

Maximum Likelihood Estimators for ARMA and ARFIMA Models: A Monte Carlo Study

Michael A. Hauser

University of Economics and Business Administration,
Department of Statistics, Vienna

First draft December 1997. This version May 1998.

Forthcoming Journal of Statistical Planning and Inference

Abstract

We analyze by simulation the properties of two time domain and two frequency domain estimators for low order autoregressive fractionally integrated moving average Gaussian models, ARFIMA (p, d, q) . The estimators considered are the exact maximum likelihood for demeaned data, EML, the associated modified profile likelihood, MPL, and the Whittle estimator with, WLT, and without tapered data, WL. Length of the series is 100. The estimators are compared in terms of pile-up effect, mean square error, bias, and empirical confidence level.

The tapered version of the Whittle likelihood turns out to be a reliable estimator for ARMA and ARFIMA models. Its small losses in performance in case of “well-behaved” models are compensated sufficiently in more “difficult” models. The modified profile likelihood is an alternative to the WLT but is computationally more demanding. It is either equivalent to the EML or more favorable than the EML. For fractionally integrated models, particularly, it dominates clearly the EML. The WL has serious deficiencies for large ranges of parameters, and so cannot be recommended in general. The EML, on the other hand, should only be used with care for fractionally integrated models due to its potential large negative bias of the fractional integration parameter. In general, one should proceed with caution for ARMA(1,1) models with almost canceling roots, and, in particular, in case of the EML and the MPL for inference in the vicinity of a moving average root of +1.

Keywords: Fractional integration, Whittle likelihood, modified profile likelihood, data taper, pile-up effect

Acknowledgments: I want to express my thanks to Peter Bloomfield, Sastry Pantula, Benedikt Pötscher, and discussants at the International Workshop on Long Range Dependence, Brisbane, January 1997, and two anonymous referees.

Short title: Maximum Likelihood Estimators for ARFIMA Models

Address of the author:

Department of Statistics, University of Economics and Business Administration,
Augasse 2-6, A-1090 Vienna, Austria.

Tel.: +43 1 31336/4759, FAX: +43 1 31336/738, E-Mail: Michael.Hauser@wu-wien.ac.at

1 Introduction

In this paper we analyze, by means of simulation, the properties of four versions of maximum likelihood estimators for fitting autoregressive fractionally integrated moving average, ARFIMA, time series models. The comparison of the small sample properties is of particular interest since different estimators may lead to substantial distinct conclusions. E.g. if the Whittle estimate or the modified profile likelihood is used for fitting a class of low order ARFIMA models to the frequently investigated postwar quarterly U.S. GNP growth rates Akaike's information criterion will choose a non fractional model. On the other hand, if the same procedure is undertaken with the exact Gaussian likelihood, it results in a fractionally integrated model with a highly significant fractional integration parameter. See Sowell(1992b) and Hauser(1995). Thus, despite the fact that the estimators are asymptotically equivalent the question remains which estimator is more reliable in small samples.

We assume that the series y_t is a sample of length n of a weakly stationary process (in discrete time) which admits a Wold representation

$$y_t = \mu + \Psi(L)u_t = \mu + (1 + \psi_1L + \psi_2L^2 + \dots)u_t.$$

where L denotes the backward shift operator. The impulse response coefficients ψ_τ satisfy $\sum_{\tau=0}^{\infty} \psi_\tau^2 < \infty$, and the u_t are Gaussian white noise with variance σ_u^2 . And, y_t obeys to an ARFIMA(p, d, q) model with $d \in [-\frac{1}{2}, \frac{1}{2})$.

$$\alpha(L)(1 - L)^d(y_t - \mu) = \beta(L)u_t,$$

where $\alpha(L) = 1 - \alpha_1L - \dots - \alpha_pL^p$ and $\beta(L) = 1 - \beta_1L - \dots - \beta_qL^q$, are polynomials in the backward shift operator and d is the fractional integration parameter. The roots of the polynomials $\alpha(z)$ and $\beta(z)$ are assumed to lie outside the unit circle. Then the process (y_t) is stationary for $d < \frac{1}{2}$ and invertible for $d > -1$ (Bloomfield, 1985).

The spectrum

$$f(\lambda) = \frac{\sigma_u^2}{2\pi} \frac{|\beta(\exp(-i\lambda))|^2}{|\alpha(\exp(-i\lambda))|^2} |1 - \exp(-i\lambda)|^{-2d},$$

is infinite at the origin for $d > 0$ and zero for $d < 0$. Long memory is associated with $d > 0$. If $d < 0$, the process is said to have intermediate memory.

Two generic maximum likelihood procedures for stationary Gaussian series for the parameter vector $\theta = (\alpha_1, \dots, \alpha_p, d, \beta_1, \dots, \beta_q)$ are available: the (approximative) spectral (Whittle) maximum likelihood, and the exact Gaussian maximum likelihood method. Both methods yield \sqrt{n} -consistent, asymptotically normal and asymptotically efficient parameter estimates. See Fox and Taqqu(1986), Dahlhaus(1989) and Giraitis and Surgailis(1990) for the Whittle estimator, and Dahlhaus(1989) and An, Bloomfield and Pantula(1992) for the exact maximum likelihood. The small sample properties of four estimators will be investigated: the exact Gaussian likelihood, the modified profile likelihood derived by An and Bloomfield(1993) according to the proposal of Cox and Reid(1987), the Whittle likelihood and two versions thereof using tapered data. Having economic applications in mind we restrict our study to low order ARMA and ARFIMA models. The criteria for comparison are the mean square error, bias, and the empirical confidence level of the 95% confidence interval. In addition, the estimators will be investigated with respect of the occurrence of the pile-up effect, which has been intensively discussed in the literature for the exact maximum likelihood method.

Our study is a significant extension to the former simulation experiments performed by Ansley and Newbold(1980), Boes et al.(1989) and Cheung and Diebold(1994). No Monte Carlo study seems to be available which compares the small sample properties of maximum likelihood estimators even for ARMA models. And, estimators using tapered data have been applied almost exclusively to AR(p) or FI(d) models. Section 2 defines the estimators and summarizes their small sample properties as far as they are known and relevant for our study. Section 3 summarizes the Monte Carlo results. Computational aspects are discussed in Section 3.1. For the Monte Carlo results of the pile-up effect see Section 3.2. The ARFIMA models chosen and the criteria for the comparison of the estimators are given in Section 3.3. The properties of the four estimators are presented in Section 3.4. Section 3.5 gives a heuristic explanation of the systematic negative bias of the exact maximum likelihood estimates of d . Section 4 concludes.

2 The estimators

The exact maximum likelihood procedure, EML

The exact Gaussian log-likelihood is given by

$$\log L_E(\mu, \theta, \sigma_u^2) = -\frac{1}{2}[n \log(2\pi) + \log \det(\Sigma) + (Y - \mu l)' \Sigma^{-1} (Y - \mu l)] \quad (1)$$

where $\Sigma = \Sigma(\theta, \sigma_u^2)$ is the $n \times n$ covariance matrix depending on θ and σ_u^2 . n denotes the sample size, $Y = (y_1, \dots, y_n)'$ and $l = (1, \dots, 1)'$. We will investigate the reduced profile likelihood where μ is replaced by the sample mean \bar{y} , and where the solution of the maximization with respect to σ_u^2 is obtained analytically. The reduced (or concentrated) likelihood is then

$$\log L_E^*(\theta) = -\frac{1}{2}[n \log(2\pi) + n \log(\sigma_u^{2*}) + \log \det(\tilde{\Sigma}) + n] \quad (2)$$

with $\sigma_u^2 = \sigma_u^{2*}$ and $\sigma_u^{2*} = \frac{1}{n}(Y - \bar{y}l)' \tilde{\Sigma}^{-1} (Y - \bar{y}l)$. $\tilde{\Sigma}$ is defined by the relation $\Sigma = \sigma_u^2 \tilde{\Sigma}$, so that $\tilde{\Sigma} = \tilde{\Sigma}(\theta)$ is a function of θ alone. The resulting estimator is denoted by EML.

Two small sample properties of the EML may be of importance: The pile-up effect arises when the underlying MA model has a root “close” to the unit circle. Then the EML estimates yield roots on the unit circle with a positive probability. (Cp. Cryer and Ledolter(1981) and Anderson and Takemura(1986).) On the other hand, it is known that the EML estimate tends to give a negatively biased parameter for pure fractionally integrated processes (Li and McLeod, 1986). This is not in effect if the true mean is known and is used for the mean correction of the data (Cheung and Diebold, 1994).

The modified profile likelihood, MPL

The modified profile likelihood is based on the idea to correct the parameter estimates of interest (\hat{a}) for second order effects due to nuisance parameters (b) in the model. (Cp. Cox and Reid(1987).) Thereby a transformation is sought which makes b orthogonal to a . For ARFIMA models the nuisance parameters are chosen as $b = (\mu, \sigma_y^2)$, and the parameters of interest as $a = \theta = (\alpha_1, \dots, \alpha_p, d, \beta_1, \dots, \beta_q)$. An and Bloomfield(1993) give the solution for the modified profile likelihood based on the Gaussian exact maximum likelihood. The orthogonal vector results as $\phi = (\mu, \gamma)$ with $\gamma = \sigma_y^2 (\det R)^{1/n}$, where $R = \Sigma/\sigma_y^2$ is the

correlation matrix. Remarkably, μ turns out to be orthogonal to θ . The MPL without constants is then (in a notation similar to An and Bloomfield)

$$\begin{aligned} \log L_M^*(\theta; \hat{\sigma}_u^2, \hat{\mu}) = \\ \left(\frac{1}{n} - \frac{1}{2}\right) \log \det(R) - \frac{1}{2} \log(l'R^{-1}l) + \left(1 + \frac{1}{2} - \frac{n}{2}\right)(Y - \hat{\mu}l)'R^{-1}(Y - \hat{\mu}l) \end{aligned} \quad (3)$$

Our actual implementation replaces $\hat{\mu}$ by the sample mean \bar{y} , and $\hat{\sigma}_y^2$ by the maximum likelihood estimator $\hat{\sigma}_y^2 = \frac{1}{n}(Y - \bar{y}l)'R^{-1}(Y - \bar{y}l)$. $\hat{\sigma}_u^2$ results from the relation $\sigma_y^2 = \sigma_u^2 \sum_i \psi_i^2$. An and Bloomfield also offer a small Monte Carlo study illustrating for some selected low order ARFIMA models that the MPL successfully eliminates the bias in the EML estimates.

The Whittle likelihood, WL

We define the Whittle likelihood as

$$\log L_W(\theta, \sigma_u^2) = - \sum_{j=1}^m \log f(\lambda_j | \theta, \sigma_u^2) - \frac{1}{2\pi} \sum_{j=1}^m \frac{I(\lambda_j)}{f(\lambda_j | \theta, \sigma_u^2)} \quad (4)$$

where $I(\lambda_j)$ denotes the periodogram at the j -th Fourier frequency, λ_j ($\lambda_j = 2\pi \frac{j}{n}$, $j = 1, \dots, m$)

$$I(\lambda_j) = \frac{1}{n} \left| \sum_{t=1}^n y_t \exp(-it\lambda_j) \right|^2 \quad (5)$$

m is the largest integer contained in $(n-1)/2$. It is the discrete time version of the Whittle function (cp. Dahlhaus(1988) and Robinson(1990)). In the ARMA case it may be interpreted e.g. as the likelihood associated with the asymptotic distribution of the periodogram (cp. Brockwell and Davis, 1991, p.347f). Other interpretations are given in Dahlhaus(1988) and Parzen(1983, Sec. 3). On the other hand, if the term $\sum_{j=1}^m \log f(\lambda_j | \theta, \sigma_u^2)$ is dropped the asymptotic properties remain the same (Fox and Taqqu, 1986), and it becomes the Yule-Walker estimator for AR(p) processes (Dzhaparidze, 1986, p.116f).

The reduced form of L_W with respect to the error variance σ_u^2 is

$$\log L_W^*(\theta) = m \log(2\pi) - m \log \left[\frac{1}{m} \sum_{j=1}^m \frac{I(\lambda_j)}{g(\lambda_j)} \right] - \sum_{j=1}^m \log g(\lambda_j) - m \quad (6)$$

with $\sigma_u^2 = \sigma_u^{2*}$

$$\sigma_u^{2*} = \frac{1}{m} \sum_{j=1}^m \frac{I(\lambda_j)}{g(\lambda_j)} \quad (7)$$

where $f(\lambda) = \sigma_u^2 g(\lambda)/(2\pi)$ with $g(\lambda) = g(\lambda | \theta)$. This estimator is denoted by WL.

The pile-up effect has not been considered for the Whittle estimate before. In the line with Anderson and Takemura(1986) we prove that the likelihood of a MA(1) process exhibits a local maximum at $\beta_1 = \pm 1$. The same holds – contrary to the EML or MPL – for AR(1) processes at $\alpha_1 = \pm 1$. The local maximum may turn out as a global one in finite samples, so that parameter estimates of ± 1 are obtained with a positive probability. These probabilities are determined by Monte Carlo simulations in Sec. 3.2.

Actually, we only investigate the first order conditions. They are for a model with a single parameter θ (θ may be α_1 or β_1) $(\partial \log L_w^*)/(\partial \theta) = 0$, or more explicitly

$$\frac{1}{\sigma_u^{2*}} \sum_j \frac{I_j}{g_j^2} \frac{\partial g_j}{\partial \theta} - \sum_j \frac{1}{g_j} \frac{\partial g_j}{\partial \theta} = 0$$

with $g_j = g(\lambda_j)$ and $I(\lambda_j) = I_j$. After some manipulations we obtain

$$\sum_j \frac{I_j}{g_j} \left[m \frac{\partial g_j / \partial \theta}{g_j} \right] = \sum_j \frac{I_j}{g_j} \left[\sum_k \frac{\partial g_k / \partial \theta}{g_k} \right].$$

So, if the terms in brackets are constant and equal, the first order condition holds.

In case of MA(1) processes $g(\lambda) = 1 - 2\beta_1 \cos(\lambda) + \beta_1^2$, so $\partial g / \partial \beta_1 = 2[\beta_1 - \cos(\lambda)]$, and $\frac{\partial g / \partial \beta_1}{g} = \pm 1 \forall \lambda$ if $\beta_1 = \pm 1$. Both bracket terms are for $\beta_1 = \pm 1$ either $+m$ or $-m$, and may be canceled. The first order condition holds for $\beta_1 = \pm 1$.

In case of AR(1) processes $g(\lambda) = 1/[1 - 2\alpha_1 \cos(\lambda) + \alpha_1^2]$, $\partial g / \partial \alpha_1 = (-2)[\alpha_1 - \cos(\lambda)] / [1 - 2\alpha_1 \cos(\lambda) + \alpha_1^2]^2$, and again $\frac{\partial g / \partial \alpha_1}{g} = \pm 1 \forall \lambda$ if $\alpha_1 = \pm 1$. And, the same holds as above.

It is known that Yule-Walker estimates are rather bad for short series, and if the roots of the corresponding characteristic equation are close to the unit circle (Priestley, 1981, p.351). So, one may expect similar properties to hold for the Whittle estimates. On the other hand, Hauser(1995) reports an advantage of the WL with respect to the EML for pure fractionally integrated models and some higher order ARFIMA models in small samples. The WL gives essentially unbiased estimates and smaller mean square errors for \hat{d} in most cases.

The Whittle likelihood for tapered data, WLT

Dahlhaus(1988) shows that tapering reduces the leakage effect of the periodogram as estimate of the true spectrum. He finds that the new estimate competes well with the Burg estimate for an AR(14) model where roots of the characteristic equation are complex and close to the unit circle. A tapered series is defined by

$$y_t^T = h_t y_t$$

where h_t is the data taper, in our case the Tukey-Hanning taper

$$h_t = \begin{cases} \frac{1}{2}[1 - \cos\{\pi(t - \frac{1}{2})/l\}] & t = 1, \dots, l \\ 1 & t = l + 1, \dots, n - l \\ \frac{1}{2}[1 - \cos\{\pi(n - t + \frac{1}{2})/l\}] & t = n - l + 1, \dots, n \end{cases} \quad (8)$$

The proportion of the data which is altered by this taper is $2\rho = 2l/n$. We choose the variable taper of Zhang(1991) who proposes to use $\rho = 2/\sqrt{n}$. The tapered time series is then used to construct modified periodogram ordinates, $I^T(\lambda_j)$,

$$I^T(\lambda_j) = \frac{1}{H_2 n} \left| \sum_{t=1}^n h_t (y_t - \bar{y}) \exp(-it\lambda_j) \right|^2 \quad (9)$$

with $H_2 = \sum_t h_t^2$. Replacing $I(\lambda_j)$ in the WL function by $I^T(\lambda_j)$ yields the Whittle likelihood for tapered data, denoted as WLT.

There are several studies concerned with the use of data tapers dealing essentially only with AR models: Pukkila and Nyquist(1985), Kang(1987), Hurvich(1988), or Zhang(1991), among others. The general conclusions are that the type of taper is not of much importance, nor does the amount of tapering affect the results considerably. However, some tapering should be performed. Cheung and Diebold(1994), on the other hand, investigate the small sample properties of the approximative Whittle estimate (without the log f -term) for pure FI(d) models. They find it to be slightly inferior to the EML regardless whether tapered or non tapered data are used.

3 Small sample behavior of the estimators

3.1 Computational aspects

All simulated series throughout the paper are generated via the Durbin-Levinson algorithm. Thereby the true autocovariances (cp. Sowell, 1992a) and Gaussian innovations

with unit variance (Hoermann and Derflinger, 1990) are used. This method is exact. For the generation of the uniformly distributed input variates the TT800, a twisted GFSR generator (Matsumoto and Kurita, 1994) is used. It has a period of $2^{800} - 1$ and excellent equidistribution properties up to dimension 25. Due to numerical problems of the EML and MPL when the parameters approach the non stationarity region we have to restrict the estimated AR and d parameters. We set $\max |1/z_{AR}| < 0.9965$, where z_{AR} is the root of the characteristic equation for the AR polynomial. The maximal d is 0.4965. The program used for calculating the estimates of the EML, and so also of the MPL, is based on the FORTRAN code supplied by F. Sowell. The spectral estimates, on the other hand, are not bounded away from the unit circle. If estimated roots lie inside the unit circle, the inverse of the roots are used to make the models stationary and invertible. The confidence intervals of the EML and MPL estimates are calculated only when an interior maximum is found, i.e. if $|1/\hat{z}_{AR}| \leq 0.99$ and $\hat{d} \leq 0.49$.

3.2 The pile-up effect

The probability that parameter estimates of ± 1 are observed when fitting MA(1) or AR(1) models to small data sets are obtained by Monte Carlo simulation. Table 1 gives the relative frequencies (with respect to 10000 replications) that the estimates differ from $+1$ or -1 no more than 0.0035 (cp. the numerical restrictions given above) for series of the length of 25, 50 and 100. E.g., $P(\hat{\beta}_1 = 1)$ stands for $P(\hat{\beta}_1 \in [0.9965, 1])$.

** include TABLE 1 **

The results for the EML(μ), the EML where the true mean is used to demean the data, compare to the theoretical results of Cryer and Ledolter(1981). The EML and the MPL results are very similar to the theoretical ones in case of the MA(1) model with $\beta_1 = -0.9$. However, for $\beta_1 = 0.9$ the pile-up effect is smaller, especially for the MPL. The WL, which is independent of the mean correction, exhibits small frequencies for $\beta_1 = -0.9$, and is comparable to the EML for $\beta_1 = 0.9$. For the AR(1) processes the pile-up effect is less pronounced. The probability of estimates of the WLT are somewhat larger than those for the WL.

3.3 Models and criteria

We will compare four estimators for low order ARFIMA models with respect to the mean square error, MSE, of the parameter estimates, the bias, and the percentage that the true parameter lies in the 95% confidence interval based on the asymptotic normal distribution. This percentage is denoted by ECL, empirical confidence level. The values given for the MSE are scaled by a factor of 100. The length of the series considered is 100. 1000 replications are performed.

*** INCLUDE somewhere below FIGURES 1 to 10 and TABLE 2

The models considered in the simulation study are as follows: AR(1) with parameter values $\alpha_1 = -0.99, -0.95, \dots, 0.95, 0.99$. The results are summarized in Figure 1 not including the WL estimator. The WL is neither presented graphically nor in tables (except in Fig. 3) but described verbally at the end of this section due to its limited applicability. MA(1) with $\beta_1 = -1.00, -0.99, -0.95, \dots, 0.95, 0.99, 1.00$ (Figure 2). FI(d) with d values ranging from -0.50 to $0.45, 0.49$ (Figure 3). AR(2) models with real roots, $z_{1,2}$, close together as in Zhang(1991). For $z_1 < 0$ the relation $1/z_1 + 0.05 = 1/z_2$ is assumed, for $z_1 > 0$ $1/z_1 - 0.05 = 1/z_2$, with $1/z_1 = -0.95, -0.90, \dots, 0.90, 0.95$. See Figure 4. The same parameter values are implemented in MA(2) models yielding troughs in the spectrum where the AR models have peaks (Figure 5). These models are of interest, since the small sample behavior of the Whittle estimator using tapered data has essentially been applied only to AR models. In addition to real roots, we also consider complex roots, $z_{1,2} = r e^{\pm i\lambda}$, in AR(2) models (Figure 6). We choose a constant modulus close to the unit circle of $r = 1/0.95$, and uniformly distributed frequencies λ over $[0, \pi]$: $\lambda = j \frac{\pi}{10}$ with $j = 0, \dots, 10$. Both ends, $j = 0, 10$, of the interval correspond to double real roots. Due to the similarity of the behavior of both involved parameter estimates and to space restrictions only the MSE of the second parameter estimate are given in Fig. 4 to 6. Further, ARMA(1,1) processes are investigated (Figure 7). The α_1 and β_1 parameters are chosen to take the values $-0.99, -0.95, \dots, 0.95, 0.99$. For β_1 the list of values is augmented by ± 1.00 . The minimal distance between the AR and MA parameter is chosen to be 0.05, $|\alpha_1 - \beta_1| \geq 0.05$. Additional models with almost canceling roots are specified as

pairs $(\alpha_1, \beta_1 \pm 0.05)$ and $(\alpha_1 \pm 0.05, \beta_1)$ for $\alpha_1, \beta_1 = \pm 0.80, \pm 0.60, \dots, 0.00$. The contour lines of the MSE surfaces in Fig. 7.a are drawn to indicate the 5.0, 10.0, \dots levels. (The computations were performed with the triangle contour plot algorithm of Preusser(1984).) In order to accentuate a bias of null the corresponding contour line is drawn fat. The other bias levels are dotted indicating levels of $\pm 0.05, \pm 0.10, \dots$. In a similar way the empirical confidence levels are marked. The fat lines indicate a level of 95%, the dotted ones 100, 90, 85, \dots . Simulation results for ARFIMA(1, d ,0) processes are summarized in Figure 8. The values for the fractional integration parameter are chosen between -0.50 and 0.45 . The autoregressive parameter varies between -0.95 and 0.95 . However contrary to Fig. 7, the contour levels in the MSE plots are 2.5, 5.0, \dots . Finally, the results for ARFIMA(0, d ,1) processes with parameters d also ranging from -0.50 to 0.45 , and β_1 from -0.95 to 0.95 are given. See Figure 10. In addition to the graphical representation, the maximal and minimal values of each criterion for models with 2-dimensional parameter spaces, together with the points where they are assumed, are collected in Table 2.

Since an estimator may exhibit the largest maximum in a small area but performs excellent for most other models, while another estimator may be rather bad over the whole class of models, the overall maximum MSE may be no adequate indicator. So we construct a measure of net advantage of estimator A over estimator B based on the maximal differences of the MSE surfaces. We define it as

$$\max_{MSE_B > MSE_A} |MSE_A(\theta) - MSE_B(\theta)| - \max_{MSE_A > MSE_B} |MSE_A(\theta) - MSE_B(\theta)|.$$

$MSE_A(\theta)$ denotes the MSE of the parameter estimate of θ obtained by estimator A . The maximal advantage (measured in MSE) using estimator A instead of B is so related to the maximal advantage of estimator B . If our measure is positive, estimator A is to be preferred over estimator B . If it is negative, B is to be preferred. Table 3 summarizes the results for invertible ARMA(1,1), ARFIMA(1, d ,0) and ARFIMA(0, d ,1) models.

3.4 Small sample behavior of the estimators

Our numerical results compare - as far as values are available - well to previous studies like Zhang(1991) and Cheung and Diebold(1994). However, they differ to some extent from Ansley and Newbold(1980). They demean the data by the population mean while

we use the sample mean – as is commonly done – leading to a more or less pronounced asymmetric behavior of the time domain ML estimates at parameter values associated with real roots close to $+1$ on one hand, and -1 on the other.

In the following we survey the small sample properties of pairs of estimators by first stating the main properties and then backing up the conclusion by referencing to the figures and tables provided. We compare EML and MPL, MPL and WLT, give a verbal description of the WL, and discuss the effects of tapering the data.

For non fractionally integrated models the MPL turns out to be either equivalent to or more favorable than the EML. In case of AR(1) models with negative α_1 the behavior of EML and MPL are essentially identical. For larger positive α_1 the MSE and bias of the MPL estimates are smaller, and the ECL is closer to the theoretical 95% level. The same relation holds for the MA(1) models (with the exception of β_1 close to $+1$). In case of the AR(2) (Fig. 4 and 6) and the MA(2) models (Fig. 5) differences between both estimators are not visible. For the ARMA(1,1) model the minimal MSE of the MPL is the same as that of the EML (0.06). The maximum of 54.67, however, is clearly below the corresponding value of the EML, 67.07. The criterion of the net advantage between MPL and EML (see Tab. 3) rates both estimators as essentially equivalent.

For fractionally integrated models the EML is clearly dominated by the MPL. This holds particularly for mixed autoregressive or moving average fractionally integrated models. Compare Fig. 3, 8 to 10 and Tab. 2. The systematic negative bias in the d estimate of the EML is enlarged by the inclusion of the AR and MA parameter. E.g., in case of the ARFIMA(1, d ,0) model with true values of $(d, \alpha_1) = (0.15, 0.20)$ the median estimated model is $\text{med}(\hat{d}) = -0.105$ and $\text{med}(\alpha_1) = 0.426$! The associated true and estimated spectral densities are plotted in Figure 9. It turns out that the EML cannot detect the true sign of the d parameter for a rather large set of models. Those models are marked by crosses in the bias plots. Compare both Fig. 8.b and Fig. 10.b. The large bias of the EML estimate is reflected in MSE and ECL. The MPL, on the other hand, is not affected by that systematic bias and has good MSE and acceptable ECL properties. The criterion of net advantage with respect to the MSE confirms our conclusions. See Tab. 3.

The WLT, the Whittle likelihood applied to tapered data, exhibits for “well-behaved”

models slightly larger MSE than the MPL. However, for more “difficult” models this disadvantage seems to be compensated sufficiently. The WLT turns out to be a reliable estimator. The favorable properties of the WLT do not come into effect for the single parameter models AR(1) and FI(d). In those cases the MSE is larger than that of the MPL over the whole parameter region. But bias and ECL properties are essentially equivalent. See Fig. 1 and 3. For most β_1 values this relation also holds in the MA(1) models. However close to $\beta_1 = 1.0$, MSE, bias and ECL are more favorable with respect to the MPL (Fig. 2). Similarly, as in the MA(1) case the WLT dominates the MPL close to $1/z_1 = -0.95$ for the AR(2) models in Fig. 4, close to $1/z_1 = 0.95$ for the MA(2) models in Fig. 5, and close to $\lambda = \pi$ for the AR(2) models with conjugate complex roots in Fig. 6. For ARMA(1,1) models the advantage of the WLT is evident (Fig. 7 and Tab. 2, 3). Models with almost canceling polynomials can be estimated with the WLT more precisely. The spread of the bias and the maximal MSE are considerably smaller for the WLT for both $\hat{\alpha}_1$ and $\hat{\beta}_1$, than for the MPL. The net advantage as given in Tab. 3 is also in favor of the WLT. The ECL of the MPL for $\hat{\beta}_1$ at the true parameters $(\alpha_1, \beta_1) = (-0.99, 1.00)$ is extremely low (6.9), while the minimal value for the WLT is 55.3. Fig. 8.c also shows that the confidence intervals based on MPL or EML estimates for β_1 values close to +1 are too small. (We have also investigated the sensitivity of our conclusion due to the inclusion of models with possibly too strong canceling effects in relation to the length of the series. However, the advantage of the WLT remains if the model class is restricted to $|\alpha_1 - \beta_1| \geq 0.20$. In this case the maximal MSE values of MPL and WLT are rather the same, but net advantage, bias and ECL are still in favor of the WLT. The corresponding values are not tabulated.) According to visual inspection of Fig. 8, minimal and maximal MSE, the net advantage with respect to the MSE (cp. Tab. 3), and spread of the bias (see Tab. 2), the MPL is to be preferred for the ARFIMA(1, d ,0) models. However, its ECL for \hat{d} may be rather low for models close to $(d, \alpha_1) = (-0.50, -0.95)$. On the contrary, in case of the ARFIMA(0, d ,1) models all criteria apart from the net advantage are in favor of the WLT.

The WL estimates may exhibit serious deficiencies when the roots of the models are “close” to real unit roots, both for AR and MA polynomials. Unacceptable high MSE values are observed together with large biases in relative large parameter ranges. So, the

WL cannot be recommended in general, and a detailed graphical presentation is omitted. The most important simulation results are documented in the following in a verbal way. The defects of the WL may hardly be observed for AR(1) models. For MA(1) models, AR(2) and MA(2) models with real roots, however, the MSE curves adopt a pronounced trough shape with a rather flat bottom. In the AR(2) models with complex roots a U-shape is observed (with a minimal value at $\lambda = \pi/2$ which competes well with the MPL). The MSE surface for $\hat{\beta}_1$ of the ARMA(1,1) models exhibits a W-like profile with the ridge generated by the models with almost canceling roots in the middle. (The MSE surface for $\hat{\alpha}_1$ is not affected and similar to the one observed for the WLT.) The WL estimate of the fractional integration parameter, contrary to the EML, does not exhibit a systematic bias. This is most evident for the FI(d) models, where the WL is the second best estimator being only slightly inferior to the MPL, and more favorable than the WLT. See Fig. 3. Including an AR parameter in the FI(d) model the MSE of both parameters become rather high at $(d, \alpha_1) = (0.45, 0.95)$, while for the ARFIMA (0, d ,1) the overall properties are inferior to the WLT but compete with the MPL. Tapering the data increases the MSE but not the bias for “well-behaved” models and lowers both, MSE and bias, impressively for all problematic models investigated.

3.5 An explanation for the bias in the EML estimate of d

The potential negative bias in the d estimate obtained by the EML, in contrast to the WL where no systematic bias is observed, seems worth to be investigated more closely. A heuristic explanation for this behavior is offered below for fractionally integrated processes with $d > 0$. We observe that both estimators differ in the treatment of the mean, the frequency zero in the spectral representation respectively. While frequency zero is excluded in the WL estimation explicitly, it is implicitly included in the EML through the autocovariance function $\gamma(h)$.

$$\gamma(h) = \int_{-\pi}^{\pi} f(\lambda) e^{i\lambda h} d\lambda \quad (10)$$

where the covariance matrix is $\Sigma = \Sigma(\theta) = [\gamma(|i-j|)]_{i,j}$ with $\theta = (\alpha, \dots, \alpha_p, d, \beta_1, \dots, \beta_q)$. In case of demeaned data the periodogram as estimator of the spectrum is null at frequency zero independently on the process. However, for $d > 0$ the spectrum is infinite at zero. As the d parameter essentially describes only the slope of the spectrum close to frequency

zero, it is very sensitive to changes in the low frequency region (but insensitive to high frequency effects). Now consider the replacement of f in (10) by the periodogram for demeaned data. Then the EML estimate of d will try to model the upward slope of the true spectrum for low frequencies, but will have to take into account the value of null at zero. So, a negative bias will result. Not surprisingly, the negative bias in case of the ARFIMA(1, d ,0) and (0, d ,1) models may be considerably larger than in pure FI models. The additional ARMA parameter (which models the spectrum at low and high frequencies) offers more flexibility to \hat{d} to capture the sudden decrease at frequency zero. The consequence is also a considerable bias in $\hat{\alpha}_1, \hat{\beta}_1$ respectively, due to a compensating effect in the estimated spectrum at low frequencies (cp. Fig. 9).

Our explanation is in line with the negative bias in the sample autocovariance function as observed by Newbold and Agiakloglou(1993). It is also compatible with the findings of Cheung and Diebold(1994) that the exact maximum likelihood estimator for data corrected with the true mean yields unbiased d estimates.

4 Summary

We investigated the pile-up effect for the exact maximum likelihood method, the modified profile likelihood, the Whittle estimator and the Whittle estimator for tapered data. Some conclusions are: The pile-up effect of the EML applied to demeaned data is somewhat smaller than the theoretical values for known mean. Contrary to the EML, the WL exhibits the pile-up effect also for the AR(1) models, but to a lesser extent.

An extensive simulation study is performed to investigate the properties of the estimators in low order ARMA and ARFIMA models. The length of the series is 100. Despite the asymptotic equivalence of the estimators we find that their small sample properties may differ substantially. The exact maximum likelihood exhibits a rather bad performance in mixed ARFI or FIMA models. The d estimates tend to be seriously negatively biased leading to large mean square errors and low empirical confidence levels. Also, for ARMA(1,1) models with almost canceling roots the use of the EML may lead to extremely large mean square errors and biases. The modified profile likelihood exhibits for ARMA models slightly better properties than the exact maximum likelihood estimator,

and dominates it clearly for fractionally integrated ones. The Whittle likelihood, while yielding no systematically biased d estimates, has serious deficiencies for large parameter ranges especially “close” to real roots of $+/-1$, and so cannot be recommended in general. Based on a comparison of the EML and WL we offer a heuristic explanation for the bias in the EML estimates of d in long memory models.

The Whittle likelihood with tapered data performs well not only for AR models but also for MA, ARMA(1,1) and fractionally integrated models. It turns out to be an overall reliable estimator. The small losses in performance in case of “well-behaved” models seem to be compensated sufficiently in more “difficult” models. Its computational simplicity is also attractive. The computationally more demanding alternative with certain advantages is the modified profile likelihood. Caution is requested for ARMA(1,1) models with almost canceling roots in general, and, in particular, in case of the EML and the MPL for inference in the vicinity of a moving average root of $+1$.

References

- An, S., P. Bloomfield and S. Pantula, 1992, Asymptotic properties of the MLE in fractional ARIMA models, Institute of Statistics Mimeograph Series No. 2228, North Carolina State University.
- An, S. and P. Bloomfield, 1993, Cox and Reid's modification in regression models with correlated errors, Department of statistics, North Carolina State University, Raleigh.
- Anderson, T.W. and A. Takemura, 1986, Why do noninvertible estimated moving averages occur?, *Journal of Times Series Analysis*, 7, 235-254.
- Ansley, C.F. and P. Newbold, 1980, Finite sample properties of estimators for autoregressive moving average models, *Journal of Econometrics*, 13, 159-183.
- Bloomfield, P., 1985, On series representations for linear predictors, *The Annals of Probability*, 13, 226-233.
- Boes, D.C., R.A. Davis and S.N. Gupta, 1989, Parameter estimation in low order fractionally differenced ARMA processes, *Stochastic Hydrology and Hydraulics*, 3, 97-110.
- Brockwell, P.J. and R.A. Davis, 1991, *Time series: Theory and methods* (Springer, New York).
- Cheung, Y.-W. and F.X. Diebold, 1994, On maximum-likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean, *Journal of Econometrics*, 62, 301-316.
- Cox, D.R. and N. Reid, 1987, Parameter orthogonality and approximate conditional inference (with discussion), *Journal of the Royal Statistical Society Series B*, 49, 1-39.
- Cryer, J. and J. Ledolter, 1981, Small-sample properties of the maximum likelihood estimator in the first-order moving average model, *Biometrika*, 68, 691-694.
- Dahlhaus, R., 1988, Small sample effects in time series analysis: A new asymptotic theory and a new estimate, *The Annals of Statistics*, 16, 808-841.
- Dahlhaus, R., 1989, Efficient parameter estimation for self-similar processes, *The Annals of Statistics*, 17, 1749-1766.
- Dzhaparidze, K., 1986, *Parameter estimation and hypothesis testing in spectral analysis of stationary time series* (Springer, New York).
- Fox, R. and M.S. Taquq, 1986, Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series, *The Annals of Statistics*, 14, 517-532.
- Giraitis, L. and D. Surgailis, 1990, A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate, *Probability Theory and Related Fields*, 86, 87-104.

- Hauser, M.A., 1995, Long range dependence in international output series: A reexamination, (World Congress of the Econometric Society, Tokyo)
- Hoermann, W. and G. Derflinger, 1990, The ACR method for generating normal random variables, *OR Spektrum*, 12, 181-185.
- Hurvich, C.M., 1988, A mean squared criterion for time series data windows, *Biometrika*, 75, 485-490.
- Kang, H.-J., 1987, The tapering estimation of the first-order autoregressive parameters, *Biometrika*, 74, 643-645.
- Li, W.K. and A.I. McLeod, 1986, Fractional time series modelling, *Biometrika*, 73, 217-221.
- Matsumoto, M. and Y. Kurita, 1994, Twisted GFSR generators, II, *ACM Transactions on Modelling and Computer Simulation*, 4, 1994, 254-266.
- Newbold, P. and Ch. Agiakloglou, 1993, Bias in the sample autocorrelation of fractional noise, *Biometrika*, 80, 698-702.
- Parzen, E., 1983, Autoregressive spectral estimation. In *Handbook of Statistics* (Brillinger, D.R. and Krishnaiah, P.R., eds.) 3, 221-247, (North-Holland, Amsterdam).
- Priestley, M.B., 1981, *Spectral analysis of time series, Volume 1: Univariate Series* (Academic Press, London).
- Preusser, A., 1984, Triangle contour plotting, *ACM-Trans. Math. Software*, 10, 178-189. (Algorithm 626 Collected algorithms from ACM).
- Pukkila, T., and H. Nyquist, 1985, On the frequency domain estimation of the innovation variance of a stationary univariate time series, *Biometrika*, 72, 317-323.
- Robinson, P.M., 1990, Time series with strong dependence, Paper presented at the World Congress of the Econometric Society in Barcelona, August 1990.
- Sowell, F., 1992a, Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics*, 53, 165-188.
- Sowell, F., 1992b, Modeling long-run behavior with the fractional ARIMA model, *Journal of Monetary Economics*, 29, 277-302.
- Whittle, P., 1962, Gaussian estimation in stationary time series, *Bulletin of the International Institute of Statistics*, 39, 105-129.
- Zhang, H.-C., 1991, Reduction of the asymptotic bias of autoregressive and spectral estimators by tapering, *Journal of Time Series Analysis*, 13, 451-469.

TABLE 1: Pile-up effect in MA(1) and AR(1) models for $n=25, 50, 100$. The relative frequency that $\hat{\theta}$ ($\theta = \alpha_1, \beta_1$) is in the one-sided interval $[-1.00, -0.9965]$ or $[0.9965, 1.00]$. The estimators are EML, MPL, WL, WLT, and EML(μ), the exact Gaussian ML with data corrected for the true mean. The theoretical values obtained by Cryer and Ledolter (1981) are denoted by EML-th. The number of replications for our simulations is 10000.

MA(1):	$P(\hat{\theta} = -1)$ $\beta_1 = -0.90$			$P(\hat{\theta} = 1)$ $\beta_1 = 0.90$		
n	25	50	100	25	50	100
EML	0.482	0.315	0.125	0.342	0.219	0.109
MPL	0.502	0.323	0.127	0.077	0.050	0.017
WL	0.185	0.228	0.088	0.353	0.222	0.095
WLT	0.260	0.326	0.157	0.426	0.324	0.162
EML-th	0.513	0.333	0.136	0.513	0.333	0.136
EML(μ)	0.494	0.322	0.128	0.501	0.325	0.130
AR(1):	$\alpha_1 = -0.90$			$\alpha_1 = 0.90$		
n	25	50	100	25	50	100
WL	0.146	0.169	0.016	0.310	0.167	0.019
WLT	0.216	0.239	0.029	0.358	0.203	0.032

Remark: Distances from the theoretical values which exceed 0.013, 0.012 or 0.009 depending on the theoretical values 0.513, 0.333 and 0.136 are significant at the 1% level.

TABLE 2: Minima and maxima of MSE, bias and empirical confidence level for the estimators EML, MPL, WLT in ARMA(1,1), ARFIMA(1,d,0) and ARFIMA(0,d,1) models, and where they are assumed.

ARMA(1,1)												
	$\hat{\alpha}_1$						$\hat{\beta}_1$					
	MSE											
	$(\alpha_1,$	$\beta_1)$	min	$(\alpha_1,$	$\beta_1)$	max	$(\alpha_1,$	$\beta_1)$	min	$(\alpha_1,$	$\beta_1)$	max
EML	-0.99	0.90	0.06	0.95	0.90	67.07	0.95	-0.99	0.06	0.95	0.90	64.09
MPL	-0.99	0.90	0.06	0.95	1.00	54.67	0.95	-0.99	0.06	0.95	1.00	55.95
WLT	-0.99	0.80	0.13	-0.60	-0.65	35.06	0.99	-0.95	0.28	-0.60	-0.65	36.85
	BIAS											
EML	0.95	0.90	-0.556	-0.85	-0.80	0.439	-0.95	0.90	-0.536	-0.85	-0.80	0.457
MPL	0.95	1.00	-0.494	-0.85	-0.80	0.394	0.95	1.00	-0.485	-0.85	-0.80	0.404
WLT	0.95	1.00	-0.265	-0.95	-1.00	0.316	0.95	1.00	-0.250	-0.95	-1.00	0.301
	ECL											
EML	0.55	0.60	55.5	-0.99	-0.80	99.5	-0.99	1.00	7.1	-0.40	-0.95	99.8
MPL	-0.05	0.00	63.5	-0.99	-0.80	99.7	-0.99	1.00	6.9	0.95	-0.95	100.0
WLT	-0.40	-0.45	60.1	-0.95	-0.60	99.1	-0.99	1.00	55.3	0.00	0.95	99.1
ARFIMA(1,d,0)												
	\hat{d}						$\hat{\alpha}_1$					
	MSE											
	$(d,$	$\alpha_1)$	min	$(d,$	$\alpha_1)$	max	$(d,$	$\alpha_1)$	min	$(d,$	$\alpha_1)$	max
EML	0.40	0.95	0.65	0.45	0.20	24.78	-0.50	-0.95	0.14	0.45	0.00	22.59
MPL	0.45	0.95	0.64	0.30	0.20	8.06	-0.50	-0.95	0.13	0.30	0.20	7.90
WLT	-0.30	-0.90	0.98	0.45	0.90	10.02	-0.50	-0.95	0.32	0.40	0.00	9.60
	BIAS											
EML	0.45	0.20	-0.407	-0.50	-0.95	0.103	-0.40	0.95	-0.096	0.45	0.20	0.362
MPL	0.45	0.20	-0.133	-0.50	-0.95	0.126	-0.50	0.80	-0.130	0.45	0.20	0.113
WLT	0.20	0.20	-0.113	0.45	0.95	0.206	0.45	0.90	-0.120	0.40	0.00	0.090
	ECL											
EML	0.45	0.20	53.2	-0.15	0.95	97.6	0.45	0.40	48.3	0.00	0.95	99.0
MPL	-0.50	-0.95	57.8	0.10	0.95	99.2	0.45	0.20	78.1	0.15	0.95	100.0
WLT	0.45	0.95	83.1	-0.50	-0.40	94.7	0.30	0.20	79.8	0.20	-0.90	97.1
ARFIMA(0,d,1)												
	\hat{d}						$\hat{\beta}_1$					
	MSE											
	$(d,$	$\beta_1)$	min	$(d,$	$\beta_1)$	max	$(d,$	$\beta_1)$	min	$(d,$	$\beta_1)$	max
EML	-0.40	-0.95	0.88	0.45	0.80	23.73	0.45	-0.95	0.15	0.45	0.80	23.58
MPL	0.45	-0.95	0.61	-0.50	0.40	14.78	-0.45	-0.95	0.16	-0.50	0.20	13.48
WLT	-0.40	-0.95	0.99	-0.50	0.40	12.36	-0.40	-0.95	0.31	-0.50	0.20	11.97
	BIAS											
EML	0.45	0.80	-0.403	-0.50	0.60	0.136	0.45	0.80	-0.350	-0.50	0.40	0.119
MPL	0.45	0.80	-0.154	-0.50	0.40	0.288	-0.50	0.95	-0.150	-0.50	0.40	0.252
WLT	-0.45	0.95	-0.173	-0.50	0.40	0.176	-0.45	0.95	-0.192	-0.50	0.40	0.163
	ECL											
EML	0.45	0.80	45.1	-0.30	0.95	96.3	-0.40	0.90	43.1	-0.30	-0.95	100.0
MPL	-0.50	0.40	64.9	-0.20	0.95	97.5	-0.50	0.40	59.1	0.30	-0.95	100.0
WLT	-0.50	0.40	73.5	0.20	-0.40	94.7	-0.40	0.40	68.0	-0.40	-0.95	99.1

TABLE 3: The measure of net advantage in 100 MSE of estimator A over estimator B for invertible ARMA(1,1), ARFIMA(1, d ,0) and ARFIMA(0, d ,1) models:

$$\max_{MSE_B > MSE_A} |MSE_A(\theta) - MSE_B(\theta)| - \max_{MSE_A > MSE_B} |MSE_A(\theta) - MSE_B(\theta)|$$

with $MSE_A(\theta)$ as the MSE of $\hat{\theta}$ using the estimator A . The estimators considered are: EML, MPL and WLT.

ARMA(1,1)					
		$\theta = \alpha_1$		$\theta = \beta_1$	
$A \backslash B$		MPL	WLT	MPL	WLT
EML		0.09	-47.88	3.08	-43.62
MPL			-27.03		-26.64
ARFIMA(1, d ,0)					
		$\theta = d$		$\theta = \alpha_1$	
$A \backslash B$		MPL	WLT	MPL	WLT
EML		-13.73	-6.68	-12.49	-8.65
MPL			7.09		3.34
ARFIMA(0, d ,1)					
		$\theta = d$		$\theta = \beta_1$	
$A \backslash B$		MPL	WLT	MPL	WLT
EML		-10.91	-6.72	-11.42	-5.00
MPL			4.99		7.23

Remark: A positive sign indicates an advantage of estimator A over estimator B , a negative an advantage of estimator B over A .