

Modified Durbin Method for Accurate Estimation of Moving-Average Models

Piet M. T. Broersen

Abstract—Spectra with narrow valleys can accurately be described with moving-average (MA) models by using only a small number of parameters. Durbin's MA method uses the estimated parameters of a long autoregressive (AR) model to calculate the MA parameters. Probably all the pejorative remarks on the quality of Durbin's method in the literature are based on suboptimal or wrong choices for the method of AR estimation or for the order of the intermediate AR model. Generally, the AR order should considerably be higher than the order of the best predicting AR model, and it should grow with the sample size. Furthermore, the Burg estimates for the AR parameters give the best results because they have the smallest variance of all the AR methods with a small bias. A modified Durbin MA method uses a properly defined number of AR parameters, which was estimated with Burg's method, and outperforms all the other known MA estimation methods, asymptotically as well as in finite samples. The accuracy is generally close to the Cramér–Rao bound.

Index Terms—Autoregressive (AR) models, model error, order selection, spectral analysis, time series, triangular bias.

I. INTRODUCTION

THE ESTIMATED spectra or autocorrelation functions are widely used to describe the character of stationary stochastic observations. Modified periodograms can be used for the nonparametric spectral estimates. Unfortunately, the modifications depend on the preferences of the user and cannot give the spectral quality that can be attained with the time-series models [1]. Likewise, it is a historical misconception that the nonparametric lagged-product estimates are based on firm statistical concepts [2]. It has been proved that most of the lagged-product autocorrelation estimates are not statistically efficient [3]. This means that those estimators do not reach the best possible statistical accuracy for the given data. Furthermore, none of the lagged-product estimates is efficient for data that can be described with a moving-average (MA) process [3]. This implies that the computation of q MA parameters from q estimated lagged-product autocorrelation estimates with a nonlinear algorithm [4] cannot produce efficient estimates for the MA parameters. It has been shown in extensive simulation studies that many MA estimators possess anomalous properties [5].

Manuscript received June 19, 2008; revised October 6, 2008. First published January 19, 2009; current version published April 7, 2009. The Associate Editor coordinating the review process for this paper was Dr. John Sheppard.

The author is with the Department of Multi Scale Physics, Delft University of Technology, 2628 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.2008.2009183

An accurate parametric spectral or autocorrelation estimation uses three types of time-series models to obtain parsimonious models for all types of stationary data [6]. Processes with spectral poles or narrow peaks are preferably described with autoregressive (AR) models. The MA models are suitable to describe processes with spectral zeros or narrow valleys with only a few parameters. The AR models would require many more parameters to approximate a spectrum with deep valleys. Finally, the combined ARMA models may be the best type for processes with a combination of spectral poles and zeros [6], [7].

The MA estimation has many difficulties in practice [5]. The maximum likelihood estimates have convergence problems. Often, the estimated zeros are precisely found on the unit circle. This stimulated the search for a robust algorithm to determine the MA parameters. The constrained optimization of the likelihood function has been tried [8] as well as many other attempts to find approximate maximum likelihood estimators for the MA models [9]. Those methods rely on the asymptotic properties that can be successful in large samples [10]. However, the approximate nonlinear maximum likelihood estimators failed to converge in some situations [11]. Recently, the drawbacks of all the existing methods for MA estimation have been summarized [12]. For a long time, there has been no decisive answer in the literature about which MA method is to be preferred.

Durbin's method for MA estimation replaces a nonlinear estimation problem by two stages of linear estimation [13]. First, the parameters of a long AR model are estimated from the data. Afterward, the MA parameters are computed by using the lagged products of those estimated AR parameters [13]. The MA method of Durbin is based on the theoretical and asymptotical equivalence of $AR(\infty)$ and $MA(q)$ processes. In practice, estimates of the finite-order AR models have to be used. A common choice has been to use the parameters of the best predicting AR model order, or an AR model order that depends on the number of MA parameters that is estimated [12]. The asymptotic consequences of using a finite-order AR model have been investigated for a $MA(1)$ process [14]. It turned out that the minimal sufficient statistic has the dimension of the sample size. Furthermore, omitting higher-order contributions has little effect on the estimation variances, but it has considerable effects on the finite-sample bias [14]. Durbin [13] concluded that the asymptotic behavior of the AR estimates is independent of the estimation method, and both the least-square AR estimates and the Yule–Walker (YW) estimates could be used in his MA method. Many comments on the old method of Durbin [13] are based on the choices that he made in his original contribution. With proper choices for the AR estimation method

and the AR model order, Durbin's method for MA estimation can give accurate outcomes, with all the zeros inside the unit circle [1].

Different AR methods and intermediate AR orders have been used in Durbin's method [12], [15], [16]. The best value to be chosen for the intermediate AR order has been investigated [15]. The simulations showed that the best AR order is the one that gives the best approximation for the sum of the squared parameters of the extended AR model [16]. Those simulations used the method of Burg [17] to estimate the AR parameters. The popular YW method of AR estimation is biased [18], and its use is strongly discouraged [19]. It has been concluded that the results of the method of Durbin can be close to the minimum achievable Cramér–Rao bound [16]. The other authors used as the intermediate AR order four times the number of MA parameters that have to be estimated [12]. They and many other researchers reached the wrong conclusion that the method of Durbin gives rather poor results and that the other MA methods should be developed [12]. The conference paper [16] clearly showed that some of the poor results of Durbin's method in the examples in the literature are due to the wrong choices of the intermediate AR order.

This paper prescribes the AR estimation method and the AR order that have to be used in the modified Durbin method for MA estimation [16]. A new derivation and theoretical support is given for the choices that have to be made. It gives an unambiguous answer to the quality of the modified Durbin method, asymptotically and in finite samples. Examples from the literature [12] where Durbin's method seemed to be inaccurate have already been used to investigate the influence of the method of AR estimation and of the intermediate AR order [16]. Measures for the quality of spectra are discussed. An objective error measure that is based on the logarithm of the spectral density is used to compare the estimated MA models with the minimum achievable Cramér–Rao bound. If the accuracy is close to that bound, no better methods can be found. The MA processes with a range of zeros at equal radii show the performance. This example is known to be a severe test for MA estimation if the zeros are close to the unit circle.

II. TIME-SERIES MODELS

A parametric spectral or autocorrelation estimator is computed as a function of the estimated parameters of a time-series model. The time-series theory has three different model types, i.e., AR, MA, and combined ARMA. An ARMA(p, q) process x_n can be written as [1], [6], [7]

$$x_n + a_1x_{n-1} + \dots + a_px_{n-p} = \varepsilon_n + b_1\varepsilon_{n-1} + \dots + b_q\varepsilon_{n-q} \quad (1)$$

where ε_n is a purely random white noise process with zero mean and variance σ_ε^2 . This ARMA(p, q) process becomes AR for $q = 0$ and MA for $p = 0$. In theory, any stationary stochastic process with a continuous spectral density can exactly be written as an AR(∞) process. This is the theoretical fundament for Durbin's MA algorithm [13]. In practice, only finite-order

models can be used, and no examples have been found where this is not sufficient. The roots of the AR and MA polynomials $A(z)$ and $B(z)$ are denoted, respectively, by the poles and zeros of the ARMA(p, q) process as

$$\begin{aligned} A(z) &= 1 + a_1z^{-1} + \dots + a_pz^{-p} \\ B(z) &= 1 + b_1z^{-1} + \dots + b_qz^{-q}. \end{aligned} \quad (2)$$

Here, z is known as the shift operator [7]. The processes are called stationary if all the poles are strictly within the unit circle, and they are invertible if all the zeros are within the unit circle.

The parametric power spectrum $h(\omega)$ of the ARMA(p, q) model is computed for $-\pi < \omega \leq \pi$ with [1], [6], [7]

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

All the lags of the infinitely long true autocorrelation function of an AR(p) process are determined by p true AR parameters with the YW relations that are given by

$$\begin{aligned} \rho(k) + a_1\rho(k-1) + \dots + a_p\rho(k-p) &= 0, \quad k \geq 1 \\ \rho(-k) &= \rho(k) \end{aligned} \quad (4)$$

where the normalized autocorrelation $\rho(k)$ at lag k is the expectation $r(k)/r(0)$ of $x_n x_{n+k}/\sigma_x^2$. The Levinson–Durbin recursion is a computationally efficient algorithm to derive the p AR parameters from the first p YW equations with the first p autocorrelations [6]. Equation (4) is used for indexes greater than p to extrapolate the autocorrelation function. The length with nonzero values of the autocorrelation function of an AR(p) process is infinite [7].

The autocorrelation function of an MA(q) process has the finite length q [7]. The parametric description is given by

$$\begin{aligned} \rho(k) &= \sum_{i=0}^{q-k} b_i b_{i+k} / \sum_{i=0}^q b_i^2, \quad 0 \leq k \leq q \\ \rho(k) &= 0, \quad k > q \\ \rho(-k) &= \rho(k). \end{aligned} \quad (5)$$

Furthermore, explicit parametric expressions for the autocorrelation function of an ARMA(p, q) process have been derived [1].

The lagged-product estimator for the autocorrelation function of N observations x_n is for lag k given by

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

The lagged-product estimator for the autocovariance has a triangular bias $1 - k/N$ of

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

None of the lagged-product estimates (6) is a statistically efficient estimator for the autocorrelation function of an MA process [3]. Moreover, the lagged-product estimates are not sufficient statistics, which means that they do not contain all the relevant information in the data with respect to the MA parameters [20]. Theoretically, this means that using the lagged-product estimates (6) to estimate the MA parameters [4] cannot be efficient either. None of the many attempts to use them as basic ingredients for an MA estimator has produced an efficient result in practice [10]. This fact and the observed poor finite-sample behavior of the maximum likelihood MA estimates [8] are good reasons to investigate the possible improvements of Durbin’s method [13] that is based on the AR estimates, which have no theoretical problems with statistical efficiency or sufficiency.

III. SPECTRAL ACCURACY MEASURES

Two classes of spectral measures can be distinguished, i.e., absolute and relative. The absolute measures are based on the spectral density itself and the relative measures on its logarithm. The relative measures are related to the prediction error of the time-series models and to the cepstral distance between the autocorrelation functions [1]. The absolute spectral errors have no special theoretical background for the discrete-time signals. They are related to the sum of the squared differences between the spectral and autocorrelation estimates.

The absolute errors are concentrated in strong parts of the spectral density. The difference of between 1 and 1.01 contributes the same to an absolute measure as the difference of between 0.001 and 0.011. Relative errors attribute weight to all spectral errors, absolute measures primarily to errors in strong parts. Substituting the value zero for the weakest spectral parts produces a small absolute error but, at the same time, an infinite relative error. The continuous-time processes have an infinite frequency range, with a decreasing spectral density at high frequencies. In that case, the absolute measures could be favorable because they give emphasis to the interesting low-frequency part.

Generally, the estimation in time-series models attempts to minimize the squared residuals in some sense. Roughly speaking, that is realized if the correlation between the residuals is removed. This implies that the spectrum of the residuals is whitened by the time-series estimation over the whole discrete-time frequency range. The weak and strong spectral ranges are equally important, which requires a relative quality measure.

The spectral distortion (SD) is a *relative* integral spectral error measure that has been defined as [1]

$$\begin{aligned} \text{SD} &= \frac{0.5}{2\pi} \int_{-\pi}^{\pi} \left\{ \ln h(\omega) - \ln \hat{h}(\omega) \right\}^2 d\omega \\ &= \frac{0.5}{2\pi} \int_{-\pi}^{\pi} \left\{ \ln \frac{h(\omega)}{\hat{h}(\omega)} \right\}^2 d\omega \end{aligned} \tag{8}$$

where h is the true spectral density, and \hat{h} denotes the estimated spectral density. In the weak spectral parts, the SD is much more informative than the average *squared* spectral error, but this latter measure is still often used for discrete-time processes in the literature [12].

The usual accuracy measure for the time-series models is the squared error of prediction PE, which can be computed by using the estimated parameters to predict a fresh realization of x_n . The fresh or new data are the data that have not been used to estimate the parameters. Both the true and the estimated parameters have *a priori* known or computed values in the simulations. For those special situations of the PE, a theoretical derivation has been given [1]. The PE can be computed with (3) without generating new data by using [1]

$$\begin{aligned} \text{PE} &= \frac{\sigma_\varepsilon^2}{2\pi} \int_{-\pi}^{\pi} \frac{h(\omega)}{\hat{h}(\omega)} d\omega \\ &= \frac{\sigma_\varepsilon^2}{2\pi} \int_{-\pi}^{\pi} \left| \frac{B(e^{j\omega})}{A(e^{j\omega})} \right|^2 \left| \frac{\hat{A}(e^{j\omega})}{\hat{B}(e^{j\omega})} \right|^2 d\omega \end{aligned} \tag{9}$$

where \hat{A} and \hat{B} denote the estimated time-series polynomials. If the quotient of h and \hat{h} is close to 1, then the logarithm of $1 + \delta$ in (8) can be approximated by δ . This indicates that there is a strong relation between (8) and (9). The differences are mainly in scaling and an additional constant. In particular, for models with small disturbances, the use of PE and SD is almost equivalent.

This PE measure can be used to compare the accuracies of the AR and MA models of the same data by substituting $\hat{A} = 1$ if the MA models are considered and $\hat{B} = 1$ for the AR models. The model error ME has been defined by scaling the PE, subtracting a constant, and multiplying with the sample size as

$$\text{ME} = \text{ME} \left[\frac{\hat{B}}{\hat{A}}, \frac{B}{A} \right] = N \left(\frac{\text{PE} - \sigma_\varepsilon^2}{\sigma_\varepsilon^2} \right). \tag{10}$$

The arguments of the ME are left if no confusion is possible. The minimum value of the expectation of the ME for the efficiently estimated unbiased models of orders p' and q' greater than or equal to the true ARMA orders p and q is equal to $p' + q'$. The variance of each estimated parameter contributes at least 1 to the minimal ME expectation.

The Cramér–Rao lower bound for the parameters of the time-series models can easily be expressed in the relative ME

measure. It only applies to unbiased models. This requires, for an estimated ARMA(p' , q') model, an ARMA(p , q) process where $p' \geq p$ and $q' \geq q$. Therefore, the Cramér–Rao lower bound for the ME of an MA(q) process is given by q . It follows with (9) that this Cramér–Rao bound is relevant for the spectral accuracy. The invariance property of the maximum likelihood estimation is the reason that the unbiased model with the best parameter accuracy is at the same time the best model for the spectral and autocorrelation accuracies. No derivation of a lower accuracy bound expressed in an absolute measure is known.

IV. THEORY FOR INTERMEDIATE AR ORDER

The influence of the AR model order M on the accuracy of the MA models in Durbin's method has been studied before, with the simulations as a function of M [1], [15]. The true order of the AR process $C_\infty(z)$ that is exactly equivalent with an MA(q) process is always ∞ . The infinite number of parameters c_i can in principle be computed with the polynomial representation of (2) from the equivalence

$$C_\infty(z) = 1/B_q(z). \quad (11)$$

The infinite order of $C_\infty(z)$ gives the exact representation if the polynomial $B_q(z)$ is known.

This is not applicable if the AR parameters have to be estimated from the measured data. Generally, the best approximating time-series model depends on the intended application [21]. Furthermore, depending on the purpose, at least two different optimal model orders can be defined for the long *estimated* polynomial approximations in the linear regression theory [22]. The first is the order to obtain the best possible predictions, and the other is the order with the best parameter accuracy. For many applications, the order with the most accurate prediction of future observations will be an optimal choice. For the time-series models, it is the model order with the smallest PE or ME, which would be selected by the order selection criteria in ideal circumstances. With the spectral accuracy measures (8) and (9), it follows that a model with a small PE has at the same time a good relative spectral accuracy.

However, for the application in Durbin's method, it is the equivalence of the polynomials $B_q(z)$ and $C_\infty(z)$ in (11) that is important. This completely depends on the parameter values in the estimated polynomials. Therefore, the best order for an estimated long AR model in Durbin's method is the order with the best parameter accuracy.

Suppose that the true parameters c_i of the polynomial $C_\infty(z)$ in (11) are regularly decreasing for higher orders. In that case, it is easy to determine which model order L can be expected to give the best predictions with the estimated parameters \hat{c}_i . It is the highest order for which

$$c_L^2 \geq 1/N. \quad (12)$$

According to the mathematical expectations, that order L will be selected from the measured data with the Akaike information criterion (AIC) for order selection [23]. The estimation variance of the small parameters is equal to $1/N$ [1]. The expectation of

the squared value of an estimated parameter \hat{c}_L^2 is given by the true value c_L^2 with the additional variance contribution $1/N$. Parameters with squared true values greater than $1/N$ are to be included. The penalty factor for the additional parameters is 2 in AIC [23]. This agrees with selecting parameters for which the estimated values obey

$$\hat{c}_L^2 \geq 2/N \quad (13)$$

with an additional $1/N$ in comparison with (12) for the variance. This derivation uses the asymptotic expression for the variance of the estimated parameters. If the order L is greater than $N/10$, it would become more accurate to apply the finite-sample theory [1], [18].

In Durbin's MA method, the theoretical equivalence (11) between the MA(q) and AR(∞) models is based on the parameter values and not on the model predictions. Therefore, it is evident that the model order with the best parameter approximation is the best choice in this case. This order can be computed for a process $C_\infty(z)$ with known true parameters [22]. Using asymptotical theory, it is the lowest order for which the remaining sum of the squared true parameters is less than $1/N$. Therefore, that order depends on N . The parameter accuracy of an estimated AR model $\hat{C}_M(z)$ of the MA process is defined as the infinite sum of the squared differences between the true and estimated parameters. For the model $\hat{C}_M(z)$, only the first M parameters \hat{c}_i are estimated, and the higher-order parameter estimates are set equal to zero. This gives the squared parameter error (SPE)

$$\text{SPE} [\hat{C}_M(z)] = \sum_{i=0}^M (c_i - \hat{c}_i)^2 + \sum_{i=M+1}^{\infty} c_i^2. \quad (14)$$

The best order M for $\hat{C}_M(z)$, with the smallest SPE for a given true MA polynomial $B_q(z)$, has theoretically been derived, and it is asymptotically given by [22]

$$\sum_{i=M+1}^{\infty} c_i^2 < 1/N \quad (15)$$

where N is the number of observations. It has been verified in the simulations for a variety of MA processes with different sample sizes that (15) determines the order with the best parameter accuracy SPE of (14) [1]. Furthermore, it was also the best intermediate AR order for the modified method of Durbin for a given MA(q) process [1]. This is precise if the order M is less than $N/10$, because the asymptotical theory for the variance of the estimated parameters is accurate then [1].

It is easy to approximate the sum of squares in (15) for the given true values of c_i . However, the asymptotical variance of each estimated parameter \hat{c}_i is in the theory given by $1/N$ for the small true values of c_i . The variance of one estimated parameter above order M is equal to the total limiting sum of all the squared true values in (15). This critical value of $1/N$ for omitting an infinite number of very small parameters is equal to the estimation variance of each individual estimated parameter. Therefore, it will not be possible to derive the limiting value M in (15) for the sum of squares from the estimated parameters,

which all already contribute their variance $1/N$ to the sum of squares.

To find out what the best choice could be, different intermediate AR orders have been used in the simulations to estimate the MA(q) model [1], [15]. The result was that a rather flat variation of the MA(q) accuracy is obtained as a function of the AR order, which is around the optimal AR order M that gives the best accuracy. However, taking much lower or much higher intermediate AR orders instead of M considerably gives less accuracy in the finally estimated MA(q) model [1]. The ME can then become much higher, particularly if too low orders are used. A comparison of (12) and (15) shows that the best order M cannot be less than K , but it can be equal or much higher.

A compromise M^* for the best AR order in Durbin's method for practical data has been suggested before [1]. It should depend on N and be greater than the best selected predicting AR order that is denoted by K and also greater than the MA order q . However, many authors used a fixed model order, e.g., $M^* = 4q$ [12]. The best AR order theoretically depends on N in (15). This casts doubt on many conclusions in the literature about the quality of Durbin's MA method. Fixed model orders like $M^* = 4q$ can at best be optimal in a certain example for one specific value of N . Other sample sizes give other values of M in (15). All of the rules for the choice of the intermediate AR order that uses fixed values for M^* are inevitably suboptimal. Moreover, the use of (15) supposes the *a priori* knowledge of the true MA order q , and it cannot be used for the analysis of the MA data with an unknown model order.

It is known from the simulations that the accuracy of the MA(q') model is not very sensitive to the intermediate AR order [1], [15]. From (12) and (15), it follows that the AR order must always at least be as high as the best AR order for the prediction. Furthermore, it would be difficult or impossible to estimate q' MA parameters from less than q' AR parameters. Based on trying different values like $K + q'$, $2K + q'$, and $3K + q'$, the value

$$M_{\text{sw}} = 2K + q' \quad (16)$$

has been suggested for the intermediate AR order [1]. It is under all of the circumstances at least as high as the order of the estimated MA model and higher than the AR order. The order M_{sw} is called a sliding window order because it varies with the sample size and with the order q' of the MA model that is computed.

The best order K for the prediction with an AR model can be selected with a variety of selection criteria, but the combined information criterion (CIC) is preferred because of its performance in finite samples [1]. This practical value (16) for M_{sw} is used in a freely available automatic Matlab program for the spectral estimation with the time-series models [24]. The AR models are computed with the method of Burg [17], and the order is selected with CIC. For N that is less than 2000, the program computes the MA models from the order $q' = 1$ until $N/5$, with the sliding window order given by (16). For a greater N , the maximum MA candidate order is limited to 400. Afterward, it automatically selects the order of the best-fitting MA model without the interference of the user.

V. MODIFIED DURBIN ALGORITHM

The original derivation of Durbin [13] used the maximum likelihood theory for the normally distributed data to derive the MA algorithm. Here, the ME will be used for a new formulation of the modified Durbin method. Define $\hat{C}_{M_{\text{sw}}}(z)$ as the polynomial with the estimated parameters \hat{c}_i of a high-order AR(M_{sw}) model of the data. It will be considered as the true AR process whose spectrum should be approximated by an MA(q') model. Hence, the estimates of the method of Durbin for the MA(q') parameters for an arbitrary MA order q' can now be written as the solution of (9) that gives the smallest value for

$$\text{PE} = \frac{\sigma_\varepsilon^2}{2\pi} \int_{-\pi}^{\pi} \left| \frac{1}{\hat{C}_{M_{\text{sw}}}(e^{j\omega})} \right|^2 \left| \frac{1}{B_{q'}(e^{j\omega})} \right|^2 d\omega. \quad (17)$$

The smallest possible value for the monic polynomials $\hat{C}_{M_{\text{sw}}}(z)$ and $B_{q'}(z)$ is given by σ_ε^2 [20], which is found for the denominator product

$$\hat{C}_{M_{\text{sw}}}(z)B_{q'}(z) = 1. \quad (18)$$

The equality in (18) can only exactly be obtained for the infinite model orders. It can be approximated either by letting $B_{q'}(z)$ be an approximation of $1/\hat{C}_{M_{\text{sw}}}(z)$ or by looking for an approximation $1/B_{q'}(z)$ of $\hat{C}_{M_{\text{sw}}}(z)$. The first formulation computes the MA parameters for a given long AR model. This is a difficult and nonlinear calculation. The second formulation can be interpreted as looking for the best AR model with the parameter polynomial $B_{q'}(z)$ to approximate a given long MA model with the parameter polynomial $\hat{C}_{M_{\text{sw}}}(z)$. The roles of the AR and MA polynomials are interchanged here for computational reasons. This interchange is possible because the product of the polynomials should only be as close as possible to 1 in (18). It is well known that the computation of the AR parameters is a simple linear operation, which is in contrast with the MA parameters that require nonlinear iterations [4]. Therefore, the second formulation is preferred. In terms of the ME of (9), the solution can be written as

$$\hat{B}_{q'}(z) = \arg \min_{\tilde{B}_{q'}(z)} \left\{ \text{ME} \left(\frac{1}{\tilde{B}_{q'}(z)}, \frac{\hat{C}_{M_{\text{sw}}}(z)}{1} \right) \right\}. \quad (19)$$

This can computationally be interpreted as looking for an AR(q') model that gives the best fit to the MA(M_{sw}) model with the parameters of $\hat{C}_{M_{\text{sw}}}(z)$. Those parameters \hat{c}_i are substituted in an adapted version of the MA autocorrelation relation (5) as

$$\hat{\rho}(k) = \sum_{i=0}^{M_{\text{sw}}-k} \hat{c}_i \hat{c}_{i+k} / \sum_{i=0}^{M_{\text{sw}}} \hat{c}_i^2, \quad 0 \leq k \leq q'. \quad (20)$$

Finally, the q' MA parameter estimates are found with a modified YW polynomial (4) as the solutions of

$$\begin{aligned} \hat{\rho}(k) + \hat{b}_1 \hat{\rho}(k-1) + \dots + \hat{b}_{q'} \hat{\rho}(k-p) &= 0, & 1 \leq k \leq q' \\ \hat{\rho}(-k) &= \hat{\rho}(k). \end{aligned} \quad (21)$$

This algorithm guarantees that the roots of the estimated MA polynomial are always within the unit circle [1].

The modified Durbin method has the following prescribed stages for AR estimation, AR order selection, and estimation of the MA parameters of an arbitrary order q' .

- 1) Compute the AR models of orders $0, 1, \dots, N/2$ with Burg's method [17]. The highest AR model order may be limited to 1000 for $N > 2000$.
- 2) Select the AR order K with the minimum of the order selection criterion CIC [1].
- 3) Take as the intermediate AR order $M_{sw} = 2K + q'$.
- 4) Use (20) and (21) to compute the MA(q') parameters.

This modified Durbin method has been developed and tested in many simulation studies and with real-life data [1], [15]. It is numerically robust in finite samples and has excellent properties if more data are available. Then, the results are always close to the Cramér–Rao lower bound, which is characterized by $ME = q$ for a true MA(q) process. No counter examples have been found where the modified Durbin method performed poorly as long as the required sample size M in (15) is less than about $N/4$. M may become greater if some zeros of the MA process are close to the unit circle for the given sample size. The influence of N will be investigated in the next section.

In practice, the true MA order q is not known *a priori*. It must be selected from the data. Then, several values of q' are used to compute the MA candidate models. The AIC can be used to select a model order for the MA data. However, an asymptotical generalized information criterion (GIC) with penalty function 3 for order selection gives better results for MA data [1]. This criterion is obtained by changing the constant 2 in the AIC criterion [23] to the value of 3.

So far, the modified Durbin method has not been used by the other authors, and all the remarks about the quality of Durbin's method in the literature do not apply to the new modified Durbin method. The modified version has been compared [16] with the two newly developed MA methods [12] for a critical example where the original method of Durbin [13] performed poorly because wrong choices were applied. It has been shown that the results of the modified Durbin method were rather close to the Cramér–Rao bound [16]. Those results will not be repeated here, but new examples will be evaluated. Simulations are required to verify the applicability of the approximations that have been applied in the derivation of the modified Durbin algorithm.

VI. SIMULATIONS WITH ZEROS CLOSE TO UNIT CIRCLE

Processes with zeros that are close to the unit circle are challenging examples for the MA estimation methods [12]. Poor results for Durbin's method have been reported for some examples [12], but a better choice of the order of the AR polynomial in the modified Durbin method considerably improved the accuracy. It has been shown in the MA(3) processes that using the intermediate AR order M , as defined in (15), produced the best results for the MA models [16].

The Monte Carlo simulations will be reported here with an MA(5) process where the five zeros are all at the same radius $r = 0.95$. The data are generated with a MA(5) poly-

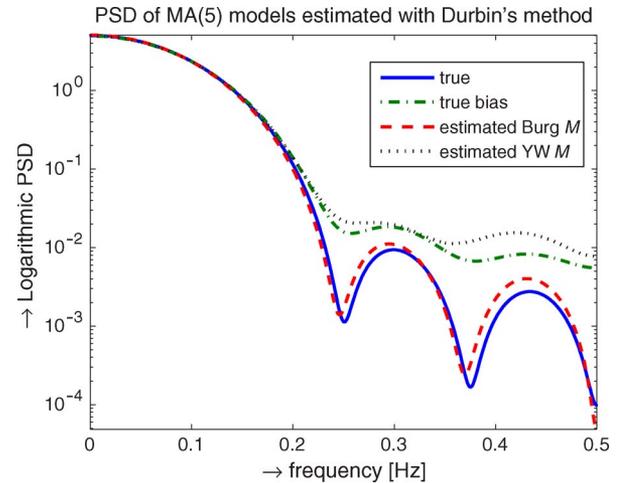


Fig. 1. True spectrum, true spectrum with the triangular YW bias, and two spectra estimated with the Durbin method from 100 MA(5) observations with $r = 0.95$. The AR estimates of the Burg and YW methods have been used, with the value $M = 35$ computed with (15).

nomial that is constructed from the zeros at $r \exp \pm(j\pi/2)$, $r \exp \pm(j3\pi/4)$, and $r \exp - (j\pi)$.

First, the accuracies of the MA models of the MA(5) data have been studied for the intermediate YW [6] and Burg [17] estimates for the AR parameters. High-order AR models $\hat{C}_{M_{sw}}(z)$ are the basic elements for Durbin's method. It may be expected that more bias in the AR models has a negative influence on the accuracy of the resulting MA models. The Burg AR models have a negligible bias, but the YW models can have a significant bias if the zeros for the MA data are not far from the unit circle. The results of using YW and Burg estimates for the intermediate AR models were only close as long as r is smaller than 0.5 for N greater than about 100. Otherwise, Burg gives better results. For $r = 0.9$, N should be greater than 7500 to have no significant bias contributions in the YW estimates. The AR YW estimates are asymptotically unbiased, and the differences between the AR estimation methods of Burg and YW will eventually become smaller and disappear for increasing sample sizes [18] for all values of r .

Fig. 1 shows the true and estimated spectra of the MA(5) models. The AR models estimated with the YW method have a strong bias in the weak high-frequency part of the spectrum for this example with $r = 0.95$. The true biased spectrum is computed as the Fourier transform of the autocorrelation function multiplied with the triangular bias $1 - k/N$ (7). The accuracies for the MA(5) models are the following: true biased spectrum without estimation, $ME = 44.3$; using AR Burg M , $ME = 3.9$; and finally using AR YW M , $ME = 66.1$. The estimated model with the YW method is close to the true biased expectation because both have the triangular bias. In this example, the MA(5) spectrum obtained with an intermediate AR YW model is poor. This is clearly visible in the logarithmic plot, but it would hardly be noticed in a linear presentation of the power spectral density. Burg M is close to the Cramér–Rao bound, which is 5 for an MA(5) process. Due to statistical variations, a single realization can give an ME value that is even smaller than the expectation 5. It has been verified that the estimated ARMAse1 model [24], which was obtained from

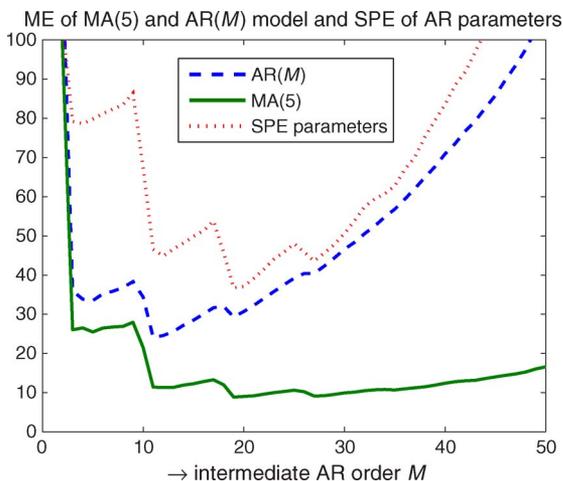


Fig. 2. Average ME of the AR models of order M , MA(5) models estimated from the AR(M) model, and scaled SPE for 1000 runs of 100 observations of the MA(5) process with $r = 0.95$ for all zeros.

the data without using any *a priori* knowledge, is close to the Burg M and to the true spectrum. The ME was 5.4 for the MA(5) model that was automatically selected by the ARMA Sel algorithm from many MA and AR candidate models [1].

The large bias for the AR YW estimates has been described before for the MA data [16] and for the AR data [18]. Furthermore, the bias asymptotically disappears, and it will become negligible for very large sample sizes or for a small radius r . However, in those situations, the accuracy of the final MA model is about the same as the accuracy of the model obtained from the Burg AR estimates. Summarizing, the Burg method for the intermediate AR model can give the same accuracy as the YW method for the finally obtained MA models, or Burg gives a better accuracy. It is never worse. Therefore, YW is not further studied, and Burg is used for all the data.

Fig. 2 used the same process parameters as in Fig. 1. The ME of the AR models reaches its minimum at order 11, which is the outcome of (12) for this example. The value of the AR order with the minimum of the SPE as computed with (14) would be $M = 35$. There is a slight local bend in the line at that order. However, the order should be less than about $N/10$ to allow the asymptotical theory to be accurate. The finite-sample result in Fig. 2 gives the best MA(5) accuracy for the AR order 18 for $N = 100$. It is clear that the accuracy of the MA models is better than that of the AR models because the ME of the MA models is smaller. Models of the MA type are required to compute the accurate spectra for this type of data. The accuracy of the MA models is poor if the AR order is too low, and it becomes slightly worse if too high AR orders are used. The minimum of the SPE and the MA(5) accuracy are reached for the same order 18. Both curves have a local minimum at order 26. Using the sliding window order M_{sw} , as defined in (16), mostly results in AR order 27 for the MA(5) model, which is a good value in this example. The results like in Fig. 2 heavily depend on the sample size and the radius of the zeros. They demonstrate the importance of a good choice of the AR order in Durbin’s MA method. For larger sample sizes, the accuracy ME of the MA(5) model will be about 5 for a rather wide

range of intermediate AR orders, and the appearance of the MA accuracy as a function of the AR order will become smoother.

VII. MA(5) SIMULATIONS

Monte Carlo simulations have been carried out with the MA(5) processes, where four zeros are at the same radius r , and the fifth zero is at the constant -0.95 . The last fixed radius is chosen rather close to the unit circle to make sure that the presence of at least one significant zero gives the overall spectrum an MA character for every value of r . The MA(5) parameters are computed from zeros at $r \exp \pm(j\pi/2)$, $r \exp \pm(j3\pi/4)$, and at -0.95 . The MA polynomials $B(z)$ are invertible if $r < 1$. All of the data have been generated with $\sigma_\varepsilon^2 = 1$. The radius r and the sample size N have been varied. The behavior of the estimates as a function of the AR model order is rather irregular in Fig. 2. Therefore, the results for the fixed AR and MA model orders that give the optimal accuracy will be compared with the results that are completely automatically obtained with the AR and MA orders selected from the available observations. This will give an indication about the additional error that is made if no *a priori* information whatsoever is used in the spectral estimation. The model orders and the model type are selected from the given data with the automatic algorithm ARMA Sel [24].

The average results for 100 observations of the fixed true order estimated MA(5) processes are presented in Table I, in the columns denoted Burg. As a comparison, the ME of the completely automatically estimated and selected ARMA Sel model is given [24]. The ARMA Sel model is estimated and automatically selected from many AR, MA, and ARMA candidates with the ARMA SA Matlab toolbox [24]. The intermediate AR order, the model type, and the model order of the ARMA Sel model are automatically selected from the data at hand without any influence of the data analyst. In most runs, the MA(5) model with the true model type and order was selected from the 50 AR, 20 MA, and ten ARMA candidate models that were computed from the 100 observations in each simulation run. The column-modified Durbin gives the average ME of the modified Durbin method, as described in Section V, including the effect of selection of the MA order. In all of the simulations, the selection from only MA candidates gives a smaller average ME than the free selection from all of the model types with ARMA Sel. The reason is that sometimes other model types are selected for the rather small sample size $N = 100$ that was used here.

The other columns in Table I used the knowledge of the true process to compute the best intermediate AR order M with (13) and L with (12), and to estimate the MA(5) process. In practice, L was computed with the AIC criterion [23], with the penalty factor 1 applied to the truncated true AR models of increasing order and with the parameters of $C_\infty(z)$. This can deal with the irregular behavior of the AR models as a function of the model order, as presented in Fig. 2. The Cramér–Rao bound for an MA(5) process is 5. This is never attained for $N = 100$, but it would be found for much greater sample sizes. The accuracy here is less than twice higher, which is still considered to be reasonably accurate. The results in the column with the Burg

TABLE I
AVERAGE ME IN 1000 SIMULATION RUNS OF THE ARMAse1 MODEL WITH SELECTED ORDER AND TYPE, THE MODIFIED DURBIN MODEL WITH SELECTED MA ORDER, AND THE MA(5) MODELS ESTIMATED FROM THE BURG AR(M), AR(L), AND AR($N/2$) MODELS, AS A FUNCTION OF THE RADIUS r OF THE MA(5) PROCESS FOR $N = 100$. THE FINAL COLUMNS GIVE THE VALUES M OF (15) AND L OF THE BEST PREDICTING AR MODEL

r	ARMAse1	Modified Durbin	Burg M	Burg L	Burg $N/2$	M	L
0.25	9.3	8.6	9.5	10.9	22.8	23	9
0.5	11.4	10.7	9.7	11.1	23.1	23	8
0.7	13.0	12.0	9.3	10.6	22.9	21	6
0.8	12.1	10.0	8.3	9.3	20.7	20	4
0.9	13.8	10.9	7.9	7.4	19.9	20	11
0.95	17.3	13.8	11.1	11.9	17.0	35	11
0.98	24.5	20.5	31.6	19.2	40.6	67	12

TABLE II
AVERAGE ME IN 1000 SIMULATION RUNS OF THE ARMAse1 MODEL WITH SELECTED ORDER AND TYPE, THE MODIFIED DURBIN MODEL WITH SELECTED MA ORDER, AND THE MA(5) MODELS ESTIMATED FROM THE BURG AR(M), AR(L), AND AR($N/2$) MODELS, AS A FUNCTION OF N . FURTHER, M AND L ARE GIVEN. THE RADIUS r OF THE MA(5) PROCESS IS 0.95

N	ARMAse1	Modified Durbin	Burg M	Burg L	Burg $N/2$	M	L
50	20.5	17.0	22.2	11.5	27.8	27	11
100	18.4	15.6	11.8	12.2	18.5	35	11
200	11.5	9.9	8.2	9.2	22.0	42	19
500	8.6	7.8	6.2	8.1	24.5	51	27
1000	6.8	6.7	5.8	6.8	28.5	58	35
2000	6.2	6.2	5.3	7.6	15.9*	61	36
5000	6.2	6.1	5.2	6.8	15.8*	74	44

estimates for the long AR(M) model are almost always the most accurate, except for $r = 0.9$, where Burg L was slightly better. This type of small sample effects can be expected if the curves as a function of the AR order are as irregular as in Fig. 2. The result for $r = 0.98$ shows that the analysis in Section IV loses its theoretical accuracy if the best parameter order is greater than $N/2$. However, the results for the modified Durbin column with selected MA order stay close to the best MA estimates that are found with the prescribed model order. It is clear that this example requires more observations to obtain accurate spectral estimates.

The results in Table I demonstrate the finite-sample behavior of the modified Durbin method. It includes the effects of automatically choosing the AR order from the data and of selecting the MA order. Table II gives the results for increasing the sample sizes.

For $N > 500$, the fixed-order MA(5) column of Burg M is close to the Cramér–Rao bound 5. In addition, it has been verified that it converges in many different examples with various true orders. It is clear that use of the order M gives better results than L or $N/2$. In the last two rows, the maximum AR order in the column $N/2$ has for computational reasons been limited to 661 and 740 for $N = 2000$ and 5000, respectively. This explains the lower values in that column and is indicated by a * sign in Table II. The ME results for ARMAse1 and the

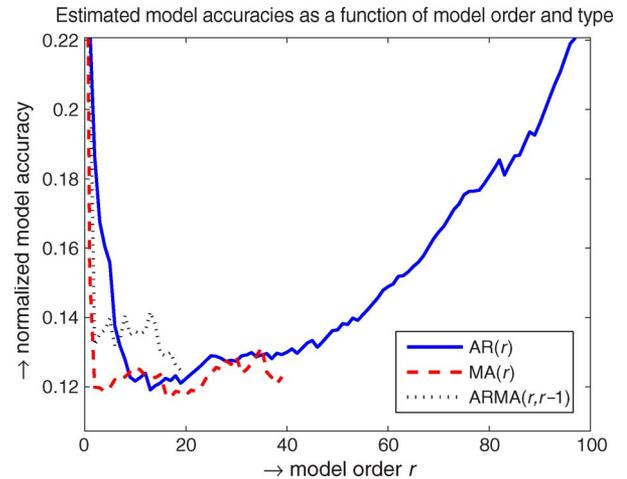


Fig. 3. Estimated mean squared errors of prediction of the 197 chemical process concentration readings [25] after two times differencing of the data. The MA models of low orders are most suitable for these data, but the MA models with about 20 parameters also give good models. The ARMA and low-order AR models are less accurate here.

modified Durbin converge to the same values. The reason is that all of the AR and ARMA models have too much bias for a greater N in comparison with the MA models, and they are never selected. Therefore, selection from the MA candidates only and from all the possible candidates converges to the same average ME value. The difference between the modified Durbin column and the Burg M column can be attributed to the cost of order selection. This cost is about 1 if only candidates of the true model type and too high model order are serious competitors [1]. Hardly any difference can be seen between the true spectrum and the estimated spectra for N greater than 1000. The asymptotical loss in the average ME value due to selecting the AR and MA orders from the data is about 1, independent of the true MA process. This has been verified in various examples with true orders of between 1 and 13.

VIII. CHEMICAL CONCENTRATIONS

The modified Durbin algorithm has been tested on real-life data. An example with 197 chemical process concentration readings from the literature [25] is presented here. The ARMAse1 [24] selected the ARMA(2,1) model for the data, the MA(1) model was at the boundary of statistical significance for the data after differencing once, and the MA(4) model was selected after twice differencing of the data. Taking the difference $x_n - x_{n-1}$ is often used to improve the stationarity of the economic data [25], but it has a strong influence on the spectrum. The estimated model accuracies of all the models that have been estimated by the ARMAse1 program for the data after differencing twice are shown in Fig. 3. The squared error of prediction of all the estimated models has a strong relation with the selection of the model order and type in ARMAse1 [1].

The MA(4) model is selected here, but Fig. 3 shows that many other MA models are also good candidates for the data, with a small estimated model accuracy. The AR(13) model is the best of all the AR candidates, but it is not selected here because it requires more parameters to obtain almost the

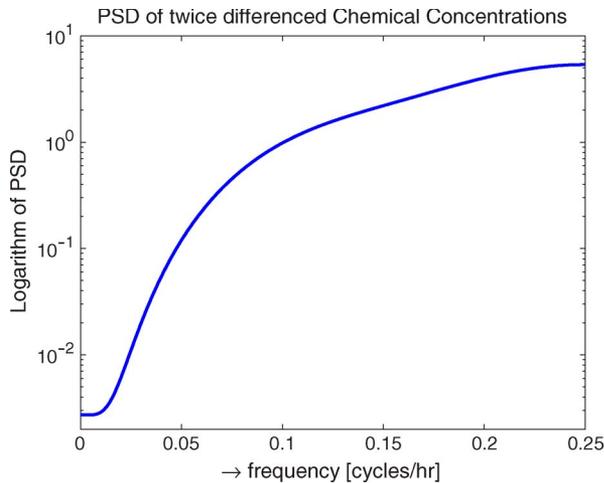


Fig. 4. Estimated power spectral density of the MA(4) model of the 197 chemical process concentration readings [25] after two times differencing of the data. The sampling period was 2 h.

same accuracy. The ARMA models are less accurate for these data. Although the AR models are often the most accurate for practical data, this example shows that sometimes the MA models can be preferable.

The power spectral density of the MA(4) model is shown in Fig. 4. Taking the differences of data removes most of the low-frequency part of the spectrum. This explains why the spectrum in Fig. 4 increases with the frequency.

IX. CONCLUDING REMARKS

To obtain good results with Durbin’s MA method, an intermediate AR model has to be estimated with Burg’s method. The intermediate AR model order must be high enough and can automatically be selected for the given data. The theoretical optimal AR order can be computed for the known MA processes as the order with the best parameter accuracy.

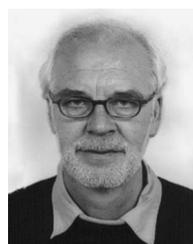
The modified Durbin method prescribes the use of twice the selected AR order plus the number of MA parameters that has to be estimated as the intermediate sliding window order. All the unfavorable judgments of Durbin’s method in the literature that are known to the author are based on the wrong choices for the intermediate AR order. No examples have been found or encountered with a poor performance if the AR order is properly chosen. The final MA results are always close to the Cramér–Rao bound for the MA model of the true order if enough data are available for the estimation.

It is very unfortunate that the simple and very robust MA method of Durbin, with the proper implementation, is not generally known to be close to the Cramér–Rao bound. No better methods for the MA estimation in finite samples have been described in the spectral estimation literature.

REFERENCES

[1] P. M. T. Broersen, *Automatic Autocorrelation and Spectral Analysis*. London, U.K.: Springer-Verlag, 2006.
 [2] P. M. T. Broersen, “Historical misconceptions in autocorrelation estimation,” *IEEE Trans. Instrum. Meas.*, vol. 56, no. 4, pp. 1189–1197, Aug. 2007.

[3] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
 [4] G. Wilson, “Factorization of the covariance generating function of a pure moving average process,” *SIAM J. Numer. Anal.*, vol. 6, pp. 1–7, 1969.
 [5] J. E. H. Davidson, “Problems with the estimation of moving average processes,” *J. Econom.*, vol. 16, no. 3, pp. 295–310, Aug. 1981.
 [6] S. M. Kay and S. L. Marple, “Spectrum analysis—A modern perspective,” *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, Nov. 1981.
 [7] M. B. Priestley, *Spectral Analysis and Time Series*. London, U.K.: Academic, 1981.
 [8] D. R. Osborn, “Maximum likelihood estimation of moving average processes,” *Ann. Econ. Soc. Meas.*, vol. 5, pp. 75–87, 1976.
 [9] A. M. Walker, “Large-sample estimation of parameters for moving-average models,” *Biometrika*, vol. 48, no. 3/4, pp. 343–357, 1961.
 [10] E. J. Godolphin, “An invariance property for the maximum likelihood estimator of the parameters of a Gaussian moving average process,” *Ann. Stat.*, vol. 8, no. 5, pp. 1093–1099, 1980.
 [11] E. J. Godolphin and J. G. de Gooijer, “On the maximum likelihood estimation of the parameters of a Gaussian moving average process,” *Biometrika*, vol. 69, no. 2, pp. 443–451, 1982.
 [12] P. Stoica, T. McKelvey, and J. Mari, “MA estimation in polynomial time,” *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 1999–2012, Jul. 2000.
 [13] J. Durbin, “Efficient estimation of parameters in moving-average models,” *Biometrika*, vol. 46, no. 3/4, pp. 306–316, 1959.
 [14] R. P. Mentz, “Estimation in the first-order moving average model through the finite autoregressive approximation,” *J. Econom.*, vol. 6, no. 2, pp. 225–236, Sep. 1977.
 [15] P. M. T. Broersen, “Autoregressive model orders for Durbin’s MA and ARMA estimators,” *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2454–2457, Aug. 2000.
 [16] P. M. T. Broersen, “Accurate estimation of moving average models with Durbin’s method,” in *Proc. IMTC*, Victoria, BC, Canada, 2008, pp. 21–26.
 [17] J. P. Burg, “Maximum entropy spectral analysis,” in *Proc. 37th Meeting Soc. Exploration Geophysicists*, 1967, p. 6.
 [18] P. M. T. Broersen, “Finite-sample bias in the Yule–Walker method of autoregressive estimation,” in *Proc. IMTC*, Victoria, BC, Canada, 2008, pp. 342–347.
 [19] 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 modeling,” *Ann. Nucl. Energy*, vol. 23, no. 15, pp. 1219–1228, Oct. 1996.
 [20] D. S. G. Pollock, *A Handbook of Time-Series Analysis, Signal Processing and Dynamics*. San Diego, CA: Academic, 1999.
 [21] D. F. Findley, “On some ambiguities associated with the fitting of ARMA models to time series,” *J. Time Ser. Anal.*, vol. 5, no. 4, pp. 213–225, Jul. 1984.
 [22] R. R. Hocking, “The analysis and selection of variables in linear regression,” *Biometrics*, vol. 32, pp. 1–49, 1976.
 [23] H. Akaike, “A new look at the statistical model identification,” *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
 [24] P. M. T. Broersen, *ARMASA Matlab toolbox*, 2002. [Online]. Available: www.mathworks.com/matlabcentral/fileexchange
 [25] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day, 1976.



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 the estimated time-series models. This provides a practical and accurate solution for the spectral and autocorrelation analysis

of stochastic data. The software for the automatic estimation of spectra and autocorrelation functions of equidistant random data, for missing data problems, and for irregularly sampled observations is available online. His main research interest is in the automatic and unambiguous identification of the character of stationary random measurement data.