Quaderni di Statistica
Vol. 4, 2002

Model selection for long-memory models

Luisa Bisaglia
Dipartimento di Statistica, Università Ca' Foscari, Venezia
E-mail: bisaglia@unive.it

Summary: For an ARFIMA(p, d, q) model it is not possible to identify the order of the short memory polynomials by using autocorrelation and partial autocorrelation functions as for the ARMA(p, q) model. Indeed when both long and short memory components are present in the data, their behavior is hard to distinguish and the model selection becomes very difficult. In this paper, by means of a large-scale simulation study, we asses the performance of some information criteria, such as AIC, AICC, BIC and HIC, when the true model is an ARFIMA(p, d, q) and the alternatives of interest are ARFIMA(p^*, d^*, q^*). The probability of successful identification increases with the sample size and depends on the values of the short memory parameters. Moreover, it varies across the different selection criteria.

Key words: Fractional ARIMA processes, Long-range dependence, Model selection, Information criteria, Whittle estimator.

1. Introduction

Several methods to estimate the long-memory parameter d of an autoregressive fractionally integrated moving-average model, ARFIMA, are available. Many of them are described, for example, in Beran (1994). Among these methods, only the likelihood or pseudo-likelihood methods allow us to estimate at once the long-memory parameter d and the other parameters (Sowell, 1992, Fox and Taqqu, 1986). Moreover, under the hypothesis of correct model specification, these methods display a better performance than the other parametric or semiparametric procedure (Bisaglia and Guegan, 1998). Nevertheless, to obtain consistent estimators we need to know exactly the data generating process. In case of misspecification the estimators could be very biased and consequently the forecast will not be optimal in the sense of mean square prediction error (Bisaglia and Bordignon, 2002).

This naturally leads to the problem of identifying the orders of the short memory components for an ARFIMA(p, d, q) process. This task is very complicated because it is impossible to identify p and q by a simple analysis of the autocorrelation and partial autocorrelation functions as it is usually done for an ARIMA process.

Schmidt and Tcherning (1995), Crato and Ray (1996) and Smith *et al.* (1997) consider various information criteria and assess, by Monte Carlo simulations, the performance of these criteria when the true model is fractionally integrated and the alternatives of interest are both ARMA and ARFIMA models. Their results suggest that when the data generating process is an ARFIMA(p, d, q) the identification of the true model may not even be assured for small or moderate sample size, although the success probability increases substantially along with the lenght of the considered series. Beran *et al.* (1998) investigated model choice criteria for a class of models that includes classical and fractional, stationary and nonstationary autoregressive process. They show, by Monte Carlo experiments, that the convergence speed to the true order of the model seems to depend on the values of the long-memory and autoregressive parameters. All these authors find that the bayesian criteria, BIC, performs better than the others.

Most of the previous contributions focus on simple fractional noise models and/or low sample sizes. In this paper we analyse systematically, through large Monte Carlo simulations, the performance of the Akaike information criteria (AIC, AICC), the Bayesian information criteria (BIC) and the Hannan-Quinn criteria (HIC). We use these criteria together with the Whittle estimator. This method is computationally faster than other exact maximum likelihood methods and easier to implement. Moreover, it allows us to estimate the parameters all together. In this way we try to improve the understanding of the finite sample properties of the automatic criteria and to offer some guidance to select which of them is superior.

Although none of the criteria works well in all the considered cases, our study shows that the frequency of correct specification increases substantially along with the sample size. Further, the Bayesian and Hannan-Quinn criteria are to be preferred in practice because they have a more consistent behavior than the Akaike criteria.

The plan of the paper is the following. Section 2 summarizes briefly the main properties of the ARFIMA(p, d, q) processes and the Whittle estimator. Section 3 presents the identification criteria that we use. Section 4 presents the Monte Carlo study and the results. Section 5 concludes.

2. ARFIMA model and Whittle estimator

The fractionally autoregressive integrated moving-average process was independently introduced by Granger and Joyeux (1980) and Hosking (1981). This process generalizes the ARIMA(p, d, q) process by relaxing the assumption that d is an integer.

The ARFIMA(p, d, q) process $\{X_t,\ t = 0, \pm 1, \ldots\}$ is defined by the difference equation

$$\Phi(B)\,\Delta(B)\,(X_t - \mu) = \Theta(B)\,\epsilon_t,$$

where $\Phi(\cdot)$ and $\Theta(\cdot)$ are polynomials in the backward shift operator B of degrees p and q respectively and ϵ_t is a white noise process having $E[\epsilon_t^2] = \sigma^2$. Furthermore, we set $\Delta(B) = (1 - B)^d = \sum_{j=0}^{\infty} \pi_j B^j$, with $\pi_j = \Gamma(j - d)/[\Gamma(j+1)\Gamma(-d)]$, where $\Gamma(\cdot)$ denotes the gamma function. When the roots of $\Phi(B) = 0$ and $\Theta(B) = 0$ lie outside the unit circle and $\mid d \mid < 0.5$, the process is stationary, causal and invertible. We will assume these conditions to be satisfied.

When $p = q = 0$, the process $\{X_t,\ t = 0, \pm 1, \ldots\}$ is called Fractionally Integrated Noise. When $d \in (0, 1/2)$ the autocorrelation function of the process decays to zero hyperbolically at a rate $O(k^{2d-1})$, where k denotes the lag. In this case, we say that the process exhibits a long-memory behavior. When $d \in (-1/2, 0)$ the ARFIMA(p, d, q) process is said to be anti-persistent. In the following we will concentrate on ARFIMA(p, d, q) processes with $d \in (0, 1/2)$, i.e. on processes that possess long-range dependence. We will also assume, for convenience and without loss of generality, that $\mu = 0$ and $\sigma^2 = 1$.

To estimate the parameters of an ARFIMA(p, d, q) process we used the Whittle estimator, that is the frequency domain approximate maximum likelihood method proposed by Fox and Taqqu (1986). This estimator extends the results of Hannan (1970) who applied Whittle's method to the estimation of the parameters of ARMA models. Fox and Taqqu's result, later generalized by Dahlhaus (1989) to the exact maximum likelihood estimator, is the basis of one of the most used methods of estimating the long (and short, if they are present) memory parameters in Gaussian time series. Giraitis and Surgailis (1990) generalized the result of Fox and Taqqu and proved the asymptotic normality of Whittle estimator without the Gaussianity assumption.

The exact maximum likelihood estimator derived in time domain has the drawback of a large computational burden. Also computational problems might arise in the evaluation of the autocovariances necessary to the computation of the likelihood function (Sowell, 1992). These difficulties do not occur when we use the Whittle estimator, which has the further advantage of not requiring the estimation of the mean of the series (generally unknown in practice). Besides, under some regularity assumptions (Fox and Taqqu, 1986, Dahlhaus, 1989) fulfilled by ARFIMA(p, d, q) processes, it is possible to prove that the Whittle estimator has the same asymptotic distribution as the exact maximum likelihood estimator and it converges to the true values of the parameters at the usual rate of $n^{-1/2}$. Finally, for Gaussian processes the Whittle estimator is asymptotically efficient in the sense of Fisher.

If the Whittle approximation to the log-likelihood function is used, the parameter vector $\boldsymbol{\theta} = (d, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ is estimated by minimizing the estimated variance of the underlying white noise process with respect to $\boldsymbol{\theta}$:

$$\hat{\sigma}^2(\boldsymbol{\theta}) = \frac{1}{2\pi} \sum_{j=1}^{n'} \frac{I_n(\lambda_j)}{f(\lambda_j, \boldsymbol{\theta})} ,$$

where n' is the integer part of $(n-1)/2$, $I_n(\lambda_j)$ denotes the periodogram of the series, defined at the Fourier frequencies $\lambda_j = 2\pi j/n$, and $f(\lambda_j, \boldsymbol{\theta})$ represents the spectral density of the ARFIMA(p, d, q) process at the Fourier frequency λ_j.

Table 1. Estimation results of different ARFIMA(p,d,q) models, when the true model is an ARFIMA(0,d,0), $d = 0.2$. Number of replications: 1000.

Model		**(0,d,0)**	**(1,d,0)**	**(0,d,1)**	**(1,d,1)**	**(2,d,0)**	**(2,d,1)**	**(0,d,2)**	**(1,d,2)**	**(2,d,2)**
N=100	ϕ_1		0.087 (0.154)		-0.017 (0.521)	0.094 (0.170)	0.010 (0.577)		-0.007 (0.571)	0.022 (0.746)
	ϕ_2					-0.004 (0.122)	0.004 (0.154)			-0.211 (0.602)
	d	0.165 (0.089)	0.103 (0.121)	0.121 (0.122)	0.103 (0.135)	0.096 (0.143)	0.088 (0.142)	0.108 (0.134)	0.080 (0.131)	0.098 (0.129)
	θ_1			0.073 (0.152)	0.111 (0.511)		0.099 (0.568)	0.087 (0.163)	0.125 (0.567)	0.075 (0.774)
	θ_2							0.025 (0.122)	0.024 (0.158)	0.239 (0.647)
N=1000	ϕ_1		0.016 (0.052)		-0.032 (0.509)	0.021 (0.065)	0.017 (0.552)		0.017 (0.542)	-0.014 (0.742)
	ϕ_2					0.004 (0.041)	-0.003 (0.050)			-0.272 (0.626)
	d	0.195 (0.025)	0.184 (0.043)	0.187 (0.041)	0.171 (0.077)	0.179 (0.059)	0.167 (0.089)	0.184 (0.053)	0.166 (0.088)	0.167 (0.087)
	θ_1			0.013 (0.050)	0.060 (0.487)		0.016 (0.527)	0.016 (0.059)	0.017 (0.519)	0.046 (0.738)
	θ_2							0.005 (0.039)	0.000 (0.047)	0.282 (0.618)

In parentheses the mean of the estimated standard errors.

The disadvantage of this estimator is that it assumes the parametric form of the spectral density to be known *a priori*. If the spectral density function is not correctly specified (as it is often the case) the estimated parameters may be biased and consequently the forecast may not be optimal in the sense of minimizing the mean square prediction error. This naturally leads to the problem of identifying the orders of the short memory components of an ARFIMA(p, d, q) process.

2.1. Misspecification and the Whittle estimator

In finite samples the Whittle estimation of long-memory parameter d can be very misleading if the model is misspecified, even for time series of remarkable lenght. Since the parameter d describes completely the long-range behaviour of the series, it is very important to identify correctly the true order of the AR and/or MA components of the process. Bisaglia and

Table 2. Estimation results of different ARFIMA(p,d,q) models, when the true model is an ARFIMA(0.9,0.2,0.1). Number of replications: 1000.

Model		**(0,d,0)**	**(1,d,0)**	**(0,d,1)**	**(1,d,1)**	**(2,d,0)**	**(2,d,1)**	**(0,d,2)**	**(1,d,2)**	**(2,d,2)**
N=100	ϕ_1		0.818 (0.084)		0.865 (0.077)	1.066 (0.175)	0.943 (0.569)		0.872 (0.075)	0.937 (0.616)
	ϕ_2					-0.175 (0.146)	-0.067 (0.504)			-0.062 (0.549)
	d	0.499 (0.000)	0.293 (0.134)	0.416 (0.093)	0.133 (0.129)	0.078 (0.134)	0.090 (0.129)	0.497 (0.019)	0.084 (0.111)	0.062 (0.111)
	θ_1			0.800 (0.250)	0.152 (0.144)		0.113 (0.544)	0.661 (0.120)	0.192 (0.152)	0.148 (0.632)
	θ_2							0.315 (0.108)	0.049 (0.122)	0.047 (0.177)
N=1000	ϕ_1		0.836 (0.049)		0.897 (0.044)	1.061 (0.131)	1.012 (0.415)		0.905 (0.046)	0.947 (0.518)
	ϕ_2					-0.143 (0.085)	-0.099 (0.353)			-0.040 (0.461)
	d	0.499 (0.000)	0.342 (0.070)	0.427 (0.075)	0.183 (0.099)	0.132 (0.125)	0.115 (0.131)	0.499 (0.005)	0.150 (0.118)	0.121 (0.133)
	θ_1			0.565 (0.023)	0.112 (0.065)		0.066 (0.342)	0.686 (0.035)	0.139 (0.083)	0.125 (0.477)
	θ_2							0.339 (0.029)	0.022 (0.045)	0.012 (0.074)
In parentheses the mean of the estimated standard errors.										

Guegan (1998) showed that the Whittle estimator performs better than other non parametric or semiparametric estimators when the true model is known. But we do not know which is the behaviour of this estimator when the model has not been correctly specified.

In Table 1 we see that when the true model is an ARFIMA$(0, 0.2, 0)$ the problem is not so serious, because even when the process is misspecified the estimation of parameter d converges to its true value. Instead, when the true model is an ARFIMA$(1, d, 1)$ the situation changes radically depending on the values of the AR and/or MA components. The Table 2 and 3 show the estimation results when the true process is an ARFIMA$(0.9, 0.2, 0.1)$ and ARFIMA$(-0.9, 0.2, 0.5)$.[1] It is possible to see that, as the sample size increases, the parameters always converge to their true values if the true model is chosen, in particular we obtain good estimates of the long-memory parameter d. But if the model is misspeci-

[1]we have chosen $d = 0.2$, but the results are valid for each value of the parameter d.

Table 3. Estimation results of different ARFIMA(p,d,q) models, when the true model is an ARFIMA(-0.9,0.2,0.5). Number of replications: 1000.

Model		**(0,d,0)**	**(1,d,0)**	**(0,d,1)**	**(1,d,1)**	**(2,d,0)**	**(2,d,1)**	**(0,d,2)**	**(1,d,2)**	**(2,d,2)**
N=100	ϕ_1		-0.610 (0.146)		-0.828 (0.122)	-0.289 (0.129)	-0.607 (0.328)		-0.775 (0.204)	-0.767 (0.561)
	ϕ_2					0.310 (0.125)	0.161 (0.209)			-0.013 (0.482)
	d	0.004 (0.017)	0.324 (0.123)	0.197 (0.148)	0.158 (0.106)	0.047 (0.091)	0.069 (0.109)	0.031 (0.079)	0.098 (0.122)	0.080 (0.128)
	θ_1			-0.427 (0.224)	0.457 (0.174)		0.339 (0.285)	-0.265 (0.148)	0.482 (0.234)	0.499 (0.571)
	θ_2							0.322 (0.109)	0.108 (0.156)	0.138 (0.381)
N=1000	ϕ_1		-0.715 (0.037)		-0.895 (0.019)	-0.375 (0.053)	-0.862 (0.122)		-0.899 (0.023)	-0.809 (0.514)
	ϕ_2					0.334 (0.046)	0.027 (0.098)			0.076 (0.463)
	d	0.000 (0.001)	0.396 (0.034)	0.224 (0.038)	0.195 (0.031)	0.122 (0.050)	0.179 (0.050)	0.023 (0.035)	0.185 (0.045)	0.166 (0.091)
	θ_1			-0.452 (0.033)	0.499 (0.044)		0.483 (0.086)	-0.274 (0.050)	0.507 (0.055)	0.442 (0.492)
	θ_2							0.356 (0.032)	0.013 (0.051)	-0.025 (0.257)

In parentheses the mean of the estimated standard errors.

fied, the estimation of the long-memory parameter can be heavily biased. In fact, as we can see from Tables 2 and 3, the means of the estimates of d vary between 0.115 and 0.499 for the $ARFIMA(0.9, 0.2, 0.1)$ and between 0.000 and 0.396 for the $ARFIMA(-0.9, 0.2, 0.5)$. [2] Thus, the correct identification of the data generating process is essential to preserve the superiority of the Whittle estimator.

3. Information criteria

All heuristic methods proposed in the literature to determine the order of an ARMA process, such as those based on autocorrelations and partial autocorrelations, on the R- and S-arrays, on the corner method (see de Gooijer *et al.*, 1985, for a review), cannot be used to determine the

[2] We have constrained the long-memory parameter d to lie in the interval (0,0.5).

order of an ARFIMA process. In fact when $d > 0$, the autocorrelations and the partial autocorrelations decay to zero at a slower rate of convergence than the ARMA models and this makes impossible to recognize the short-memory components. But the knowledge of the generating process is important in order to estimate the model with likelihood or pseudo-likelihood methods and make predictions.

In this section we compare the performances of different automatic selection criteria in the presence of long-memory. We consider only ARFIMA models since we can perform the identification in two steps. Firstly, we estimate the long-memory parameter d through one of the non parametric or semiparametric methods. Then, if the series presents long-memory, we focus our attention only on the order of the short-memory parameters of the fractional process.

We consider the information criteria that are usually used in the identification of an ARMA process: the Akaike, the modified Akaike, the Bayesian information and the Hannan-Quinn Criterion. These criteria are expressed in terms of the approximate maximum value of the log-likelihood function of a Gaussian ARMA(p, q) process plus a penalty for the number of parameters used. This penalty counteracts the over-fitting tendency of these criteria: this is what makes the criteria different each other. In fact it is well known in the literature (see for example de Gooijer *et al.,* 1985) that the Akaike and Akaike modified Information Criteria tend to overestimate the true order p and q with positive probability, while the Schwartz and Hannan-Quinn Criteria provide consistent estimates of p and q. The best model is that with the smallest value of a specific information criterion. Since to estimate the parameters of the ARFIMA(p, d, q) model we use the Whittle estimator, we express the selection criteria in terms of the implied white noise variance $\hat{\sigma}^2_N$. These criteria are:

1. Akaike Information Criteria: AIC

$$AIC = N \ln(\hat{\sigma}^2_N) + 2(p + q + 1)$$

2. modified Akaike Information Criteria: AICC

$$AICC = N \ln(\hat{\sigma}^2_N) + \frac{2N(p + q + 1)}{(N - p - q - 2)}$$

3. Bayesian Information Criteria: BIC

$$BIC = N\ln(\hat{\sigma}_N^2) + (p+q+1)\log N$$

4. Hannan-Quinn: HIC

$$HIC = N\ln(\hat{\sigma}_N^2) + 2(p+q+1)c\log\log(N)$$

where $c = 1.0001$.

The use of these automatic selection criteria has not yet been supported by an extensive analysis of their applicability to time series generated by ARFIMA process. We want to investigate here how succesfully these information criteria will identify the true $ARFIMA(p,d,q)$ generating process.

4. Monte Carlo experiments

We have generated 1000 independent time series driven by Gaussian ARFIMA$(1,d,1)$ model:

$$(1-\phi B)(1-B)^d X_t = (1+\theta B)\epsilon_t.$$

Since we are interested only in long-memory models, we take $d = (0.1, 0.2, 0.3, 0.4)$, and we set the AR and MA parameters to the values: -0.9, $-0.7, -0.5, -0.3, -0.1, 0.0, 0.1, 0.3, 0.5, 0.7, 0.9$, so that lower order models are considered as special cases. The sample sizes considered are $N = 100, 250, 500, 1000$. The processes are generated by using the recursive Durbin-Levinson algorithm (Brockwell and Davis, 1991). To estimate the parameters we constrained the AR and MA parameters to lie in the stationary and invertibility region, and we restrict $d \in (0, 1/2)$. The four criteria are used to select one of the nine models $ARFIMA(p,d,q)$ up to $p,q = 2$. We studied the number of correct choices obtained by varying the parameters d, ϕ and θ and assuming $N = 100$. For the values of AR and MA parameters $\phi = (-0.9, -0.5, -0.1, 0.0, 0.1, 0.5, 0.9)$

and $\theta = (-0.9, -0.5, -0.1, 0.0, 0.1, 0.5, 0.9)$ we have studied how the selection performance of the criteria varies with the sample size.[3]

First of all, the performance of the four automatic information criteria is almost the same when d varies between 0.1 and 0.4. The performance of AIC and AICC criteria is almost everywhere worse than BIC and HIC, even if the sample size increases. Moreover, AIC and AICC criteria converge more slowly than the BIC and HIC criteria. Usually the use of BIC criteria consistently leads to the highest success percentage in detecting the true fractional model. Anyway, as we can see in Tables 4, 5, 6 and 7, the greater difference in the frequencies of success depends on the values of AR and/or MA coefficients of the data generating process. The ARFIMA$(0, d, 0)$, ARFIMA$(1, d, 0)$ and ARFIMA$(0, d, 1)$ processes are the most easily recognized, even if the sample size of the generating process is very small (N=100). Only the AIC and AICC criteria work badly even if $N = 1000$, this confirms the fact that AIC and AICC criteria are not consistent. Usually the BIC criterion works better. The only exceptions are when the AR or MA coefficient of the data generating process is very small (± 0.1). In these cases the criterion that works better is the HIC criterion, but even when $N = 1000$ these models are very difficult to be detected. We suppose that we need very long time series to identify the order of the true generating process in these cases. Moreover, in the Tables it is possible to see that there is a sort of asymmetry between ARFIMA$(0.1, d, 0)$ or ARFIMA$(0, d, 0.1)$ and ARFIMA$(-0.1, d, 0)$ or ARFIMA$(0, d, -0.1)$ models. In fact, when the AR or MA parameter is egual 0.1 we obtain a greater number of correct choices than when the AR or MA parameter is egual -0.1, specially for the MA parameter. When the true data generating process is an ARFIMA$(1, d, 1)$, the task becomes more complicated. Also in this case, the criteria that work better are BIC and HIC, so in the following the behaviour of AIC and AICC will not be commented upon.[4] Even if, the percentage of successes increases along with N, there are models that are very difficult to be detected. Gener-

---

[3]We report only the results for $d = 0.3$. The results obtained for the other cases are analogous and are available from the author upon request.

[4]In the Tables we do not consider the case $\mid \phi \mid = - \mid \theta \mid$ because, obviously, this case corresponds to the ARFIMA$(0, d, 0)$ model.

ally, if both the values of AR and/or MA parameters are different from ± 0.1, it is not difficult to identify the real order of the data generating process, specially with the BIC criterion. For example, if we consider the ARFIMA$(-0.5, d, 0.9)$ model, we have that for $N = 1000$ the percentage of success with BIC criteria is 97.9% when $d = 0.1,\ 98.9\%$ when $d = 0.2,\ 98.9\%$ if $d = 0.3$ and 97.8% if $d = 0.4$. But if we consider the ARFIMA$(-0.1, d, -0.1)$ and ARFIMA$(0.1, d, 0.1)$ models, even if $N = 1000$, the criteria tend to select other models. In general, all models where the AR and/or the MA parameter is equal to ± 0.1 are difficult to identify. These are the only cases where AIC and AICC criteria work better than BIC and HIC. Thus, it could be interesting to find out which are most chosen models when the true model is an $ARFIMA(-0.1, d, -0.1)$ or an $ARFIMA(0.1, d, 0.1)$. In fact, as we can see in Tables $4 - 7$, differently from the other cases the selection frequencies of the true model decreases as N increases. When the true model is the $ARFIMA(-0.1, d, -0.1)$ and the sample size is very small ($N = 100$) BIC and HIC criteria choose more often the $ARFIMA(0, d, 0)$ specification. On the other hand, with a larger sample size the most chosen models are the $ARFIMA(1, d, 0)$ and $ARFIMA(0, d, 1)$ whatever criterion being used. The same happens when the true model is the $ARFIMA(0.1, d, 0.1)$. This means that, in these situations, the probability of choosing the correct model is mistaken. Note that all criteria underestimate the order of the true model.

5. Conclusions

In conclusion, we think that, if we wish to identify correctly the true ARFIMA generating process, the BIC and HIC criteria are the ones to be used. These criteria have in fact a more consistent behaviour than AIC and AICC criteria, specially if the true process has only one AR or MA parameter. However, when the true data generating process is an ARFIMA$(1, d, 1)$ none of the used criteria perform well in all the cases, except when the short-memory components are strong. Finally, the different values of the long-memory parameter d, do not have effect on the choice of the model especially when sample size is large.

Table 4. Selection frequencies of the correct specification when the DGP is a FARIMA(1,0.3,1), for N=100. Number of replications: 1000.

Criteria	$\frac{AR}{MA}$	**-0.9**	**-0.7**	**-0.5**	**-0.3**	**-0.1**	**0.0**	**0.1**	**0.3**	**0.5**	**0.7**	**0.9**
AIC	**-0.9**	68.3	70.9	62.9	34.7	7.6	60.9	12.4	32.6	43.2	35.1	-
AICC		71.7	73.3	65.2	35.5	7.8	64.0	12.2	33.2	44.4	36.1	-
BIC		91.0	91.8	74.6	23.8	2.2	92.3	5.2	25.3	31.2	8.6	-
HIC		82.9	82.5	71.8	32.7	5.3	79.1	9.5	32.4	44.3	27.3	-
AIC	**-0.7**	56.2	52.2	39.2	15.1	3.1	42.6	8.8	20.7	16.9	-	19.5
AICC		58.9	54.4	39.8	14.6	2.7	46.0	9.2	21.0	16.7	-	18.7
BIC		68.8	64.2	40.0	7.4	0.5	77.7	5.0	7.1	2.6	-	5.4
HIC		66.5	61.5	41.9	12.6	2.0	61.6	8.5	16.1	9.7	-	13.7
AIC	**-0.5**	31.1	26.9	15.0	3.1	2.6	21.7	8.7	7.5	-	6.2	21.3
AICC		31.2	27.4	14.6	2.5	2.9	23.1	9.0	7.3	-	5.7	21.5
BIC		24.3	18.8	5.7	0.4	1.4	30.8	1.9	1.4	-	1.1	13.5
HIC		30.5	24.5	10.8	0.9	2.8	28.8	6.3	3.8	-	2.2	18.9
AIC	**-0.3**	10.1	6.5	3.8	2.2	5.5	6.0	5.7	-	5.5	7.6	5.3
AICC		9.6	6.5	3.1	1.9	5.5	6.2	5.7	-	5.0	7.1	5.3
BIC		4.1	1.4	0.5	0.5	1.0	4.2	0.9	-	1.4	1.8	2.1
HIC		7.6	4.6	1.3	0.8	3.1	7.3	3.0	-	2.9	5.1	3.1
AIC	**-0.1**	4.6	4.8	3.3	3.5	4.2	3.2	-	6.0	4.3	4.7	2.0
AICC		4.1	4.4	3.4	3.6	4.3	3.4	-	6.0	4.0	4.5	2.0
BIC		0.8	0.9	0.9	0.5	0.7	1.3	-	2.5	0.9	1.0	0.6
HIC		2.4	2.5	2.0	2.3	2.2	2.6	-	4.3	2.3	2.6	1.5
AIC	**0.0**	61.1	56.7	44.7	18.4	5.9	39.4	17.7	33.5	41.9	42.7	59.2
AICC		63.5	59.1	47.3	19.7	6.2	43.6	18.7	35.2	44.1	45.2	62.4
BIC		91.0	87.4	71.8	20.7	2.0	88.8	9.2	38.6	67.5	80.5	92.2
HIC		76.7	73.7	60.1	22.2	3.8	68.0	15.4	39.6	57.1	62.7	78.4
AIC	**0.1**	7.7	8.4	4.7	4.7	-	10.1	7.7	6.0	10.4	14.4	8.1
AICC		7.4	8.3	4.8	4.3	-	10.6	7.8	5.3	9.7	13.8	8.0
BIC		3.9	4.3	1.6	1.1	-	10.2	3.5	2.5	2.5	8.2	4.4
HIC		5.8	6.8	3.5	2.4	-	11.5	6.4	4.3	7.4	12.3	6.4
AIC	**0.3**	21.3	13.9	6.3	-	13.3	28.1	8.5	16.5	28.5	30.3	20.8
AICC		21.5	13.6	5.6	-	12.9	29.5	8.0	15.9	29.6	31.4	21.4
BIC		17.0	7.2	1.2	-	6.6	38.1	4.2	4.6	18.8	29.4	16.2
HIC		20.4	11.2	3.2	-	10.4	35.6	5.8	11.5	28.1	33.7	19.1
AIC	**0.5**	32.6	10.2	-	20.6	12.0	46.0	14.4	28.9	43.2	44.4	32.7
AICC		33.2	9.9	-	20.4	12.0	48.6	14.9	29.0	44.8	46.6	34.0
BIC		35.0	2.5	-	11.8	8.1	74.8	4.8	17.1	45.8	57.4	35.6
HIC		37.1	6.7	-	17.7	10.9	61.6	9.8	27.2	50.2	53.7	36.0
AIC	**0.7**	24.0	-	34.2	24.9	12.9	50.2	19.0	34.7	47.8	49.1	38.8
AICC		24.4	-	35.4	25.4	12.5	52.9	19.1	34.7	49.9	52.2	40.5
BIC		22.3	-	22.2	23.5	9.4	88.5	8.1	27.9	60.5	72.5	46.0
HIC		25.7	-	33.4	25.2	12.0	71.2	14.6	34.8	58.2	64.1	45.5
AIC	**0.9**	-	60.1	52.9	34.5	14.7	61.4	21.6	39.6	55.6	51.7	42.6
AICC		-	61.6	54.7	36.2	14.3	64.5	20.6	40.3	58.2	54.5	43.5
BIC		-	53.3	61.6	35.4	6.2	90.4	8.1	32.5	67.0	77.5	53.0
HIC		-	65.3	61.9	38.9	11.6	80.5	15.1	40.0	66.0	67.4	50.2

Table 5. Selection frequencies of the correct specification when the DGP is a FARIMA(1,0.3,1), for N=250. Number of replications: 1000.

Criteria	$\frac{AR}{MA}$	**-0.9**	**-0.5**	**-0.1**	**0.0**	**0.1**	**0.5**	**0.9**
AIC	**-0.9**	70.7	71.3	66.3	65.4	12.2	55.1	-
AICC		72.6	72.1	67.2	66.5	11.9	56.0	-
BIC		97.2	94.3	95.5	97.1	1.6	45.4	-
HIC		88.8	87.2	84.7	87.6	5.1	62.2	-
AIC	**-0.5**	57.1	42.3	3.2	44.5	3.6	-	45.0
AICC		57.5	43.0	3.1	45.2	3.5	-	45.2
BIC		59.1	32.5	0.0	63.2	0.3	-	55.9
HIC		62.2	41.2	0.8	58.7	1.3	-	53.7
AIC	**-0.1**	8.7	7.5	3.6	4.4	-	5.4	2.7
AICC		8.6	7.2	3.7	4.5	-	5.3	2.8
BIC		3.0	0.8	0.3	1.2	-	1.4	0.0
HIC		5.7	3.5	1.5	3.1	-	3.7	1.1
AIC	**0.0**	60.9	54.4	8.0	39.5	15.7	56.5	66.5
AICC		62.6	55.6	8.1	41.4	15.8	56.9	67.5
BIC		95.1	90.8	3.1	94.4	11.1	84.7	97.7
HIC		84.1	75.5	6.7	78.6	16.4	74.1	86.4
AIC	**0.1**	10.3	9.9	-	11.8	6.9	17.7	13.1
AICC		10.2	9.9	-	12.1	6.2	17.5	13.1
BIC		3.7	2.7	-	7.2	1.7	5.3	4.8
HIC		7.3	5.9	-	12.2	3.6	11.1	10.0
AIC	**0.5**	61.1	-	9.8	58.2	15.8	62.0	57.5
AICC		62.2	-	9.8	59.1	16.1	62.6	58.1
BIC		77.4	-	2.1	92.0	6.4	73.0	69.8
HIC		73.6	-	5.2	79.1	13.1	71.5	67.3
AIC	**0.9**	-	60.4	10.7	57.8	19.8	66.9	61.1
AICC		-	61.6	10.7	59.2	20.1	68.0	62.2
BIC		-	84.5	5.8	94.8	9.6	86.9	83.2
HIC		-	77.4	7.9	82.6	15.8	80.6	78.1

Table 6. Selection frequencies of the correct specification when the DGP is a FARIMA(1,0.3,1), for N=500. Number of replications: 1000.

Criteria	$\frac{AR}{MA}$	**-0.9**	**-0.5**	**-0.1**	**0.0**	**0.1**	**0.5**	**0.9**
AIC	**-0.9**	65.3	70.0	10.9	60.3	28.7	60.6	-
AICC		65.8	70.4	10.8	60.9	28.7	61.4	-
BIC		97.8	98.5	2.1	97.2	6.8	81.4	-
HIC		86.6	90.9	7.2	87.0	17.9	80.3	-
AIC	**-0.5**	70.7	63.1	11.4	58.3	6.2	-	57.5
AICC		71.0	63.3	11.2	58.6	6.1	-	57.9
BIC		82.4	70.6	1.8	87.0	0.4	-	77.2
HIC		79.3	71.3	5.4	78.3	1.6	-	70.9
AIC	**-0.1**	12.3	8.2	4.3	9.5	-	10.0	3.0
AICC		12.2	8.0	4.3	9.4	-	10.1	3.0
BIC		3.4	1.4	0.2	3.6	-	1.7	0.0
HIC		6.8	4.5	1.8	7.5	-	5.8	0.4
AIC	**0.0**	66.2	63.7	11.7	45.4	19.4	65.9	69.4
AICC		66.4	63.9	11.8	46.0	19.7	66.7	69.8
BIC		97.5	96.1	4.7	97.2	11.5	92.2	98.9
HIC		88.1	85.3	10.7	83.2	21.7	84.4	91.0
AIC	**0.1**	18.3	8.5	-	15.5	5.3	23.7	15.8
AICC		18.5	8.6	-	15.7	5.3	23.7	15.7
BIC		6.9	1.7	-	11.0	0.7	7.3	5.5
HIC		13.0	5.1	-	18.3	3.3	17.1	11.6
AIC	**0.5**	71.3	-	9.3	64.0	17.2	69.1	68.8
AICC		71.6	-	9.4	65.0	17.2	69.7	68.9
BIC		92.8	-	1.6	97.3	6.1	87.4	89.6
HIC		87.2	-	5.4	86.5	12.4	83.1	82.8
AIC	**0.9**	-	69.0	14.8	63.7	21.6	75.8	71.7
AICC		-	69.7	14.8	64.2	21.7	76.0	72.6
BIC		-	96.0	5.6	98.0	9.4	92.2	96.0
HIC		-	88.1	11.4	88.1	17.5	87.5	89.6

Table 7. Selection frequencies of the correct specification when the DGP is a FARIMA(1,0.3,1), for N=1000. Number of replications: 1000.

Criteria	$\frac{AR}{MA}$	**-0.9**	**-0.5**	**-0.1**	**0.0**	**0.1**	**0.5**	**0.9**
AIC	**-0.9**	61.1	70.5	14.3	57.9	33.9	64.9	-
AICC		61.3	71.3	14.4	58.7	34.0	65.2	-
BIC		97.8	99.5	3.6	97.7	9.5	95.5	-
HIC		85.6	90.9	9.7	84.5	26.1	86.9	-
AIC	**-0.5**	72.9	71.0	20.4	62.7	12.1	-	65.6
AICC		73.2	71.1	20.6	63.1	12.0	-	65.7
BIC		94.5	91.8	6.8	94.3	2.0	-	86.1
HIC		89.1	87.8	15.1	86.1	6.3	-	79.8
AIC	**-0.1**	20.1	14.1	4.1	14.9	-	13.1	6.8
AICC		20.2	14.1	1.2	15.0	-	13.0	6.8
BIC		7.4	2.5	0.1	7.5	-	3.2	0.6
HIC		15.5	6.9	1.6	16.5	-	9.7	2.6
AIC	**0.0**	65.2	68.3	21.3	45.1	25.8	68.9	64.1
AICC		65.5	68.5	21.3	45.4	26.1	69.3	64.4
BIC		98.4	98.3	14.0	98.3	18.5	96.1	98.5
HIC		90.1	90.7	22.6	86.5	28.4	89.0	89.2
AIC	**0.1**	26.5	13.1	-	23.6	3.3	30.5	19.2
AICC		26.5	13.1	-	23.8	3.3	30.6	19.1
BIC		12.4	1.7	-	17.5	0.1	15.3	7.7
HIC		22.1	6.6	-	27.9	1.2	29.0	15.4
AIC	**0.5**	69.9	-	14.0	63.8	19.1	77.1	71.5
AICC		69.9	-	14.0	63.9	19.1	77.8	71.8
BIC		97.4	-	1.8	97.8	5.3	97.1	98.2
HIC		88.6	-	7.4	87.6	13.9	92.8	89.7
AIC	**0.9**	-	76.5	25.7	64.3	29.8	79.0	74.0
AICC		-	76.7	25.7	64.5	29.8	79.1	74.3
BIC		-	98.9	11.1	98.5	15.0	98.6	98.9
HIC		-	92.9	22.2	89.2	26.7	93.9	91.6

References

Beran J. (1994) *Statistics for Long-Memory Processes*, Chapman and Hall, London.

Beran J., Bhansali R.J. and Ocker D. (1998) On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes, *Biometrika*, 85, 921-934.

Bisaglia L. and Bordignon S. (2002) Mean square prediction error for long-memory processes, *Statistical Papers*, 43, 161-175.

Bisaglia L. and Guegan D. (1998) A comparison of techniques of etimation in long-memory processes, *Computational Statistics & Data Analysis*, 27, 61-81.

Brockwell P. J. and R. A. Davis (1991) *Time series: Theory and Methods*, Springer, New York.

Crato N. and Ray K.R. (1996) Model Selection and Forecasting for Long-Range Dependent Processes, *Journal of Forecasting*, 15, 107-125.

Dahlhaus R. (1989) Efficient parameter estimation for self-similar processes, *Annals of Statistics*, 17, 1749-1766.

Fox R. and Taqqu M.S. (1986) Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series, *The Annals of Statistics*, 14, 517-532.

Giraitis L. and Surgailis D. (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its applications to asymptotical normality of Whittle's estimate, *Probability Theory and Related Fields*, 86, 87-104.

de Gooijer J., Abraham B., Gould A. and Robinson L. (1985) Methods for Determining the Order of an Autoregressive-Moving Average Process: A Survey. *International Statistical Review*, 53, 301-329.

Granger C.W.J. and Joyeux R. (1980) An introduction to long-range time series models and fractional differencing, *Journal of Time Series Analysis*, 1, 15-30.

Hannan E. (1970) *Multiple time series*, Wiley, New York.

Hosking J.R.M. (1981) Fractional Differencing, *Biometrika*, 68, 165-176.

Schmidt C.M. and Tcherning R. (1995) The Identification of Fractional ARIMA Models, Discussion Paper of Humboldt-Universitat zu Berlin.

Smith J., Taylor N. and Yadav S. (1997) Comparing the bias and misspecification in ARFIMA models, *Journal of Time Series Analysis*, 18, 507-527.

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

Whittle P. (1951) *Hypothesis Testing in time Series Analysis*, Hafner, New York.