This conditional distribution is independent of the original nuisance parameters and has Neyman structure with respect to the sample variances. The one-sided tests are UMP-similar.

Simulations confirm that this similar test is independent of the degree of correlation and, therefore, has controllable size and controllable type I errors. In contrast, the standard $F$-test has a true size smaller than the design size and are consequently more likely to cause cause type II errors.

## ACKNOWLEDGMENT

The authors thank a reviewer for suggesting the transform to the beta distribution.

## REFERENCES

[1] D. R. Cox and D. V. Hinkley, *Theoretical Statistics*. London, U.K.: Chapman & Hall, 1974.

[2] R. L. Kirlin, H. Brügel, C. Ying, and D. F. Groutage, "Maximum likelihood real part of coherence between two noisy sensor outputs with zero-delay signal inputs," *IEEE Trans. Signal Processing*, vol. 42, pp. 1852–1855, July 1994.

[3] N. R. Goodman, "Statistical analysis based on a certain multivariate Gaussian distribution (an introduction)," *Ann. Math. Stat.*, vol. 34, pp. 152–177, Mar. 1963.

[4] W. S. Burdic, *Underwater Acoustic System Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[5] G. P. Beaumont, *Intermediate Mathematical Statistics*. New York: Methuen, 1980.

[6] M. Kendall and A. Stuart, *The Advanced Theory of Statistics: Vol. 2, Inference and Relationship*, 4th ed. New York: McMillan, 1979.

[7] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.

[8] D. E. Amos, "Additional percentage points for the incomplete beta distribution," *Biometrika*, vol. 50, p. 449, 1963.

[9] M. Kendall and A. Stuart, *The Advanced Theory of Statistics: Vol. 1, Distribution Theory*, 4th ed. New York: McMillan, 1979.

---

# Autoregressive Model Orders for Durbin's MA and ARMA Estimators

P. M. T. Broersen

***Abstract*—Durbin's methods for moving average (MA) and autoregressive-moving average (ARMA) estimation use the parameters of a long AR model to compute the MA parameters. Linear regression theory is applied to find the best AR order. This yields two different orders: one for the best predicting AR model and another one for the long AR model with the best parameter accuracy, as intermediate for Durbin's estimates. Both orders increase with the sample size and have no finite limiting value.**

## I. INTRODUCTION

Maximum likelihood estimation of moving average (MA) parameters in pure MA models or as part of autoregressive-moving average (ARMA) models is a nonlinear problem, with possible difficulties in convergence and invertibility of solutions [1], [2]. Nonlinear methods sometimes will not converge to invertible models for small sample sizes $N$ unless the algorithm uses constrained minimization or takes reciprocals of estimated noninvertible roots. If estimated poles and zeros can be inside or outside the unit circle, they can also be close to it. The single realizations where invertible solutions do have zeros very close to the unit circle, with distance less than $1/N$, dominate the simulation average of quality measures for estimated models [1]. Theoretically optimal ARMA estimators are based on iterative procedures whose global convergence is not guaranteed [3]. However, the statistical accuracy of computationally simple practical ARMA estimators may be poor in some cases [3].

The best achievable accuracy of the simple methods of Durbin is unknown so far because of the unknown influence of the order of the intermediate AR model. Durbin's methods for MA [4] and for ARMA estimation [5] replace a nonlinear estimation problem by two stages of linear estimation. First, the parameters of a long AR model are estimated from the data. Afterwards, MA parameters are computed by using the sequence of estimated AR parameters as input signal for the Yule–Walker algorithm, which is usually applied in AR estimation. This MA method is based on the asymptotical theoretical equivalence of AR$(\infty)$ and MA$(q)$ processes. Practice and simulations, however, have shown that the best AR order in estimation is finite and depends on the true process parameters and on the number of observations [6]. A practical choice for the order of the long AR model to be used for MA processes has been described [7]. Earlier studies just took an order to be large enough to make the AR approximation useful for the purpose of estimation [8]. Therefore, the accuracy of Durbin's ARMA method if the best long AR order is used is still an open question. Hence, the search is for a theoretical concept that can be used to find a best order for the long AR model as an intermediate stage in the estimation of a MA model or of the MA part of an ARMA model. Examples have been given wherein the best approximate time series model is determined by the true process as well as by the application [9]. Therefore, the best AR order for approximate or estimated models is known to depend on the purpose of the model.

The subject of order selection is extensively treated in linear regression with deterministic regressors. The best order for an estimated subset regression model depends on the intended application, which, among others, can be prediction with the model or the accuracy of its parameters [10]. In this correspondence, the theory for linear regression will be applied to AR order selection problems encountered in MA and ARMA processes. The first application is characterized by using the AR model itself for prediction, and the second one uses the parameters of an estimated long AR model to compute the MA parameters. The different optimal AR model orders for both types of application will be described for *known* ARMA processes. Simulations show that those two different, theoretically optimal, orders are also the best in practice. The accuracy of the resulting ARMA models is close to the Cramér-Rao lower bound for integrated spectra.

## II. MODEL ORDERS IN LINEAR REGRESSION

The subject of order selection is known as subset selection or as the selection of variables in linear regression. The best order for the estimated subset model depends on the intended application of the model [10], which can be prediction with the estimated model or a small mean squared error of the estimated parameters. Suppose that a regression equation for noisy measurements is given by

$$\mathbf{Y} = \mathbf{X}_{\mathbf{K}}\beta_{\mathbf{K}} + \varepsilon \tag{1}$$

---

Manuscript received November 16, 1998; revised March 21, 2000. The associate editor coordinating the review of this paper and approving it for publication was Dr. Sergio Barbarossa.

The author is with Department of Applied Physics, Delft University of Technology, Delft, The Netherlands (e-mail: broersen@tn.tudelft.nl).

Publisher Item Identifier S 1053-587X(00)05982-1.

where

| | |
|---|---|
| $Y$ | $N \times 1$ vector of observations; |
| $X_K$ | $N \times K$ vector of deterministic independent variables as regressors; |
| $\beta_K$ | $K \times 1$ true parameter vector; |
| $\varepsilon$ | $N \times 1$ random vector of independent identically normally distributed variables with zero mean and variance $\sigma^2$. |

The true process order $K$ can be infinite, e.g., for the description of an exponential function with polynomials. The regressors $\mathbf{X_K}$ can be split into two groups: $\mathbf{X_p}$ for regressors that are included in a subset model and $\mathbf{X_r}$, which are not. Therefore

$$\mathbf{X_K}\beta_\mathbf{K} = \mathbf{X_p}\beta_\mathbf{p} + \mathbf{X_r}\beta_\mathbf{r}, \quad \mathbf{K} = \mathbf{p} + \mathbf{r}. \tag{2}$$

For hierarchical models where the sequence of regressors is fixed, $X_p$ always contains the first $p$ regressors. Generally, arbitrary regressors can be included in subset models, but regressors can always be renumbered such that the first $p$ are in the subset. This gives a considerable simplification in the notation without loss of generality. The question is whether estimating only the parameters

$$\hat{\mathbf{b}}_\mathbf{p} = \left(\mathbf{X_p^T X_p}\right)^{-\mathbf{1}} \mathbf{X_p^T Y} \tag{3}$$

of the subset in (2) can give a better model than estimation of all $K$ parameters of the complete model; $X^T$ denotes the transpose of $X$. The statistical significance of higher model orders can be measured by the influence on the residual sum of squares (RSS) defined as

$$\mathbf{RSS_p} = (\mathbf{Y} - \mathbf{X_p}\hat{\mathbf{b}}_\mathbf{p})^\mathbf{T}(\mathbf{Y} - \mathbf{X_p}\hat{\mathbf{b}}_\mathbf{p}). \tag{4}$$

The increase of RSS$_p$, by excluding the parameters for the last $r$ regressors from the model, is given by [10]

$$\mathbf{RSS_p} - \mathbf{RSS_{p+r}} = \hat{\mathbf{b}}_\mathbf{r}^\mathbf{T}\mathbf{B_{rr}^{-1}}\hat{\mathbf{b}}_\mathbf{r} \tag{5}$$

where $\hat{\mathbf{b}}_\mathbf{r}^\mathbf{T}$ is the transpose of the $r \times 1$ parameter that belongs to $\mathbf{X_r}$, and $\sigma^2\mathbf{B_{rr}}$ is the appropriate $r \times r$ submatrix of the total covariance matrix $\sigma^2(\mathbf{X_K^T X_K})^{-1}$ for estimating $K = p + r$ parameters. The statistical expectation of (5) is given by the sum of a bias and a variance term

$$\mathbf{E}\left\{\hat{\mathbf{b}}_\mathbf{r}^\mathbf{T}\mathbf{B_{rr}^{-1}}\hat{\mathbf{b}}_\mathbf{r}\right\} = \beta_\mathbf{r}^\mathbf{T}\mathbf{B_{rr}^{-1}}\beta_\mathbf{r} + \mathbf{r}\sigma^\mathbf{2}. \tag{6}$$

Therefore, the expected increase in the residual sum of squares by leaving $r$ regressors from the subset is minimally the variance contribution $r\sigma^2$ if there is no bias and all true values of $\beta_\mathbf{r}$ are zero.

Requirements for optimality can be formulated either for true parameters values $\beta_\mathbf{r}$ or for estimated values $\hat{\mathbf{b}}_\mathbf{r}$ in (6) by adding $r\sigma^2$. Depending on the precise use of the estimated model, two different requirements for $\beta_\mathbf{r}^\mathbf{T}\mathbf{B_{rr}^{-1}}\beta_\mathbf{r}$ have been defined [10] that make the complete model of order $K = p + r$ better than the subset model of order $p$. The first requirement for the inclusion of $r$ extra parameters is given by

$$\beta_\mathbf{r}^\mathbf{T}\mathbf{B_{rr}^{-1}}\beta_\mathbf{r} > \mathbf{r}\sigma^2 \tag{7}$$

if accurate prediction with the model is the purpose. Therefore, the reduction with the bias term must be greater than the expected contribution $r\sigma^2$ of the variance term. Prediction forms the basis of almost all existing order selection criteria. Regressors have to be excluded from the best estimated subset model, and order $p$ is preferred above order $p + r$ if the $r$ *true* process parameters that belong to those regressors give a reduction $\beta_\mathbf{r}^\mathbf{T}\mathbf{B_{rr}^{-1}}\beta_\mathbf{r}$ that is less than $r\sigma^2$ or, equivalently, with (5) and (6), that is, less than $2r\sigma^2$ in RSS$_p$. This factor 2 can be recognized in the subset selection criterion $C_p$ in regression [10] and in the AIC selection criterion of Akaike for time series [11].

To find the subset with the smallest mean square error of all parameters, the minimal requirement for statistical significance of $r$ parameters has been derived as [10]

$$\beta_\mathbf{r}^\mathbf{T}\mathbf{B_{rr}^{-1}}\beta_\mathbf{r} > \sigma^\mathbf{2}. \tag{8}$$

For the smallest value of the mean squared error of the estimated parameters, any group of $r$ estimated parameters is better excluded if the bias component of the reduction in RSS$_p$, $\beta_\mathbf{r}^\mathbf{T}\mathbf{B_{rr}^{-1}}\beta_\mathbf{r}$ is less than $\sigma^2$. A comparison of (7) and (8) shows that smaller parameter values will remain included in the best subset for parameter accuracy; therefore, the subset size for the best parameter accuracy is always greater than or equal to the best subset size for prediction.

## III. AR ORDERS FOR MA AND ARMA PROCESSES

Some notation is required to apply those results to time series. An ARMA$(p, q)$ process can be written as

$$\mathbf{A(z)x_n} = \mathbf{B(z)}\varepsilon_\mathbf{n} \tag{9}$$

with $A(z) = 1 + a_1 z^{-1} + \cdots + a_p z^{-p}$, $B(z) = 1 + b_1 z^{-1} + \cdots + b_q z^{-q}$, and $z^{-1}x_n = x_{n-1}$. $\varepsilon_n$ represents a series of independent, identically distributed stochastic variables, which is a white noise process that generates the data; $\sigma_x^2$ is the variance of the process, and $\sigma_\varepsilon^2$ is the variance of the innovations $\varepsilon_n$. The roots of $A(z)$ and $B(z)$ should be within the unit circle to guarantee a stationary and invertible process, respectively. The theory of regression analysis will be applied to the long AR model of a *known* ARMA process (9), which is defined by

$$\mathbf{C(z)x_n} = \varepsilon_\mathbf{n} \tag{10}$$

with the parameter polynomial given by $C(z) = A(z)/B(z)$ and with the same innovation variance $\sigma_\varepsilon^2$ as (9). $C(z)$ has order $\infty$. A *finite* number of parameters of $C(z)$ can be determined with any desired accuracy by computing the theoretical covariance function of the ARMA process [7], [12] until that finite order and by transforming those covariances to AR parameters or to reflection coefficients $k_i$ with the Levinson–Durbin algorithm [3]. The theoretical RSS$_{\rm th}(m)$ of all AR$(m)$ models from order AR(1) to AR$(\infty)$ for a *known* ARMA process can be defined for a sample size of $N$ observations as

$$\begin{aligned}\mathbf{RSS_{th}(m)} &= \mathbf{N}\sigma_\mathbf{x}^\mathbf{2}\prod_{\mathbf{i=1}}^{\mathbf{m}}\left(\mathbf{1} - \mathbf{k_i^2}\right) \\ &= \mathbf{N}\sigma_\varepsilon^\mathbf{2} \bigg/ \prod_{\mathbf{i=m+1}}^{\infty}\left(\mathbf{1} - \mathbf{k_i^2}\right)\end{aligned} \tag{11}$$

with RSS$_{\rm th}(0) = N\sigma_x^2$. The true values $k_i$, for $i > m$, describe the bias contribution in all finite-order AR models. The asymptotical expression RSS$_{\rm th}(\infty)$ becomes $N\sigma_\varepsilon^2$.

Applying the subset theory of Section II, it can be derived with (5)–(7) and (11) that the best AR order $L$ for prediction is found, for given $N$, as the order $L$ with the property that for arbitrary values of $r > 0$:

$$\begin{aligned}\mathbf{RSS_{th}(L+r)} &> \mathbf{RSS_{th}(L)} - \mathbf{r}\sigma_\varepsilon^\mathbf{2} \\ \mathbf{RSS_{th}(L-r)} &> \mathbf{RSS_{th}(L)} + \mathbf{r}\sigma_\varepsilon^\mathbf{2}.\end{aligned} \tag{12}$$

Therefore, the reduction of RSS$_{\rm th}$ is greater than $r\sigma_\varepsilon^2$ for $r$ orders below $L$ and smaller than $r\sigma_\varepsilon^2$ above $L$ simultaneously for all values of $r$. This unique order $L$ can be found easily for a *known* ARMA process and given $N$ with the order selection criterion AIC by using the penalty 1 instead of 2 [7].

The AR order with the best parameter accuracy is found as the order for which the theoretical $\mathrm{RSS}_{\mathrm{th}}(M)$, due to bias of incomplete models, is only $\sigma_\varepsilon^2$ greater than $\mathrm{RSS}_{\mathrm{th}}(\infty)$; therefore, the greatest $M$ with

$$\mathbf{RSS_{th}(M)} = \mathbf{N}\sigma_{\mathbf{x}}^{\mathbf{2}} \prod_{\mathbf{i=1}}^{\mathbf{M}} \left(\mathbf{1} - \mathbf{k_i^2}\right) > (\mathbf{N}+\mathbf{1})\sigma_\varepsilon^{\mathbf{2}} \qquad (13)$$

for $N$ observations, or the highest order with as residual variance more than $(1 + 1/N)\sigma_\varepsilon^2$, which is the same. This shows that the best order obviously increases with the number of observations $N$. Inspection of (12) and (13) shows that $M \geq L$. This theory has been applied successfully to Durbin's method for MA processes [4], where a similar order has been derived, and simulations have corroborated the theoretical derivation [7]. The order $M$ is important in Durbin's ARMA method because the *parameters* of the long estimated AR model $\hat{\mathbf{C}}(\mathbf{z})$ are used as intermediate to estimate the MA parameters. Simulations with known ARMA processes are necessary to validate all approximations that have been made because the application of subset results of regression theory with *deterministic* regressors to *stochastic* time series observations requires a verification.

## IV. Best AR Orders in ARMA Simulations

In evaluating the statistical accuracy of estimation methods, it is necessary to have an objective measure for the quality of models. A logical choice for time series models is the prediction error PE, which is defined as the variance of the one step ahead prediction in applying a model to new data. The Cramér–Rao lower boundary for unbiased ARMA$(p', q')$ models is $\sigma_\varepsilon^2\{1 + (p' + q')/N\}$. Models can be unbiased if $p' \geq p$ and $q' \geq q$. The quality measure for ARMA modeling [the model error (ME)] has been defined [12] as a scaled transformation

$$\mathbf{ME} = \mathbf{N}\left(\mathbf{PE}/\sigma_\varepsilon^2 - \mathbf{1}\right). \qquad (14)$$

Scaling with $\sigma_\varepsilon^2$ and subtraction of 1 gives the minimum value zero for ME if process and model are identical. Multiplication with $N$ in (14) makes the numerical value of ME largely independent of $N$. An efficient algorithm for computation of ME is available [12]. The Cramér-Rao boundary for ME of ARMA$(p', q')$ models is $p' + q'$.

The best AR orders for MA$(q)$ and ARMA$(p, q)$ processes have been established in numerous simulations, with different $p$ and $q$ and with $N$ between 20 and 50 000. Fig. 1 gives the sum of squared errors of parameters as a function of the AR order for AR models of an ARMA$(3, 2)$ process given by

$$\begin{aligned}\mathbf{x_n} + \mathbf{0.2x_{n-1}} - \mathbf{0.4x_{n-2}} + \mathbf{0.3x_{n-3}} \\ = \varepsilon_{\mathbf{n}} - \mathbf{0.4}\varepsilon_{\mathbf{n-1}} - \mathbf{0.5}\varepsilon_{\mathbf{n-2}}.\end{aligned} \qquad (15)$$

The theoretically computed optimal model orders for $N = 100$ are seven for $L$ with (12) and 19 for $M$ with (13) for prediction and parameter accuracy, respectively. The accuracy of the AR parameters for model order $i$ is approximated by using the AR$(100)$ approximation as the true representation of the ARMA process

$$\sum_{\mathbf{j=0}}^{\infty}(\mathbf{a} - \hat{\mathbf{a}})^2 \approx \sum_{\mathbf{j=1}}^{\mathbf{i}}(\mathbf{a_j} - \hat{\mathbf{a}}_{\mathbf{j}})^2 + \sum_{\mathbf{j=i+1}}^{100}\mathbf{a_j^2}. \qquad (16)$$

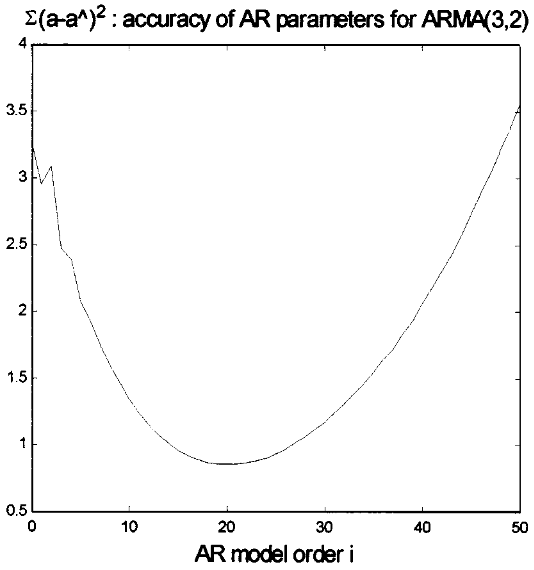

Fig. 1. Average mean squared parameter error $\Sigma(a - \hat{a})^2$ between the AR$(i)$ model and the true parameters of the AR$(100)$ approximation of the ARMA$(3, 2)$ process as a function of the AR model order $i$. $N = 100$, average of 10 000 simulation runs.

Fig. 1 shows that the minimum of this criterion is found at order 20, which is very close to the theoretical value of 19 for $M$ that follows from ordinary regression theory. In addition, for many different examples, it has been verified that the best accuracy for parameters with $\Sigma(a - \hat{a})^2$ is found at an order close to $M$ of (13). Durbin's methods for MA parameters use the first $q + 1$ terms of the covariance of estimated AR parameters $\sum \hat{\mathbf{a}}_{\mathbf{j}}\hat{\mathbf{a}}_{\mathbf{j+k}}, k = 1, \ldots, q$ to compute $q$ MA parameters. Hence, it may also be expected that the best result for Durbin's MA and ARMA models is obtained for an AR order with a small value of $\Sigma(a - \hat{a})^2$.

Fig. 2 shows the ME of AR models of the ARMA process (15). The minimum is found for AR order 7, which is equal to the theoretical result $L$ of (12). No examples have been found in simulations with AR, MA, or ARMA processes where the results were very different from those orders $L$ and $M$; the calculated AR order $L$ is the best order for AR models used for prediction, and order $M$ gives the best parameter accuracy.

To find initial parameters for an ARMA model with Durbin's estimator, a long AR model $\hat{\mathbf{C}}(\mathbf{z})$ is used to reconstruct residuals $\hat{\varepsilon}_{\mathbf{n}}$ [5]. Previous values $x_{n-1}, x_{n-2}, \ldots x_{n-p}$ and $\hat{\varepsilon}_{\mathbf{n-1}}, \ldots, \hat{\varepsilon}_{\mathbf{n-q}}$ are used to estimate the $p + q$ parameters of the ARMA$(p, q)$ model with least squares, as in linear regression. The ME results of this initial step are not reported here because several simulation runs produced MA models that are not invertible. Moreover, the ME is always improved by the next stage of Durbin's method, where MA and AR parameters are updated sequentially [5]. By filtering out the initial estimate of the AR parameters from the long estimated AR model $\hat{\mathbf{C}}(\mathbf{z})$, an improved estimate of the MA parameters can be computed. After filtering out this estimated MA model from the observations, an improved AR estimate can be computed. Durbin's method always gives invertible MA models because AR parameters are used as data in the Yule–Walker algorithm to find the MA model, which ensures invertibility. Likewise, the AR parameters are obtained with Burg's method [3], which ensures stationary models. In Fig. 2, the intermediate AR order is the order of $\hat{\mathbf{C}}(\mathbf{z})$, which is varied between 5 and 50. The best accuracy of the ARMA$(3, 2)$ model is found with the intermediate AR order $i = 20$, which is close to the theoretical value 19 that is found for $M$ with (13). Small differences in the precise location of the minimum in this type of simulation can be expected and explained because the minimum ME of the ARMA$(3, 2)$ model as a function of the AR order is generally flat, and the variance of simulation results may shift it somewhat. The common phenomena with many other examples are a clear minimum in the fit of AR models for prediction and a flat minimum for the best intermediate long AR order for ARMA models. Moreover, taking lower order AR models in Durbin's method can be much worse than taking

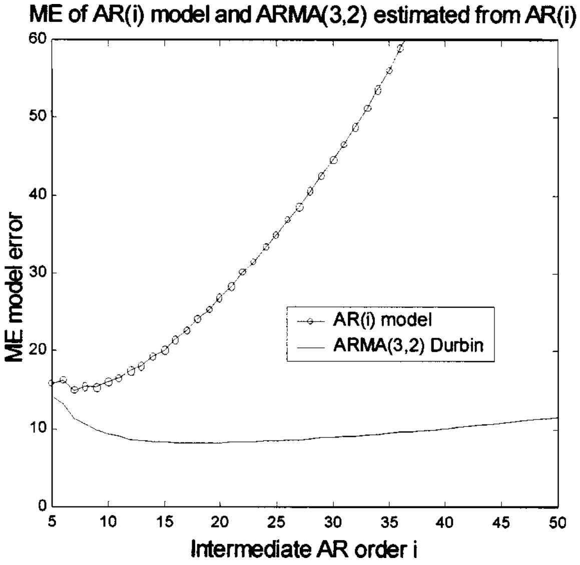

Fig. 2. Average ME of the AR($i$) model and of the ARMA$(3, 2)$ model estimated from AR($i$) with Durbin's method, as a function of the AR model order $i$. $N = 100$, average of 10 000 simulation runs.

TABLE I
LONG AR ORDERS IN THEORY AND IN SIMULATIONS AND THE MODEL ERROR ME FOR THE ARMA$(3, 2)$ MODEL AS A FUNCTION OF $N$. ARMA$(3, 2)$ IS ESTIMATED USING THE BEST LONG AR ORDER $M$ AS INTERMEDIATE ORDER. DATA HAVE BEEN GENERATED WITH THE ARMA$(3, 2)$ PROCESS OF (15)

| N | 50 | 500 | 5000 | 50000 |
|---|---|---|---|---|
| L theory | 5 | 16 | 32 | 49 |
| L' AR prediction | 5 | 15 | 32 | 49 |
| M theory | 14 | 30 | 47 | 64 |
| M' AR parameters | 12 | 32 | 51 | 68 |
| M" ARMA Durbin | 12 | 32 | 51 | 64 |
| ME of ARMA(3,2) | 9.0 | 6.6 | 5.6 | 5.3 |

higher orders. Durbin's method gives considerably better models with the new AR order $M$ defined in (13) than with the usual AR order $L$ of (12). According to simulation studies, the approximations that have been made to derive the best AR orders for prediction in (12) and for parameter accuracy in (13) are justified for AR models of ARMA processes. The (higher) order for best parameter accuracy should be used for ARMA estimation with Durbin's method.

The accuracy of the AR orders and the ME of the ARMA$(3, 2)$ models for (15) have been studied in simulations for different sample sizes. Table I gives the results. The theoretically best orders $L$ and $M$ depend on the number of observations $N$. $L'$ and $M'$, which are the best values for AR models in the simulations, are very close to $L$ and $M$, respectively, for all $N$. In addition, the best value $M''$ found for ARMA$(3, 2)$ models is close to $M$. The Cramér–aRao bound for the ME value of an ARMA$(3, 2)$ process is 5, which is equal to the number of estimated parameters [12]. It is seen that the ME of Durbin with long AR order $M$ is rather close to that lower bound. The result in this example shows that this lower bound is almost reached for $N = 50\,000$. MA examples have also been used in simulations [7], as well as many different ARMA processes. Every time, two different optimal AR approximations were found, with orders given by (12) and (13). For increasing $N$, the results of Durbin's MA method [4] and his ARMA method [5] are close to the optimal efficiency of the Cramér–Rao bound.

## V. CONCLUDING REMARKS

Theoretical values for two different best orders of a long AR model for MA and ARMA processes have been derived. The first one is the AR order that gives the best predicting AR model, and the second is characterized as the order yielding parameter estimates with the smallest mean square error. The second order is also the best intermediate AR model order in Durbin's MA and ARMA algorithms. Simulations show that MA and ARMA models calculated from the parameters of that long AR model are quite efficient. The use of the well-known best AR order for prediction as intermediate for Durbin's methods yields inefficient MA and ARMA models.

## REFERENCES

[1] J. E. H. Davidson, "Problems with the estimation of moving average processes," *J. Econometr.*, vol. 16, pp. 295–310, 1981.
[2] B. S. Choi, *ARMA Model Identification*. New York: Springer-Verlag, 1992.
[3] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.
[4] J. Durbin, "Efficient estimation of parameters in moving average models," *Biometrika*, vol. 46, pp. 306–316, 1959.
[5] ——, "The fitting of time series models," *Rev. Inst. Int. Stat.*, vol. 28, pp. 233–243, 1960.
[6] P. M. T. Broersen and H. E. Wensink, "Small sample MA estimation with Durbin's method," in *Proc. Ninth IFAC-IFORS Symp. Ident. Syst. Parameter Estimation*, Budapest, Hungary, 1992, pp. 983–988.
[7] P. M. T. Broersen, "The best order of long autoregressive models for moving average estimation," in *Signal Process. VIII, Proc. Eusipco Conf.*, Trieste, Italy, 1996, pp. 799–802.
[8] R. P. Mentz, "Estimation of first-order moving average model through the finite autoregressive approximation," *J. Econometr.*, vol. 6, pp. 225–36, 1977.
[9] D. F. Findley, "On some ambiguities associated with the fitting of ARMA models to time series," *J. Time Ser. Anal.*, vol. 5, pp. 213–25, 1984.
[10] R. R. Hocking, "The analysis and selection of variables in linear regression," *Biometrics*, vol. 32, pp. 1–49, 1976.
[11] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, 1974.
[12] P. M. T. Broersen, "The quality of models for ARMA processes," *IEEE Trans. Signal Processing*, vol. 46, pp. 1749–1752, June 1998.