PRACTICAL ASPECTS OF MOVING AVERAGE ESTIMATION

P.M.T. Broersen and H.E. Wensink

Department of Applied Physics, Delft University of Technology
P.O.Box 5046, 2600 GA Delft, The Netherlands
phone xx 31 15 786419, fax xx 31 15 786081, email broersen@duttncb.tn.tudelft.nl wensink@duttncb.tn.tudelft.nl

RÉSUMÉ

Le papier traite l'estimation des modèles MA à la base d'un nombre de données fini. Une nouvelle mesure est présentée, permettant d'evaluer la qualité des modèles MA et ARMA, estimés par des algorithmes différents. Les algorithmes non-lineaires produisent assez souvent des résultats non-interprétables. Seulement la méthode de Durbin livre toujours des résultats utilisables. Cette méthode estime les paramètres MA à l'aide d'un modèle AR intermédiaire. L'ordre optimal de ce modèle AR est discuté.

ABSTRACT

Non-linear algorithms for the estimation of MA models sometimes fail to produce useful solutions. This paper describes some reasons for problems in statistical MA estimation. The possibilities of order selection and a new measure to assessing the model accuracy for MA and ARMA modelling are discussed. Only Durbin's MA estimation method through a finite AR approximation can produce useful MA models under all circumstances. The order selection and the estimation method for the intermediate AR model is discussed.

I. INTRODUCTION

The statistical behaviour of AutoRegressive (AR) estimates in finite samples deviates considerably from the asymptotical theory [1]. Numerous simulations with Moving Average (MA) models show that also in this class of models considerable discrepancies exist between the asymptotical theory and results in finite samples.

The original mathematical problem of MA estimation comes down to a non-linear minimization of the likelihood function [2,3]. Several non-linear algorithms exist, that all fail to produce reliable results in finite samples. Some problems are [4-8]:
- non-invertible models are found: zero outside the unit circle
- non-linear optimization doesn't converge to the global solution
- some orders give a solution, others don't for the same data.

Successful results have been reported for various non-linear algorithms, when a sufficient number of observations is available to produce the desired results. For all example processes, however, a smaller sample size doesn't always produce acceptable results in simulations.

A prerequisite for an estimation method to be useful in the context of *statistical* signal processing is that it solves the parameters of models of *any* specified order. The method of Durbin [9] uses the asymptotic equivalence of MA(q) models with AR(∞) to estimate MA parameters. The advantage is that the non-linear estimation problem has been transformed into two sequential linear estimations, which always gives a solution.

In this paper, a new measure to the model accuracy is introduced: the model error ME. This is needed for an objective comparison of the performance of estimation methods and order selection criteria. It measures the quality of estimated ARMA models and can also be applied to AR or MA models. Furthermore, a histogram of optimal parameters shows peculiarities of non-linear estimation. Finally, some variants of the MA method of Durbin with an intermediate AR model are shown.

II. MODEL ERROR

To calibrate the performance of estimation and selection algorithms, an objective measure to the quality of an estimated or selected model is needed. In most applications, the model that predicts future values of the time series with the highest accuracy will be the best model. Those applications include: prediction in the time domain, parametric spectral modelling, error or change detection, classification of data into categories and quantization of parameters for speech coding. The Model Error is now derived for this purpose.

Suppose that a true ARMA(p,q) process is given by: $A(z)\ X_n = B(z)\ \varepsilon_n$, where ε_n is Gaussian white noise with variance σ_ε^2 and A(z) is defined as:

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + ... + a_p z^{-p}. \qquad (1)$$

The process is stationary if A(z) has no poles outside the unit circle. Estimation from observations X_n yields a model with parameters A'(z) and B'(z) of orders p' and q' respectively, not necessarily equal to p and q. Substituting those parameters into new data Y_n, the result can be written as: $B'(z)\eta_n = A'(z)\ Y_n$, where the errors η_n can be seen as the output of the estimated model with Y_n as input, B'(z) as AR part and A'(z) as MA part. The squared error of prediction PE(p',q') is defined as the expectation $E[\ \eta_n^2\ ]$, if Y_n is independent of the signal X_n that had been used to estimate the parameters.

Now suppose that the data Y_n have been generated with the process given by: $Y_n = B(z)/A(z)\ \varepsilon_n$. The output η_n of the

model with A'(z) and B'(z) can be expressed as $\eta_n = A'(z)/B'(z) Y_n$, yielding $\eta_n = [A'(z)\ B(z)]/[A(z)B'(z)]\varepsilon_n$. So the relation between the prediction error of the model and the innovations that generated the process is given by the ARMA(p+q', p'+q) process:

$$\begin{aligned} &\eta_n + c_1\ \eta_{n-1} + \ldots + c_{p+q'}\ \eta_{n-p-q'} = \\ &\varepsilon_n + d_1\ \varepsilon_{n-1} + \ldots + d_{p'+q}\ \varepsilon_{n-p'-q}\ , \end{aligned} \qquad (2)$$

where the parameters of C(z) are given by the product A(z)B'(z) and D(z) equals A'(z)B(z). By considering (2) as a filtering operation [10], the prediction error PE is the variance of η_n. Åström [11] has given a solution for a similar problem.

The output of the ARMA process in (2) can be found by separating it into 2 consecutive filters with an intermediate signal v_n as:

$$\begin{aligned} &v_n + c_1\ v_{n-1} + \ldots + c_{p+q'}\ v_{n-p-q'} = \varepsilon_n \\ &\eta_n = v_n + d_1\ v_{n-1} + \ldots + d_{p'+q}\ v_{n-p'-q}\ . \end{aligned} \qquad (3)$$

The covariance $R_{vv}(k)=E[v_n v_{n+k}]$ can be computed with the well-known Yule-Walker relations for autoregressive processes [11]. Afterwards, the variance of η_n follows as $E[\eta_n^2]$, so

$$\sigma_\eta^2 = (1 d_1 .. d_{p'+q}) \begin{pmatrix} R_{vv}(0) & R_{vv}(1) & . . & R_{vv}(p'+q) \\ R_{vv}(1) & R_{vv}(0) & . . & R_{vv}(p'+q-1) \\ . & . & . . & . \\ . & . & . . & . \\ R_{vv}(p'+q) & R_{vv}(p'+q-1) & . . & R_{vv}(0) \end{pmatrix} \begin{pmatrix} 1 \\ d_1 \\ . \\ . \\ d_{p'+q} \end{pmatrix} \qquad (4)$$

If the order of D(z) is greater than that of C(z), the extra terms of $R_{vv}(k)$ can always be computed with the first line of (3), extrapolating the autoregressive covariance function [10].

Finally, the Model Error ME is defined with (4) as:

$$ME = \sigma_\eta^2 \,/\, \sigma_\varepsilon^2 - 1\ . \qquad (5)$$

This model error is a scaled version of the prediction error $E[\eta_n^2]$. ME will be positive if the poles of C(z) are inside the unit circle. It can only be zero if all parameters in C(z) and D(z) are exactly equal, otherwise it is greater. It is a useful measure to the model accuracy, with significance in time and frequency domain. It simplifies to NATFE [1] when MA parts are absent and it can also be used for the fit of MA models by using zeros for the AR parameters. This distance measure can easily be computed in simulation studies where the true parameters of the generating process are known.

III. Computational Aspects and Least Squares

In applying the non-linear methods for MA estimation to small samples, several problems arise [7]; we discuss here the model fit, the location of estimated zeros and the irregular behaviour of the residual variance.

a) Fit of MA models to AR data and AR models to MA data.

The theoretically best fitting AR(p') model to a given MA process is found by computing the theoretical covariances of the MA process and to use the first p' true covariances for computation of the AR parameters, with the Yule-Walker equations [10]. The best fitting MA(q') model to an AR process is found by a non-linear minimization of the Model Error as a function of the MA parameters. No simpler way to compute that MA model has been found. Numerical computations show that AR(p) models fit better to MA(1) processes than vice versa. However, this cannot be generalized to higher order process. Therefore, simulations with MA(1) processes [5,6] are too limited and this paper gives simulation results for MA(2) and MA(3) processes.

b) Zeros and the unit circle

The poles of an AR polynomial are all inside the unit circle if the reflection coefficients or partial correlation coefficients are less than 1 in magnitude [10]. The zeros of a MA polynomial can be studied in the same way, although, of course, the interpretation in terms of partial correlations is lost. When zeros are estimated that are close to the unit circle, it is possible that the true process had such a zero and consequently the model has a small ME. However, it is also possible that the estimated zero was caused by the statistical inaccuracy, which leads to an extremely high ME. As a consequence, estimation methods that can yield estimated models with zeros on the unit circle for processes without such zeros are not useful in practice.

c) Residual variance

The residual variance measures the fit of a model to given data. Unfortunately, its behaviour in finite samples has some peculiarities, as will be shown. Several definitions are available [2], with zero residuals or with backforecasted residuals before the data interval, used in Conditional (CLS) and Unconditional (ULS) Least Squares respectively [5]. Backforecasting the residuals with index 0 to -q'+1 is carried out by running the estimated filter backwards over the data [2]. With those initial backforecasted residuals, the fit of an arbitrary MA(q') model to given data x_i is computed as:

$$\begin{aligned} &\hat{\varepsilon}_{-q'+1} \ldots \hat{\varepsilon}_0 \quad \textit{backforecasted} \\ &\hat{\varepsilon}_i = x_i - b_1' \hat{\varepsilon}_{i-1} - \ldots - b_{q'}' \hat{\varepsilon}_{i-q'}\ , \qquad i = 1,\ldots,N\ ; \\ &\hat{\sigma}_\varepsilon^2(q') = \frac{1}{N} \sum_{i=1}^{N} \{ x_i - b_1' \hat{\varepsilon}_{i-1} - \ldots - b_{q'}' \hat{\varepsilon}_{i-q'} \}^2\ . \end{aligned} \qquad (6)$$

Simulations have shown that backforecasting gives a better reconstruction of the true excitation signal than with zeros in front. Fig.1 gives the residual variance as a function of the model parameter for a MA(1) process, with and without backforecasting. Based on this figure and many others, it may be concluded that, at least for small samples, say N less than 50 or 100

- the behaviour of the residual variance as a function of the model parameter can be quite irregular

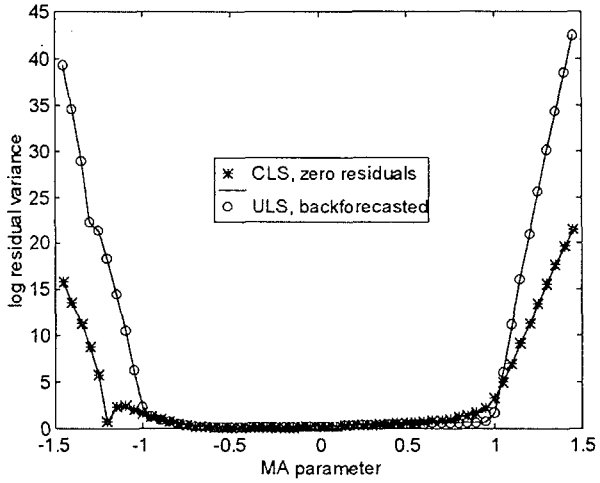

Figure 1 Logarithm of residual variance of a MA(1) process as a function of the MA parameter, N=30, b= - 0.7.

- the global minimum may be found outside the invertible region
- minima may be found on or near -1 or +1, although the model error ME becomes ∞ for a model with zeros on the unit circle.
- sometimes no non-linear solution can be found within a reasonable number of iterations.
- examples have been found with the global minimum within the interval -1 to +1, with only a local minimum or without any minimum. Constraining the solutions to parameters within the unit circle will then give exactly that constraint as minimum, so that technique doesn't provide sensible models.

Other non-linear methods, like the method of moments [2] or Godolphin's method [3] also give problems [7]. We didn't find a non-linear method without difficulties when applied to data records with less than 100 observations.

The behaviour of the residuals has been investigated further by determining the parameter value for which the minimum of the residual variances is found, for 2000 simulation runs in Fig.2.

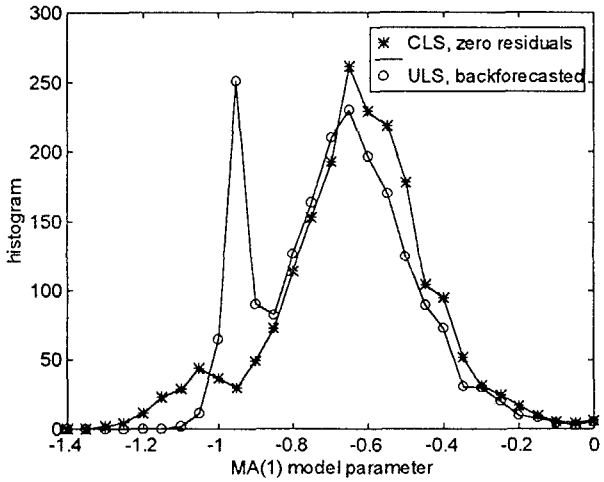

Figure 2 Histogram of estimated MA (1) parameters for the same MA (1) process of fig.1, with N = 30, b = - 0.7.

The histogram shows the known tendency of backforecasted residuals to produce a minimum near the unit circle [5]. Moreover, the variance as computed with the histogram is much greater than the asymptotical theoretical result for the variance being $(1-b^2)/N$ [2]. The variance for the parameter value of a white noise process estimated with ULS has been determined in simulation experiments for different sample sizes; the theoretical variance is 1/N. It was 14.2/N for N=5, 5.6/N for N=10, 3.1/N for N=20, 1.7/N for N=30, 1.1/N for N=60 and the asymptotical 1/N was found for N>100. So the variance obtained in small sample simulations is much greater than predicted by the asymptotical theory. This explains partially why MA parameters produced by non-linear methods will often lie outside the invertibility region.

IV. DURBIN's METHOD

Durbin's method for the estimation of MA models [9] starts with the estimation of a long intermediate AR model from the data. Afterwards, the estimated AR parameters are used as data in a second AR estimation procedure to find the MA parameters. The method is based on the asymptotical equivalence of AR(∞) and MA(q) processes.

Current issues of research are:
- the best estimation method for the intermediate AR model
- the best intermediate AR model order
- selection criteria for the MA model order.

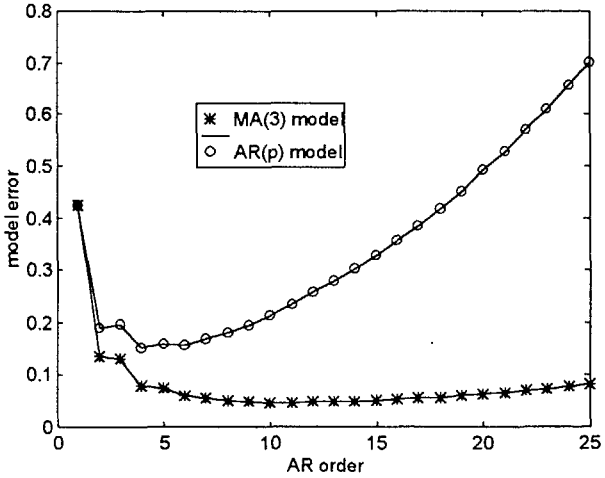

Figure 3 Average Model Error ME of 1000 AR(p) and MA(3) models of a MA (3) process, as function of p. N = 60.

Figure 3 shows the ME for AR(p) models and for MA(3) models that are estimated based on the parameters of that AR(p) model. It is clear that the best fitting AR model with smallest ME is of order 4. The best MA model, however, is found from the AR(11) model, that has nothing special in the figure. When AR models with higher or lower orders are used to estimate the MA(3) model, the resulting ME is higher. It has been demonstrated before [7] that the best intermediate AR model orders depend on the process that originally generated the data. Similar simulations show that the best intermediate AR order for a given MA process depends furthermore on the sample size. So the AR order has to be selected.

Simulations can give a good survey of the performance of methods. As MA(1) processes are not in every respect representative, we will give the results of simulations with various MA(2) processes, with parameters $b_1=\beta(1+\beta)$ and $b_2=\beta$, which is equivalent to two equal "reflection coefficients" of magnitude β, for $-1<\beta<1$. Autoregressive model orders have been selected with the newly developed finite sample information criterion, FSIC, that outperforms all other AR order selection criteria [12]. MA orders have been selected with AIC(3), which is the usual AIC criterion with penalty 3 instead of 2, because that penalty gives a good compromise between selecting too many and too few parameters in the model, at least for AR models [13]. The MA model order has been selected from the range 0 to 6, the AR model order from the range 0 to N/2, where the number of observations N was 60 in the simulations. It is possible to consider AR orders as high as N/2 for selection if the FSIC criterion is used for selection of the intermediate AR order. For other AR selection criteria, the highest possible AR order has to be constrained in advance, which is in contrast with the theoretical equivalence of MA(p) and AR(∞). The average model error of a number of simulations is given for estimated models with 10 different methods.

The following methods and models are compared in Table 1:

A MA model from AR[MA order +1× AR(FSIC)] Burg model
B MA model from AR[MA order + 2× AR(FSIC)] Burg model
C MA model from AR[MA order + 3× AR(FSIC)] Burg model
D non-linear model of with method **B** selected MA order
E MA model from AR[MA order + 2× AR(AIC)] Burg model

F MA model from AR[MA order+2×AR(FSIC)] Yule-Walker
G MA model estimated from AR(6) Burg model
H MA model estimated from AR(N/2) Burg model
I AR(FSIC) model computed with Yule-Walker
J AR(FSIC) model with the Burg method

AR models have been included in **I** and **J** for a comparison of the accuracy of AR and MA modelling. The sliding window technique for selecting an intermediate order [7] has been used in **A, B, C, E,** and **F**. It firstly selects the AR model order; the order of the MA model that is currently computed is added [7]. The sum of the current MA order plus one (**A**), two (**B,E,F**) or three (**C**) times the selected AR order is the intermediate AR order that has been used for the sliding window technique. The current MA order is added to be sure that the intermediate AR order is at least equal to that order; factors 1, 2 and 3 are investigated for the selected AR order because it is known from figure 3 that the best intermediate AR order in MA estimation may be higher than the AR order with smallest ME. The Selection Error SE is defined as N×ME from (5).

Table I *Average Selection Error SE of 10 different estimated models from 60 MA(2) observations as a function of the MA parameter β; $b_1 = \beta(1+\beta)$ and $b_2 = \beta$.*

β	-.9	-.6	-.3	0	.3	.6	.9
A	9.60	4.58	4.78	*1.04*	4.85	5.63	9.82
B	***7.54***	**4.75**	**5.10**	**1.45**	**5.28**	**5.34**	**8.17**
C	7.78	5.75	5.67	1.33	5.89	5.88	*7.98*
D	52.1	47.0	29.9	26.0	23.5	49.9	25.1
E	9.39	8.23	6.61	1.41	6.96	7.40	8.81
F	17.6	4.64	4.70	1.05	4.87	4.88	9.16
G	8.56	*3.49*	*4.63*	1.20	*4.40*	*4.71*	10.2
H	10.2	12.2	10.9	2.19	9.81	9.75	9.11
I	24.6	8.95	4.71	2.14	5.29	10.8	19.3
J	19.3	9.65	5.38	2.29	5.83	11.9	20.7

Taking the average of a certain method over the different values of β, the sequence in decreasing quality gives:

G - B - C - A - F - E - H - J - I - D.

Non-linearly estimated MA models **D** and AR models **I** and **J** have the highest SE. Fixed AR orders 6 and 30 in **G** and **H** show that a fixed order may be a good or a bad choice, depending on the (in practice unknown) true process parameters. **E** is always worse than **B**, so FSIC should be used for AR order selection. Intermediate AR models found with Yule-Walker in **F** are seen to perform badly for high absolute values of β, so if the zeros of the true MA process are close to the unit circle. Many more simulations are necessary to give a definite answer which method is to be preferred. So far, we never found a single example where the sliding window methods **A, B** and **C** performed badly; for all other methods such examples have been found and an explanation can be given. For statistical processing, it is important to have a method that will always, for all types of data, perform with an acceptable quality. The price to be paid is that this method will not always be the very best for one given process and sample size.

V. Conclusions

The model error is a good measure for the evaluation of the quality of estimated models in simulations, AR, MA or ARMA. It provides an objective measure to compare the performance of estimation methods and order selection criteria.

Non-linear least squares gives no good models in simulations with N=60 (method **D**), even if the order is selected previously with another method or if only MA(2) models are considered.

MA models, estimated with one of the variants of Durbin's method have a smaller model error ME than AR models estimated from the same MA(2) processes (methods **I** and **J**).

The Burg method **B** for AR estimation is advisable for intermediate AR models. The Yule-Walker method **F** gives often a slightly better result, but the model error becomes very great when the zeros of the generating process are close to the unit circle, for β=0.9.

For the second stage in Durbin's method, estimating the MA parameters from the AR model, only the Yule-Walker technique can be used if one requires invertible models.

References

[1] Broersen, P.M.T. and H.E. Wensink, "On finite sample theory for autoregressive model order selection", *IEEE Trans. on Signal Processing*, Vol 41, p 194-204, 1993.

[2] Box, G.E.P. and G.M. Jenkins, *Time Series Analysis, Forecasting and Control.* Revised edition, San Fransisco: Holden Day, 1976.

[3] Godolphin, E.J. and J.G. de Gooijer, "On the maximum likelihood estimation of the parameters of a Gaussian moving average process", *Biometrika*, 69, p 443-451, 1982.

[4] Davidson, E.H., "Problems with estimation of moving average processes", *Journal of Econometrics*, 16, p 295-310, 1981.

[5] Nelson, C.R., "The first-order moving average process", *Journal of Econometrics*, 2, p 121-141, 1974.

[6] Cryer, J.D. and J. Ledolter, "Small-sample properties of the maximum likelihood estimator in the first-order moving average model", *Biometrika*, 68, p 691-694, 1981.

[7] Broersen, P.M.T. and H.E. Wensink, *"Small sample MA estimation with Durbin's method"*, Selected papers from the 9-th IFAC-IFORS Symposium on Identification and System Parameter Estimation, Budapest 1991, Oxford: Pergamon, p 983-98, 1992.

[8] McClave, J.T., "A comparison of moving average estimation procedures", *Comm.in Statistics*, 3, p 865-883, 1974.

[9] Durbin, J., "Efficient estimation of parameters in moving average models", *Biometrika*, 46, p306-316, 1959.

[10] Priestley, M.B., *Spectral Analysis and Time Series*, New York: Academic Press, 1981.

[11] Åström, K.J., E.I. Jury and R.G. Agniel, "A numerical method for the evaluation of complex integrals", *IEEE Trans. on Automatic Control*, AC-15, p 468-471, 1970.

[12] Wensink, H.E. and P.M.T. Broersen, *"Estimating the Kullback-Leibler information for autoregressive model order selection in finite samples"*, Signal Processing VII: Theories and Applications, Proc. Eusipco Conf., Edinburgh, p 1847-1850, 1994.

[13] Broersen, P.M.T. and H.E. Wensink, "The influence of the penalty factor in order selection criteria", *Proc. IMTC conference*, Hamamatsu, p 1459-1462, 1994.