

Windowed Periodograms and Moving Average Models

Piet M.T. Broersen and Stijn de Waele

Department of Applied Physics, Delft University of Technology
P.O.Box 5046, 2600 GA Delft, The Netherlands
Phone +31 15 2786419 Fax +31 15 2784263 email broersen@tn.tudelft.nl

ABSTRACT

A windowed and tapered periodogram can be computed as the Fourier transform of an estimated covariance function of tapered data, multiplied by a lag window. Covariances of finite length can also be modeled as moving average (MA) time series models. The direct equivalence between periodograms and MA models is shown in the method of moments for MA estimation. A better MA representation for the covariance and the spectral density is found with Durbin's improved MA method. That uses the parameters of a long autoregressive (AR) model to find MA models, followed by automatic selection of the MA order. A comparison is made between the two MA model types. The best of many MA models from windowed periodograms is compared to the single selected MA model obtained with Durbin's method. The latter typically has a better quality.

Keywords: spectral estimation, order selection, spectral distance, spectral window, spectral error

1. INTRODUCTION

Time series analysis or parametric spectral estimation is a modern perspective for the non-parametric approach with tapered and windowed periodogram estimates [1,2,3]. Time series models can be subdivided into three model types: AR or AutoRegressive models, MA or Moving Average models and the combined ARMA models [1]. Theoretically, most stationary stochastic processes can be expressed as an unique AR(∞) or MA(∞) process [1]. In practice, finite order MA or AR models for those infinite order processes are often sufficiently accurate, because the true parameters of higher orders decrease rapidly for most processes. The gain in a reduced bias of the model by including such small parameters is less than the increased variance by estimating them. A positive semi-definite covariance with maximum shift Q can be described exactly with a MA(Q) model [2]. Therefore, finite covariance functions estimates with the periodogram as their Fourier transform can be expressed exactly as covariances of finite order MA models.

Maximum likelihood estimation of MA parameters is non-linear [1,4]. The likelihood function is symmetric with respect to the unit circle and exact maximum likelihood estimation has a tendency to find roots of the MA polynomial exactly on the unit circle [5]. The method of moments [2] is a simple non-linear MA algorithm that is not efficient. It relates a positive semi-definite covariance estimate of length q to a MA(q) model. In this way, the periodogram as the

Fourier transform of the covariance can be considered as coming from a MA model. It is known that this representation of the covariance is not a sufficient estimator for the MA parameters. A robust MA algorithm exists which estimates the model directly from a long AR model of the data. Durbin's method [6] never has problems with convergence. It estimates always invertible models by using the parameters of a long autoregressive model in a linear MA estimation procedure; invertible models have all zeros inside the unit circle. This MA method has been evaluated as inefficient in the literature in the past, and it was almost forgotten. However, use of the recently discovered optimal length for the intermediate AR model for Durbin's methods [7] significantly improved the quality of the estimated MA models. In simulations, the performance is now close to the Cramér-Rao lower bound for the achievable accuracy.

The prediction error as a quality measure allows a mutual comparison of AR, MA and ARMA models. The Model Error (ME) [8] is a scaled version of this prediction error. It is an objective quality measure for time series models. It can also be used for the quality of windowed periodograms, by expressing the periodogram as a MA model.

The paper summarizes known properties of estimated periodograms and of MA models. An example shows that tapering in periodogram estimation has the advantage that the bias at frequencies with low power is diminished. MA models obtained from periodogram estimates with different window lengths are compared to directly estimated MA models of the same length, and also to a single MA model with an automatically selected model order.

2. PERIODOGRAMS

Two classical spectral estimation methods are based on Fourier analysis of the data. The first is the periodogram, which is the squared absolute value of the Fourier transform; the second is the spectral estimate obtained by tapering the data, followed by windowing the covariance or by smoothing the raw periodogram [3]. Those non-parametric approaches together will be called (tapered and windowed) periodograms in the sequel. The periodogram is unbiased asymptotically, but the Fejer kernel gives a *bias* in finite sample *spectral* estimates that is also present as the *triangular* bias of the covariance in the *time domain* [1]. It is caused by dividing the covariance sum of $x_n x_{n+k}$ by N instead of $N-k$ to obtain a positive semi-definite covariance estimate. The bias is also called spectral leakage and its magnitude is of order $O(\log N/N)$ for N observations [1]. The variance of the raw periodogram is approximately

equal to the square of the expectation and not diminishing with increasing N [1]. Estimates at distances greater than $2\pi/N$ are statistically almost independent [1].

To diminish leakage, a taper or data window can be applied to the data. Instead of using the data x_n they are multiplied by the taper w_n to give $x_n w_n$. A popular taper for stochastic processes is a raised cosine, which equals 1 in the middle 80 % of the data and uses a cosine as envelop from 1 to 0 in the first and last 10 % of the data. The *theoretical* treatment of the effect of tapers is limited. More attention is given to the variance reduction obtained by spectral windows or lag windows, applied to the raw periodogram or equivalently to the covariance estimate [1], known as smoothing. Different window types have been described [1]. The optimal length of a window is a compromise between additional bias and reduction of the variance. A serious drawback is that an optimal compromise between bias and variance of windows requires the *a priori* knowledge of the true spectrum. For unknown spectra, some subjective choice must be made for the window length. Both choosing the window length as a fixed proportion of N and “window closing”, varying the window size until the estimate is considered to be smooth enough, are not optimal. However, those techniques are often used in practice for want of something better [1].

The periodogram bias, caused by the triangular window on the estimated covariance, is also present in the Yule-Walker (or autocorrelation) method of AR parameter estimation, which uses the same biased covariance [9]. If the absolute value of the p^{th} reflection coefficient is $1-p/N$, then all reflection coefficients of order higher than p are estimated at least a factor 2 wrong [9]. In AR estimation, this is a sufficient reason to prefer other estimation algorithms with a much smaller bias. It is known that derivatives of the covariance at zero shift have a relation with especially higher frequencies in the spectrum. The triangular bias in periodograms introduces a discontinuity in the first derivative of the estimated covariance. The use of tapers suppresses this bias. However, tapers give a distortion of the measured data. Direct estimation of MA models from data or from a long AR model (if estimated with Burg’s method) does not suffer from this distortion.

3. MA MODELS

Time series models are parametric models for stationary random signals. Both the spectrum and the covariance can be computed from the parameters of a time series model, which can be seen as a parametric alternative for periodograms. The most general time series model is ARMA; a theoretical ARMA(p,q) process is written as:

$$\mathbf{A}(z)\mathbf{x}_n = \mathbf{B}(z)\boldsymbol{\varepsilon}_n, \quad (1)$$

with $\mathbf{A}(z)$ as a shorthand notation for $1+a_1z^{-1}+\dots+a_pz^{-p}$ and $\mathbf{B}(z)$ for $1+b_1z^{-1}+\dots+b_qz^{-q}$, with $z^{-1}x_n=x_{n-1}$; $\boldsymbol{\varepsilon}_n$ represents a series of independent identically distributed stochastic variables with variance σ_ε^2 : a white noise process. This ARMA process reduces to an AR process for $q=0$ and to a pure MA process for $p=0$.

Maximum likelihood estimation of the MA model is non-linear. Many MA algorithms have convergence problems, or they converge to non-invertible models with some zeros of $\mathbf{B}(z)$ in (1) estimated on or outside the unit circle [4,5]. Durbin’s method for MA estimation gives all zeros inside [4]. By carefully selecting an order for the long AR model and by using the Burg estimation method, the performance of Durbin’s algorithm has been much improved recently [7, 10]. It now yields satisfactory MA models, also for finite or small samples.

AR modeling is the backbone of time series analysis in practice. For a few observations, the best AR model order is often greater than $0.1N$, where N is the number of observations. This requires care, because the outcome of AR estimation depends on the algorithms used in those finite sample conditions where the order is not small in comparison to N [11]. The AR method of Burg estimates reflection coefficients for increasing model orders [2], thus making sure that the model will be stationary, with all roots of $\mathbf{A}(z)$ in (1) within the unit circle. Asymptotical AR order selection criteria will often select wrong orders if candidate orders are higher than $0.1N$ [11]. Therefore, the order K^* of AR models computed with Burg’s method is selected with a finite sample order selection criterion $\text{CIC}(p)$ [12].

The long AR order, chosen with a sliding window algorithm as basis for $\text{MA}(q^*)$ estimation with Durbin’s method, is twice the (with $\text{CIC}(p)$ selected) AR order K^* plus the number q^* of MA parameters that is to be estimated [7]: $\text{MA}(q^*)$ is computed from $\text{AR}(2K^*+q^*)$. MA models are computed for a large number of orders q^* between 1 and a maximum order L . The order q is selected as the order with the minimum of the asymptotical selection criterion $\text{GIC}(q,3)$ [13] which is defined as:

$$\text{GIC}(q,3) = \ln\{\text{RES}(q)\} + 3q / N. \quad (2)$$

$\text{RES}(q)$ is the residual variance after estimating the $\text{MA}(q)$ model. The penalty factor α in $\text{GIC}(q,\alpha)$ handles the compromise between *biased* underfit models and overfit models with increased *variance*. The compromise 3 for the penalty is based on asymptotical arguments. It has a good performance for practical data. This improved MA algorithm for estimation and order selection gives good results when applied to different types of data [6,7,10,13].

The method of moments is an old MA method, based on elementary statistical rules, but it is known to be inefficient [2]. The first q correlations do not yield a sufficient statistic for the $\text{MA}(q)$ model. The correlation function of a $\text{MA}(q)$ process is given by:

$$\rho(r) = \sum_{s=|r|}^q b_s b_{s-|r|} / \sum_{s=0}^q b_s^2, \quad 0 \leq |r| \leq q; \rho(r) = 0, \quad |r| > q. \quad (3)$$

The method of moments uses the inverse relation, finding q estimated parameters b_i from an estimated correlation or covariance of length q . An iterative algorithm solves the factorization of a covariance function to a moving average model [14]. If a periodogram is computed for at least $2N-1$ frequencies, the biased covariance estimate can be recovered exactly by an inverse Fourier transform [1].

Windowed or smoothed spectra can be obtained by multiplying (a part of) the covariance with a lag window. With the method of moments, this windowed covariance can be represented with the parameters of an MA model. This inefficient MA model obtained from some *covariance* estimates will be compared to MA models, obtained from a long *AR model of the data* with the method of Durbin.

The prediction error PE is defined as the squared one step ahead prediction error in applying a model to new and independent data [11]. It is a measure for the accuracy of time series models. Its asymptotical expectation for unbiased MA(q) models is given by:

$$E[PE] = \sigma_\varepsilon^2 (1 + q / N), \quad (4)$$

where q is the number of MA parameters. The model error ME is a scaled version of the prediction error:

$$ME = N (\text{PE} / \sigma_\varepsilon^2 - 1). \quad (5)$$

The ME can be computed with an efficient expression in the time domain [8]. For unbiased models, the asymptotical expectation of ME equals the number of estimated parameters, independent of the number of observations.

4. SIMULATIONS

Simulations have been carried out with some ARMA(5,4) processes, with all poles and all zeros at the same radius β , but at different phase angles. The 5 AR parameters are calculated with the Levinson-Durbin recursion by taking $(-\beta)^i$ for the 5 AR reflection coefficients and the 4 MA parameters by taking 4 reflection coefficients $(\beta)^i$ for $B(z)$.

Fig.1 compares periodograms estimated with and without taper. As taper, a raised cosine has been used, leaving the middle 80 % undisturbed. As lag or spectral window, a Parzen window [1] has been used, with length $N/2$. It is clear in Fig.1 that a taper strongly reduced the influence of the triangular bias on the periodogram estimate. Using a Parzen window of length $N/2$ on the *true* spectrum has almost no influence on the quality of the spectrum: the ME would only be 0.07. The triangle on the true covariance (equivalent with the Fejer kernel on the spectrum) together with a Parzen window gives a poor

Power spectrum of ARMA(5,4) process with $\beta = 0.8$, $N = 512$

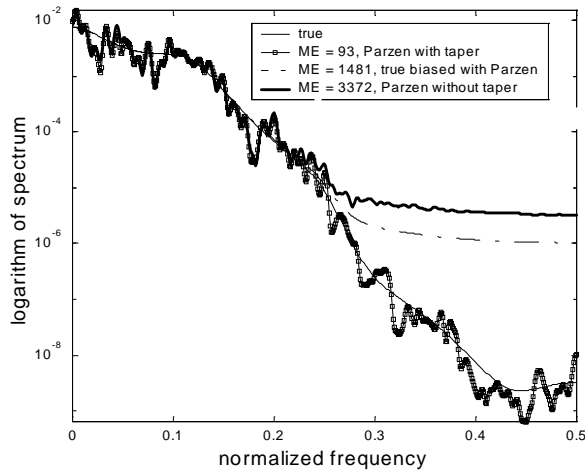


Fig.1 Influence of a taper on a windowed periodogram.

accuracy and the high ME value 1481 in Fig.1, as the broken line for “true biased with Parzen”. Without the use of a taper, the estimated periodogram denoted “Parzen without taper” follows more or less the same distortion with $ME=3372$. With a taper, the estimated spectrum “Parzen with taper” remains close to the true spectrum with $ME=93$. It shows a large variance but a small bias. The triangular bias or leakage is important in spectra with a large dynamic range, but it has much less influence in less pronounced spectra. For $\beta=0.5$, the ME of the true spectrum with triangular bias becomes only 0.13 instead of the value 1481 that was found for $\beta=0.8$.

The simulations have been carried out with two types of spectral window: the Parzen window and the Bartlett triangular window. In all results with different β and N , the quality obtained with the Parzen window was much better than with Bartlett. The ME value for a triangular window of length N is 1481 in Fig.1, the same ME value as given for the true biased covariance, including the Parzen window. This shows that the Parzen window of length $N/2$ or N itself contributes almost zero to the ME for $N=512$. A triangle of length $N/2$ instead of the Parzen window would give $ME=2599$. This was found in theoretical computations and agrees with simulation results. The more favorable ME results with Parzen windows will mostly be used in a comparison with MA models that are estimated from the data with the improved algorithm of Durbin [6,7,10,13].

Windowed periodograms can be formulated as MA models calculated with the method of moments, applied to windowed covariances. The ME of those MA models in Table 1 is compared with the quality of more efficiently estimated MA models in Table 2. By taking different values of β , different types of spectra are created, from almost white noise to the spectrum of Fig.1 and from emphasis on low frequencies to a strong high frequency part. The processes with different β are also characterized by the ratio $\sigma_x^2 / \sigma_\varepsilon^2$ which is 228.2 for $|\beta|=0.8$, 3.8 for $|\beta|=0.5$ and 1.04 for $|\beta|=0.1$.

Table1: The ME of windowed periodograms for $N=2048$ as a function of the window length L for various values of β in an ARMA(5,4) process

| β / L | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 512 | 1024 |
|-------------|-------|------------|------|------------|------|-------------|------|-----|------|
| -0.8 | 72290 | 17643 | 5000 | 1349 | 260 | 62.1 | 89.5 | 173 | 350 |
| -0.5 | 731 | 74.8 | 9.8 | 9.5 | 19.0 | 38.9 | 79.4 | 164 | 340 |
| .1 | 7.7 | 2.5 | 4.3 | 9.1 | 18.7 | 38.5 | 79.6 | 165 | 343 |
| .5 | 733 | 75.8 | 10.0 | 9.4 | 18.7 | 38.4 | 79.4 | 165 | 343 |
| .8 | 72154 | 17605 | 5003 | 1347 | 249 | 49.0 | 78.4 | 164 | 343 |

Table2: The ME of estimated MA models with Durbin's method for $N=2048$ as a function of the fixed model order q for various values of β in an ARMA(5,4) process

| β / q | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 512 | 1024 | sel |
|-------------|------------|------------|-------------|------|------|-----|-----|-----|------|-------------|
| -0.8 | 3670 | 38.3 | 19.5 | 36.2 | 71.8 | 141 | 286 | 627 | 1806 | 25.1 |
| -0.5 | 8.9 | 8.1 | 16.4 | 33.1 | 65.3 | 132 | 270 | 598 | 1771 | 7.1 |
| .1 | 4.1 | 8.1 | 16.1 | 32.5 | 64.6 | 131 | 268 | 596 | 1761 | 2.6 |
| .5 | 8.8 | 8.1 | 16.2 | 32.9 | 65.4 | 132 | 272 | 603 | 1801 | 6.9 |
| .8 | 3660 | 36.6 | 17.6 | 34.1 | 68.6 | 137 | 280 | 619 | 1815 | 21.0 |

For the models estimated with Durbin's method, the quality of the model *selected* with (2) is given in the last column of Table 2, denoted *sel*. The smallest value for ME in Table 2 is always lower than the smallest value in Table 1 for the same β . The exception is for $\beta=0.1$, but here the smallest ME for Durbin's method, $ME = 2.0$, is found for $q=2$, which is not presented in the Tables. In other words, using Durbin's MA method outperforms the MA method of moments, that is equivalent with windowed periodograms. Still more important, the column *sel* shows that it is possible with Durbin's MA method to select automatically a model order with has almost the same accuracy as the best fixed order in the previous columns in Table 2. In contrast, no automatic or objective choices exist for periodogram based estimates. Only rules based on the a priori knowledge of the spectral density exist [1]. "Window closing" by looking at the influence of the window length on the spectra is highly subjective and misleading. The accuracy for periodograms with high L in Table 1 is better than for MA(q) models with the same high order in Table 2. This is a consequence of the reduction of the *variance* in Table 1 by the lag window in situations where *bias* is no longer a problem. But only the accuracy of the best model order or window length really matters. For small values of q, the bias is much greater for the windowed periodograms

For greater L and q, both Tables show ME values that are almost independent of β . Bias is no problem then, and the ME is determined by the variance. For q less than N/4, the ME of Durbin's method is close to the model order q. So the asymptotical theory applies for the ME of (5). Only for model orders q as high as N/2, the ME becomes considerably greater than q. Finite sample effects have been described for AR processes where the deviation from the asymptotical theory is noticeable for orders above N/10 and special order selection criteria are necessary. This is not the case for MA(q) models where the asymptotically based order selection criterion (2) can be used until $q=N/4$.

Table 3 shows the behavior for two window types as a function of the sample size. For the periodogram estimates with Parzen and Bartlett windows, the best choice between the windows of length N/2, N/4, N/8 until smallest length 4 is given in Table 3, for each N. Always, the Parzen window is better than Bartlett and Parzen requires a shorter window length for the best accuracy. The differences are more pronounced for greater values of β . The selected MA model with Durbin's MA method is for $\beta < 0.5$ comparable with the best possible choice for windowed periodograms. The best MA order increases with the sample size, as might be expected for MA models of an ARMA process. The

Table3: The ME and the *length* of the best periodograms with Parzen and Bartlett window and of fixed and selected order MA models as a function of N, $\beta=0.5$ in ARMA(5,4)

| N | 32 | | 128 | | 512 | | 2048 | |
|------------|------|-----|------|-----|------|-----|------|-----|
| Parzen | 5.3 | 8 | 8.1 | 16 | 5.8 | 16 | 9.4 | 32 |
| Bartlett | 10.0 | 16 | 14.3 | 32 | 25.9 | 64 | 42.9 | 64 |
| MA Durbin | 6.8 | 2 | 5.2 | 4 | 5.3 | 4 | 8.1 | 8 |
| sel Durbin | 8.2 | 2.5 | 7.6 | 3.7 | 6.4 | 4.5 | 6.9 | 4.8 |

advantage of automatic MA order selection for Durbin's method is not seen in this table. Estimated MA models are better than periodograms to find an accurate description of the spectral density and the covariance. In the examples given, the best estimated covariance is found by estimating a MA model rather than by using shifted data products, which belongs to the periodogram estimates. Using an AR or MA or ARMA model, with *selected* model type [13], will generally give a still better model than only MA.

5. CONCLUDING REMARKS

Windowed and tapered periodograms can be expressed as MA models, estimated with the method of moments. The model error ME can be used for an objective comparison of periodograms and MA models. Typically, the accuracy of the spectral density of the automatically selected MA model, computed with Durbin's improved method, is better than the best of many windowed and tapered periodogram estimates. For accurate estimates of the spectral density or the covariance function, time series models are better than periodograms or shifted data products respectively.

REFERENCES

- [1] M.B. Priestley, "Spectral Analysis and Time Series", London, Ac. Press, 1981.
- [2] S.M. Kay and S.L. Marple, "Spectrum Analysis-A Modern Perspective", *Proc. IEEE*, vol. 69, pp. 1380-1419, 1981.
- [3] P. Stoica and R. Moses, "Introduction to Spectral Analysis", Upper Saddle River, New Jersey, Prentice Hall, 1997.
- [4] P.M.T. Broersen and H.E. Wensink, "Small Sample MA Estimation with Durbin's Method", *Selected Papers from the 9th IFAC/IFORS Symposium on Identification and System Parameter Estimation*, Oxford: Pergamon, pp. 983-988, 1992.
- [5] P.M.T. Broersen and S. de Waele, "Empirical Time Series Analysis and Maximum Likelihood Estimation", *Proc. 2nd IEEE Benelux Signal Proc. Symp.* SPS-2000, pp. 1-4, 2000.
- [6] J. Durbin, "Efficient Estimation of Parameters in Moving Average Models", *Biometrika*, vol. 46, pp. 306-316, 1959.
- [7] P.M.T. Broersen, "The Best Order of Long Autoregressive Models for Moving Average Estimation", *Signal Processing VIII, Proc. Eusipco Conf., Trieste, Italy*, pp. 799-802, 1996.
- [8] P.M.T. Broersen, "The Quality of Models for ARMA Processes", *IEEE Trans. on Signal Process.*, vol. 46, pp. 1749-1752, 1998.
- [9] J.S. Erkelens and P.M.T. Broersen, "Bias Propagation in the Autocorrelation Method of Linear Prediction", *IEEE Trans. on Speech and Audio*, vol. 5, pp. 116-119, 1997.
- [10] P.M.T. Broersen, "Robust Algorithms for Time Series Models", *Proc. ProRISC/IEEE CSSP98 Conf.*, pp. 75-82, 1998.
- [11] P.M.T. Broersen and H.E. Wensink, "On Finite Sample Theory for Autoregressive Model Order Selection", *IEEE Trans. on Signal Processing*, vol. SP-41, pp. 194-204, 1993.
- [12] P.M.T. Broersen, "The ABC of Autoregressive Order Selection Criteria", *Preprints Sysid '97 Conf., Kitakyushu, Japan*, pp. 231-236, July 1997.
- [13] P.M.T. Broersen, "Facts and Fiction in Spectral Analysis of Stationary Stochastic Processes", *Signal Processing IX, Proc. Eusipco Conf., Rhodes, Greece*, pp. 61-64, 1998.
- [14] G. Wilson, "Factorization of the covariance generating function of a pure Moving Average process", *SIAM Journal Numer. Anal.*, vol. 6, pp. 1-7, 1969.