Journal of Forecasting, Vol. 15, 107–125 (1996)

Model Selection and Forecasting for Long-Range Dependent Processes

NUNO CRATO
Stevens Institute of Technology, Hoboken, NJ, USA

BONNIE K. RAY
New Jersey Institute of Technology, Newark, NJ, USA

ABSTRACT

Fractionally integrated autoregressive moving-average (ARFIMA) models have proved useful tools in the analysis of time series with long-range dependence. However, little is known about various practical issues regarding model selection and estimation methods, and the impact of selection and estimation methods on forecasts. By means of a large-scale simulation study, we compare three different estimation procedures and three automatic model-selection criteria on the basis of their impact on forecast accuracy. Our results endorse the use of both the frequency-domain Whittle estimation procedure and the time-domain approximate MLE procedure of Haslett and Raftery in conjunction with the AIC and SIC selection criteria, but indicate that considerable care should be exercised when using ARFIMA models. In general, we find that simple ARMA models provide competitive forecasts. Only a large number of observations and a strongly persistent time series seem to justify the use of ARFIMA models for forecasting purposes.

KEY WORDS ARFIMA models; forecasting; long-memory; model selection; persistance

In this paper, we discuss the use of fractionally integrated autoregressive moving-average (ARFIMA) models for forecasting time series that present long-memory characteristics. Fractionally integrated models have proved useful tools in the analysis of time series with long-range dependence. Since their introduction by Granger and Joyeux (1980) and Hosking (1981), these models have been extensively used in various applications, such as the analysis of geophysical phenomena (e.g. Noakes *et al.*, 1988; Bloomfield, 1992), econometric modeling (e.g. Diebold and Rudebush, 1989 Sowell, 1992), financial time series analysis (e.g. Shea, 1991; Cheung, 1993a), and long-range forecasting (e.g. Geweke and Porter-Hudak, 1993; Ray, 1993b; Sutcliffe, 1993). However, selection and estimation of fractionally integrated models is more difficult than that of standard ARMA models. When both long- and short-range components are present in the data, their behaviour is hard to distinguish, making model selection difficult. Additionally, exact likelihood estimation techniques present computational problems in calculating the autocovariances needed to evaluate the likelihood function (Sowell, 1992). Other 'semi-parametric' techniques result in significant finite sample biases and large

CCC 0277–6693/96/020107–19
Received July 1995
Revised October 1995

variances (see Agiakloglou, Newbold, and Wohar, 1993). Approximate likelihood techniques, based on the conditional sum-of-squares function or on the Whittle spectral likelihood, are available (see Beran, 1994) but perform well only for large sample sizes. Consequently, the use of fractionally integrated models can present difficulties.

In particular, little is known about the practical usefulness of the estimated models, especially for forecasting: few results are available on the success of different model selection criteria in choosing an ARFIMA model, few studies have discussed the quality of the different estimation procedures in practice, and no systematic study has been conducted on the impact of the estimation procedures and model selection criteria on the accuracy of the predictions. The problems are magnified in light of theoretical results showing that the misspecification of fractional models can have adverse consequences for long-range forecasting (Crato, 1992; Ray, 1993a).

The process (X_t) is said to be an ARFIMA(p, d, q) process if it is a solution of the difference equations

$$\phi(B)(1-B)^d X_t = \theta(B)\varepsilon_t \tag{1}$$

where $\phi(\cdot)$ and $\theta(\cdot)$ represent the polynomials $\phi(z) = 1 - \phi_1 z \ldots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ in the backwards shift operator B: $B^j X_t = X_{t-j}$. Fractional powers of the differencing operator are defined through its binomial expansion $(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k$. The innovations (ε_t) are assumed to be white noise, i.e. a zero-mean uncorrelated process with constant variance σ_ε^2.

We will assume that $\phi(z)$ and $\theta(z)$ have all their roots outside the unit circle and no common roots. With these conditions, and with $d \in (-1/2, 1/2)$, the process (X_t) is stationary and invertible. In the particular case $d = 0$, (X_t) is an autoregressive moving-average (ARMA) model. When $d \neq 0$, the autocorrelation function, $\rho(k)$, has a slow hyperbolic decay, $\rho(k) \sim C|k|^{2d-1}$ as $|k| \to \infty$. In the case $0 < d < 1/2$, the autocorrelations are not summable and the process has *long memory* or *persistence*. In the case $-1/2 < d < 0$, the process may be thought of as *anti-persistent*. See Brockwell and Davis (1991) for a general discussion of the properties of fractionally integrated processes.

In this paper, we assess the effects of model selection and estimation procedures on forecasts of long-range dependent processes made from estimated ARMA and ARFIMA models via a large-scale simulation study. We simulate processes generated by various models: five different ARFIMA models, a Fractional Gaussian Noise (FGN) model (Mandelbrot and Van Ness, 1968), a Fractional EXP (F-EXP) model (Beran, 1993), a long autoregressive model, and an ARMA(1,1) model having the autoregressive coefficient slightly larger than the moving-average coefficient. All of these processes exhibit some type of long-memory characteristics. We fit ARMA models to the simulated series using the time-domain approximate MLE procedure of conditional least-squares, and ARFIMA models to the simulated series using three estimation procedures: the frequency domain Whittle-type approximate maximum likelihood procedure developed by Fox and Taqqu (1986)—FT, the periodogram regression procedure suggested by Geweke and Porter-Hudak (1983)—GPH, and the time-domain approximate MLE procedure of Haslett and Raftery (1989)—HR. No simulation studies that we are aware of have assessed the performance of the HR procedure, and we believe the results we have obtained in this regard are of interest in their own right.

We select a model using three automatic selection criteria and assess the performance of the selection criteria in the presence of long memory. Criteria such as the Akaike Information Criterion (AIC) and the information criterion of Schwarz (SIC), have been used in various applications for the selection of ARFIMA models (e.g. Hosking, 1984; Sowell, 1992; Cheung, 1993a). However, no theoretical results have been derived concerning the efficiency and/or bias of these criteria when applied to long-memory processes. Recently, Schmidt and Tschernig (1993) conducted a small simulation study to investigate the performance of these criteria, along with the bias-corrected version of AIC (AICc) of Hurvich and Tsai (1989), the BIC criterion, and the Hannan–Quinn (HQ) criterion, in selecting the true generating model in the particular case of ARFIMA$(0, d, 0)$ processes. They found that the BIC criterion performed poorly compared to the other criteria, and that the HQ criterion performed slightly worse than the SIC criterion. For details concerning these criteria, see Brockwell and Davis (1991) or de Gooijer *et al.* (1985). Only the Whittle procedure was used for model estimation. In our study, we investigate the performance of the AIC, AICc and the SIC criteria. The AIC and SIC are included because of their common use in practice. The AICc is included to investigate whether it performs better in small samples than AIC for long memory processes.

Specifically, we address the following issues. First, we investigate the number of times the selected model matches the generating model. This allows an evaluation of the success of the different criteria, used with different estimation procedures, in the detection of the generating model. Note that the set of generating models is larger than the set of fitted models. Thus this exercise in simulation also provides information regarding another, more practical, issue: which ARFIMA models are likely to be selected for general long-memory alternatives. This situation more closely approximates actual practice; in real situations, the time series do not usually present the exact characteristics of the estimated models.

Second, we evaluate the performance of the selection criteria on the basis of the selected model's forecasts. As is well known in applied work, the model that provides the best fit is not necessarily the one that yields the best multi-step-ahead forecasts. Previous studies concerning long-range dependent processes have not addressed the model selection issue from the standpoint of forecasting. It is not known which criteria, if any, should be used for the selection of a model for forecasting and under what conditions the different criteria can be expected to perform well.

Third, we investigate the use of ARMA models for forecasting long-memory processes. As we have noted, it has been shown theoretically that this type of misspecification can induce large forecasting errors in the long run. However, since the detection of long memory is difficult, a question of practical interest is the likelihood of selecting an ARMA model in the presence of long-memory behaviour. Additionally, the computational problems associated with estimation of fractional models make it of practical interest to know how well the estimated ARMA models compare with estimated ARFIMA models in terms of forecast accuracy.

The remainder of the paper is organized as follows. The next section briefly describes the design of the simulation study. The third section discusses the performance of the model selection criteria in terms of their ability to select the generating model, or an approximating model having long-range dependence if the true model is outside the selection set. The fourth section discusses the performance of the model-selection criteria in terms of the selected model's forecast accuracy. The fifth section discusses the selection and forecast accuracy of ARMA models. A final section presents conclusions.

A complete presentation of the all simulation results is available as a working paper (Ray and Crato, 1994) and can be accessed via the World Wide Web at http://chaos.njit.edu/~borayx/lrdfore.ps.

THE LAYOUT OF THE SIMULATION STUDY

We have generated time series driven by nine different models as follows. In all cases, the simulated generating noise was standard Gaussian:

(1) ARFIMA(0, −0.3, 0), which is an anti-persistent simple fractional noise with all autocorrelation coefficients negative for lags $k \neq 0$.
(2) ARFIMA(0, 0.1, 0), which is a persistent fractional noise with all autocorrelation coefficients positive. The autocorrelations decay at a slow hyperbolic rate but are all relatively small, thus the long-memory properties are not very prominent.
(3) ARFIMA(0, 0.4, 0), which is a persistent fractional noise with all autocorrelation coefficients positive. The autocorrelation function displays positive and relatively large values, even for large lags. The long-memory properties are very prominent.
(4) ARFIMA(1, 0.3, 0) with $\phi = 0.65$, a process with persistent long-memory and a short-memory autoregressive component. The combination of a large AR component with a long-memory component is notoriously hard to estimate (see, for example, Agiakloglou, Newbold, and Wohar, 1993). In the autoregressive representation of the process, the first AR coefficient is $(d + \phi) \simeq 1$. Thus estimation of an AR(1) model may point towards a near unit root.
(5) ARFIMA(0, 0.3, 1) with $\theta = -0.65$, a process with persistent long-memory and a short-memory moving-average component. In the moving-average representation of the process, the first MA coefficient is $(d + \theta) = -0.35$ and all the other coefficients have very small values. Thus, the estimation may point towards an MA(1) model.
(6) ARMA(1, 1) with $\phi = 0.87$ and $\theta = -0.59$, a short-memory process that displays properties similar to persistence when the autoregressive coefficient is slightly larger than the moving-average component (O'Connell, 1974).
(7) ARMA(10, 0), a long autoregression capable of displaying some type of persistence. The coefficients in the polynomial $\phi(B)$ were set as those resulting from the AR(10) estimation of a single simulated ARFIMA(0, 0.4, 0) process.
(8) Fractional Gaussian Noise (FGN) with $H = 0.9$. The value of the self-similarity parameter makes the asymptotic behaviour of the autocorrelation function of the process similar to that of a fractional noise with $d = 0.4$.
(9) F-EXP model, defined by the spectral density function

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2\times 0.3915} \exp(-1.343 - 2.856|\lambda| - 0.428|\lambda|^2) \tag{2}$$

This model cannot be represented as a finitely parameterized ARFIMA. The particular parameters are taken from Beran (1993). It is a persistent long-memory model, since its spectrum diverges at the zero frequency, displaying the following behaviour:

$$f(\lambda) \sim C|\lambda|^{-2\times 0.3915} \text{ as } \lambda \to 0 \tag{3}$$

Series from Models (1)–(7) were generated using the algorithm of Hosking (1984) as implemented in the Haslett and Raftery (1989) FORTRAN subroutines (these programs are available via anonymous ftp from *statlib.cmu.edu*), with modification to allow $-1/2 < d < 0$. Series from Model (8) were generated using the Cholesky decomposition of the exact covariance matrix for the process. Series from Model (9) were generated using the discrete inverse Fourier transform of the spectral density given in equation (2). For each model, we generated 500 time series, each having length $N = 396$, and estimated models for two different samples sizes n: (1) $n = 120$, which corresponds to 10 years of monthly observations, and (2) $n = 360$, which corresponds to 30 years of monthly observations. These sample sizes are similar to the ones used in other simulation studies (e.g. Cheung and Diebold, 1994) and are most common with business and economic data. In each case, we used 36 additional data points for out-of-sample forecast assessment. This allows us to see how modeling the long-memory characteristic aids in long-term forecasting.

We used the following estimation procedures for the ARFIMA models:

(1) **GPH**—Geweke and Porter-Hudak (1983) periodogram regression method. This is a two-step procedure. First, the parameter d is estimated by a regression based on the following relation, which holds with u_j approximately i.i.d. for low-order Fourier frequencies $\omega_j = 2\pi j/n$:

$$\ln I_n(\omega_j) = a - 2d \ln\left(2 \sin \frac{\omega_j}{2}\right) + u_j \qquad j = 1, 2, \ldots, [n^{1/2}] \tag{4}$$

where $I_n(\cdot)$ denotes the periodogram of the series. Second, the series is approximately fractionally differenced using $\hat{d}$ and the ARMA coefficients estimated by approximate time-domain maximum likelihood. The mean of the process, used in forecasting, is estimated as the average of the series. Details about the GPH procedure can be found in Geweke and Porter–Hudak (1983) or in Brockwell and Davis (1991). It is well known that this two-step method may yield large finite sample biases when short-memory ARMA components are also present (Agiakloglou, Newbold, and Wohar, 1993). Cheung (1993b) shows that the GPH procedure may also yield positively biased estimates of d when the underlying process has infrequent shifts in mean, but is robust to ARCH effects. It continues to be widely used in applied work, as in Diebold and Rudebush (1989), Baillie and Pecchenino (1991), Cheung (1993a), Cheung and Lai (1993), and Hassler and Wolters (1995).

(2) **HR**—Haslett and Raftery (1989) approximate time-domain maximum likelihood procedure. This procedure uses the Durbin–Levinson algorithm with approximate partial autocorrelation coefficients to evaluate the conditional means and variances of an ARFIMA process, which are then used to calculate the one-step-ahead predictors needed in the innovations representation of the likelihood function. A mean term is estimated explicitly as the solution of the log-likelihood normal equations with estimated ARFIMA parameters. The partial autocorrelation coefficients are calculated exactly up to lag L and then are approximated. In our study, we set $L = \min\{n, 200\}$. This technique essentially corresponds to minimization of the squared residuals of the process (see Beran, 1992, p. 409). The original Haslett and Raftery algorithm maximizes the log-likelihood in the range $0 < d < 1/2$ using a grid search. We modified the algorithm to maximize the log-likelihood in the range $-1/2 < d < 1/2$. The modified algorithm can be obtained from the authors upon request.

(3) **FT**—Fox and Taqqu (1986) frequency-domain approximate maximum likelihood method. Using a Whittle approximation to the log-likelihood function (Brockwell and Davis, 1991, p. 529, equation (13.2.26)), the function

$$\frac{2}{n} \sum_{j=1}^{[n/2]} \frac{I_n(\omega_j)}{f^*(\omega_j)} + \frac{2}{n} \sum_{j=1}^{[n/2]} \log f^*(\omega_j) \tag{5}$$

is minimized, where $f^*(\omega_j)$ is the spectrum of the ARFIMA model being estimated and $I_n(\omega_j)$ is the periodogram of the series. The mean of the process is estimated as the average of the series.

We did not use exact maximum likelihood estimation due to its large computational burden. The autocovariance function for the estimated model must be computed and the variance-covariance matrix inverted at each iteration. Moreover, evidence suggests that Sowell's exact maximum likelihood and the FT approximate procedure yield very similar results in the realistic case of unknown mean. See Cheung and Diebold (1994) for a comparison of exact and Whittle approximate maximum likelihood estimation methods for ARFIMA$(0, d, 0)$ processes with estimated mean term.

All ARMA models were estimated using only one method—the time-domain method of Haslett and Raftery modified so that d is restricted to be zero. Again, this essentially corresponds to minimizing the sum of squared residuals, a standard technique for ARMA model estimation. In this case, the estimation of μ reduces to $\mu = \bar{x}$. We set $L = \min\{n, 200\}$, as for the ARFIMA models.

For each series, we estimated nine ARMA(p, q) models with p, $q = 0, 1, 2$ and nine ARFIMA(p, d, q) models with $-1/2 < d < 1/2$ and p, $q = 0, 1, 2$. Optimal models were selected using three common automatic selection criteria, the AIC, the AICc, and the SIC. We also look at one other statistic relevant to the model selection, that denoted by **All**. This represents the intersection of all three criteria. We think it is important to know to what extent the three criteria choose the same model. The small range of p and q values reflects the range of ARMA models usually estimated in practice. Additionally, results in Ray (1991) concerning the use of AR models for modeling long-memory processes indicate that the AIC criterion generally selects low-order AR models ($p \leqslant 4$) for long-range dependent processes, even when the amount of persistence is strong and p is allowed to range up to 20. Longer AR models or higher-order ARMA models do a reasonable job of emulating long-memory behaviour, but can produce unstable parameter estimates, which may affect the model's forecast accuracy. For a very few of the generated series, one (or more) of the estimation methods did not converge or gave estimated parameters outside the stationarity/invertibility range for some models. We report results only for series in which all models were successfully estimated by all estimation methods.

Exact k-step-ahead out-of-sample forecasts were computed using the innovations algorithm (Brockwell and Davis, 1991, p. 171). The autocovariances needed for the algorithm were computed using the method of Sowell (1992).

To compare the forecasting performance of the selected models for the different estimation procedures, we computed the MSE $(h_1 - h_2)$ statistic—the comparative Mean Square Error statistic defined over the forecast horizon $n + h_1$, $n + h_1 + 1, \ldots, n + h_2$ for various values of h_1, h_2 and with reference to the forecasts obtained by the true model with known parameters. This allows for comparisons across different estimation procedures:

$$\mathrm{MSE}(h_1 - h_2) := \frac{\sum_{i=h_1}^{h_2} (\hat{X}_{n+i|n} - X_{n+i})^2 - \sum_{i=h_1}^{h_2} (\hat{X}^g_{n+i|n} - X_{n+i})^2}{\sum_{i=h_1}^{h_2} (\hat{X}^g_{n+i|n} - X_{n+i})^2} \tag{6}$$

where X_{n+i} represents the ith out-of-sample observation after the last observation (nth) used for estimation, $\hat{X}_{n+i\,|\,n}$ represents the forecast made by the estimated model for time $n + i$ with the knowledge of observations up to time n, and $\hat{X}^g_{n+i\,|\,n}$ represents the same forecast but made by using the generating model with the generating parameters. This statistic is obviously zero when the generating model with known parameters is used, is positive when the estimated model behaves worse than the generating model, and is negative in the opposite case.

We also computed a similar statistic with reference to the forecasts made using the true model with estimated parameters. This allows for comparisons across different selection criteria within the same estimation method. The corresponding results are presented in the working paper (Ray and Crato, 1994) and they essentially corroborate the conclusions extracted from the MSE $(h_1 - h_2)$.

CAN AUTOMATIC SELECTION CRITERIA FIND THE GENERATING LONG-MEMORY MODEL?

We will divide our discussion regarding the performance results of the different criteria according to the three main types of simulated models.

Fractional noise models

For the generated models and sample sizes considered, we find that the SIC criterion performs considerably better than any of the other criteria. This confirms results reported in Schmidt and Tschernig (1993) based only on the Whittle estimation method.

Figure 1 shows the results for the fractional noise model with $d = 0.4$. AIC and AICc almost always choose the same model, and are less successful in selecting the generating ARFIMA $(0, 0.4, 0)$ model. AIC and AICc correctly selected models are essentially a subset of the correctly selected models by application of SIC, as can be seen from the bars labelled 'All' in

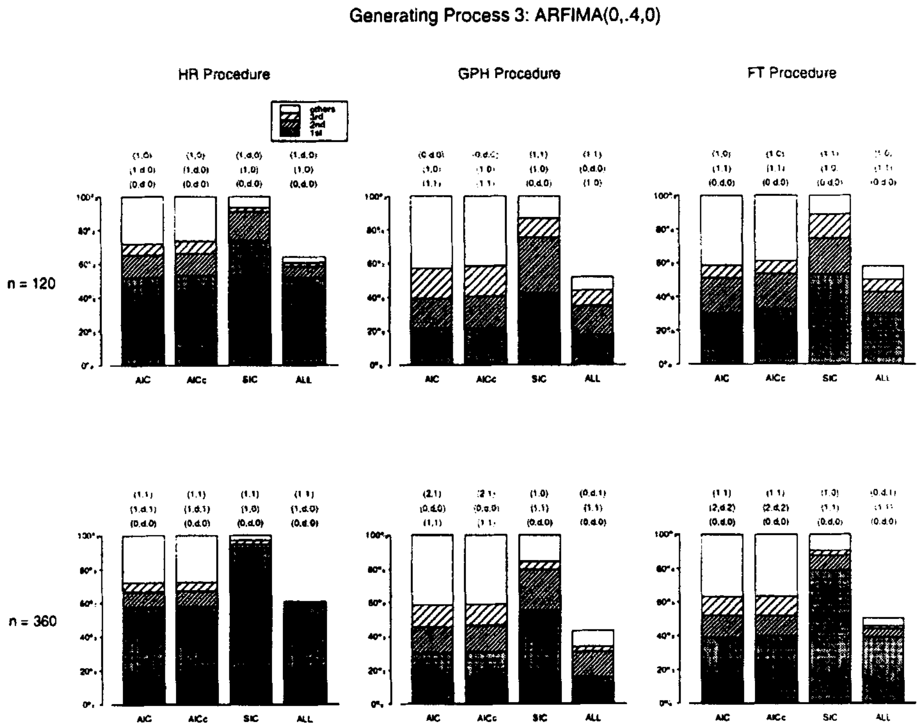

Figure 1. Percentage of times a model was selected using the different selection criteria. The models selected most often are read from *bottom to top* above each bar, with the model selected the most listed directly above each bar, the second most selected model listed second from each bar and the third most selected model listed third from each bar.

Figure 1. Successful model selection was relatively low for all estimation procedure and selection combinations, in general below 80% when $n = 120$. For the larger sample size, $n = 360$, the percentage of cases in which the SIC chose the correct model increased. However, the performances of the AIC and AICc were only slightly better than when $n = 120$, with the most improvement coming in conjunction with the FT estimation procedure.

The most difficult process to identify was the ARFIMA(0, 0.1, 0). This is not surprising, considering that the autocorrelations are all very small. A white-noise model was the model most frequently chosen. Interestingly, the anti-persistent fractional noise ($d = -0.3$) was easier to identify than the persistent one ($d = 0.4$) with each criterion.

Mixed ARFIMA models

For both generated ARFIMA processes, and for each estimation procedure and automatic selection criterion, the percentage of successful identifications was extremely small. Overwhelmingly, short-memory ARMA models were selected when $n = 120$, indicating that a short-memory component and long-range dependence are very difficult to distinguish in small samples.

With the two mixed ARFIMA processes, AIC and AICc performed better than the SIC and yielded better results when the FT estimation procedure was used. In all cases, a criterion's performance increased with the sample size. The SIC performed worse than AIC and AICc in these models because it more heavily penalizes the parameterization, pointing towards a single-

Generating Process 4: ARFIMA(1,.3,0)

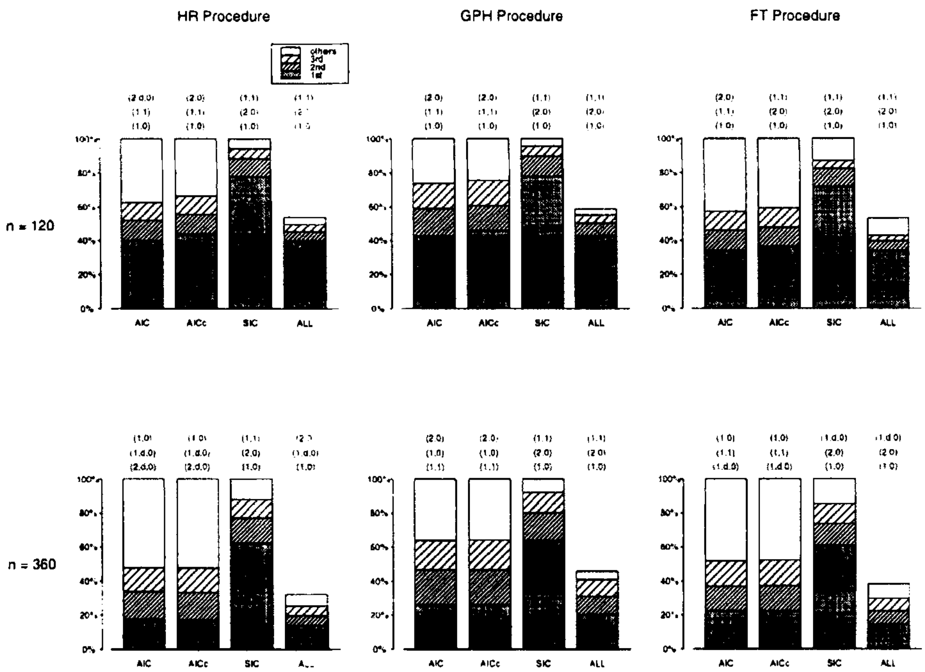

Figure 2. Percentage of times a model was selected using the different selection criteria. The models selected most often are read from *bottom to top* above each bar, with the model selected most listed directly above each bar, the second most selected model listed second from each bar and the third most selected model listed third from each bar

parameter model, either an AR(1) or an MA(1). The fact that none of the criteria chose the correct model in most cases further suggests that, in small samples, the selection criteria are biased in the choice between short- and long-memory models.

Figure 2 shows the results for Model 4, the ARFIMA (1, 0.3, 0).

Non-ARFIMA long-memory models

The results for processes in this category were very disappointing. For the FGN and the F-EXP processes studied here, we found that a long-memory model is not likely to be selected from among ARMA-ARFIMA models. Roughly, only about 20% of the time was a fractionally integrated model selected for the FGN process, with a fractionally integrated model selected for the F-EXP process only about 10% of the time. ARMA models were overwhelmingly chosen. Figure 3 shows the results for the FGN (Model (8)).

For the long autoregression, however, the selection of an ARFIMA model, in particular the pure fractional noise, occurred quite frequently. The result is not surprising, since this model can match well the autocorrelation structure using a very parsimonious model.

The analysis of the different selection criteria does not point in a clear direction regarding which selection criterion to use. However, it provides an indication that selection criteria may behave differently with different estimation procedures. For a more complete picture of the results for all the generated processes included in the selection set, Table I shows the percentage of times each criteria selected the correct model for each generating process.

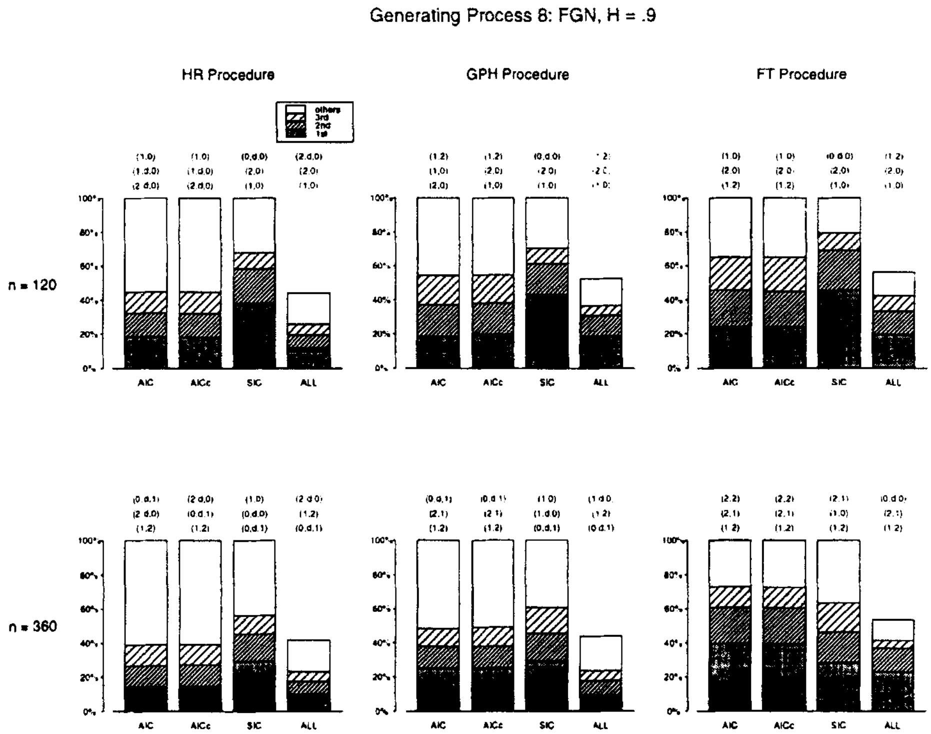

Figure 3. Percentage of times a model was selected using the different selection criteria. The models selected most often are read from *bottom to top* above each bar, with the model selected the most listed directly above each bar, the second most selected model listed second from each bar and the third most selected model listed third from each bar

Table I. Correct (%) model selection table for all generating processes in the selection set

Generating processes and sample sizes	HR procedure			GPH procedure			FT procedure		
	AIC	AICc	SIC	AIC	AICc	SIC	AIC	AICc	SIC
ARFIMA(0, −0.3, 0), $n = 120$	69.0	69.6	86.0	35.4	37.8	48.0	49.2	52.2	78.4
ARFIMA(0, −0.3, 0), $n = 360$	64.4	65.4	94.8	29.2	29.6	55.8	48.0	49.6	87.4
ARFIMA(0, 0.1, 0), $n = 120$	16.4	16.6	9.6	4.4	4.2	2.6	19.8	21.0	13.6
ARFIMA(0, 0.1, 0), $n = 360$	36.6	37.0	31.6	9.4	9.6	11.4	35.8	36.6	38.0
ARFIMA(0, 0.4, 0), $n = 120$	51.8	53.2	73.6	17.4	18.0	42.2	30.0	32.2	53.4
ARFIMA(0, 0.4, 0), $n = 360$	57.6	58.0	92.2	15.6	15.8	55.4	39.0	39.8	78.8
ARFIMA(1, 0.3, 0), $n = 120$	3.0	2.8	0.2	2.2	2.4	0.2	8.6	8.4	4.0
ARFIMA(1, 0.3, 0), $n = 360$	16.0	16.2	9.2	10.0	10.0	6.4	21.8	21.8	11.6
ARFIMA(0, 0.3, 1), $n = 120$	2.0	2.0	0.4	2.8	2.8	1.4	11.8	12.0	3.6
ARFIMA(0, 0.3, 1), $n = 360$	24.2	24.4	21.8	20.8	21.2	17.6	43.4	43.6	32.6
ARMA(1,1), $n = 120$	13.2	12.8	7.2	34.0	34.2	25.8	14.8	14.8	10.4
ARMA(1,1), $n = 360$	19.6	19.4	25.0	55.8	56.6	63.6	32.8	34.2	32.2

For each process and sample size, the table shows the percentage of times the correct model was chosen by the particular selection criteria.

The table clearly singles out the HR and the FT procedures as more likely to lead to a correctly selected model, with the best procedure dependent on the generating process. The GPH trailed the other procedures for the pure fractional noise processes, but was competitive with the HR procedure for the mixed ARFIMA models. The differences in the number of correct selections for HR, GPH, and FT are most likely attributable to differences in the performance of the estimation techniques in estimating the parameters of the model. Looking at equation (4), we see that GPH is at a disadvantage from a parameter efficiency standpoint, in that only $[n^{1/2}]$ points are used to estimate d, while the other methods use all n data points. However, GPH was designed as a semi-parametric procedure of estimating d without knowledge of the form of the spectral density except at frequencies close to zero. Thus it should not be expected that GPH perform as well as the HR and FT methods for parameter estimation and, consequently, model selection. The performance of the GPH procedure for forecasting is discussed in the next section. As an example of the estimation performance of each method, the average parameter values and standard errors of the averages for the ARFIMA(0, 0.4, 0) and ARFIMA(1, 0.3, 0) models are given in the Appendix.

As to how the different procedures perform for different models, Table I shows HR working better for pure fractional noise processes, with FT working better for mixed ARFIMA. Thus a general recommendation regarding model identification is difficult to formulate. GPH yields the worst results, in general, for samples of the sizes studied here. The HR and FT procedures performed similarly, perhaps with some global advantage for the FT, especially for mixed models. Moreover, the FT is a very straightforward method, which can be implemented quite easily.

Regarding the selection criteria, we suggest SIC and AIC (or AICc) should be used simultaneously. In the presence of a pure fractional noise, SIC seems preferable. In case of a mixed ARFIMA model, AIC seems superior. Since, in practice, it is not possible to know in advance whether a pure fractional noise or a mixed ARFIMA is present, this recommendation of using both criteria has a limited utility. However, as SIC and AIC seem to overlap for the correct identification, if both criteria point in one particular direction—an ARFIMA$(0, d, 0)$ or some ARFIMA(p, d, q)—that result can be taken as a relatively reliable indication of the plausibility of a class of models.

It is interesting to note that for the ARMA(1, 1) model, a short-memory model, both the HR and FT procedures preferred ARFIMA models over ARMA models. This example points out the dangers of including ARFIMA models in the selection set by default. Crato and Ray (1995) provide an additional discussion of this point. The GPH procedure selected the correct model more often in this case, perhaps due to the extremely poor fit obtained when ARFIMA $(0, d, 0)$ models were estimated using GPH.

WHEN IS THE SELECTED MODEL THE BEST FORECASTING MODEL?

Forecasting is an important component of time series analysis, and the forecasting performance of a selection criterion or an estimation procedure needs to be assessed independently of other characteristics of the criterion and the estimation method. Table II indicates the average value of the comparative MSE statistic over all models for each estimation method and selection criterion.

Since the MSE statistic is computed relative to the forecasts obtained using the true model with known parameters, we can compare the forecasting accuracy across estimation methods.

There are a few aspects that deserve to be treated separately.

Estimation procedures

As far as forecasting performance is concerned, the global results shown in Table II indicate that all methods perform similarly. HR produced lower mean-squared errors in the first steps when $n = 120$ but higher for long-range forecasts and higher at all forecast steps when $n = 360$. This might indicate that HR produces worse estimates of the processes' mean. GPH and FT performed very similarly at all steps.

Globally, all procedures are in a virtual tie, including GPH. Even when ARMA components are present—models (4), (5), (6), and (7)—the GPH procedure performed very competitively

Table II. Comparative MSE (%) for all generating processes

Sample size	Steps ahead	HR procedure			GPH procedure			FT procedure		
		AIC	AICc	SIC	AIC	AICc	SIC	AIC	AICc	SIC
$n = 120$	1–6	19.73	19.56	19.56	26.93	26.90	30.02	28.33	28.27	28.77
	7–12	28.98	28.61	27.87	19.31	13.23	19.76	19.18	19.13	19.44
	13–24	15.77	15.72	15.97	11.78	11.80	12.19	11.84	11.79	12.09
	25–36	16.15	16.13	16.16	11.75	11.74	11.92	10.93	10.90	10.75
$n = 360$	1–6	12.08	12.11	18.36	7.35	7.31	14.08	6.84	6.78	11.41
	7–12	28.76	28.78	28.43	16.16	16.11	16.59	16.46	16.41	16.63
	13–24	13.28	13.33	13.34	8.01	8.04	8.43	9.40	9.36	9.35
	24–36	13.12	13.20	12.89	10.74	10.67	10.15	10.66	10.65	10.45

in terms of forecast accuracy. For the non-ARFIMA long-memory models studied here, GPH was, for the most part, the best procedure to use. This fact is somewhat surprising, since GPH has been shown to present significant biases in parameter estimation. However, it could be explained by considering the autoregressive forecasting equations for the estimated model. If the biased parameter estimates combine to give an autoregressive representation with approximately unbiased AR coefficients, then the resulting forecasts will be competitive with the forecasts obtained using the correct model.

Selection criteria

Based on the aggregate results, the MSE performances of all three criteria are rather similar when $n = 120$. When $n = 360$, SIC appears to perform better for longer forecasting horizons (seven to twenty-four steps ahead), and AIC and AICc appear to perform better than SIC at one to six steps ahead. However, when the forecast results are examined for each model individually, we find that results from model (5), the ARFIMA(0, 0.3, 1) process, unduly influence the aggregate results. The SIC criterion tends to choose (incorrectly) an MA(1) model for this process, whereas the AIC and AICc tend to choose higher-order ARMA models or ARFIMA models, although not necessarily the correct model. The forecasts resulting from the SIC selected model are very poor relative to those of the true model, especially for one to six steps ahead. At larger forecast horizons, neither the selected model or the true model produce very accurate forecasts, thus the relative difference in forecast performance of the two models is small. Exceptions to these generalizations were found for some processes, but the essential conclusion of the study is that the selection criteria are generally equivalent in terms of long-range forecast performance. As examples, Tables III and IV show the results for model (2), the weakly persistent ARFIMA(0, 0.1, 0) process, and model (4), the strongly persistent ARFIMA (1, 0.3, 0) process.

As mentioned in the section describing the layout of the simulation study, the MSE statistic was also computed relative to the forecasts obtained using the true model with *estimated* parameters although those results are not reported here. The conclusions drawn concerning model selection for forecasting were essentially the same, although the relative MSE values were, on average, 1–2% smaller. An exception occurred for processes having short-range dependent components and small sample sizes. For these models, the relative MSE computed using the correct model with estimated parameters was significantly smaller than the relative

Table III. Comparative MSE (%) for generating process 2: ARFIMA(0, 0.1, 0); ARFIMA/ARMA estimated models

Sample size	Steps ahead	HR procedure			GPH procedure			FT procedure		
		AIC	AICc	SIC	AIC	AICc	SIC	AIC	AICc	SIC
$n = 120$	1–6	3.83	3.76	2.59	3.71	3.71	6.16	5.79	5.46	2.79
	7–12	3.28	3.25	2.81	3.25	3.24	2.72	3.47	3.27	2.91
	13–24	2.80	2.81	2.47	2.47	2.46	2.38	2.71	2.55	2.50
	25–36	1.65	1.65	1.48	1.49	1.48	1.42	1.55	1.53	1.51
$n = 360$	1–6	2.56	2.48	1.15	1.57	1.54	1.35	2.46	2.37	1.07
	7–12	1.25	1.27	0.96	1.32	1.32	0.95	1.69	1.68	1.08
	13–24	0.73	0.71	0.56	0.66	0.66	0.55	0.83	0.82	0.54
	25–36	0.54	0.54	0.46	0.55	0.55	0.51	0.69	0.69	0.51

Table IV. Comparative MSE (%) for generating process 4: ARFIMA(1,0.3,0); ARFIMA/ARMA estimated models

Sample size	Steps ahead	HR procedure			GPH procedure			FT procedure		
		AIC	AICc	SIC	AIC	AICc	SIC	AIC	AICc	SIC
$n = 120$	1–6	19.44	18.95	14.49	16.67	16.69	13.18	17.81	17.72	14.11
	7–12	32.40	31.75	22.05	26.75	26.91	20.69	24.40	24.32	20.09
	13–24	23.25	23.00	20.63	23.02	23.13	21.20	18.97	18.81	19.78
	25–36	12.54	12.45	11.47	13.32	13.35	10.83	10.84	11.02	9.85
$n = 360$	1–6	7.58	7.63	5.43	6.61	6.62	6.20	5.22	5.07	5.93
	7–12	18.76	18.66	13.65	13.61	13.59	10.80	10.08	9.84	10.86
	13–24	10.19	10.15	8.23	10.90	10.88	9.96	10.41	10.16	9.88
	25–36	12.32	12.36	9.87	12.84	12.83	11.98	10.63	10.53	11.23

MSE computed using the correct model with known parameters. Additionally, the relative MSE values for the GPH estimated models were smaller using the estimated correct model. This result reflects the larger finite sample bias of the GPH estimates.

To obtain information about the precision of our forecast results, we used the technique of Lewis and Orav (1989, Ch. 9). We divided the 500 forecast MSE values into 10 sections of 50 replications each and found the standard error of the average forecast MSE value for each of the 10 sections. This value is different for each sample size, estimation procedure, selection criterion, and forecast step range, but in general, we found that the standard error was less than 4% when $n = 120$ and less than 1% or 2% (sometimes as low as 0.5%) when $n = 360$. An exception to this general rule occurred for model (5), the ARFIMA(0, 0.3, 1) model, when the SIC was used for model selection and forecast errors from one to six steps ahead were of interest. As mentioned above, the SIC tended to select an MA(1) model in that case, leading to poor forecasts and large variability in forecast MSE.

HOW WELL DO ARMA MODELS PREDICT LONG-MEMORY PROCESSES?

The theoretical disadvantages of misspecifying long-memory processes as regular ARMA models have been pointed out by various authors, such as Brockwell and Davis (1991) and Ray (1993a). Other studies, such as Byers and Peel (1993) and Tiao and Tsay (1994), have suggested that, in practice, *estimated* ARMA models may perform competitively with *estimated* ARFIMA models. Our simulation study substantially bolsters this claim. Additionally, ARMA models have significant advantages in the estimation process, although Beran (1994) shows that when estimation of d is taken into account for standard ARIMA models, precision of the parameter estimates is the same as for ARFIMA models.

For each process, sample size, and replication, we recorded the models selected in the restricted set of ARMA (p, q) processes and their out-of-sample forecasting performance. For ARFIMA processes, the forecasts from the ARMA models chosen by the three selection criteria are competitive for the first forecasting steps (one to six) in most cases, with an exception being the non-persistent ARFIMA model (model (1)) when $n = 120$, for which an MA(1) was often selected from among ARMA models. For larger forecasting horizons, the ARMA models remain competitive when $n = 120$ but tend to perform slightly worse than the true model at long forecast horizons when $n = 360$ and the amount of persistence is strong.

Tables V and VI exemplify the typical findings, showing results for model (2), the weakly persistent ARFIMA (0, 0.1, 0) process, and model (4), the strongly persistent ARFIMA(1,0.3,0) process.

In all the other cases, ARMA forecasts are equivalent or better than the best results attained by ARFIMA models with all selection criteria and for all forecasting horizons. The best procedure/criteria combination among ARMA/ARFIMA choices varies substantially according to the generating model and the sample size, leading to the dominance of the HR/SIC or the HR/AIC combinations in many cases and to the dominance of the FT/SIC combination in many others. Combinations that in certain cases are very successful, perform very poorly in certain other circumstances. The ARMA forecasts, however, always perform very competitively with the best ARMA/ARFIMA selection-estimation combination.

Concerning the converse question, the accuracy of forecasts generated from ARFIMA models when the generating process is ARMA, results for model (6), the ARMA(1, 1) model,

Table V. Comparative MSE (%) for generating process 2: ARFIMA (0, 0.1, 0); ARMA estimated models

Sample size	Steps ahead	AIC	AICc	SIC
$n = 120$	1–6	3.24	3.23	6.29
	7–12	3.01	3.00	2.74
	13–24	2.55	2.54	2.44
	25–36	1.50	1.50	1.45
$n = 360$	1–6	1.31	1.27	1.59
	7–12	1.17	1.17	0.98
	13–24	0.65	0.65	0.57
	25–56	0.54	0.54	0.52

Table VI. Comparative MSE (%) for generating process 4: ARFIMA (1, 0.3, 0); ARMA estimated models

Sample size	Steps ahead	AIC	AICc	SIC
$n = 120$	1–6	16.61	16.43	13.26
	7–12	25.46	25.43	20.75
	13–24	21.07	21.07	20.27
	25–36	11.23	11.30	11.27
$n = 360$	1–6	6.70	6.43	6.52
	7–12	14.10	13.91	12.45
	13–24	10.87	10.99	10.56
	25–36	12.88	12.93	13.36

supply a partial answer. From Table I, we see that the correct ARMA (1, 1) model is selected less than 15% of the time when $n = 120$ and less than 35% of the time when $n = 360$ and the HR or FT methods are used for estimation. In fact, ARFIMA (p, d, q) models were chosen in the majority of cases. From Tables VII and VIII, we see that the relative forecast errors are about 1–2% higher when ARFIMA models are included in the selection set than when the selection set is restricted to ARMA models. This indicates that using ARFIMA models to predict ARMA processes results in a small loss of forecast accuracy. We note, however, that the particular ARMA (1, 1) model investigated in our study was chosen to emulate the behavior of a long memory process. Forecast performance of ARFIMA models for other ARMA processes may be worse than that represented here.

We believe that our results strongly endorse the use of ARMA models for forecasting possible ARFIMA processes of the length normally found in business and economic applications and recommend the use of ARFIMA forecasting models only for clearly persistent time series and when a relatively large number of observations is available.

Table VII. Comparative MSE (%) for generating process 6: ARMA(1, 1); ARFIMA/ARMA estimated models

Sample size	Steps ahead	HR procedure			GPH procedure			FT procedure		
		AIC	AICc	SIC	AIC	AICc	SIC	AIC	AICc	SIC
$n = 120$	1–6	10.10	9.85	9.11	11.05	11.07	11.92	11.59	11.76	10.37
	7–12	11.11	10.58	10.19	12.06	11.90	12.40	11.53	11.20	10.43
	13–24	11.21	11.24	10.76	11.69	11.78	12.74	12.02	11.88	10.85
	25–36	8.82	8.73	8.63	10.06	10.15	10.39	8.95	8.82	8.62
$n = 360$	1–6	3.65	3.64	3.73	2.17	2.23	2.75	4.61	4.53	3.50
	7–12	5.12	5.17	4.69	2.74	2.76	3.74	5.43	5.41	3.79
	13–24	3.25	3.30	4.05	2.35	2.32	2.62	3.76	3.58	3.30
	25–36	4.18	4.27	4.45	3.76	3.76	3.63	4.36	4.33	4.11

Table VIII. Comparative MSE (%) for generating process 6: ARMA (1, 1); ARMA estimated models

Sample size	Steps ahead	AIC	AICc	SIC
$n = 120$	1–6	9.42	9.62	12.74
	7–12	8.90	8.74	9.26
	13–24	9.36	9.33	9.68
	25–36	7.29	7.30	7.40
$n = 360$	1–6	2.14	2.09	2.39
	7–12	3.25	3.26	3.40
	13–24	2.64	2.65	2.73
	25–36	2.84	2.84	2.84

CONCLUSIONS

By means of a simulation study, we have assessed various issues related to the practical use of long-memory fractionally integrated models. First, our results indicate that the selection of an ARFIMA model is not an easy task. Automatic selection criteria were very unreliable in distinguishing between long and short memory and in choosing the correct model. In the presence of ARFIMA processes, AIC and AICc behaved very similarly and frequently chose ARMA or overparameterized ARFIMA models, and the SIC criteria frequently chose ARMA models. In the presence of non-ARFIMA persistent models all criteria tended to overlook the presence of long memory and to point in the direction of ARMA processes.

When a pure fractional noise was present—ARFIMA$(0, d, 0)$—SIC could be recommended; when a mixed model was present—ARFIMA(p, d, q)—all criteria had a very low success rate. However, *inside each class of models* the success rate of the criteria was relatively encouraging. AIC and AICc performed better than SIC for a mixed ARFIMA model and SIC performed much better than both AIC and AICc for a pure fractional noise. Thus, when both type of criteria point either in the direction of the ARMA class of short-memory models or in the direction of the ARFIMA class of long-memory models, the resulting selected model should be reasonably plausible. This conclusion suggests that the most difficult problem in this model identification setting is the choice between short-memory ARMA models and long-memory ARFIMA models and not an order identification inside each type of models.

Second, our results show that success in the use of selection criteria is not independent of the estimation procedure. On the whole, the HR procedure—the time-domain maximum likelihood method of Haslett and Raftery (1989)—and the FT procedure—spectral domain Whittle likelihood method suggested by Fox and Taqqu (1986)—performed similarly. But the success of the different combinations of the estimation procedures with the selection criteria varied according to the type of process. In case of pure fractional noise models, the HR/SIC combination yielded the best results. In the case of mixed ARFIMA models, the FT/AIC combination provided the best results. The GPH procedure has been shown to present significant estimation biases and our simulation results confirm these results for model identification.

Third, our results suggest that the selection of a model for forecasting purposes can be different from the selection and estimation of a model for analytic purposes. GPH emerged as a very reasonable estimation procedure and no one method systematically dominated for forecasting, especially at long lead times.

Fourth, our results indicate that estimated ARFIMA models do not perform well in forecasting. In general, estimated ARMA models, even for long-memory processes, outperform or have a competitive performance. This seems to caution against the use of ARFIMA forecasting models. Theoretically, ARFIMA models should give the best forecasts for ARFIMA processes. However, the low success rate in the selection of the right ARFIMA model, along with the large variance in the estimation of the parameters, seem to indicate that ARFIMA models should only be used when strong persistence is present and when a large number of observations allows, in principle, reliable estimates.

Future research will focus on a theoretical investigation of why the standard automatic model selection criteria do not perform well for long-memory processes and development of an improved model selection tool when the goal is long-range forecasting. However the results presented here, found via simulation, can be of immediate use to practitioners involved in modeling and forecasting data displaying long-range dependent behaviour.

APPENDIX

The tables below give the average estimated parameter values for the ARFIMA (0, 0.4, 0) process and the ARFIMA (1, 0.3, 0) process. The standard error of the average is given in parentheses. For the ARFIMA (0, 0.4, 0) model, the FT estimate of d has a smaller bias and standard error than the other estimates when $n = 120$. For $n = 360$, the GPH estimate is slightly less biased than the FT estimate, but has a larger standard error, as can be expected from the method's use of only $[n^{1/2}]$ sample points for estimation of d. The standard errors of the averages of the 500 HR and FT estimates of d are close to the asymptotic standard error of $(6/500\pi^2 n)^{1/2}$.

For the ARFIMA (1, 0.3, 0) model with $\phi = 0.65$, the HR method results in extremely poor estimates, with d underestimated and ϕ overestimated. The two-step GPH procedure also performs poorly, with d overestimated at the first step and ϕ consequently underestimated at the second step. The FT procedure performs the best in terms of bias for both small and large sample sizes. Thus from both an estimation standpoint, as well as a model selection and forecasting standpoint, the FT estimation method is recommended.

Mean and standard error of the estimated ARFIMA (0, 0.4, 0) model parameter

	$d = 0.40$	
Method	$n = 120$	$n = 360$
HR	0.3754 (0.0035)	0.4302 (0.0022)
GPH	0.3610 (0.0079)	0.3812 (0.0055)
FT	0.4130 (0.0037)	0.4217 (0.0026)

Mean and standard error of the estimated ARFIMA (1, 0.3, 0) model parameters

	$d = 0.30$		$\phi = 0.65$	
Method	$n = 120$	$n = 360$	$n = 120$	$n = 360$
HR	0.1046 (0.0055)	0.1917 (0.0050)	0.7720 (0.0040)	0.7258 (0.0043)
GPH	0.4255 (0.0059)	0.3879 (0.0055)	0.4999 (0.0098)	0.5519 (0.0050)
FT	0.3360 (0.0072)	0.3094 (0.0060)	0.5827 (0.0069)	0.6247 (0.0058)

ACKNOWLEDGEMENTS

We wish to thank two anonymous referees and the department editor for insightful comments which helped us improve the paper. We also thank John Perez for computational assistance. The research of Bonnie Ray was supported, in part, by NSF under grant #DMS-9409273. A working paper version of this paper is accessible via the World Wide Web at http://chaos.njit.edu/borayx/lrdfore.ps.

REFERENCES

Agiakloglou, C., Newbold, P. and Wohar, M., 'Bias in an estimator of the fractional difference parameter', *Journal of Time Series Analysis*, **14** (1993), 235–46.

Baillie, R. T. and Pecchenino, R. A., 'The search for equilibrium relationships in international finance: The case of the monetary model', *Journal of International Money and Finance*, **10** (1991), 582–93.

Beran, J., 'Statistical methods for data with long-range dependence (with discussion)' *Statistical Science*, **7** (1992), 404–27.

Bloomfield, P., 'Trends in global temperature', *Climate Change*, **21** (1992), 1–16.

Beran, J., 'Maximum likelihood estimation of the differencing parameter for invertible short- and long-memory ARFIMA models', preprint, University of Zurich, 1994.

Beran, J., 'Fitting long-memory models by generalized linear regression', *Biometrika*, **80** (1993), 817–22.

Brockwell, P. J. and Davis, R. A., *Time Series: Theory and Methods*, 2nd edition, New York, Springer-Verlag, 1991.

Byers, J. D. and Peel, D. A. 'Forecasting with fractionally integrated models', Department of Economics, University of Wales, 1993.

Cheung, Y.-W., 'Long-memory in foreign exchange rates', *Journal of Business and Economic Statistics*, **11** (1993a), 93–101.

Cheung, Y.-W., 'Test for fractional integration: A Monte Carlo investigation', *Journal of Time Series*, **14** (1993b), 331–46.

Cheung, Y.-W. and Diebold, F. X., 'On the maximum-likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean', *Journal of Econometrics*, **62** (1994), 301–16.

Cheung, Y.-W. and Lai, K. S., 'A fractional cointegration analysis of purchasing power parity', *Journal of Business and Economic Statistics*, **11** (1993), 103–12.

Crato, N., 'Long-memory time series misspecified as nonstationary ARIMA', *American Statistical Association, Business and Economic Statistics Section Proceedings*, (1992), 82–7.

Crato, N. and Ray, B. 'Some problems in the overspecification of ARMA and ARIMA models using ARFIMA models', *Proceedings of the 3rd Congress of the Portuguese Statistical Society*, June 1995, forthcoming.

de Gooijer, J. G., Abraham, B., Gould, A. and Robinson, L., 'Methods of determining the order of an autoregressive-moving average process: a survey', *International Statistical Review*, **53** (1985), 301–29.

Diebold, F. X. and Rudebush, G. D., 'Long memory and persistence in aggregate output', *Journal of Monetary Economics*, **24** (1989), 189–209.

Fox, R. and Taqqu, M., 'Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series' *Annals of Statistics*, **14** (1986), 517–32.

Geweke, J. and Porter-Hudak, S., 'The estimation and application of long-memory time series models', *Journal of Time Series Analysis*, **4** (1983), 221–38.

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

Hassler, U. and Wolters, J., 'Long memory in inflation rates: International evidence', *Journal of Business and Economic Statistics*, **13** (1995), 37–46.

Haslett, J. and Raftery, A. E., 'Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource', *Applied Statistics*, **38** (1989), 1–50.

Hosking, J. R. M., 'Fractional differencing', *Biometrika*, **68** (1981), 165–76.

Hosking, J. R. M., 'Modeling persistence in hydrological time series using fractional differencing', *Water Resources Research*, **20** (1984), 1898–908.

Hurvich, C. M. and Tsai, C. L., 'Regression and time series model selection in small samples', *Biometrika*, **76** (1989), 297–307.

Lewis, P. and Orav, E., *Simulation Methodology for Statisticians, Operations Analysts, and Engineers, Vol. I*, Belmont, CA: Wadsworth & Brooks/Cole.

Mandelbrot, B. B. and Van Ness, J. W., 'Fractional Brownian motions, fractional noises and applications', *SIAM Review*, **10** (1968), 422–36.

Noakes, D. J., Hipel, K. W., McLeod, A. I., Jimenez, C., and Yakowitz, S. 'Forecasting annual geophysical time series', *International Journal of Forecasting*, **4** (1988), 103–15.

O'Connell, P. E., *Stochastic modeling of long-term persistence in streamflow sequences*, PhD thesis, Civ. Eng. Dept., Imperial College, London, 1974.

Ray, B. K., *Fractionally differenced ARMA models: Seasonality and forecasting issues*, PhD thesis, Columbia University, 1991.
Ray, B. K., 'Modeling long memory processes for optimal long-range prediction' *Journal of Time Series Analysis*, **14** (1993a), 511–25.
Ray, B. K., 'Long-range forecasting of IBM product revenues using a seasonal fractionally differenced ARMA model', *International Journal of Forecasting*, **9** (1993b), 255–69.
Ray, B. K. and Crato, N., 'Model selection and forecasting of long-range dependent processes: Results of a simulation study', Center for Applied Mathematics and Statistics, Working Paper 50, Fall 1994, New Jersey Institute of Technology.
Shea, G. S., 'Uncertainty and implied variance bounds in long-memory models of the interest rate term structure', *Empirical Economics*, **16** (1991), 287–312.
Schmidt, C. M. and Tschernig, R., 'Identification of fractional ARIMA models in the presence of long memory', Discussion Paper, University of Munich, 1993.
Sowell, F., 'Maximum likelihood estimation of stationary univariate fractionally integrated time series models', *Journal of Econometrics*, **53** (1992), 165–88.
Sutcliffe, A., 'Time series forecasting using fractional differencing', *Journal of Forecasting*, **13** (1993), 383–92.
Tiao, G. C. and Tsay, R. S., 'Some advances in non-linear and adaptive modelling in time series', *Journal of Forecasting*, **14** (1994), 109–31.

Authors' biographies:
Nuno Crato is currently Assistant Professor in the Department of Mathematical Sciences of the Stevens Institute of Technology. His research interests are in time series analysis and forecasting.
Bonnie Ray is currently Assistant Professor in the Department of Mathematics at the New Jersey Institute of Technology. Her research interests include modelling and forecasting long-range dependent time series and non-linear time series, as well as non-parametric techniques for time series.

Authors' address:
Nuno Crato, Department of Mathematical Sciences, Stevens Institute of Technology, Hoboken, NJ 07040, USA. ncrato@stevens-tech. edu.
Bonnie Ray, Department of Mathematics and Center for Applied Maths and Statistics, New Jersey Institute of Technology, NJ 07102, USA. borayx@chaos.njit.edu.