac{1}{N + 1 - i}$$

for 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.

The finite-sample expressions for s_K^2 and PE(K) become

$$\mathrm{E}_{\mathrm{FS}}\left[s_K^2 \right] = \sigma_\varepsilon^2 \prod_{i=1}^{K} \left(1 - v_{i,\cdot}\right),\ K \geq p \tag{19}$$

$$\mathrm{E}_{\mathrm{FS}}\left[\mathrm{PE}(K) \right] = \sigma_\varepsilon^2 \prod_{i=1}^{K} \left(1 + v_{i,\cdot}\right),\ K \geq p\ . \tag{20}$$

By substituting the YW or the Burg variance coefficients for $v_{i,\cdot}$, different results are obtained for the estimation methods.

The residual variance s_q^2 is known for practical data 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)

$$\mathrm{FSC}(q) = s_q^2 \frac{\prod_{i=1}^{q} \left(1 + v_{i,\cdot}\right)}{\prod_{i=1}^{q} \left(1 - v_{i,\cdot}\right)}\ . \tag{21}$$

This approximation can be computed for N given data x_n. It has a strong relation with the final prediction error FPE of Akaike (1974) that has been used in his first order selection criterion. It 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 (19) and (20) for higher orders. The expectation of FSC(q) for higher model orders is given by E_{FS}[PE(q)]. For finite samples, it depends on the estimation method if q is greater than about N/10.

4. YULE-WALKER METHOD APPLIED TO AR(4)

To demonstrate the bias propagation, an AR(4) example with four equal reflection coefficients has been studied. This example has the advantage that all unbiased true reflection coefficients are the same and the bias for different orders can be compared easily. The triangular bias of (8) is applied to the true autocorrelation function for $N = 100$ and the biased values for the reflection coefficients are found by substituting those biased correlations in (9). As it is hardly illustrative to derive the analytical expressions for this example, numerical values are presented as a function of the true reflection coefficients. Table 1 gives the true value k, the four biased reflection coefficients and the two measures: ME($\hat{A},A$) and κ.

Table 1. True reflection coefficient k and expectations for the biased ones for an AR(4) process generated with four equal true reflection coefficients k. The ME of the biased AR(4) model as computed with (9) and the condition number κ are given for $N = 100$ and for different values of the true k.

true	biased expectations					
k	k_1	k_2	k_3	k_4	ME($\hat{A},A$)	κ
0.2	0.198	0.196	0.193	0.190	0.02	2.5
0.4	0.396	0.391	0.381	0.365	0.3	8.2
0.6	0.594	0.581	0.547	0.452	5.7	43.6
0.7	0.693	0.673	0.599	0.361	33.0	135.3
0.8	0.792	0.757	0.574	0.050	213.1	616.8
0.9	0.891	0.813	0.268	-0.296	1451.1	6896.5
.99	0.980	0.492	-0.304	0.193	$1.8\ 10^6$	$9.6\ 10^6$

The first reflection coefficient has a bias of the magnitude $1/N$ for all true values of the reflection coefficient k. 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. Finally, the true value $k = 0.99 = 1–1/N$ gives already a theoretical bias of 50 % for the second reflection coefficient k_2.

It is remarkable how the bias propagates to higher orders for values of k much smaller than $1– p/N$, which are rather far from the unit circle. Moderate values of the condition number κ lead already to strong biases and low condition numbers do not always guarantee a small bias. The example with $k = 0.7$ demonstrates that the moderate condition number 135.3 can give a significant model error ME. The demand that the ME($\hat{A},A$) is less than 4 is only met for $k < 0.578$ in this AR(4) example, with $\kappa = 35.1$ as the condition number. The sample size N should be greater than 6 000 000 to obtain the reduced bias influence ME($\hat{A},A$) < 4 for $k = 0.9$; the theoretical value for the biased k_4 would be 0.8997 then. The variance influence will become really small for very high N only. However, the bias influence is still vanishing asymptotically.

This is also demonstrated in Table 2. The values of k_4(biased) and ME($\hat{A},A$) of the AR(4) model are given for increasing N. Due to the multiplication in the ME with the sample size in (13), the influence of the bias grows for N less than 1000. It decreases proportional to $1/N$ for N greater than about 100000 in this example with 4 equal true reflection coefficients. ME becomes equal to the model order 4 for N is about 6 000 000.

Table 2. The expectation of the biased reflection coefficient of order 4 and the corresponding model error ME($\hat{A},A$) of the bias in (9) for an AR(4) process with four equal true reflection coefficients 0.9, as a function of sample size N.

N	k_4(biased)	ME($\hat{A},A$)
10^2	-0.296	1451.1
10^3	0.063	4200.9
10^4	0.720	1866.4
10^5	0.880	236.5
10^6	0.898	24.2
10^7	0.900	2.43

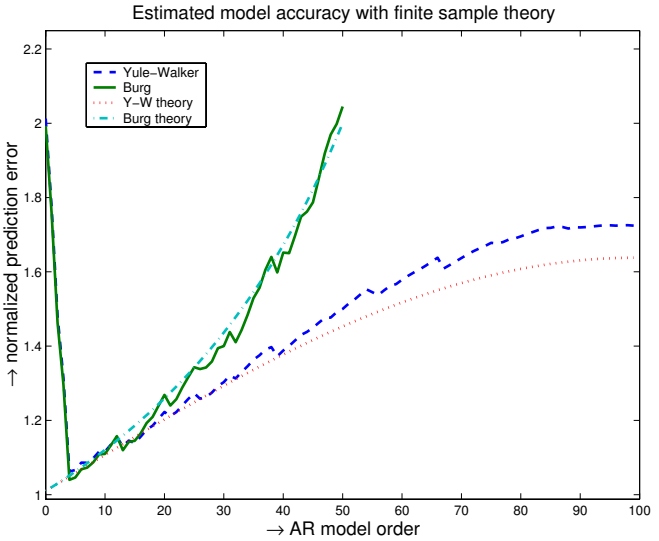

Fig. 1. The estimated accuracy FSC(q) of (21) and the theoretical accuracy E_{FS}[PE(q)] of (20) for AR models estimated from a single realization of $N = 100$ observations of an AR(4) process with 4 equal true reflection coefficients 0.4.

The model accuracy can be determined for estimated models of increasing orders. Software is available to estimate many AR models and to select the best model order automatically (Broersen, 2002). Fig. 1 gives results for an AR(4) example with equal true reflection coefficients 0.4 where the bias is not influential. Both YW and Burg AR estimates are shown. The estimated accuracies FSC(q) are computed with (21) and further the theoretical finite-sample expectations for unbiased models computed with (20) are presented for the two methods. Both methods follow their expectations. The YW lines become flattened at higher orders due to the triangular autocorrelation bias of (8) at high correlation lags, but that has no further negative influence on the accuracy because all true parameters are zero for orders greater than 4. The results for Burg and YW are very close for model orders less than $N/10$, but differences become noticeable for higher orders.

5. YULE-WALKER METHOD APPLIED TO AN 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

Table 3. The true parameters a_i of an AR(7) process, the reflection coefficients k_i and the biased expectations; $N = 100$.

order i	a_i	k_i	k_i (biased)
1	-1.90	-0.37	-0.366
2	3.46	0.85	0.830
3	-3.68	-0.33	-0.267
4	3.59	0.77	0.607
5	-2.26	-0.32	0.057
6	1.21	0.70	0.145
7	-0.30	-0.30	0.267
8		0	-0.081
9		0	-0.180
10		0	0.034

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 and their biased expectations are given in Table 3. The bias is small for the first two reflection coefficients and increases for higher orders due to the rather high value of k_2. It is shown that the bias also has influence on the parameters of orders higher than the true order 7. Due to the triangular bias, all reflection coefficients belonging to the true biased autocorrelation function are non-zero. Therefore, the triangularly biased YW model gets the theoretical order ∞ for all true p. However, the values are small for high orders and all true biased reflection coefficients above order 9 are less than 0.1 for $N = 100$; above order 16 they are less than 0.01.

The best AR model order to be selected with the AIC criterion of Akaike (1974) for estimates with the YW method can be computed theoretically. The reduction of the residual variance is given by (15) and that can be used in AIC. The best order that can be selected for biased YW models would become 9 in this true AR(7) example. The AR(8) model of one order higher would be the best order for selection with AIC if the first extra reflection coefficient would be greater than $1/\sqrt{N}$. 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 the true biased model are made for the AR(7) model. The model error ME($\hat{A}$,A) is 117.4 for this example, with $\hat{A}(z)$ and $A(z)$ as the triangularly biased polynomial of (9) and the true AR(7) polynomial, respectively. The condition number κ is 14547. This AR(7) example would require 125000 observations before the condition ME($\hat{A}$,A) = 7 is met.

The AR(7) example can be characterized by its poles, which are given in Fig. 2. The true complex conjugated poles are at the radii 0.953, 0.940 and 0.936. All biased complex poles are much further away from the unit circle, at radii 0.886, 0.859 and 0.858. Also the angles of the poles are shifted, which will cause a shift of the frequencies of spectral peaks.

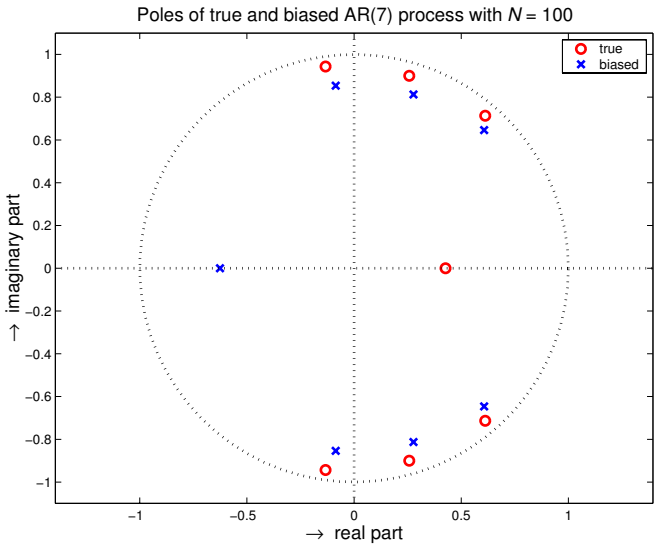

Fig. 2. Poles of true and biased AR(7) process, where the theoretical triangular AR(7) bias is computed for $N = 100$.

The large displacement of the pole on the real axis will give a biased spectrum, which is much too large at high frequencies.

The accuracy of AR models of different orders is given in Fig. 3. The estimated model accuracy of the Burg method is close to the theoretical white noise expectation for orders 7 and higher. Lower order models have a larger ME value because not all truly non-zero parameters are included. The Burg models of orders higher than 6 are unbiased. The influence of the triangular bias on 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 the estimates with triangular bias. However,

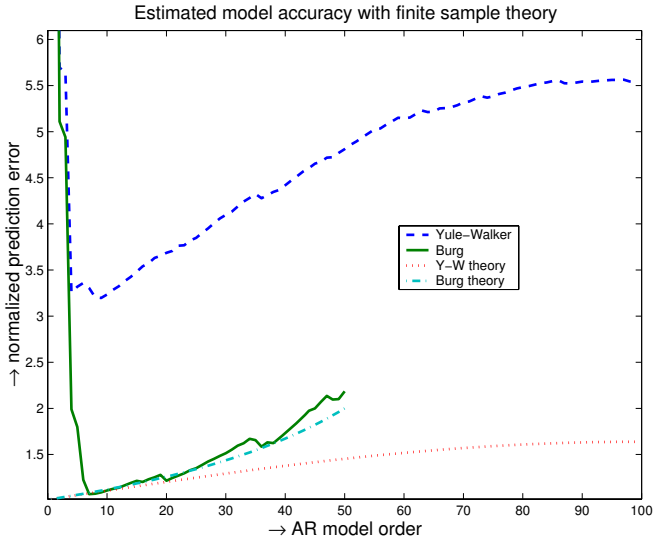

Fig. 3. The estimated accuracy FSC(q) of (21) and the theoretical accuracy E_{FS}[PE(q)] of (20) for AR models estimated from $N = 100$ AR(7) observations, as a function of the model order. The strong bias of the Yule-walker estimates gives a very large difference with the theoretical finite-sample accuracy that does not take the bias into account.

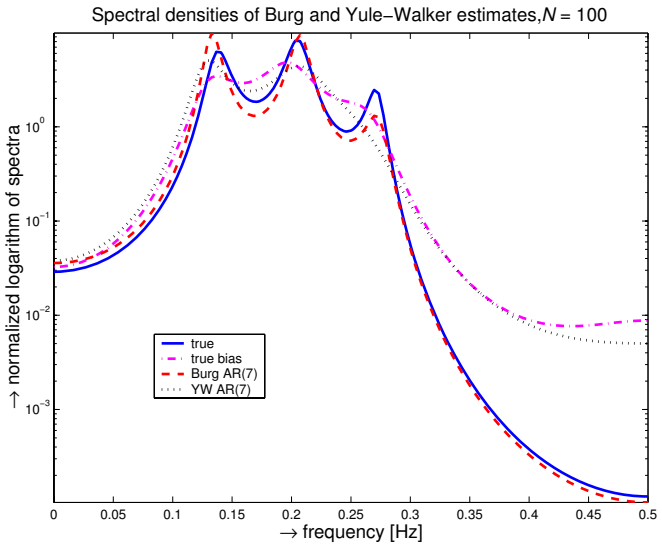

Fig. 4. The estimated and the theoretical spectra of AR models estimated from $N = 100$ AR(7) observations with the Burg and the YW method. The order 7 was selected both for Burg and for YW estimates in this simulation run.

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, because all YW models are poor. The bias propagates to all higher order models. AR models with Burg are much better.

The estimated spectra of the Burg and YW method are compared with the true and with the expectation of the biased spectrum in Fig. 4. The shift of the real pole to the left causes the distortion of the 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; the YW estimates approximate the biased true spectrum. The YW estimate of AR order $N - 1$ approximately gives the periodogram as spectral estimate. 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 are much greater. As indicated in Fig. 3, YW models of all higher orders have the bias error that was already present in the YW estimate of the true order 7.

The average model qualities ME of YW and Burg estimates, both with selected AR orders, have been determined in 1000 simulation runs. The average ME for YW was 102.3 and for Burg the ME value was 10.7. The quality of estimated Burg models with the automatically selected model orders of the ARMAsel program of Broersen (2006) is rather close to the Cramér-Rao lower bound 7 for the ME of an AR(7) process. This average quality includes the uncertainty of order selection; the average quality of AR(7) Burg models was 8.8.

Not all processes with poles close to the unit circle are very sensitive for the triangular bias of the Yule-Walker method. As an example, take the AR(7) process with all true parameters zero, except a_7 that is equal to 0.7. All 7 poles would be at the equal radius 0.95 then. 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 closer to the unit circle than the poles of Fig. 2. However, the influence of the bias is very much smaller, with ME($\hat{A}$,A) equal to 0.47 and with the condition number $\kappa = 5.7$.

6. CONCLUDING REMARKS

Biased lagged-product autocorrelation estimates are used for the Yule-Walker method of AR parameter estimation. The small triangular autocorrelation bias can cause a significant bias of the AR parameters and its spectrum in finite samples. The model error is introduced as an objective measure to quantify the significance of the triangular bias. It determines the smallest sample size where properties are asymptotic. The triangular bias modifies the true order to ∞ for all biased AR(p) processes. The best approximating model order may become higher than the true order p and depend on the sample size. The locations and the height of spectral peaks may be changed considerably by the Yule-Walker bias. Burg estimates are preferred and do not have those problems.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control*, **AC-19**, 716-723.

Broersen, P.M.T. (2002). *ARMASA Matlab toolbox*, 2002, [Online], Available at www.mathworks.com/matlabcentral/fileexchange.

Broersen, P.M.T. (2006). *Automatic Autocorrelation and Spectral Analysis*. London, U.K.: Springer-Verlag.

Broersen, P.M.T. (2007). Historical misconceptions in autocorrelation estimation. *IEEE Trans. on Instrumentation and Measurement*, **56**, 1189-1197.

Burg, J.P. (1967). Maximum entropy spectral analysis. *Proc 37th Meeting Soc. of Exploration Geophysicists,* 6 pp.

Erkelens, J.S. and P.M.T. Broersen (1997). Bias propagation in the autocorrelation method of linear prediction. *IEEE Trans. on Speech and Audio Processing*, **5**, 116-119.

Hoon, M.J.L. de, T.H.J.J. van der Hagen, H. Schoonewelle and H. van Dam (1996). Why Yule-Walker should not be used for autoregressive modelling. *Ann.Nucl. Energy*, **23**, 1219-1228.

Kay, S. and J. Makhoul (1983). On the statistics of the estimated reflection coefficients of an autoregressive process. *IEEE Trans.Acoust., Speech, Signal Process.*, **ASSP-31**, 1447-1455.

Kay, S.M. and S.L. Marple (1981). Spectrum analysis-a modern perspective, *Proc. IEEE*, **69**, 1380-1419.

Priestley, M.B. (1981). *Spectral Analysis and Time Series*. London, U.K.: Academic Press.

Shaman, P. and R.A. Stine (1988). The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association*, **83**, 842-848.

Stoica, P. and R. Moses (1997). *Introduction to Spectral Analysis*. Prentice Hall, Upper Saddle River, NJ.

Wensink, H. E. and W.J. Dijkhof (2003). On finite sample statistics for Yule-Walker estimates. *IEEE Trans. on Information Theory*, **49**, 509-516.