Automatic Identification of Time-Series Models From Long Autoregressive Models

Piet M. T. Broersen and Stijn de Waele

***Abstract*—Identification is the selection of the model type and of the model order by using measured data of a process with unknown characteristics. If the observations themselves are used, it is possible to identify automatically a good time-series model for stochastic data. The selected model is an adequate representation of the statistically significant spectral details in the observed process. Sometimes, identification has to be based on many less than N characteristics of the data. The reduced statistics information is assumed to consist of a long autoregressive (AR) model. That AR model has to be used for the estimation of moving average (MA) and of combined ARMA models and for the selection of the best model orders. The accuracy of ARMA models is improved by using four different types of initial estimates in a first stage. After a second stage, it is possible to select automatically which initial estimates were most favorable in the present case by using the fit of the estimated ARMA models to the given long AR model. The same principle is used to select the best type of the time-series models and the best model order. No spectral information is lost in using only the long AR representation instead of all data. The quality of the model identified from a long AR model is comparable to that of the best time-series model that can be computed if all observations are available.**

***Index Terms*—Autocorrelation, autocovariance function, order selection, parameter estimation, power spectral density, spectral analysis, system identification.**

I. Introduction

TIME-SERIES modeling is a parametric method to estimate the covariance and the power spectral density of stochastic processes. The parameters of the model describe the statistically significant details in the autocovariance function or in the spectral density of the data. Three model types can be used: autoregressive (AR), moving average (MA) and combined ARMA models. No well-established algorithm for practical ARMA spectral estimation has been found with asymptotic theoretical arguments [1]. Maximum likelihood estimates may display spurious peaks as a result of almost canceling pole-zero pairs. The ARMA estimator of Durbin [2] can be used for the separate estimation of the AR and the MA parameters. The quality of the original estimator of Durbin has been improved by using a better order for the intermediate AR model [3] in a new and robust algorithm that will always produce stationary and invertible models. The usual intermediate AR order was too low, while taking it too long also has a negative influence on the accuracy. The improved Durbin estimator enabled the automatic identification of a single time-series model for spectral estimation [4]. For stationary stochastic processes, a freely available automatic time-series algorithm [5] outperforms the modified, windowed, and tapered periodograms that are derived from the fast Fourier transform of the data. Identification with the ARMAsel algorithm [5] includes the automatic selection of the model type, (AR, MA, or ARMA) and the model order from a large number of candidate models.

The ARMAsel algorithm uses already a long AR model during the computation of ARMA and MA parameters, even if the data themselves are available. The measured observations themselves are only used for three purposes in the automatic time series identification algorithm [5]:

- the computation of a long AR model;
- the initial stage in ARMA estimation to separate AR and MA parts;
- the computation of the residual variances required as input for order selection.

The question addressed in this paper is whether the last two points can be evaluated with only the information of a long AR model as reduced statistics input. Then, the automatic selection of model type and model order is also feasible if the data themselves are not available, but only the reduced information of a long AR model. The purpose is to find a reduced statistics estimator with an accuracy that is always close to the best accuracy that can be attained if all data would be available.

The development of a reduced statistics estimator has several reasons. The first may be the reduction of the computing time if the number of observations N is extremely large. Another application is the reduction of the amount of raw information to be transmitted to earth by satellites from outer space, without loss of statistically significant information. Also, the identification of a time-series model for a given spectral density is possible. That was necessary to develop the signal processing for a new satellite to explore the gravity field of the earth [6]. Only the expected power spectral density of the disturbing noise signal can be specified before the launching. Its inverse Fourier transform is an autocovariance function that can be transformed into a long AR model with the Yule–Walker equations and the Levinson–Durbin recursion [7]. That yields the optimal whitening filter for the accurate estimation of the gravity. Further applications are found in missing data problems and in the analysis of segmented data. Only AR models are easily estimated from the data then. Accurate algorithms for MA and ARMA estimation are not available for those applications,

Manuscript received June 15, 2002; revised December 14, 2004.

P. M. T. Broersen is with the Department of Multi Scale Physics, Delft University of Technology, 2600 GA Delft, The Netherlands (e-mail: broersen@tnw.tudelft.nl).

S. de Waele is with the Philips Research Laboratories, 2600 GA Delft, The Netherlands.

Digital Object Identifier 10.1109/TIM.2005.853232

although those models may be the best representation. Furthermore, the statistical analysis of subbands of the frequency range has been made with a reduced-statistics algorithm.

A well-known reduced statistics ARMA estimator is the method of the extended Yule–Walker equations, which used estimated autocovariances to separate AR and MA parts [7]. The performance greatly varies with the process characteristics due to the triangular bias in estimated lagged product autocovariances. Therefore, this method cannot be applied with confidence to data with unknown characteristics. A second known reduced-statistics ARMA estimator is based on an intermediate autoregressive model [8], and a third estimator is based on the impulse response of an intermediate AR model [9]. This paper describes a fourth principle to determine ARMA estimates, based on inverse autocorrelations. Also, for pure MA estimation, Durbin introduced a reduced statistics estimator based on a long AR model [10].

None of the four reduced-statistics ARMA methods gives satisfactory results in one stage. Therefore, the AR parameters obtained from those four different reduced-statistics estimators are used as initial conditions for a second stage ARMA estimator of the MA and of the AR parameters separately. A new reduced-statistics order selection criterion is derived. It computes the fit of estimated models by comparing them with the long AR model that generally has many insignificant parameters. That selection criterion is applied for an automatic, statistically based, choice which of the four initial estimates for the ARMA parameters is used. Having obtained the best ARMA model, the same criterion can be used to select the best order for AR and MA models and for the automatic selection of the model type. The accuracy of the resulting spectral model based on the long AR model is in simulations compared to the best spectral model that could be obtained if the data would be available.

II. ARMA Models

An ARMA(p, q) process is defined as [11]

$$x_n + a_1 x_{n-1} + \cdots + a_p x_{n-p} = \varepsilon_n + b_1 \varepsilon_{n-1} + \cdots + b_q \varepsilon_{n-q} \quad (1)$$

where ε_n represents a series of independent, identically distributed, zero mean white noise observations. The power spectral density and the autocovariance function of the process are exactly determined by the parameters and by the variance of ε_n or x_n. The ARMA process can also be written with polynomials of AR and of MA parameters as

$$\begin{aligned} A_p(z)x_n &= B_q(z)\varepsilon_n \\ z^{-1}x_n &= x_{n-1} \\ A_p(z) &= 1 + a_1 z^{-1} + \cdots + a_p z^{-p} \\ B_q(z) &= 1 + b_1 z^{-1} + \cdots + b_q z^{-q}. \end{aligned} \quad (2)$$

The process is purely AR if $q = 0$ and MA if $p = 0$. ARMA models may have estimated polynomials of arbitrary orders, not necessarily equal to p and q. In this paper, AR(p'), MA(q'), and ARMA$(r', r'-1)$ models will be considered. Models are stationary if the estimated roots of $\hat{A}_{p'}(z)$ are inside the unit circle and invertible if the roots of $\hat{B}_{q'}(z)$ are inside the unit circle. Reduced-statistics estimators are derived from $C_L(z)$, a long AR(L) model estimated from the data x_n. Apart from the L parameters (or autocovariances or reflection coefficients), it is also assumed that the number of observations N is given for order selection with the reduced statistics estimators. If N is not known, it can be guessed from the long AR model by assuming that the highest order AR parameters are not significant and have the asymptotical variance $1/N$. Therefore, N can be estimated as the inverse of the average of the squares of those high-order AR parameters.

The parameters of an estimated time-series model can be used to compute the power spectral density and the autocovariance function of a stationary stochastic process [11].

III. Four First-Stage Reduced-Statistics ARMA Estimators

The simultaneous estimation of AR and MA parameters from the data with a maximum likelihood approach is computationally not attractive and it does not produce accurate models in many circumstances [1]. The solution in the ARMAsel algorithm [4], [5] uses filtering of the data with the inverse of a long AR model to have an estimate $\hat{\varepsilon}_n$ of the innovations ε_n in (1). Using $\hat{\varepsilon}_n$ and x_n, initial ARMA estimates are determined with a least squares algorithm in a first-stage ARMA estimator using the data.

To develop a reduced-statistics method, the first problem is to find an initial separation of the ARMA model in an AR part and an MA part. The first stage determines *initial estimates* for the AR part of an ARMA model. It is enough to compute ARMA$(r, r-1)$ models, for r between 2 and some maximum order, generally $N/10$, or 200 if that is less. Estimation of ARMA(p', q') models with arbitrary orders p' and q' is discouraged if the orders are not *a priori* known but have to be selected. It requires much computation time without providing a better model after automatic identification. For $r = 1$, the AR(1) model is used. The size M of the intermediate AR model that is used to calculate the $2r-1$ parameters of an ARMA$(r, r-1)$ model is given by $M = 3K^* + 2r$ if the mean is subtracted, else $M = 3K^* + 2r - 1$. Here, K^* denotes the AR order that is selected from AR models with candidate orders from AR(0) to AR(L), including white noise and all possible AR models. If L can be chosen freely, it can be taken as $N/2$ with a limit of 1000 or 2000, like in the ARMAsel algorithm [5]. If $M = 3K^* + 2r$ would become greater than L, that upper value L has to be used for M. It has been shown that using only this fraction of the totally available L AR parameters improves the accuracy of the estimated ARMA model [3]. The sliding order M depends on the number of estimated parameters and on the selected AR order K^*. Hence, a computer algorithm can choose it automatically and it requires no subjective choices of the data analyst.

- The first method, ***long AR***, has been introduced originally as a single-stage ARMA estimator [8]. It uses an intermediate AR(M) model denoted $C_M(z)$ as basis for computations, where $C_M(z)$ is determined by the first M reflection coefficients of the given long AR(L) model. M depends on the order of the ARMA model,

which is currently estimated, as defined above. The long AR method employs the relation

$$\frac{\hat{B}_{r-1}(z)}{\hat{\hat{A}}_r(z)} \approx \frac{1}{C_M(z)} \quad \text{or} \quad \hat{B}_{r-1}(z)C_M(z) \approx \hat{\hat{A}}_r(z). \tag{3}$$

A double hat in $\hat{\hat{A}}_r(z)$ and $\hat{\hat{a}}_i$ indicates the first-stage AR polynomial and parameters, respectively. A serious problem is that the relations in (3) are approximations. They give exact results for exact polynomials, but in the practice with estimated polynomials, they cannot be satisfied exactly. Standard estimation theory would require information about the statistical errors in these equations. This information is not available, and for that reason, it is not possible to derive a maximum likelihood estimator or any other efficient estimator from (3). By arbitrarily concentrating the inaccuracies in an error signal ζ_m, the second representation in (3) can be written as

$$\sum_{i=0}^{r-1} \hat{b}_i c_{m-i} = \hat{\hat{a}}_m + \varsigma_m, \qquad m = 0, \ldots, M \tag{4}$$

by equating the left- and right-hand side coefficients of z^{-m} in (3). Note that the parameters c_i are zero for $i < 0$. Now, the initial stage MA parameters of an ARMA$(r, r-1)$ model can be estimated without knowing the AR parameters, but with the knowledge that AR parameters have to be zero in an ARMA$(r, r-1)$ model for orders $m > r$. Therefore, MA parameter estimates can be found as a least squares solution from the equations in (4) with orders higher than r as

$$\sum_{i=0}^{r-1} \hat{b}_i c_{m-i} = \varsigma_m, \qquad m = r+1, \ldots, M \tag{5}$$

by minimizing the sum of ζ_m^2 for $r+1 < m < M$. A solution exists for $M \geq 2r-1$. If $M = 2r-1$, there are just enough equations to estimate the $r-1$ MA parameters with $\zeta_m = 0$. For greater M, the equations are overdetermined and need a least squares solution. This step will be followed by the calculation of the r AR parameters. It seems attractive to determine an AR solution with the first $r+1$ equations in (4). Substitution of the MA estimates from (5) yields estimates $\hat{\hat{a}}_i$ with $\zeta_m = 0$ for $m = 0, 1, \ldots, r$. However, simulations have shown that better initial first stage AR parameters can be found for a given MA model with the second-stage AR algorithm, which will be derived later. The informal derivation of this algorithm clarifies that no statistical optimality can be claimed for this first-stage solution of the initial estimates. The same lack of statistical optimality applies to the next three first-stage estimators of initial estimates. It is expected or hoped that at least one of the four estimators will be adequate for each different type of measured data. Simulations will be necessary to investigate the performance.

- The second method, denoted ***long MA***, is an alternative evaluation of the first relation in (3). It uses an estimate for the impulse response $g(z)$. This $g(z)$ can also be seen as a long MA model, which is calculated from the M parameters of $C_M(z)$. The length of the impulse response computed from $g(z)C_M(z) = 1$ can be chosen free, much greater than M or L. Suppose that the impulse response practically died out at L_0

$$\frac{\hat{B}_{r-1}(z)}{\hat{\hat{A}}_r(z)} \approx \frac{1}{C_M(z)} = g(z) \quad \text{or} \quad \hat{\hat{A}}_r(z)g(z) \approx \hat{B}_{r-1}(z). \tag{6}$$

  Knowing that the MA parameters are zero for $m > r-1$, the r AR parameters for the initial $\hat{\hat{A}}_r(z)$ follow as the least squares solution from

$$\sum_{i=0}^{r} \hat{\hat{a}}_i g_{m-i} = \varsigma_m, \qquad m = r, \ldots, L_0. \tag{7}$$

  The different expressions for ζ_m in (5) and (7) indicate once more that the application of least squares on approximate relations will not lead to theoretically optimal or efficient estimators. This long MA method has been described before [9].
- The third method to find initial conditions for the AR part $\hat{\hat{A}}_r(z)$ of the ARMA$(r, r-1)$ model is denoted ***long COV***, because it uses the covariance function as reduced statistic information [7]. The required formulas to establish the relations between the AR parameters in the long AR(L) model $C_L(z)$ and the autocovariances $R(k)$ are the Yule–Walker equations [11]

$$\sum_{i=0}^{L} c_i \hat{R}(m-i) = 0, \qquad m = 1, 2, \ldots, L.$$
$$\text{with} \qquad \hat{R}(-k) = \hat{R}(k), \qquad k > 0. \tag{8}$$

  Now the initial AR parameters $\hat{\hat{A}}_r(z)$ are calculated from Yule–Walker equations with a least squares solution

$$\sum_{i=0}^{r} \hat{\hat{a}}_i \hat{R}(m-i) = \varsigma_m, \qquad m = r, \ldots, M \tag{9}$$

  by minimizing the sum of ζ_m^2 for $r < m < M$. It is assumed that M is greater than $2r$, which will be guaranteed by the choice $M = 3K^* + 2r$.
- The fourth method, ***long Rinv***, is new and uses inverse correlations [11]. Inverse correlations and spectra are obtained by interchanging AR and MA polynomials [11]. The inverse correlation is found from the parameters of the intermediate AR model $C_M(z)$ with

$$R_{\text{inv}}(k) = \sum_{i=0}^{M-k} \hat{c}_i \hat{c}_{i+k}, \qquad k = 0, 1, \ldots, M. \tag{10}$$

  Further, $R_{\text{inv}}(k)$ is zero for shifts greater than M and it is symmetric around shift zero. The initial estimates for the

$r-1$ MA parameters are found with the overdetermined equations

$$\sum_{i=0}^{r-1} \hat{b}_i R_{\text{inv}}(m-i) \approx \varsigma_m, \qquad m = r+1, \ldots, M \tag{11}$$

by minimizing the sum of ζ_m^2 for $r+1 < m < M$. Afterwards, the second-stage AR method is used to calculate the first-stage AR model from these estimated MA parameters, like in the long AR method before.

IV. Second-Stage Reduced-Statistics ARMA Estimators

None of the four investigated reduced-statistics estimators gives a satisfactory accuracy for all types of stationary stochastic data in this initial stage; all require at least two stages. The **first stage** determines *initial estimates* for the AR part of an ARMA model; also, initial MA estimates are sometimes necessary to find that AR part, but those are not used any further. A considerable improvement in the accuracy is always obtained in the **second stage**. That stage uses new and improved variants of Durbin's ARMA methods [2], [10].

The accuracy of estimated models $\hat{A}_{p'}(z)$ and $\hat{B}_{q'}(z)$ can be characterized with the model error (ME), which is a scaled version of the squared error of prediction (PE) [12]. PE is the variance of the signal remaining after filtering data with the inverse of the estimated model

$$\text{ME}\left(\frac{\hat{B}_{q'}(z)}{\hat{A}_{p'}(z)}, \frac{B_q(z)}{A_p(z)}\right) = N\left\{\frac{\text{PE}}{\sigma_\varepsilon^2} - 1\right\}. \tag{12}$$

The ME can only be computed if the true process parameters in (1) are known. It is an excellent measure to evaluate the quality of estimation and selection algorithms in simulations. The expectation of the ME for unbiased models equals $p'+q'$, the number of estimated parameters. The quotient $\text{PE}/\sigma_\varepsilon^2$ is also defined as the power gain P_g of an ARMA process

$$\frac{\text{PE}}{\sigma_\varepsilon^2} = P_g\left(\frac{\hat{A}_{p'}(z)B_q(z)}{A_p(z)\hat{B}_{q'}(z)}\right). \tag{13}$$

This quantity can be computed as $\sigma_x^2/\sigma_\varepsilon^2$ of the ARMA$(p+q', p'+q)$ process

$$A_p(z)\hat{B}_{q'}(z)x_n = \hat{A}_{p'}(z)B_q(z)\varepsilon_n. \tag{14}$$

With this definition of the P_g, some properties of the ME are derived easily like

$$\begin{aligned}\text{ME}\left(\frac{E(z)}{F(z)}, \frac{G(z)}{H(z)}\right) &= \text{ME}\left(\frac{1}{F(z)}, \frac{G(z)}{E(z)H(z)}\right)\\ &= \text{ME}\left(\frac{1}{1}, \frac{F(z)G(z)}{E(z)H(z)}\right)\\ &= N\left\{P_g\left(\frac{F(z)G(z)}{E(z)H(z)}\right) - 1\right\}.\end{aligned} \tag{15}$$

It is well known that the best fitting AR(m) model to an arbitrary process is the solution of the Yule–Walker relations to equate the AR(m) autocovariance function to the first m autocovariances of the process [7], [11]. This formulation can be given formally with the ME of (12). The best AR(m) model of an arbitrary ARMA(p,q) process can be written as

$$\hat{A}_m(z) = \arg\min_{\tilde{A}_m(z)}\left\{\text{ME}\left(\frac{1}{\tilde{A}_m(z)}, \frac{B_q(z)}{A_p(z)}\right)\right\} \qquad \forall m. \tag{16}$$

It is not accurate to approximate some estimated autocovariance sequence by an MA model of the same length [7], and the best MA model requires a nonlinear optimization. If $F(z)G(z)$ equals exactly $E(z)H(z)$, the numerator and the denominator of P_g in (15) can be interchanged and the product AR and MA polynomials in (14) are equal. If the ME is small, about the number of estimated parameters, it turns out that interchanging the MA and the AR polynomial has only a small influence on the ME

$$\text{ME}\left(\frac{E(z)}{F(z)}, \frac{G(z)}{H(z)}\right) \approx \text{ME}\left(\frac{G(z)}{H(z)}, \frac{E(z)}{F(z)}\right). \tag{17}$$

The small error of this approximation can be illustrated by comparing the output variances of an AR(1) and of an MA(1) process that have the same small parameter δ and the same input variance σ^2. They are $\sigma^2/(1-\delta^2)$ and $\sigma^2(1+\delta^2)$, respectively. Therefore, the approximation

$$\begin{aligned}\hat{B}_m(z) &= \arg\min_{\tilde{B}_m(z)}\left\{\text{ME}\left(\frac{\tilde{B}_m(z)}{1}, \frac{B_q(z)}{A_p(z)}\right)\right\}\\ &= \arg\min_{\tilde{B}_m(z)}\left\{\text{ME}\left(\frac{A_p(z)}{B_q(z)}, \frac{1}{\tilde{B}_m(z)}\right)\right\}\\ &\approx \arg\min_{\tilde{B}_m(z)}\left\{\text{ME}\left(\frac{1}{\tilde{B}_m(z)}, \frac{A_p(z)}{B_q(z)}\right)\right\}\end{aligned} \tag{18}$$

can be derived to estimate MA models of a given ARMA(p,q) process. The final step in (18), interchanging the sequence of the arguments of estimated and true parameters in the original ME of (12), is only a close approximation for well fitting models. Nevertheless, this interchange will be applied because the first representation of (18) has no easy or computationally attractive solution that minimizes this ME. The solution would be highly nonlinear. The last representation of (18) shows a close agreement with Durbin's MA method [10].

After the four first-stage AR estimates have been computed, a **second stage** algorithm follows, consisting of two steps. It is applied to all four AR initial first-stage estimates and it requires or uses no information about initial MA parameters. In stage two, the initial AR estimates are first used to compute the final MA model. Then, this final MA model is used to improve the first-stage AR estimate.

- Divide the intermediate AR model $C_M(z)$ by the initially estimated AR(r) polynomial $\hat{\hat{A}}_r(z)$ of the first stage and use the MA method of Durbin [10] to determine the MA estimates $\hat{B}_{r-1}(z)$.

$$\begin{aligned}\hat{B}_{r-1}(z) &= \arg\min_{\tilde{B}_{r-1}(z)}\left\{\text{ME}\left(\frac{\tilde{B}_{r-1}(z)}{\hat{\hat{A}}_r(z)}, \frac{1}{C_M(z)}\right)\right\}\\ &\approx \arg\min_{\tilde{B}_{r-1}(z)}\left\{\text{ME}\left(\frac{1}{\tilde{B}_{r-1}(z)}, \frac{C_M(z)}{\hat{\hat{A}}_r(z)}\right)\right\}.\end{aligned} \tag{19}$$

In this derivation, use is made of the approximation due to interchanging the sequence of the elements of the ME, like in (17) and (18).

- Multiply the intermediate AR model $C_M(z)$ with this newly estimated MA model $\hat{B}_{r-1}(z)$. The solution for the improved AR parameters can be formulated as

$$\hat{A}_r(z) = \arg\min_{\tilde{A}_r(z)} \left\{ \mathrm{ME}\left(\frac{1}{\tilde{A}_r(z)}, \frac{1}{\hat{B}_{r-1}(z) C_M(z)} \right) \right\}. \tag{20}$$

This step yields the AR(r) model with the first r autocovariances equal to the first r autocovariances of the product model $\hat{B}_{r-1}(z)C_M(z)$. The same computation is also used to calculate $\hat{A}_r(z)$ in the first stage for the long AR and the long R_{inv} methods where the overdetermined equations yielded an initial estimate for $\hat{B}_{r-1}(z)$.

The two steps in the second stage can be iterated if desired. Iteration of the second stage will give an improved model if poor or zero initial stage AR estimates would be used. In simulations with the reduced-statistics initial estimates of this paper, up to ten iterations did not improve the quality of the estimated models for the very best of the four types of initial estimates, at least in most simulation examples. Generally, iterations tend to converge for all four methods to the same value that is slightly worse than the result of the best type of the first stage obtained after a single iteration, but it will often give an improvement for the three other types of initial estimates. In other words, poor initial estimates will mostly be improved with iterations in the second stage, but generally the best initial estimate will become worse with iterations.

V. Reduced-Statistics Order Selection

Most order selection criteria are based on the reduction of the residual variance as a function of the model order, with an additional penalty for every estimated parameter. As this paper deals with time-series models that are based on a long AR model, without knowledge of the data, no direct estimate for the residual variance is available. However, it is possible to compare the fit of different estimated AR, MA, and ARMA models to the long AR model $C_L(z)$ that represents all given information about the process. The power gain of (13) describes the fit of an estimated model to a true process. Replacing the true process polynomials by the long AR model $C_L(z)$, the fit is given as the power gain

$$P_g\left(\frac{\hat{A}_r(z)}{\hat{B}_{r-1}(z) C_L(z)} \right) = \frac{\mathrm{RES}(r, r-1)}{\sigma_\varepsilon^2}. \tag{21}$$

Thus, the power gain $P_g\{\hat{A}_r(z)/\hat{B}_{r-1}(z)C_L(z)\}$ is the residual variance of an estimated ARMA$(r, r-1)$ model *relative* to the unknown input variance. It can be calculated from the estimated parameters together with $C_L(z)$, without knowledge of the data x_n. A simple computation of the ARMA power gain is found with (14) by separating the AR and MA filter operations [12]. This relative residual variance $P_g\{\hat{A}_r(z)/\hat{B}_{r-1}(z)C_L(z)\}$ can be used in an order selection criterion that is based on the *logarithm* of the residual variance. The unknown input variance becomes a constant in the logarithm of P_g, the same constant for all estimated models. Therefore, it has no influence on the minimum of the order selection criterion. The penalty factor 3 is preferred for order selection [13]. The selection criterion for an ARMA$(r, r-1)$ model becomes

$$\mathrm{GIC}(2r, 3) = \ln\left[P_g\left(\frac{\hat{A}_r(z)}{\hat{B}_{r-1}(z) C_L(z)} \right) \right] + \frac{6r}{N} \tag{22}$$

where $6r$ equals the penalty factor three times the total number of parameters, r AR parameters, $r-1$ MA parameters, and an additional parameter for the subtraction of the estimated mean. This order selection criterion has been used to select the best order r after the second stage, for each of the four ARMA first-stage methods of Section III. Furthermore, GIC$(2r, 3)$ has also been used to select *between* those four estimators. In this way, the best first-stage algorithm is selected automatically.

The same type of criterion (22) can also be used for the selection of the order of estimated AR or MA models, by substituting in (22) $b_0 = 1$ or $a_0 = 1$ for $\hat{B}_{r-1}(z)$ and $\hat{A}_r(z)$, respectively, and by using the number of estimated parameters in the penalty function. In this way, the criterion for the MA model of order q becomes

$$\mathrm{GIC}(q, 3) = \ln\left[P_g\left(\frac{1}{\hat{B}_q(z) C_L(z)} \right) \right] + \frac{3(q+1)}{N}. \tag{23}$$

For AR order selection, a similar criterion can be used, but a finite sample criterion is to be preferred if the highest candidate order for selection is greater than $N/10$ [13].

VI. Simulations

In simulations, the quality of estimated models can be measured with the model error ME (3). The Cramér–Rao lower bound for the spectral accuracy of unbiased models, expressed in the ME, equals the number of estimated parameters. Every efficiently estimated parameter gives a variance contribution 1 to the expectation of the ME. For biased models, also an additional bias contribution is present in the ME if the model order is lower than the true process order. The bias is directly proportional to N. Generally, order selection criteria should select lower order biased models if the bias contribution to the ME is less than 1 for each parameter that is left out of the selected model. In this way, the average ME of biased estimated models could become less than the number of truly nonzero parameters. That number is the Cramér–Rao lower bound for unbiased models, with at least all truly nonzero parameters included. The expectation of the ME has been made independent of N and of the variance of the signal by using the normalization in (12).

First, simulations have been carried out with Yule–Walker and with Burg [7] estimates for the parameters of the long AR model $C_L(z)$. Using AR parameters, estimated with the Yule–Walker AR algorithm, is identical to using the biased $\sum x_n x_{n+k}/N$ for the covariance as information. This bias is known to give spectral distortions [4]. In some examples, the final ARMA results are comparable with either Burg or Yule–Walker parameter estimates for the long AR information. However, in other examples, the performance of Yule–Walker estimates for the parameters of the long AR is much worse than

TABLE I
COMPARISON OF MODELS FROM REDUCED-STATISTICS AND FROM DATA. THE MODEL ERROR ME OF SELECTED ARMA MODELS FOR THE FOUR FIRST-STAGE INITIAL ESTIMATES AFTER THE SECOND STAGE, THE AUTOMATIC SELECTION FROM THOSE FOUR, THE ARMA MODEL SELECTED FROM THE DATA, THE REDUCED–STATISTICS MODEL ARMAsel_rs WITH TYPE AND ORDER SELECTED FROM THE REDUCED-STATISTICS INPUT AND THE ARMAsel MODEL WITH TYPE AND ORDER SELECTED FROM THE OBSERVED DATA, ALL AS A FUNCTION OF THE RADIUS INDICATOR β. THE NUMBER OF OBSERVATIONS N IS 200 AND THE LENGTH LOF THE REDUCED-STATISTICS AR MODEL IS 100. THE FIRST SIX COLUMNS ARE AVERAGES OF ONLY ARMA MODELS WITH SELECTED ORDERS. IN THE LAST TWO COLUMNS, THE MODEL TYPE IS SELECTED AS WELL, FROM THE AR(100) MODEL AND DIRECTLY FROM THE DATA, RESPECTIVELY

β	*long AR*	*long MA*	*long COV*	*long Rinv*	*selected*	*ARMA data*	*ARMAsel_rs*	*ARMAsel*
-0.8	16.1	22.0	32.7	21.5	17.7	22.0	15.7	16.1
-0.6	5.60	5.75	8.42	5.50	6.49	5.69	8.58	8.69
-0.4	3.42	4.11	7.23	3.72	4.30	4.42	6.44	6.45
-0.2	2.62	2.59	2.61	2.63	2.67	2.98	3.49	3.59
0.0	1.38	1.37	1.26	1.29	1.26	0.93	1.42	1.42
0.2	2.68	2.57	2.62	2.77	2.71	2.65	3.37	3.34
0.4	3.61	3.60	6.79	4.20	4.17	3.73	5.82	5.92
0.6	5.28	5.11	7.89	5.33	5.58	5.18	7.75	7.95
0.8	25.3	41.9	42.8	29.7	25.7	31.5	27.5	33.8

Burg, while Burg never is so bad. Hence, Burg's AR results are preferred for the long AR estimator as input for automatic identification of a good time-series model.

Simulations of an available automatic reduced statistics algorithm ARMAsel_rs [14] with an ARMA$(6,5)$ process with poles and zeros on the same radius $|\beta|$ are reported here. The AR reflection coefficients are $1, \beta, \beta^2, \beta^3, \beta^4, \beta^5$, and β^6; the MA follow from $1, -\beta, (-\beta)^2, (-\beta)^3, (-\beta)^4, (-\beta)^5$ with the same Levinson recursion [7] as the usual for the AR polynomial. Some results are reported in Table I. For other processes and sample sizes, similar results are obtained in comparing models estimated from long AR with models estimated from N observations. Table I shows that the first-stage method *long COV* is a poor method for the processes used. However, examples have been reported [9] where this method produced the best first-stage initial estimates. In this example, the other three methods of initial estimates are more or less comparable, except for the largest radius, where long AR is definitely the best. If the absolute value $|\beta|$ is less than 0.8, the best estimated ARMA models are of lower order than the true ARMA$(6,5)$ orders for $N = 200$. The bias in estimated ARMA$(3,2)$ models is less than the variance increase that would occur if the full ARMA$(6,5)$ model had to be estimated. The order selection criterion selects those lower orders. The average ME is less than the Cramér–Rao lower bound for unbiased models, which is 12 for the ARMA$(6,5)$ process with subtraction of the mean.

It is remarkable that the ME of the automatically selected initial condition type in the column "selected" follows the smallest ME of the four previous columns rather closely. This simulation result is a strong support for the use of four different initial conditions as a first stage in reduced statistics ARMA modeling. The selection of the best of four reduced statistics estimators can be carried out with the selection criterion of (22). A comparison of the columns "selected" and "ARMA data" reveals that the quality of ARMA models from reduced statistics and from data is comparable in this example; each of the columns is sometimes better. Also, the quality of models with selected order and type in the last two columns is similar. The data identification algorithm ARMAsel [5] has been used in the last column. The identification of both the model order and the model type shows that a long AR model can provide about the same accuracy for identification as using the data.

TABLE II
COMPARISON OF AUTOMATICALLY SELECTED MODELS FROM DATA AND FROM LONG AR. AVERAGE ME OF MODELS OF SELECTED ORDER AND TYPE FOR AN ARMA$(6,5)$ PROCESS WITH RADIUS $\beta = 0.7$ AS A FUNCTION OF N. IDENTIFICATION IS BASED ON DATA WITH ARMAsel [5] OR ON REDUCED STATISTICS WITH ARMAsel_rs [14]

N	*ARMAsel*	*ARMAsel_rs*
25	27.9	28.0
50	21.6	16.0
100	20.2	16.5
200	17.2	16.0
400	13.9	13.9
800	12.7	13.2
1600	13.1	13.4
3200	15.2	15.4
6400	17.7	17.9

Table II compares identification of time-series models from the observations themselves and from a long AR model, respectively, as a function of the sample size N, for ARMA$(6,5)$ processes with $\beta = 0.7$. The accuracy is about the same. The selected reduced statistics model is slightly better for small N, whereas data are somewhat better for the larger N. The results are, for N about 1000, close to the Cramér–Rao lower bound, which is 12. For higher N, the average ME increases, which indicates a small persistent bias in the algorithms. Furthermore, with reduced statistics the data can still speak for themselves in selecting their best model with the algorithms that have been proposed here. Also, the selection of the best model type is possible with the order selection criteria (22) and (23) that use the fit of the estimated AR, MA, and ARMA models with respect to the very long AR model.

VII. CONCLUSION

An order selection criterion has been defined that can be derived from a long AR model of the data. That criterion can be used to select automatically the best performing of four ARMA estimators. It can also select the ARMAsel_rs model with the best model order and the best model type (AR, MA, or ARMA) for processes that are only represented by the information of a long AR model.

All ARMA models estimated with the two-stage reduced-statistics estimators are stationary and invertible. Single-stage estimators give no accurate ARMA models for most processes. The performance of four estimators of the AR part as initial estimates for a second stage depends greatly on the example. However, at least one of the four proposed estimators was quite good in each simulation example.

The accuracy of the ARMAsel_rs model based on a long AR model is in most examples comparable with the ARMAsel model that selects the best model type and model order with estimations from the observed data. Hence, using a long AR model estimated with the Burg algorithm gives no loss of spectral information. It remains possible to determine the statistical significance of details in the spectrum with the selection of order and type of time-series models if only a long AR model is given.

REFERENCES

[1] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.

[2] J. Durbin, "The fitting of time-series models," *Rev. Inst. Int. Stat.*, vol. 28, pp. 233–243, 1960.

[3] 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.

[4] ——, "Automatic spectral analysis with time-series models," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 2, pp. 211–216, Apr. 2002.

[5] ——, "ARMASA Matlab Toolbox," Matlab, http://www.mathworks.com/matlabcentral/fileexchange, 2002. [Online]. Available: with selection of the keywords signal processing, spectral analysis.

[6] R. Klees, P. Ditmar, and P.M.T. Broersen, "How to handle colored observation noise in large least-squares problems," *J. Geodesy*, vol. 76, pp. 629–640, 2003.

[7] S. M. Kay and S. L. Marple, "Spectrum analysis-a modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, Nov. 1981.

[8] D. Graupe, D. J. Krause, and J. B. Moore, "Identification of autoregressive moving-average parameters of time series," *IEEE Trans. Autom. Control*, vol. AC-20, no. 1, pp. 104–107, Feb. 1975.

[9] P. M. T. Broersen and S. de Waele, "Improving the efficiency of reduced statistics ARMA estimators," in *Proc. SYSID 2000 Symp. System Identification*, Santa Barbara, CA, 2001, pp. 1019–1024.

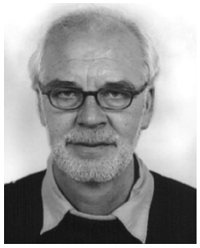
[10] J. Durbin, "Efficient estimation of parameters in moving average models," *Biometrika*, vol. 46, pp. 306–316, 1959.

[11] M. B. Priestley, *Spectral Analysis and Time Series*. London, U.K.: Academic, 1981.

[12] P. M. T. Broersen, "The quality of models for ARMA processes," *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1749–1752, Jun. 1998.

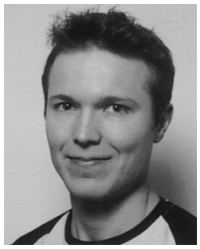
[13] ——, "Finite sample criteria for autoregressive order selection," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3550–3558, Dec. 2000.

[14] S. de Waele, Matlab, http://www.mathworks.com/matlabcentral/fileexchange, 2003. [Online]. Available: with selection of the keywords signal processing, spectral analysis.



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 the Delft University of Technology, Delft, The Netherlands, in 1968 and 1976, respectively.

He is currently with the Department of Multi Scale Physics of the Delft University of Technology. His main research interest is in automatic identification on statistical grounds by letting measured data speak for themselves. He developed a practical solution for the spectral and the autocorrelation analysis of stochastic data by the automatic selection of a suitable order and type for a time-series model of the data.



Stijn de Waele was born in Eindhoven, The Netherlands, in 1973. He received the M.Sc. degree in applied physics and the Ph.D. degree, with a dissertation entitled "Automatic model inference from finite time observations of stationary stochastic processes," from the Delft University of Technology, Delft, The Netherlands, in 1998 and 2003, respectively.

Currently, he is a Senior Scientist at Philips Research Laboratories, Eindhoven, The Netherlands, where he works in the field of digital video compression.