On the Effects of Imposing or Ignoring Long Memory when Forecasting

Michael K. Andersson*

Working Paper Series in Economics and Finance No. 225

February 1998

Abstract

Since the true nature of a time series process is often unknown it is important to understand the effects of model choice. This paper examines how the choice between modelling stationary time series as ARMA or ARFIMA processes affects the accuracy of forecasts. This is done, for first-order autoregressions and moving averages and for ARFIMA(1,d,0) processes, by means of a Monte Carlo simulation study. The fractional models are estimated using the technique of Geweke and Porter-Hudak, the modified rescaled range and the maximum likelihood procedure. We conclude that ignoring long memory is worse than imposing it, when forecasting, and that the ML estimator is preferred.

Key words: ARFIMA; Fractional integration; Periodogram regression; Rescaled range; Maximum likelihood; Forecast error

JEL Classification Code: C15; C22; C53

*Department of Economic Statistics, Stockholm School of Economics, E-mail: *stma@hhs.se*.
Suggestions by and discussions with Sune Karlsson and Johan Lyhagen are gratefully acknowledged. I also thank Lars-Erik Öller for comments on previous drafts and participants of the ISF'96, the Sixteenth International Symposium on Forecasting, Istanbul meeting, June 24-27, 1996. The usual disclamer applies.

1. Introduction

The use of erroneous models when analyzing time series may have a great impact on, for instance, the accuracy of forecasts and policy decisions. This paper investigates the forecasting performance of the usual autoregressive moving average ($ARMA$) model when the true process is fractionally integrated and the performance of the fractionally integrated ARMA ($ARFIMA$) model when the true process is non-integrated. The faulty specification is compared to the correct one.

The ARFIMA model generalizes the well-known ARIMA model by allowing for non-integer differencing powers and thereby provides a more flexible framework when examining time series. For example, fractional specifications can model data dependencies which are stronger than those allowed in stationary ARMA models, but weaker than those implied by unit root processes. This is an attractive feature when investigating economic and financial time series, which often exhibit a strong dependence between distant observations. Although the ARMA model is a special case of the ARFIMA model, its use is motivated by its simplicity and smaller bias when the differencing parameter is integer-valued.

In a previous study, Lyhagen (1997) demonstrates theoretically that ignoring long memory (i.e. a fractional differencing power) may lead to very high relative mean squared errors of prediction. However, the results of Ray (1993) show that high-order AR models forecast fractional noise series well, even in the long term. Also for fractional noise, Smith and Yadav (1994) point out that there is a potential loss from incorrectly fitting an AR model when the differencing parameter is positive. On the other hand, negative fractional differencing will bring about a performance loss of the AR model only at a horizon of one step.

This paper analyzes the forecasting performance of the respective models in terms of their mean squared prediction errors. Three popular estimation procedures for the fractional model are used, namely the one of Geweke and Porter-Hudak (1983), the modified rescaled range of Lo (1991) and maximum likelihood (Sowell, 1992), in order to compare the performance of the estimation techniques. As a consequence, we incorporate the effects of estimation bias and possible model selection mistakes. Furthermore, a test for significant differences in prediction performance is utilized.

The results suggest that the forecast errors from ARMA models are larger than those from ARFIMA models and that the maximum likelihood procedure is best.

The paper is organized as follows: the estimation techniques for the ARFIMA

model are introduced in Section 2. Section 3 contains the simulation study where the model predictions are compared and Section 4 concludes the paper.

2. Estimation of the ARFIMA Model

A time series process is said to be integrated of order d, denoted $I(d)$, if it has a stationary and invertible autoregressive moving average (ARMA) representation with uncorrelated disturbances after differencing d times, that is after applying the filter $(1-B)^d$. When d is not integer-valued (as required for ARIMA processes) the series is fractionally integrated. An ARFIMA(p,d,q) process $\{x_t\}$ is generated by

$$\phi_p(B)(1-B)^d x_t = \theta_q(B)\, a_t, \tag{2.1}$$

where the members of the sequence $\{a_t\}$ are identically and independently distributed (iid) with finite variance. If the roots of $\phi_p(B)$ and $\theta_q(B)$, the autoregressive and moving average polynomials, lie outside the unit circle and $d < 0.5$, x_t is stationary. When $d > 0$ x_t is persistent, implying that there exists a region, $d \in (0, 0.5)$, where the ARFIMA model generates stationary series ruled by long memory. This behavior cannot be mimicked with ARMA models.

The differencing filter, denoted the long-memory filter (LMF), describes the long-term dependence in the series and may be expanded as

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = 1 - dB - \frac{d}{2!}(1-d)B^2 \ldots . \tag{2.2}$$

The short-term structure is captured by the autoregressive and moving average parameters, which enables separate modelling of the long and short-run dynamics. The properties of the ARFIMA model are exhaustively described by Granger and Joyeux (1980) and Hosking (1981).

In this study the ARFIMA models are estimated by the GPH estimator of Geweke and Porter-Hudak (1983), the modified R/S or modified rescaled range $(MRR$, Lo (1991)) and the full information maximum likelihood estimator (MLE) of Sowell (1992).

2.1. The GPH Estimation Technique

Geweke and Porter-Hudak (1983) proposed a two-step procedure for the estimation of fractionally integrated models, based on a non-parametric periodogram

regression.

In a *first* step, d is estimated in the regression equation (where $I_x(\cdot)$ denotes the sample periodogram)

$$\ln\{I_x(\omega_j)\} = \beta_0 + \beta_1 \ln\left\{4\sin^2(\omega_j/2)\right\} + \eta_j \quad (\omega_j = 2\pi j/T,\ j = 1, ..., g(T)) \quad (2.3)$$

by ordinary least squares (OLS). Prior to estimation of the remaining (ARMA) parameters in a *second* step, the long-memory part of the series is filtered out using the LMF (2.2). $\hat{\beta}_1$ is under a proper choice of $g(T)$ a consistent estimator of -d. $g(T)$ is set to the integer part of T^v and the widespread choice of v is 0.5. The crucial assumption is that the spectrum of an ARFIMA(p, d, q) process is the same as the spectrum of an ARFIMA$(0, d, 0)$, for the same value of d, at low frequencies. However, Agiakloglou *et.al.* (1993) show that large positive AR and large negative MA parameters affect the spectrum at low frequencies and hence cause biased estimates.

2.2. The MRR Estimation Technique

The modified rescaled range, MRR, approach rests upon the same basic idea as the GPH procedure, that is to estimate d in a first step and the other parameters in a second one. Instead of using the log-periodogram regression, the fractional differencing parameter is estimated by the R/S statistic, which is also consistent. Lo (1991) robustifies the statistic to short-range dependence in data and this modified R/S statistic is frequently used when testing for fractional integration and for estimation of long-memory models.

The MRR statistic is defined by the ratio

$$Q_T = \frac{R_T}{\hat{\sigma}_T(k)}. \quad (2.4)$$

Following Cheung (1993), the range and standard error are estimated by

$$R_T = \max_{0<i\leq T} \sum_{t=1}^{i} (x_t - \bar{x}) - \min_{0<i\leq T} \sum_{t=1}^{i} (x_t - \bar{x}) \quad (2.5)$$

$$\hat{\sigma}_T^2(k) = \hat{\sigma}^2 + \frac{2}{T} \sum_{j=1}^{k} \sum_{i=j+1}^{T} \left(1 - \frac{j}{k}\right) (x_i - \bar{x})(x_{i-j} - \bar{x}), \quad (2.6)$$

where $\hat{\sigma}^2$ is the usual maximum likelihood variance estimate and the correction term is similar to that of Newey and West (1987). The truncation lag k depends on the short-term correlation structure of the series and is set, according to Andrews' (1991) data-dependent formula, to the integer part of $(3T/2)^{1/3}\left\{2\hat{\rho}/\left(1-\hat{\rho}^2\right)\right\}^{2/3}$, with T and $\hat{\rho}$ denoting the serial length and sample first-order autocorrelation coefficient respectively. Asymptotic results suggest that $\ln Q_T/\ln T$ approaches $1/2$ for short-range dependent processes, and thus our estimator of d is constructed as

$$\hat{d} = \frac{\ln Q_T}{\ln T} - \frac{1}{2}. \tag{2.7}$$

The MRR estimator is biased when the true process is a moving average with large negative parameter values (Cheung (1993)).

2.3. The ML Estimation Technique[1]

The maximum likelihood estimator differs from the other two; while the GPH and MRR estimators are performed in two steps the ML procedure estimates all parameters in one single step.

Given that the roots of $\phi_p(B)$ and $\theta_q(B)$ in (2.1) are all outside the unit circle and the disturbances are normally distributed with mean zero, the likelihood function to be maximized is the well-known

$$L\left(X_T, \Sigma(\varsigma)\right) = (2\pi)^{-\frac{T}{2}} \left|\Sigma(\varsigma)\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} X_T' \Sigma^{-1}(\varsigma) X_T\right\}, \tag{2.8}$$

where X_T is the $T \times 1$ data vector and autocovariance matrix $\Sigma(\varsigma)$ is a function of the unknown parameter vector $\varsigma = [d, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_a^2]'$. The log of (2.8) is maximized directly with respect to the vector ς. Explicit expressions for the elements of $\Sigma(\varsigma)$ and practical issues concerning the ML estimation are discussed by Sowell (1992).

Generally the MLE is efficient; however, it has some drawbacks are. For example, the MLE requires distinct roots, performs less well when d is close to the non-stationary area (i.e. close to 0.5) and exhibits small-sample bias when d and AR parameters are estimated jointly (Tschernig (1993)). Furthermore, mis-specification of the likelihood function will in general lead to inconsistent estimates of both the differencing parameter and the short-run parameters.

[1] A FORTRAN routine for the estimation of the ARFIMA model was generously supplied by Fallaw Sowell.

3. The Monte Carlo Study

ARFIMA(p, d, q) processes are generated by the algorithm of Diebold and Rudebusch (1991). The processes are primarily chosen to generate persistent or autoregressive series, but negative d-values and MA processes are also considered. In each Monte Carlo iteration we estimate ARMA and ARFIMA models, from which predictions are generated. For each prediction horizon h we calculate the loss functions $g^m(x_h, \hat{x}_h)$, where x_h denotes the actual (simulated) value of the process at time $T + h$ and $\hat{x}_h$ the corresponding forecast using model $m = \{ARMA, ARFIMA\}$. The accuracies are compared by the mean squared prediction error ($MSPE$) loss function

$$g^m(x_h, \hat{x}_h) = R^{-1} \sum_{r=1}^{R} (\hat{x}_{h,r} - x_{h,r})^2 , \tag{3.1}$$

where $r = 1, ..., R$ is the Monte Carlo replicate. R is selected to 1000. The mean squared prediction error of model m is expected to increase with the prediction horizon and eventually coincide with, or exceed, the variance of the process. At that particular horizon, that is when the MSPE variance ratio

$$MVR_m = g^m(x_h, \hat{x}_h) / \mathrm{V}(y) \tag{3.2}$$

is greater than or equal to one, the model forecasts are not more accurate than when using the process mean. The difference in h-step performance, expressed by the null-hypothesis of no difference in MSPE, is evaluated by a usual matched-pair t-test,

$$t_h = \frac{g_h^{ARMA} - g_h^{ARFIMA}}{\sqrt{\mathrm{V}\left[g_h^{ARMA} - g_h^{ARFIMA}\right]}}, \tag{3.3}$$

along with asymptotically normal critical values, motivated by the large number of replicates and the central limit theorem.

Results for first-order autoregressions and fractionally integrated processes of length 100 and forecasting horizons up to step 20 are reported in the figures of this paper. In addition, larger sample sizes ($T = 225$, 400 and 625) and longer prediction horizons (up to $0.2T$) have been investigated. An increase in the serial length implies more accurate estimates (in particular for the fractional procedures), and a relatively better prediction performance for the ARFIMA model. Otherwise, the conclusions drawn from the $T = 100$ case are not altered by the results obtained for the larger sample sizes.

3.1. Autoregressions

The forecasting performances for generated first-order autoregressions are examined using the specification

$$x_t = \phi x_{t-1} + a_t, \tag{3.4}$$

where $\{a_t\}$ is a sequence of $iidN(0,1)$ random deviates. In order to handle model selection uncertainty, the lag-orders p and q of the ARMA and fractional models are selected by the Bayesian information criterion, BIC, of Schwartz (1978) from the values $\{0, 1, 2, 3\}$ in each Monte Carlo iteration. The case of no differencing parameter in the ARFIMA model is ruled out.

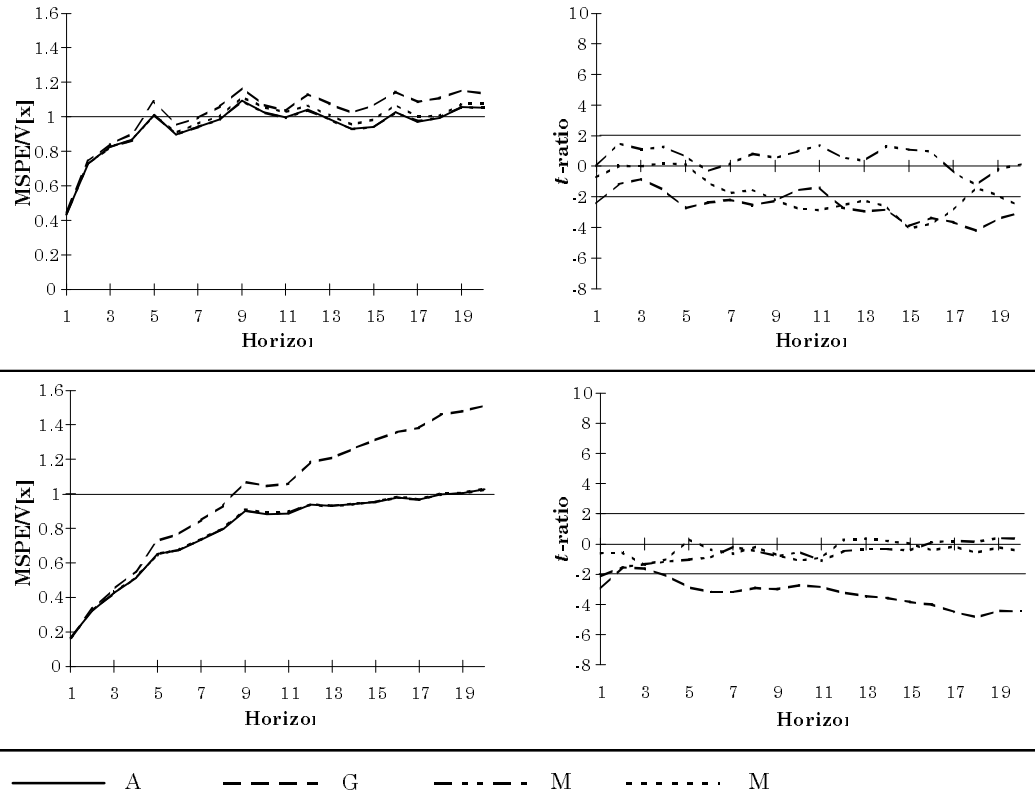
The study investigates autoregressions with parameter $\phi = 0.5$, 0.7 and 0.9. Figure 3.1 presents the MSPE variance ratio (3.2) for each of the models together with the associated t–statistic (3.3) for $\phi = 0.7$ and 0.9 processes.

The maximum possible prediction horizon increases with the size of the AR parameter. In particular, when $\phi = 0.5$ (not reported in the figures), the MVRs rapidly assume values above one, which implies short prediction horizons. For prediction steps one to three, the model mean squared prediction errors are all quite similar. However, a few significant differences (from the ARMA performance) are found for the GPH and maximum likelihood estimated models. Using the MRR procedure, the forecasts are as accurate as those produced by the ARMA model. For prediction horizons above three, forecasting is valueless.

As the parameter ϕ becomes larger, the prediction horizon (using the ARMA model) increases to a maximum of five (maybe eight) steps ahead for $\phi = 0.7$ processes and 18 steps when $\phi = 0.9$. For $\phi = 0.7$, and in particular $\phi = 0.9$, the bias of the GPH procedure is quite large and thus the forecast errors obtained are also quite large. As a consequent, the GPH performance is often significantly worse than that of the non-fractional model, that is the t-values fall below -2, for $\phi = 0.7$ processes. For almost unit root processes, the GPH estimator produces prediction errors that are usually significantly larger than those of the ARMA model. The other estimation techniques predict highly short-term dependent processes (almost) as well as the ARMA models; their t-ratios usually fall between -2 and 2.

Turning to the mean squared prediction errors, we find no support in favor of the non-fractional specification. In a forecasting context, no substantial loss is experienced when imposing long memory by using ARFIMA models on AR processes. In particular, the MRR procedure performs very well, but the ML estimator is also quite accurate.

Figure 3.1: Estimated mean squared prediction errors and t-values of the MSPE test for an AR(1) process of length 100.



The figure presents the MSPE variance ratio (3.2) (first column) and the t-statistic of the MSPE test (3.3) (second column) of the estimation procedures for horizons to 20. The rows correspond to $\phi = 0.7$ and 0.9. The data are generated according to (3.4).

3.2. Persistent Processes

The fractionally integrated series are generated as ARFIMA$(1, d, 0)$ processes,

$$(1 - \phi B)(1 - B)^d x_t = a_t, \tag{3.5}$$

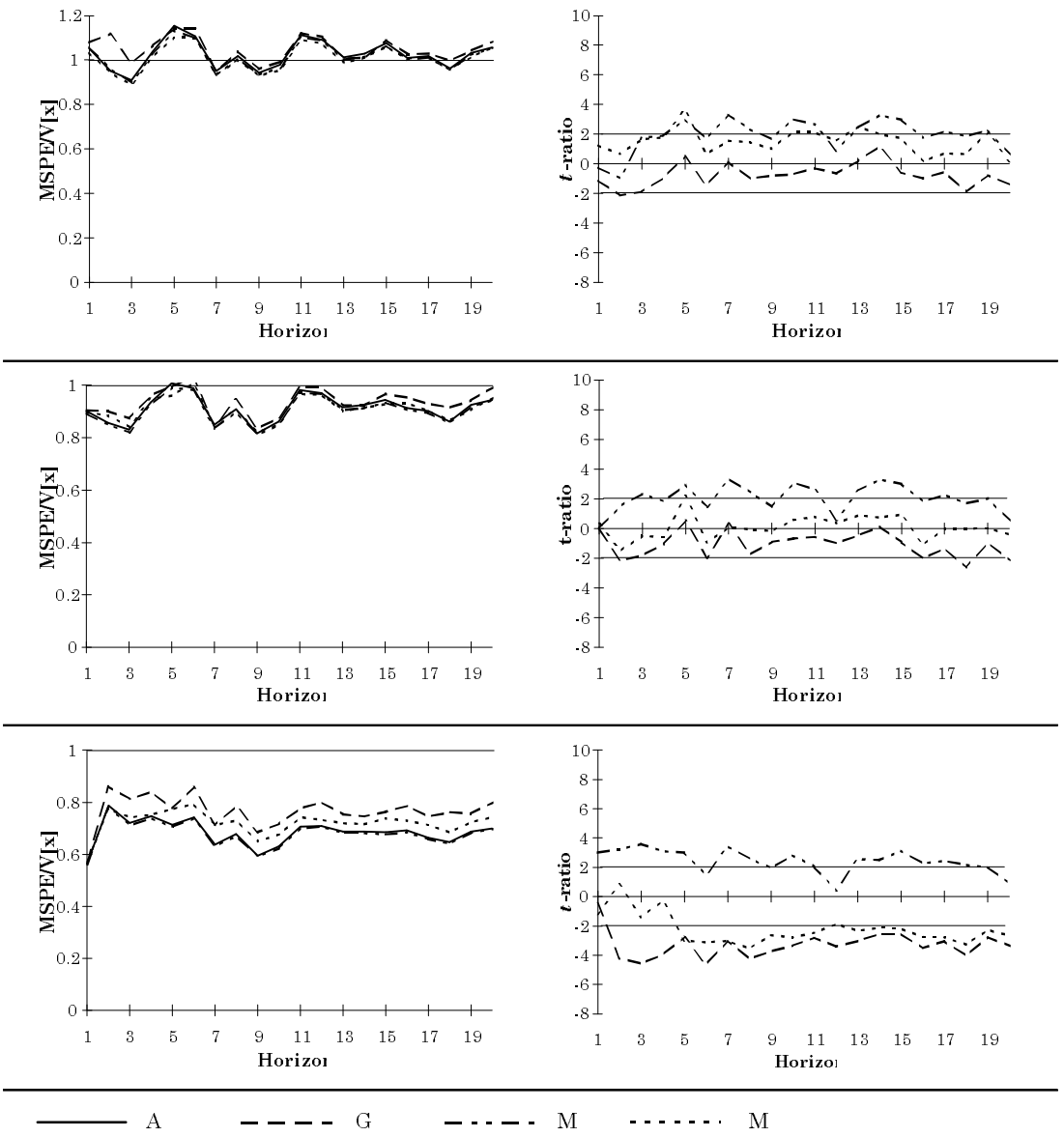
where a_t is again $iidN(0, 1)$. The process is denoted fractionally differenced white noise ($FDWN$) when ϕ equals zero. In the case of generated fractionally integrated processes, the models to be estimated are specified by the BIC, where the orders p and q are selected from the values $\{0, 1, ..., 10\}$. The MSPE variance ratios (3.2) and t-values of the MSPE test (3.3) are reported in Figures 3.2 and 3.3.

Weakly persistent FDWN processes with the parameter d equal to 0.1 or 0.2, being almost white noise, do not contain enough structure to be forecasted. Already at horizon one, the MVRs (see Figure 3.2 for $d = 0.2$) exceed one and subsequently predictions are useless, regardless of estimation technique and model choice.

Results for $d = 0.4$ in Figure 3.3 suggest that only occasional ARFIMA MVRs of unity are found for a more persistent fractional noise process. In fact, up to step 100 forecasts using the ARFIMA models (with the MLE) are possible. Furthermore, the ARFIMA-MLE is always significantly better than the ARMA model; the t-values are well above 2. The GPH estimator is better than the ARMA specification at step five and above (significantly at step seven) and the MRR at horizon eight (ten). The relatively weak performance of the MRR procedure is explained by over-adjustment for (non-existing) short-term dependence in data. Therefore, additional AR parameters, and thus mis-specified models, are as a rule required to capture all autocorrelation in the series. Judged by the mean squared errors, the performance of the MRR procedure for ARFIMA$(0, 0.4, 0)$ processes is close to that of the ARMA model, while the GPH procedure and in particular the method of maximum likelihood generates notably lower values. The results suggest, when forecasting is worthwhile, that the ARFIMA-MLE should be used for predicting fractional noise processes.

An introduction of AR-type short memory, according to equation (3.5), enhances the predictability compared to FDWN for the same value of d. Unlike the case of fractionally differenced white noise with $d = 0.2$, processes with $d = 0.2$ and $\phi = 0.2$ may be predicted several steps ahead. The performances of the ARFIMA model, with all estimation techniques, and the ARMA model are fairly similar, and the t-ratios suggest that the MRR may be better (by a close margin)

Figure 3.2: Estimated mean squared prediction errors and t-values of the MSPE test for an ARFIMA(1,0.2,0) process of length 100.



See note to Figure 3.1. The rows correspond to $\phi = 0.0$, 0.2 and 0.5 respectively. The data are generated according to (3.5).

than the other estimators. Again for $d = 0.2$, when ϕ equals 0.5 we notice that the MRR procedure produces the most accurate forecasts, followed by the non-fractional model. It appears to be the case that the 0.2 differencing power is not large enough to cause problems for the ARMA model. The GPH estimator gives the worst performance in this situation, but even this is not bad.

Proceeding with ARFIMA$(1, 0.4, 0)$ processes, when $\phi = 0.2$ the GPH predictions are better than those of the ARMA model. However, the variance of $\left(g_h^{ARMA} - g_h^{ARFIMA}\right)$ is quite large, leading to small t-ratios which are usually positive but not always significantly larger than zero. The MRR procedure works well in the presence of AR parameters, and exhibits t-ratios favoring the ARFIMA model. However significant, the differences in mean squared prediction errors are quite small. The maximum likelihood estimated ARFIMA models display the best performance and the mean squared prediction errors are notably lower than those of the other procedures, especially from horizon seven and onwards.

When $d = 0.4$ and $\phi = 0.5$, the MLE estimated models again have much smaller MSPE figures than the ARMA model. The differences are not that great for the MRR estimated ARFIMA models, but they are strongly significant. As in the case of AR processes with large values of ϕ, the GPH procedure generates poor forecasts for ARFIMA$(1, 0.4, 0)$ processes with intermediate to large positive autoregressive parameters.

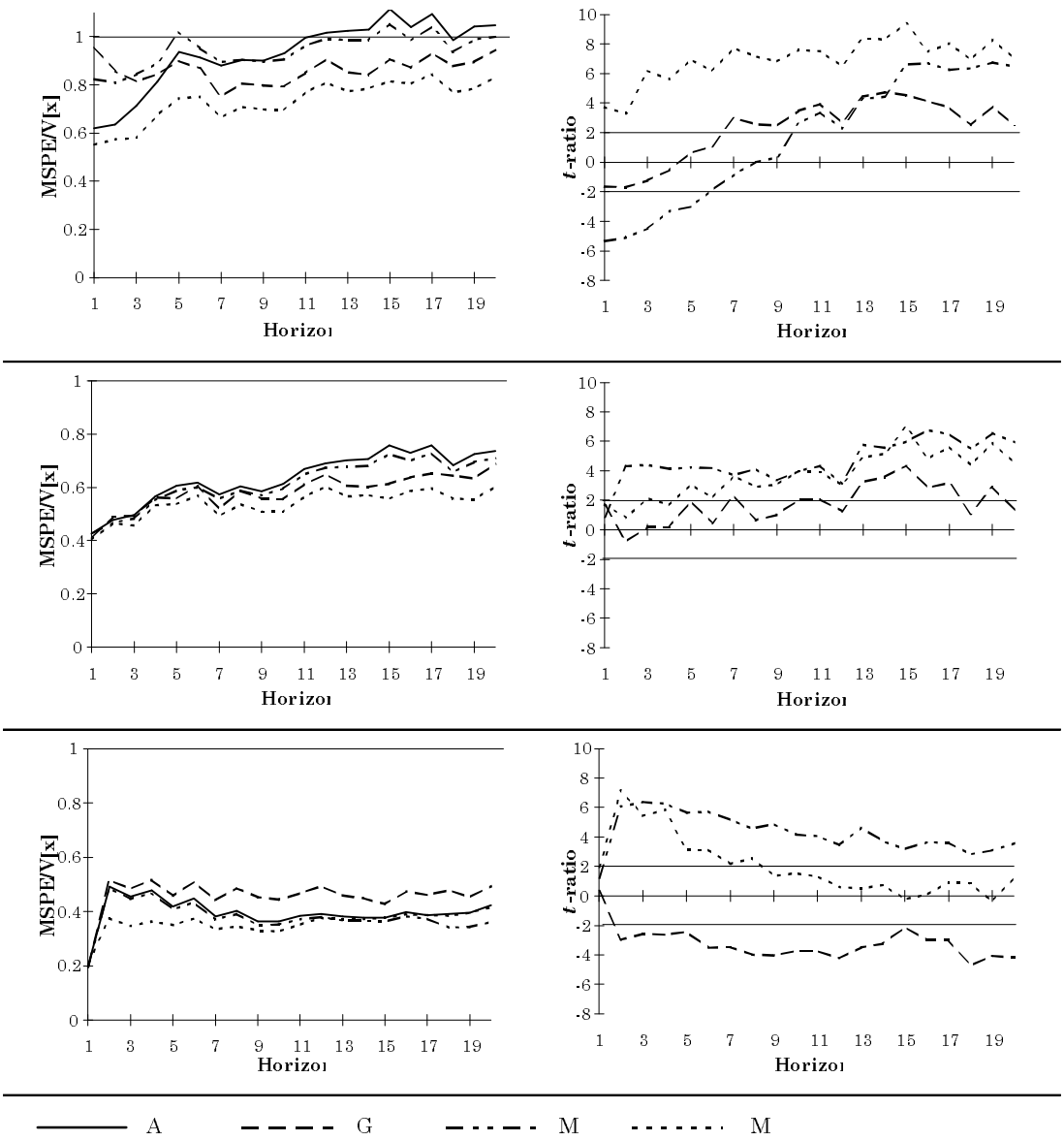
The non-fractional ARMA model works fairly well for short-range predictions when the true process is fractionally differenced, whereas the ARFIMA specifications are optimal for long-range forecasting. In general, for the fractionally integrated processes investigated, the best predictions are obtained when using the maximum likelihood ARFIMA estimation procedure.

3.3. Moving Averages and Anti-persistent Processes

For MA(1) processes the choice of model hardly matters. The ARMA model is found to have a slightly better performance, and the maximum likelihood estimated ARFIMA models are in turn marginally more accurate than those estimated by the GPH and MRR procedures. The simplicity of the non-fractional specification motivates the use of ARMA models in this case. However, the true process is unknown and the modelling choice hardly affects the forecasting performance.

For anti-persistent ($d < 0$) processes also, the difference in forecasting performance between the ARMA and ARFIMA models is very small. This is in

Figure 3.3: Estimated mean squared prediction errors and t-values of the MSPE test for an ARFIMA(1,0.4,0) process of length 100.



See note to Figure 3.1. The rows correspond to $\phi = 0.0$, 0.2 and 0.5 respectively. The data are generated according to (3.5).

agreement with the results of Smith and Yadav (1994).

4. Concluding Remarks

This paper investigates and compares the forecasting performance of ARMA and ARFIMA models. The results suggest that the modified rescaled range and maximum likelihood estimators for the fractionally integrated model generate predictions that are almost as accurate as those of the ARMA model, when the true process is a first-order autoregression or moving average. The predictions are also quite similar for negatively fractionally integrated processes. The estimator of Geweke and Porter-Hudak performs badly when the true process is an AR(1) with a large positive parameter.

For persistent processes, the GPH and MRR produce short-term predictions that are worse than the ARMA forecasts. However, the opposite is found for intermediate and long-range forecasts. The MLE is better than the ARMA model for all horizons.

The GPH technique experiences problems when combining fractional integration with an autoregression, especially for intermediate to large AR parameters. In the case of simultaneous short and long memory, the MRR and in particular the MLE procedures generate better predictions than the ARMA model.

In general, it is worse to ignore than to impose long memory when forecasting. Overall, the MLE is the best of the fractional estimators and the ARFIMA model is better than the ARMA.

References

Agiakloglou, C., P. Newbold and M. Wohar, 1993, Bias in an estimator of the fractional difference parameter, *Journal of Time Series Analysis*, 14, 235-246.

Andrews, D., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, 59, 817-858.

Cheung, Y.-W., 1993, Tests for fractional integration: a Monte Carlo Investigation, *Journal of Time Series Analysis*, 14, 331-345.

Diebold, F.X. and G.D. Rudebusch, 1989, Long memory and persistence in aggregate output, *Journal of Monetary Economics*, 24, 189-209.

Diebold, F.X. and G.D. Rudebusch, 1991a, Is consumption too smooth? Long memory and the Deaton paradox, *The Review of Economics and Statistics*, (Feb), 1-9.

Diebold, F.X. and G.D. Rudebusch, 1991b, On the power of Dickey-Fuller tests against fractional alternatives, *Economics Letters*, 35, 155-160.

Geweke, J. and S. Porter-Hudak, 1983, The estimation and application of long memory time series models, *Journal of Time Series Analysis*, 4, 221-238.

Granger, C.W.J. and R. Joyeux, 1980, An introduction to long-memory time series models and fractional integration, *Journal of Time Series Analysis*, 1, 15-29.

Hosking, J. R. M., 1981, Fractional differencing, *Biometrica*, 68, 165-176.

Lo, A. W., 1991, Long-term memory in stock market prices, *Econometrica*, 59, 1279-1313.

Lyhagen J., 1997, Essays on univariate long-memory models, *Ph. D. Thesis*, Department of Statistics, Uppsala University, Sweden.

Newey, W. and K. West, 1987, A simple positive definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703-705.

Ray, B. K., 1993, Modelling long-memory processes for optimal long-range predictions, *Journal of Time Series Analysis*, 14, 511-525.

Smith, J. and S. Yadav, 1994, Forecasting costs incurred from unit differencing fractionally integrated processes, *International Journal of Forecasting*, 10, 507-514.

Sowell, F., 1992, Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics*, 53, 165-188.

Schwartz, G., 1978, Estimating the order of a model, *The Annals of Statistics*, 2, 461-464

Tschernig, R.,1993, Tackling and understanding the small sample bias in ARFIMA models, *Unpublished manuscript*, Department of Finance, University of Limburg, Maastricht, The Netherlands.