Computational Statistics and Data Analysis 76 (2014) 116–131

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Modified information criteria and selection of long memory time series models

Richard T. Baillie [a,b,c], George Kapetanios [c], Fotis Papailias [d,*]

[a] Department of Economics, Michigan State University, USA

[b] Department of Finance, Michigan State University, USA

[c] School of Economics and Finance, Queen Mary University of London, UK

[d] Queen's University Management School, Queen's University Belfast, UK

ARTICLE INFO

Article history:
Received 2 June 2012
Received in revised form 23 April 2013
Accepted 23 April 2013
Available online 3 May 2013

Keywords:
Long memory
ARFIMA models
Modified information criteria

ABSTRACT

The problem of model selection of a univariate long memory time series is investigated once a semi parametric estimator for the long memory parameter has been used. Standard information criteria are not consistent in this case. A Modified Information Criterion (MIC) that overcomes these difficulties is introduced and proofs that show its asymptotic validity are provided. The results are general and cover a wide range of short memory processes. Simulation evidence compares the new and existing methodologies and empirical applications in monthly inflation and daily realized volatility are presented.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

There is now a substantial literature on the successful implementation of long memory processes for describing economic, financial and physical science time series. Such series generally require fractional differencing to remove their long memory component and to reduce them to a short memory series. A substantial number of articles have been concerned with general semi parametric estimation in the frequency domain to specifically estimate the long memory parameter. This work originally began with Geweke and Porter-Hudak (1983) and has now grown to include nonstationary long memory models; e.g. Phillips (2007). An early comparison of such estimation methods can be found in Bisaglia and Guegan (1998) and more recently in Haldrup and Nielsen (2007). While there have been great advances in the specific area of semi parametric estimation of the long memory parameter, there seems to be a correspondingly small literature on model selection of the short memory component of the process. This seems to be an important issue given that a typical investigator will often want to analyze both long and short memory components of a time series. For example, Baillie and Kapetanios (2008) discuss estimation of ARMA and nonlinear models for the short run component and Baillie and Kapetanios (forthcoming) consider inference for impulse response functions from long memory models with significant short run components. However, neither of these papers considers model selection issues resulting from the use of a semi parametric estimator of the long memory parameter.

This paper sets out to address this gap in the literature and specifically develops a procedure where the initial semi parametric estimator of the long memory parameter is used to obtain a modified information criterion for the purpose of model selection of the short memory parameters. The semi parametric estimators of the long memory parameter converge

* Correspondence to: Queen's University Management School, Queen's University Belfast, Riddel Hall, 185 Stranmillis Road, BT9 5EE, Northern Ireland, UK. Tel.: +44 2890974667; fax: +44 2890974201.

E-mail addresses: f.papailias@qub.ac.uk, f.papailias@quantf.com (F. Papailias).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.04.012

at a non standard rate to their limiting distribution. This feature affects the theoretical properties of the modified information criterion. The results in this paper are general and cover any short memory model as defined by Sin and White (1996), as long as it allows a semi parametric estimation of the long memory parameter. While this covers an extremely wide range of situations, the basic workhorse ARFIMA(p, d, q) model is the leading example. Hence, this model is used for assessing the relative performance of the modified information criterion, or MIC, in simulation evidence presented in this paper. Furthermore, using a consistent semi parametric estimator for nonlinear long memory models (e.g. see Dalla et al., 2005), we can extend the results of this paper to specific nonlinear models which are similar in coverage to those covered by Sin and White (1996).

On face value, it would appear that an alternative model selection strategy could be based on simultaneous estimation of all the parameters of the original series, including the long memory parameter. However, this can be problematic in the situation of nonstationary long memory series, whereas the methods developed in this paper can deal with quite easily. A further advantage with our selection approach is that it bypasses complexities arising from the existence of long memory and reduces the analysis to working on a weakly dependent process. For parsimonious reasons, the form of the MIC used in this paper is relatively simple, although the potential use of other MICs would be consistent with the same conclusions.

A well-known problem with estimators in which the long memory parameter is estimated prior to treatment of short memory components is that short and long memory components are hard to distinguish in small samples because either will project onto a model with the other element. It is important that empirical researchers are fully aware of this potential identification issue when using estimators of the long memory parameter and opt for estimators that approach this issue such as Andrews and Sun (2004).

The plan of the paper is as follows: Section 2 provides a review of some of the relevant literature on population quantities and also the various semi parametric estimation methods for long memory processes used in this paper. Section 3 presents the derivation of the MIC methodology and the next section briefly describes the implementation in the context of the linear ARFIMA model. Section 4 then describes the results of a detailed simulation study where the relative success of the standard criteria, Akaike (1974), Schwarz (1978) and Hannan and Quinn (1979), is compared with our new MIC for *each* semi parametric estimator. Overall there is strong evidence of the relative superiority of the MIC. The next section compares the success of each method when applied to the monthly CPI inflation series of Japan and the UK and the daily realized volatility series of DM–Yen. The paper ends with a short conclusions section, with the most important finding being that the MIC is easily implemented and has definite advantages for model selection.

2. Theoretical issues

A long memory, fractionally integrated process has slow hyperbolic rates of decay associated with its impulse response weights and autocorrelations. Following Granger and Joyeux (1980), Granger (1980) and Hosking (1981), a univariate time series process with fractional integration in its conditional mean is represented by,

$$(1-L)^d y_t = u_t, \quad t = 1, \ldots, T, \tag{1}$$

where L is the lag operator and u_t is a short memory, $I(0)$ process. Then, y_t is said to be a fractionally integrated process of order d, or $I(d)$. An $I(0)$ process is defined here as having partial sums that converge weakly to Brownian motion. The parameter d represents the degree of "long memory", or persistence in the series. For $-1/2 < d < 1/2$ the process is stationary and invertible; while for $1/2 \leq d \leq 1$, the process does not have a finite variance, but still has a finite cumulative impulse response function. The Wold decomposition, or infinite order moving average representation, of this process is given by,

$$y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \tag{2}$$

where ϵ_t is a martingale difference sequence such that $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = \sigma^2$. Equivalently, the infinite AR representation is given by,

$$y_t = \sum_{i=1}^{\infty} \pi_i y_{t-i} + \epsilon_t. \tag{3}$$

For large lag i, the MA coefficients decay at very slow hyperbolic rates giving $\psi_i \sim c_1 i^{d-1}$ and similarly the infinite autoregressive representation coefficients decay at the rate of $c_2 i^{-d-1}$. Finally, the autocorrelation coefficients decay at a rate of $c_3 i^{2d-1}$, where c_1, c_2 and c_3 are constants. If the short memory component is represented as an ARMA(p, q) process, then Eq. (1) becomes the well known ARFIMA(p, d, q) model,

$$\phi(L)(1-L)^d y_t = \theta(L)\epsilon_t, \tag{4}$$

where $\phi(L)$ and $\theta(L)$ are polynomials in the lag operator of orders p and q respectively, with all their roots lying outside the unit circle. If an investigator has knowledge of the process, then MLE may be used and its properties have been derived by Fox and Taqqu (1986), Sowell (1992), Hosoya (1997) and Doornik and Ooms (2003).

Following Geweke and Porter-Hudak (1983) there have been many semi parametric estimation procedures in the frequency domain proposed for the estimation of the long memory parameter. This paper considers four of the most widely used estimators and each estimator solves a minimization problem of the form $\widehat{d} = \arg_{d\in[d_1,d_2]} \min R(d)$, where d_1, d_2 are the lower and upper bounds of the values for d such that $-\infty < d_1 < d_2 < \infty$ and $R(d)$ is the relevant objective function.

The first estimator is the Local Whittle (LW) estimator, which is obtained by minimizing the objective function,

$$R^{\mathrm{LW}}(d) = \ln\left[\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d} I(\omega_j)\right] - \frac{2d}{m}\sum_{j=1}^{m}\ln(\omega_j), \tag{5}$$

with respect to d, where $\omega_j = (2\pi j)/T$ for $j = 1, 2, \ldots, T$ and $I(\omega_j)$ is the periodogram defined as,

$$I(\omega_j) = \frac{1}{2\pi T}\left|\sum_{j=1}^{T} y_t e^{i\omega_j t}\right|^2. \tag{6}$$

The estimator depends on the choice of bandwidth, m, which is commonly chosen as $m = \lfloor T^{\delta} \rfloor$ where $0 < \delta < 4/5$; $\lfloor\cdot\rfloor$ denotes the integer part. The limiting distribution of the LW estimator has been derived under various assumptions concerning the short memory process by Robinson (1995), Dalla et al. (2005) and Shimotsu and Phillips (2006).

Several important extensions of the LW estimator have been introduced in the literature. In particular, Andrews and Sun (2004) have proposed the Local Polynomial Whittle (LPW) estimator which approximates the short run component of the spectrum of the series. They approximate its logarithm by a polynomial. The relevant objective function in this case is,

$$R^{\mathrm{LPW}}(d) = \ln\left[\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d} I(\omega_j) e^{pr(\omega_j,\theta)}\right] - \frac{2d}{m}\sum_{j=1}^{m}\ln(\omega_j) - \frac{1}{m}\sum_{j=1}^{m} pr\left(\omega_j,\theta\right), \tag{7}$$

where the polynomial is given by $pr\left(\lambda_j,\theta\right) = \sum_{k=1}^{r}\theta_k\omega_j^{2k}$ and $\theta = (\theta_1,\ldots,\theta_r)'$. One issue concerns the choice of the order, r, of the polynomial, and Andrews and Sun (2004) have suggested an automated adaptive method. In the simulation experiments reported in this study we set $r = 1$ for reasons of simplicity.

Shimotsu and Phillips (2005) have proposed the Exact Local Whittle (ELW) approach using a "corrected" discrete Fourier transform of the series, where the objective function now becomes,

$$R^{\mathrm{ELW}}(d) = \ln\left[\frac{1}{m}\sum_{j=1}^{m} I_{\nabla^d y}(\omega_j)\right] - \frac{2d}{m}\sum_{j=1}^{m}\ln(\omega_j), \tag{8}$$

where $\nabla^d = (1-L)^d$.

Abadir et al. (2007) have introduced the Fully Extended Local Whittle (FELW) where $d \in (p-1/2, p+1/2]$, for $p = 0, 1, 2, \ldots,$ which has the particular attraction of covering the region of nonstationarity for long memory processes. The following simpler version of the estimator can be found in Abadir et al. (2011) and its performance is identical to the original version. The periodogram is then defined as,

$$I^{\mathrm{FELW}}(\omega_j) = \left|1-e^{i\omega_j}\right|^{-2p} I_{\nabla^p y}(\omega_j).$$

Then, the FELW is obtained by minimizing,

$$R^{\mathrm{FELW}}(d) = \ln\left[\frac{1}{m}\sum_{j=1}^{m} j^{2d} I^{\mathrm{FELW}}(\omega_j)\right] - \frac{2d}{m}\sum_{j=1}^{m}\ln(j). \tag{9}$$

The LW and LPW are only known to be consistent estimators of d in the stationary region of $-1/2 < d < 1/2$, while the ELW and FELW estimators are known to be consistent for all values of d.

It should be noted for completeness that there is a corresponding literature on semi parametric estimation in the time domain using minimum distance estimation. In particular, see Tieslau et al. (1996), Mayoral (2007) and Zevallos and Palma (2013). However, these methods have not appeared in many practical applications and hence are not considered in this study.

3. Model selection

Many previous articles have considered the issue of model selection using information criteria. The seminal work of Kullback and Leibler (1951) was founded on information theory, while Shibata (1980) used the concept of model approximations and the work of Schwarz (1978) was motivated by Bayesian considerations. However, motivating the theoretical properties of information criteria along these lines is not straightforward for long memory processes.

Much of the previous literature on the properties of information criteria was constrained to the class of weakly dependent processes and did not consider long memory processes. For example Shibata (1976) and Hannan and Quinn (1979) have studied the AR(p) model and Hannan (1980) has investigated the stationary ARMA(p, q) model. Further, Sin and White (1996) and Kapetanios (2001) have extended the scope of previous results and shown that the consistency properties extend to nonlinear models for weakly dependent processes. Hidalgo (2002) has shown that similar results are valid for regressions involving long memory regressors. Beran et al. (1998) have established some model selection criteria for stationary long memory with an autoregression and Crato and Ray (1996) present some simulation experimental evidence on model selection for long memory models. It is also worth noting that Shibata (1980) has extended some of the results on AR(p) models to cases where the lag order p may be growing with the sample size, or at least is not fixed a priori; and recently, Poskitt (2007) has extended these optimality results to stationary long memory processes.

The context considered in this paper is to establish weak consistency of information criteria in the sense of Sin and White (1996) for model selection applied to short memory component u_t of the observed process y_t. The distinctive feature of the problem is that the semi parametric estimate of the long memory parameter will be $m^{1/2}$ consistent, where $0 < m < T^{4/5}$, while the remaining parameters associated with the short memory process will be $T^{1/2}$ consistent when d is known.

The general aim is to specify a model for the short memory component, $u_t = u_{d,t}$. It is assumed that there exist two competing models for $u_{d,t}$, which are defined by their objective functions $Q_T(u_d, \theta)$ that are being minimized. Possible criteria might be obtained from MLE of a Gaussian density, or from a minimization of the conditional sum of squares. However, in general and without loss of generality,

$$Q_{iT}(u_d, \theta_i) = Q_{iT,d}(\theta_i) = \sum_{t=1}^{T} q_{it}(u_d, \theta_i) = \sum_{t=1}^{T} q_{it,d}(\theta_i), \tag{10}$$

for $i = 1, 2$, where $\theta_i \subset \Re^{p_i}$, $p_1 > p_2$ and $u_d = \left(u_{d,1}, \dots, u_{d,T}\right)'$. The only observed process is y_t and hence $u_{d,t}$ is unobservable. However, given a semi parametric estimator of d, denoted by $\widehat{d}$, the y_t series can be Feasibly Fractionally Filtered (FFF) to obtain $u_{\hat{d},t}$, which is the corresponding estimate of the short memory component. Some of the properties of the FFF transformation are discussed in Wright (1995), Baillie and Kapetanios (2007) and Baillie and Kapetanios (2008).

The information criterion defined by,

$$\mathrm{IC}\left(\hat{d}\right) = \sum_{t=1}^{T} q_{1t,\hat{d}}\left(\hat{\theta}_{\hat{d},1}\right) - \sum_{t=1}^{T} q_{2t,\hat{d}}\left(\hat{\theta}_{\hat{d},2}\right) - c_T, \tag{11}$$

can be shown to provide a weakly consistent estimate of the identity of the true model where $\hat{\theta}_{d,i}$ is defined by,

$$\hat{\theta}_{d,i} = \arg\max_{\theta_i} \frac{1}{T} \sum_{t=1}^{T} q_{it,d}(\theta_i). \tag{12}$$

In particular, Model 1 is selected if $\mathrm{IC}\left(\hat{d}\right) > 0$ and Model 2 is selected otherwise. In what follows, the true value of d is denoted by d^0 and $p\lim_{T\to\infty} \hat{\theta}_{d^0,i} = \theta_i^*$. Then, the following assumptions are made.

Assumption 1. For $i = 1, 2$,

$$\hat{\theta}_{d^0,i} - \theta_i^* = O_p\left(T^{-1/2}\right),$$

and,

$$Q_{iT,d^0}\left(\theta_i^*\right) = E\left(Q_{iT,d^0}\left(\theta_i^*\right)\right) + o_p(T).$$

Assumption 2. There exists $0 < \delta < 1/2$, such that,

$$\hat{d} - d^0 = O_p\left(T^{-\delta}\right).$$

Assumption 3. For $i = 1, 2$,

$$\lim_{\varepsilon\to 0} \sup_{\theta_i} \sup_{\bar{d}\in B(d^0,\varepsilon)} \sup_{t} \left.\frac{\partial q_{it,d}(\theta_i)}{\partial d}\right|_{d=\bar{d}} = O_p(1), \tag{13}$$

$$\lim_{\varepsilon\to 0} \sup_{\bar{\theta}\in B(\theta_i^*,\varepsilon)} \sup_{t} \left.\frac{\partial q_{it,d^0}(\theta_i)}{\partial \theta_i}\right|_{\theta_i=\bar{\theta}} = O_p(1), \tag{14}$$

and,

$$\lim_{\varepsilon\to 0}\sup_{\bar\theta\in B(\theta_i^*,\varepsilon)}\left.\frac{\partial^2 Q_{iT,d^0}(\theta_i)}{\partial\theta_i\partial\theta_i'}\right|_{\theta_i=\bar\theta}=O_p(T). \tag{15}$$

Assumption 4. Let,

$$c_T=o_p(T),$$

and,

$$T^{1-\delta}c_T^{-1}=o_p(1).$$

Assumption 1 is standard and relates to the estimation of the short memory component of the model, while Assumption 2 relates to the estimation of d. It should be noted that all the estimation methods described in the previous section satisfy Assumption 2. Generally, the consistency of estimators holds for a wide class of data generating sequences. For example, following Abadir et al. (2007), FELW is consistent for signal plus noise processes, transformations of stationary Gaussian sequences and for EGARCH type volatility models. Also, see Andrews and Sun (2004), Shimotsu and Phillips (2005) and Abadir et al. (2011). Assumption 3 is essentially a set of high level regularity conditions that underlie the proof of the main theorem and should be verified for specific models. The application of these conditions to a leading case is shown in Corollary 1. Finally, Assumption 4 presents sufficient conditions for the penalty term of a consistent information criterion. The above assumptions lead to the following theorem.

Theorem 1. *Under Assumptions 1–4, and if,*

$$\liminf_T\left[T^{-1}E\left(Q_{1T,d^0}\left(\theta_1^*\right)\right)-T^{-1}E\left(Q_{2T,d^0}\left(\theta_2^*\right)\right)\right]>0,$$

$$\lim_{T\to\infty}\Pr\left(\mathrm{IC}\left(\hat d\right)>0\right)=1. \tag{16}$$

Further, if,

$$Q_{1T,d^0}\left(\theta_1^*\right)-Q_{2T,d^0}\left(\theta_2^*\right)=O_p(1),$$

$$\lim_{T\to\infty}\Pr\left(\mathrm{IC}\left(\hat d\right)\le 0\right)=1. \tag{17}$$

It is clear that Assumption 2 and especially Eq. (13) are high level that may be difficult to establish in some situations. The following theorem shows how this condition can be simplified when the model has a particular form, that has similarities to, but is essentially more general than an autoregressive model.

Theorem 2. *Let Assumptions 1–2, 4 and Eqs. (14)–(15) hold. Let,*

$$q_{it}\left(u_d,\theta_{d,i}\right)=q_{it}\left(u_{d,t}^{(p)},\theta_{d,i}\right), \tag{18}$$

where $u_{d,t}^{(p)}=\left(u_{d,t},\ldots,u_{d,t-p}\right)'$ *for some finite constant* p*, for* $i=1,2$*. Let,*

$$\sup_{\theta_i}\sup_z\sup_t\frac{\partial q_{it}(u,\theta_i)}{\partial u}<\infty,\quad i=1,2.$$

Then, Theorem 1 holds.

As an example of how the high level Assumption 3 is proven when the model cannot be written in the AR type form given by Assumption 2, it is useful to note the following corollary that covers MA models.

Corollary 1. *Let* $z_{d^0,t}$ *follow an invertible* MA(p) *model. Then, Assumption 3 holds.*

It is important at this point to discuss the empirical implementation of the modified criterion. Given the original time series, $y_t\sim I(d_0)$, the estimate of the long memory parameter is $\widehat{d}$, which is obtained from using one of the four methods described in Section 2. The FFF series is constructed by truncating the infinite AR representation in Eq. (3). The FFF series is then used to estimate the parameters of the short memory component. In the case of an AR process this simply means that the Yule–Walker equations are solved from the FFF series. For more complicated models such as an ARMA(p, q), estimation is achieved by the minimization of a conditional sum of squares function, which assumes Gaussianity of the process.

4. Monte Carlo results

This section discusses the results from detailed simulation experiments concerning the properties of the information criteria, AIC, BIC, HQ and MIC, when the long memory parameter is estimated from one of the semi parametric methods, LW, LPW, FELW and ELW. For all the reported experiments a sample size of $T = 500$ observations is used and the simulation results are based on 1000 replications. The various information criteria are,

$$\text{AIC} = \ln\left(\sigma_\varepsilon^2\right) + \frac{2k}{T}, \tag{19}$$

$$\text{BIC} = \ln\left(\sigma_\varepsilon^2\right) + \frac{k\ln T}{T}, \tag{20}$$

$$\text{HQ} = \ln\left(\sigma_\varepsilon^2\right) + \frac{2kc\ln\ln T}{T}, \tag{21}$$

where $c = 1.0001$. The modified information criterion is given by,

$$\text{MIC} = \ln\left(\sigma_\varepsilon^2\right) + \left(\frac{k}{T}\right)T^{1-\delta+\xi} \tag{22}$$

where for any small ξ, the δ parameter is defined as $\delta = 0.5\alpha$, α is the parameter associated with the bandwidth selection used in the estimation, $m = \lfloor T^\alpha \rfloor$, and $\lfloor\cdot\rfloor$ denotes the integer part. The values $\alpha = \{0.4, 0.6\}$ are considered as general rule of thumb choices in applications for various semi parametric estimators. The choice used in this study is to take $\alpha = 0.4$; see Henry (2001) for an automated approach. Hence, the exponent of T in the modified penalty term coefficient becomes $(1 - \delta) = 0.8$.

It should be noted that all results involving simulations with LPW are based on the choice of $r = 1$, so that the low frequency ordinates of the spectrum are approximated by a linear, first order polynomial. Higher order adjustments were not found to lead to any noticeable improvement.

From Table 1 to Table 5 we report the results based on: (i) ARFIMA$(1, d, 0)$ data generating process with the AR parameter $\phi_1 = 0.9$, (ii) ARFIMA$(0, d, 1)$ data generating process with $\theta_1 = 0.9$ and (iii) ARFIMA$(1, d, 1)$ data generating process with $\phi_1 = 0.5$ and $\theta_1 = 0.9$. In all experiments d takes the following values $d = \{0, 0.2, 0.4, 0.49, 0.8, 1.2\}$. The sample size is set to $T = 500$ observations. All tables report the rate of success, i.e. the percentage of occasions when certain models were selected by the various criteria. Furthermore, the average AR order selected is denoted by $\widehat{p}$ and similarly the average MA order is denoted by $\widehat{q}$.

Each table can be read in the following manner: different estimation methods are seen on the top horizontal bar. The second column for each estimation method provides the rate of success for each criterion. The first and third columns count the number of times that the criterion chose a lower or higher order than the true one. Columns $\widehat{p}$ and $\widehat{q}$ report the average order selected across all replications. The top four panels of results are concerned with the different values of d. The bottom panel reports the average model selected across the top four panels for each criterion given the estimation method.

Tables 1 and 2 are concerned with the ARFIMA$(1, d, 0)$ with $\phi_1 = 0.9$, and their difference is simply the maximum order allowed for the information criteria. In Table 1, when there is no search for the order of an MA model, then the maximum orders are set as $\widehat{p}_{\max} = 8$ and $\widehat{q}_{\max} = 0$. However in Table 2 when a search is conducted across both AR and MA terms we set $\widehat{p}_{\max} = 8$ and $\widehat{q}_{\max} = 8$. From Table 1 it can be seen that the MIC succeeds in selecting the true AR order on 87% of the occasions, while the percentage of success for the AIC, BIC and HQ are 31%, 72% and 53% respectively. Moving to Table 2, where both the maximum autoregressive order and the maximum moving average order are set to 8, all the success rates are reduced. However, the MIC is still the superior method. Looking at the LW we see that there is an average rate of success of 54% for the MIC across all values of d, while the relevant rates for the other criteria are 2%, 32% and 18% for AIC, BIC and HQ respectively. Even in the nonstationary cases it can be seen that the MIC using ELW and FELW chooses the true order more than 50% of the times.

Tables 3 and 4 present the results using the ARFIMA$(0, d, 1)$ design for $\theta_1 = 0.9$. In Table 3 the restriction that $\widehat{p}_{\max} = 0$ is imposed while $\widehat{q}_{\max} = 4$. Then the AIC, BIC and HQ tend to misinterpret the short run dynamics of the series and generally overestimate the order of the model. In fact, the average $\widehat{q}$ is always close to the maximum lag of 4. However, MIC chooses the true order more than 50% of the times across all values of d and all the various estimators. The qualitative result is very similar to the case where $\widehat{p}_{\max} = 4$ and $\widehat{q}_{\max} = 4$. As seen in Table 4 the success rates are now decreased. This is due to the fact that the MIC also attributes the persistence of the series to the presence of AR parameters; it should be noted that in all cases the average orders selected are 0.5.

In the final reported experiment with a design of an ARFIMA$(1, d, 1)$ and with parameters of $\phi_1 = 0.5$ and $\theta_1 = 0.9$, the maximum order for the criteria is fixed to $\widehat{p}_{\max} = 8$ and $\widehat{q}_{\max} = 8$. It is clear that the MIC selects the true order more than 90% of the occasions across all d values and over all estimation methods. The average orders across all d using LW are 0%, 20%, 10% and 92% of the occasions for AIC, BIC, HQ and MIC respectively. Similarly, when using LPW it is found that the MIC success rate is 85% compared to BICs which is 14%. On using the ELW and FELW, it is clear that MIC is successful in selecting the correct order on 91% of occasions, compared to the next best alternative of BIC with 21%.

In summary, the results presented in this section are indicative of the fact that the standard Akaike (AIC), Bayesian (BIC) and Hannan–Quinn (HQ) criteria are not consistent when the long memory exponent is semi parametrically estimated. These criteria tend to introduce a large upward bias in the estimated order of the short memory component of the process.

Table 1
ARFIMA(1, d, 0) simulation results without MA lags.

ARFIMA(1, d, 0), $\phi_1 = 0.9$, $\widehat{p}_{\max} = 8$, $\widehat{q}_{\max} = 0$, $\alpha = 0.4$, $T = 500$

d	*Crit.*	LW					LPW1					ELW					FELW				
		Orders																			
		$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$
0	AIC	0.00	0.32	0.68	2.62	0.00	0.01	0.21	0.78	3.00	0.00	0.00	0.32	0.68	2.63	0.00	0.00	0.32	0.68	2.61	0.00
	BIC	0.01	0.74	0.26	1.31	0.00	0.02	0.48	0.50	1.66	0.00	0.01	0.73	0.26	1.30	0.00	0.01	0.73	0.26	1.30	0.00
	HQ	0.00	0.54	0.46	1.68	0.00	0.01	0.36	0.63	2.06	0.00	0.01	0.53	0.46	1.68	0.00	0.01	0.53	0.47	1.68	0.00
	MIC	0.12	0.88	0.00	0.88	0.00	0.13	0.73	0.14	1.01	0.00	0.13	0.87	0.00	0.87	0.00	0.12	0.87	0.00	0.88	0.00
0.2	AIC	0.00	0.29	0.71	2.77	0.00	0.01	0.23	0.76	3.05	0.00	0.00	0.28	0.72	2.81	0.00	0.00	0.28	0.72	2.81	0.00
	BIC	0.00	0.72	0.28	1.32	0.00	0.01	0.48	0.50	1.65	0.00	0.01	0.71	0.29	1.32	0.00	0.00	0.72	0.28	1.32	0.00
	HQ	0.00	0.53	0.47	1.71	0.00	0.01	0.38	0.61	2.04	0.00	0.00	0.53	0.47	1.70	0.00	0.00	0.53	0.47	1.71	0.00
	MIC	0.11	0.89	0.00	0.89	0.00	0.12	0.76	0.12	1.00	0.00	0.14	0.86	0.00	0.87	0.00	0.12	0.88	0.00	0.88	0.00
0.4	AIC	0.00	0.32	0.68	2.63	0.00	0.00	0.23	0.76	2.88	0.00	0.00	0.30	0.69	2.64	0.00	0.00	0.32	0.68	2.63	0.00
	BIC	0.00	0.71	0.29	1.32	0.00	0.01	0.52	0.48	1.62	0.00	0.00	0.70	0.30	1.34	0.00	0.01	0.71	0.29	1.32	0.00
	HQ	0.00	0.53	0.47	1.65	0.00	0.01	0.39	0.60	1.97	0.00	0.00	0.51	0.48	1.67	0.00	0.01	0.53	0.47	1.66	0.00
	MIC	0.13	0.86	0.00	0.87	0.00	0.10	0.78	0.13	1.03	0.00	0.12	0.87	0.00	0.88	0.00	0.13	0.87	0.00	0.88	0.00
0.49	AIC	0.00	0.32	0.68	2.58	0.00	0.01	0.22	0.78	2.96	0.00	0.00	0.31	0.69	2.60	0.00	0.00	0.31	0.69	2.61	0.00
	BIC	0.00	0.73	0.27	1.29	0.00	0.02	0.51	0.48	1.61	0.00	0.01	0.71	0.28	1.32	0.00	0.01	0.72	0.27	1.30	0.00
	HQ	0.00	0.52	0.47	1.71	0.00	0.01	0.38	0.61	2.00	0.00	0.00	0.52	0.48	1.70	0.00	0.00	0.53	0.47	1.69	0.00
	MIC	0.13	0.87	0.00	0.87	0.00	0.13	0.74	0.13	1.00	0.00	0.12	0.88	0.00	0.88	0.00	0.13	0.87	0.00	0.87	0.00
0.8	AIC											0.00	0.31	0.69	2.66	0.00	0.00	0.30	0.70	2.69	0.00
	BIC											0.00	0.74	0.26	1.29	0.00	0.00	0.71	0.29	1.33	0.00
	HQ											0.00	0.53	0.47	1.67	0.00	0.00	0.50	0.49	1.72	0.00
	MIC											0.13	0.88	0.00	0.88	0.00	0.12	0.88	0.01	0.89	0.00
1.2	AIC											0.00	0.25	0.75	2.75	0.00	0.00	0.26	0.74	2.72	0.00
	BIC											0.00	0.70	0.30	1.34	0.00	0.00	0.70	0.30	1.33	0.00
	HQ											0.00	0.48	0.52	1.77	0.00	0.00	0.50	0.50	1.75	0.00
	MIC											0.11	0.89	0.00	0.89	0.00	0.11	0.89	0.00	0.89	0.00
All d	AIC	0.00	0.31	0.69	2.65	0.00	0.01	0.22	0.77	2.98	0.00	0.00	0.29	0.70	2.68	0.00	0.00	0.30	0.70	2.68	0.00
	BIC	0.00	0.72	0.27	1.31	0.00	0.01	0.50	0.49	1.63	0.00	0.00	0.71	0.28	1.32	0.00	0.00	0.71	0.28	1.32	0.00
	HQ	0.00	0.53	0.47	1.69	0.00	0.01	0.38	0.61	2.02	0.00	0.00	0.52	0.48	1.70	0.00	0.00	0.52	0.48	1.70	0.00
	MIC	0.12	0.87	0.00	0.88	0.00	0.12	0.75	0.13	1.01	0.00	0.12	0.87	0.00	0.88	0.00	0.12	0.88	0.00	0.88	0.00

Table 2
ARFIMA(1, d, 0) simulation results with MA lags.

ARFIMA(1, d, 0), $\phi_1 = 0.9$, $\widehat{p}_{\max} = 8$, $\widehat{q}_{\max} = 8$, $\alpha = 0.4$, $T = 500$

d	*Crit.*	LW					LPW1					ELW					FELW				
		Orders																			
		$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$
0	AIC	0.00	0.04	0.96	5.73	4.93	0.00	0.03	0.97	5.98	5.15	0.00	0.03	0.97	5.85	5.00	0.00	0.04	0.96	5.74	4.93
	BIC	0.22	0.36	0.41	1.82	1.34	0.13	0.26	0.61	2.57	1.83	0.21	0.35	0.44	1.93	1.44	0.22	0.37	0.41	1.75	1.25
	HQ	0.10	0.21	0.69	2.79	2.03	0.05	0.15	0.80	3.47	2.57	0.11	0.21	0.68	2.84	2.02	0.10	0.21	0.68	2.77	1.96
	MIC	0.37	0.62	0.01	0.67	0.29	0.23	0.61	0.16	0.97	0.22	0.39	0.60	0.02	0.68	0.34	0.38	0.61	0.01	0.66	0.29
0.2	AIC	0.01	0.03	0.96	5.94	5.13	0.00	0.02	0.98	6.07	5.38	0.01	0.03	0.97	5.90	5.19	0.00	0.03	0.97	5.95	5.09
	BIC	0.24	0.33	0.43	1.81	1.44	0.17	0.21	0.62	2.54	1.92	0.26	0.32	0.42	1.83	1.51	0.25	0.34	0.41	1.81	1.43
	HQ	0.12	0.17	0.71	2.83	2.14	0.08	0.12	0.80	3.49	2.65	0.13	0.17	0.70	2.89	2.25	0.13	0.17	0.70	2.85	2.15
	MIC	0.43	0.56	0.01	0.62	0.34	0.30	0.54	0.17	0.92	0.26	0.44	0.55	0.01	0.60	0.32	0.44	0.56	0.01	0.57	0.30
0.4	AIC	0.00	0.02	0.98	6.05	5.38	0.00	0.01	0.99	6.03	5.44	0.00	0.02	0.98	6.03	5.45	0.01	0.02	0.98	6.01	5.35
	BIC	0.31	0.26	0.44	1.77	1.58	0.17	0.20	0.63	2.60	2.08	0.31	0.27	0.42	1.76	1.57	0.30	0.26	0.43	1.79	1.58
	HQ	0.15	0.17	0.68	2.75	2.24	0.08	0.10	0.82	3.56	2.85	0.15	0.17	0.68	2.86	2.36	0.16	0.16	0.69	2.80	2.27
	MIC	0.51	0.48	0.01	0.52	0.41	0.31	0.53	0.16	0.87	0.25	0.54	0.45	0.01	0.48	0.45	0.51	0.47	0.01	0.54	0.45
0.49	AIC	0.00	0.01	0.99	6.08	5.48	0.00	0.01	0.98	6.14	5.42	0.00	0.01	0.99	6.07	5.50	0.00	0.01	0.99	6.07	5.42
	BIC	0.29	0.32	0.39	1.66	1.43	0.17	0.24	0.60	2.53	1.94	0.29	0.30	0.41	1.76	1.49	0.29	0.30	0.41	1.72	1.44
	HQ	0.15	0.15	0.70	2.81	2.28	0.09	0.12	0.79	3.48	2.81	0.16	0.16	0.69	2.81	2.28	0.15	0.15	0.71	2.86	2.33
	MIC	0.49	0.50	0.01	0.54	0.40	0.33	0.54	0.14	0.85	0.27	0.51	0.48	0.01	0.54	0.41	0.50	0.49	0.01	0.54	0.39
0.8	AIC											0.01	0.02	0.97	5.94	5.21	0.01	0.02	0.98	6.03	5.35
	BIC											0.23	0.38	0.39	1.77	1.33	0.22	0.36	0.43	1.97	1.53
	HQ											0.12	0.20	0.69	2.87	2.11	0.11	0.19	0.70	2.97	2.22
	MIC											0.44	0.56	0.01	0.59	0.34	0.42	0.57	0.01	0.62	0.33
1.2	AIC											0.00	0.05	0.95	5.72	4.67	0.00	0.05	0.95	5.57	4.64
	BIC											0.15	0.44	0.41	2.02	1.31	0.14	0.44	0.42	1.90	1.25
	HQ											0.08	0.26	0.66	2.88	1.93	0.08	0.27	0.65	2.69	1.86
	MIC											0.32	0.68	0.01	0.71	0.21	0.31	0.68	0.00	0.69	0.20
All d	AIC	0.00	0.02	0.97	5.95	5.23	0.00	0.02	0.98	6.06	5.35	0.00	0.03	0.97	5.92	5.17	0.00	0.03	0.97	5.89	5.13
	BIC	0.27	0.32	0.42	1.77	1.45	0.16	0.23	0.61	2.56	1.94	0.24	0.34	0.42	1.84	1.44	0.24	0.34	0.42	1.82	1.41
	HQ	0.13	0.18	0.69	2.79	2.17	0.08	0.12	0.80	3.50	2.72	0.12	0.19	0.68	2.86	2.16	0.12	0.19	0.69	2.82	2.13
	MIC	0.45	0.54	0.01	0.59	0.36	0.29	0.55	0.15	0.90	0.25	0.44	0.55	0.01	0.60	0.35	0.43	0.56	0.01	0.60	0.33

Table 3
ARFIMA(0, d, 1) simulation results without AR lags.

ARFIMA(0, d, 1), $\theta_1 = 0.9, \widehat{p}_{\max} = 0, \widehat{q}_{\max} = 4, \alpha = 0.4, T = 500$																					
d	*Crit.*	LW					LPW1					ELW					FELW				
		Orders																			
		$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$
0	AIC	0.00	0.00	1.00	0.00	4.00	0.00	0.01	0.99	0.00	3.94	0.00	0.00	1.00	0.00	4.00	0.00	0.00	1.00	0.00	4.00
	BIC	0.00	0.00	1.00	0.00	3.95	0.00	0.01	0.99	0.00	3.86	0.00	0.00	1.00	0.00	3.95	0.00	0.00	1.00	0.00	3.96
	HQ	0.00	0.00	1.00	0.00	3.99	0.00	0.01	0.99	0.00	3.91	0.00	0.00	1.00	0.00	3.99	0.00	0.00	1.00	0.00	3.99
	MIC	0.48	0.52	0.00	0.00	0.52	0.41	0.59	0.00	0.00	0.59	0.47	0.53	0.00	0.00	0.53	0.46	0.54	0.00	0.00	0.54
0.2	AIC	0.00	0.00	1.00	0.00	4.00	0.00	0.01	0.99	0.00	3.90	0.00	0.00	1.00	0.00	4.00	0.00	0.00	1.00	0.00	4.00
	BIC	0.00	0.00	1.00	0.00	3.97	0.00	0.02	0.98	0.00	3.85	0.00	0.00	1.00	0.00	3.97	0.00	0.00	1.00	0.00	3.97
	HQ	0.00	0.00	1.00	0.00	4.00	0.00	0.01	0.99	0.00	3.89	0.00	0.00	1.00	0.00	4.00	0.00	0.00	1.00	0.00	3.99
	MIC	0.45	0.55	0.00	0.00	0.55	0.42	0.58	0.00	0.00	0.58	0.48	0.53	0.00	0.00	0.53	0.46	0.54	0.00	0.00	0.54
0.4	AIC	0.00	0.00	1.00	0.00	4.00	0.00	0.01	0.99	0.00	3.93	0.00	0.00	1.00	0.00	4.00	0.00	0.00	1.00	0.00	4.00
	BIC	0.00	0.00	1.00	0.00	3.96	0.00	0.01	0.99	0.00	3.86	0.00	0.00	1.00	0.00	3.96	0.00	0.00	1.00	0.00	3.96
	HQ	0.00	0.00	1.00	0.00	3.98	0.00	0.01	0.99	0.00	3.90	0.00	0.00	1.00	0.00	3.99	0.00	0.00	1.00	0.00	3.98
	MIC	0.49	0.51	0.00	0.00	0.51	0.45	0.55	0.00	0.00	0.56	0.50	0.51	0.00	0.00	0.51	0.49	0.51	0.00	0.00	0.51
0.49	AIC	0.00	0.00	1.00	0.00	4.00	0.00	0.01	0.99	0.00	3.93	0.00	0.00	1.00	0.00	4.00	0.00	0.00	1.00	0.00	4.00
	BIC	0.00	0.00	1.00	0.00	3.97	0.00	0.02	0.99	0.00	3.86	0.00	0.00	1.00	0.00	3.97	0.00	0.00	1.00	0.00	3.97
	HQ	0.00	0.00	1.00	0.00	3.99	0.00	0.01	0.99	0.00	3.91	0.00	0.00	1.00	0.00	3.99	0.00	0.00	1.00	0.00	3.99
	MIC	0.51	0.49	0.00	0.00	0.49	0.45	0.55	0.00	0.00	0.55	0.49	0.51	0.00	0.00	0.51	0.49	0.51	0.00	0.00	0.51
0.8	AIC											0.00	0.00	1.00	0.00	4.00	0.00	0.00	1.00	0.00	4.00
	BIC											0.00	0.00	1.00	0.00	3.96	0.00	0.00	1.00	0.00	3.96
	HQ											0.00	0.00	1.00	0.00	3.99	0.00	0.00	1.00	0.00	3.99
	MIC											0.47	0.53	0.00	0.00	0.53	0.50	0.50	0.00	0.00	0.50
1.2	AIC											0.00	0.00	1.00	0.00	4.00	0.00	0.00	1.00	0.00	4.00
	BIC											0.00	0.00	1.00	0.00	3.96	0.00	0.00	1.00	0.00	3.96
	HQ											0.00	0.00	1.00	0.00	3.99	0.00	0.00	1.00	0.00	3.99
	MIC											0.48	0.52	0.00	0.00	0.52	0.47	0.53	0.00	0.00	0.53
All d	AIC	0.00	0.00	1.00	0.00	4.00	0.00	0.01	0.99	0.00	3.92	0.00	0.00	1.00	0.00	4.00	0.00	0.00	1.00	0.00	4.00
	BIC	0.00	0.00	1.00	0.00	3.96	0.00	0.01	0.99	0.00	3.86	0.00	0.00	1.00	0.00	3.96	0.00	0.00	1.00	0.00	3.96
	HQ	0.00	0.00	1.00	0.00	3.99	0.00	0.01	0.99	0.00	3.90	0.00	0.00	1.00	0.00	3.99	0.00	0.00	1.00	0.00	3.99
	MIC	0.48	0.52	0.00	0.00	0.52	0.43	0.57	0.00	0.00	0.57	0.48	0.52	0.00	0.00	0.52	0.48	0.52	0.00	0.00	0.52

Table 4
ARFIMA(0, d, 1) simulation results with AR lags.

ARFIMA(0, d, 1), $\theta_1 = 0.9$, $\widehat{p}_{\max} = 4$, $\widehat{q}_{\max} = 4$, $\alpha = 0.4$, $T = 500$																					
d	*Crit.*	LW					LPW1					ELW					FELW				
		Orders																			
		$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$
0	AIC	0.00	0.00	1.00	2.59	3.00	0.00	0.00	1.00	2.59	3.05	0.00	0.00	1.00	2.57	3.00	0.00	0.00	1.00	2.57	3.03
	BIC	0.00	0.00	1.00	1.16	1.48	0.00	0.00	1.00	1.23	1.74	0.00	0.00	1.00	1.15	1.48	0.00	0.00	1.00	1.15	1.48
	HQ	0.00	0.00	1.00	1.59	2.00	0.00	0.00	1.00	1.69	2.23	0.00	0.00	1.00	1.62	2.04	0.00	0.00	1.00	1.63	2.07
	MIC	0.42	0.11	0.47	0.48	0.52	0.41	0.08	0.51	0.54	0.53	0.43	0.11	0.46	0.47	0.51	0.43	0.11	0.46	0.46	0.50
0.2	AIC	0.00	0.00	1.00	2.53	3.00	0.00	0.00	1.00	2.61	3.13	0.00	0.00	1.00	2.53	3.01	0.00	0.00	1.00	2.54	2.96
	BIC	0.00	0.00	1.00	1.13	1.49	0.00	0.00	1.00	1.23	1.80	0.00	0.00	1.00	1.16	1.50	0.00	0.00	1.00	1.13	1.50
	HQ	0.00	0.00	1.00	1.51	1.97	0.00	0.00	1.00	1.63	2.25	0.00	0.00	1.00	1.53	1.99	0.00	0.00	1.00	1.54	1.96
	MIC	0.39	0.12	0.49	0.50	0.55	0.40	0.09	0.51	0.52	0.52	0.40	0.12	0.49	0.51	0.55	0.40	0.12	0.48	0.49	0.54
0.4	AIC	0.00	0.00	1.00	2.58	3.05	0.00	0.00	1.00	2.58	3.03	0.00	0.00	1.00	2.60	3.06	0.00	0.00	1.00	2.60	3.04
	BIC	0.00	0.00	1.00	1.16	1.54	0.00	0.00	1.00	1.26	1.81	0.00	0.00	1.00	1.15	1.50	0.00	0.00	1.00	1.18	1.53
	HQ	0.00	0.00	1.00	1.58	2.04	0.00	0.00	1.00	1.69	2.25	0.00	0.00	1.00	1.60	2.05	0.00	0.00	1.00	1.61	2.07
	MIC	0.43	0.10	0.47	0.48	0.50	0.41	0.07	0.53	0.55	0.52	0.43	0.10	0.47	0.48	0.51	0.44	0.10	0.46	0.47	0.50
0.49	AIC	0.00	0.00	1.00	2.56	3.01	0.00	0.00	1.00	2.58	3.01	0.00	0.00	1.00	2.59	3.05	0.00	0.00	1.00	2.57	3.02
	BIC	0.00	0.00	1.00	1.13	1.53	0.00	0.00	1.00	1.26	1.77	0.00	0.00	1.00	1.11	1.51	0.00	0.00	1.00	1.11	1.53
	HQ	0.00	0.00	1.00	1.51	2.01	0.00	0.00	1.00	1.65	2.21	0.00	0.00	1.00	1.51	2.02	0.00	0.00	1.00	1.52	2.04
	MIC	0.45	0.10	0.45	0.46	0.49	0.40	0.09	0.51	0.53	0.52	0.46	0.09	0.46	0.46	0.48	0.46	0.10	0.45	0.45	0.49
0.8	AIC											0.00	0.00	1.00	2.51	3.01	0.00	0.00	1.00	2.52	3.00
	BIC											0.00	0.00	1.00	1.15	1.44	0.00	0.00	1.00	1.14	1.48
	HQ											0.00	0.00	1.00	1.51	1.93	0.00	0.00	1.00	1.50	1.97
	MIC											0.44	0.11	0.45	0.45	0.50	0.46	0.11	0.44	0.44	0.48
1.2	AIC											0.00	0.00	1.00	2.55	2.98	0.00	0.00	1.00	2.56	2.97
	BIC											0.00	0.00	1.00	1.12	1.52	0.00	0.00	1.00	1.14	1.54
	HQ											0.00	0.00	1.00	1.54	2.02	0.00	0.00	1.00	1.56	2.03
	MIC											0.44	0.10	0.46	0.47	0.47	0.45	0.10	0.46	0.46	0.46
All d	AIC	0.00	0.00	1.00	2.56	3.02	0.00	0.00	1.00	2.59	3.06	0.00	0.00	1.00	2.56	3.02	0.00	0.00	1.00	2.56	3.00
	BIC	0.00	0.00	1.00	1.14	1.51	0.00	0.00	1.00	1.24	1.78	0.00	0.00	1.00	1.14	1.49	0.00	0.00	1.00	1.14	1.51
	HQ	0.00	0.00	1.00	1.55	2.00	0.00	0.00	1.00	1.67	2.24	0.00	0.00	1.00	1.55	2.01	0.00	0.00	1.00	1.56	2.02
	MIC	0.42	0.11	0.47	0.48	0.51	0.40	0.08	0.52	0.54	0.52	0.43	0.10	0.46	0.47	0.50	0.44	0.11	0.46	0.46	0.49

Table 5
ARFIMA(1, d, 1) simulation results.

ARFIMA(1, d, 1), $\phi_1 = 0.5$, $\theta_1 = 0.9$, $\widehat{p}_{\max} = 8$, $\widehat{q}_{\max} = 8$, $\alpha = 0.4$, $T = 500$																					
d	*Crit.*	LW					LPW1					ELW					FELW				
		Orders																			
		$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$	$\{\widehat{p},\widehat{q}\} < \{p,q\}$	$\{\widehat{p},\widehat{q}\} = \{p,q\}$	$\{\widehat{p},\widehat{q}\} > \{p,q\}$	$\widehat{p}$	$\widehat{q}$
0	AIC	0.00	0.00	1.00	4.97	6.20	0.00	0.00	1.00	5.00	6.21	0.00	0.00	1.00	4.98	6.21	0.00	0.00	1.00	4.99	6.16
	BIC	0.00	0.20	0.80	3.94	5.18	0.00	0.14	0.87	4.05	5.24	0.00	0.20	0.80	3.89	5.16	0.00	0.23	0.78	3.77	5.02
	HQ	0.00	0.10	0.90	4.36	5.60	0.00	0.06	0.94	4.40	5.60	0.00	0.09	0.91	4.33	5.60	0.00	0.11	0.89	4.27	5.47
	MIC	0.00	0.92	0.08	1.21	1.36	0.06	0.85	0.09	1.16	1.30	0.01	0.92	0.08	1.21	1.35	0.01	0.93	0.06	1.18	1.31
0.2	AIC	0.00	0.00	1.00	4.90	6.10	0.00	0.00	1.00	5.06	6.19	0.00	0.00	1.00	4.83	6.20	0.00	0.00	1.00	4.81	6.11
	BIC	0.00	0.19	0.81	3.93	5.08	0.00	0.14	0.86	3.97	5.07	0.00	0.22	0.79	3.73	5.06	0.00	0.18	0.82	3.86	5.16
	HQ	0.00	0.08	0.92	4.32	5.53	0.00	0.07	0.93	4.37	5.53	0.00	0.11	0.89	4.14	5.50	0.00	0.10	0.90	4.15	5.50
	MIC	0.01	0.91	0.09	1.21	1.37	0.05	0.86	0.10	1.16	1.28	0.01	0.91	0.09	1.20	1.38	0.01	0.90	0.10	1.24	1.44
0.4	AIC	0.00	0.00	1.00	5.06	6.21	0.00	0.00	1.00	5.08	6.21	0.00	0.00	1.00	4.97	6.32	0.00	0.00	1.00	4.98	6.20
	BIC	0.00	0.20	0.80	3.95	5.14	0.00	0.13	0.87	4.04	5.21	0.00	0.21	0.79	3.81	5.16	0.00	0.21	0.79	3.87	5.11
	HQ	0.00	0.10	0.91	4.36	5.57	0.00	0.07	0.93	4.36	5.55	0.00	0.11	0.89	4.20	5.63	0.00	0.10	0.90	4.27	5.57
	MIC	0.01	0.92	0.07	1.21	1.32	0.05	0.85	0.10	1.21	1.28	0.01	0.91	0.09	1.25	1.43	0.01	0.91	0.09	1.23	1.39
0.49	AIC	0.00	0.01	0.99	4.93	6.21	0.00	0.00	1.00	5.08	6.18	0.00	0.00	1.00	4.91	6.27	0.00	0.00	1.00	4.90	6.17
	BIC	0.00	0.21	0.79	3.83	5.09	0.00	0.15	0.85	3.92	5.02	0.00	0.19	0.81	3.86	5.25	0.00	0.18	0.82	3.90	5.14
	HQ	0.00	0.11	0.89	4.16	5.51	0.00	0.07	0.93	4.34	5.49	0.00	0.10	0.90	4.18	5.63	0.00	0.09	0.91	4.24	5.55
	MIC	0.00	0.93	0.07	1.19	1.34	0.05	0.85	0.10	1.18	1.31	0.00	0.92	0.08	1.21	1.37	0.00	0.92	0.08	1.22	1.37
0.8	AIC											0.00	0.00	1.00	4.92	6.23	0.00	0.00	1.00	5.03	6.26
	BIC											0.00	0.22	0.78	3.74	5.07	0.00	0.23	0.77	3.77	5.08
	HQ											0.00	0.12	0.88	4.10	5.49	0.00	0.12	0.88	4.21	5.53
	MIC											0.00	0.92	0.08	1.19	1.35	0.00	0.91	0.09	1.24	1.43
1.2	AIC											0.00	0.00	1.00	4.96	6.14	0.00	0.00	1.00	5.08	6.09
	BIC											0.00	0.19	0.81	3.94	5.17	0.00	0.19	0.81	3.95	5.08
	HQ											0.00	0.08	0.92	4.38	5.60	0.00	0.09	0.91	4.38	5.41
	MIC											0.02	0.89	0.09	1.24	1.39	0.02	0.90	0.08	1.23	1.38
All d	AIC	0.00	0.00	1.00	4.96	6.18	0.00	0.00	1.00	5.05	6.20	0.00	0.00	1.00	4.93	6.23	0.00	0.00	1.00	4.96	6.16
	BIC	0.00	0.20	0.80	3.91	5.12	0.00	0.14	0.86	3.99	5.14	0.00	0.21	0.79	3.83	5.14	0.00	0.20	0.80	3.86	5.10
	HQ	0.00	0.10	0.90	4.30	5.55	0.00	0.07	0.93	4.37	5.54	0.00	0.10	0.90	4.22	5.58	0.00	0.10	0.90	4.25	5.51
	MIC	0.01	0.92	0.08	1.20	1.35	0.05	0.85	0.10	1.17	1.29	0.01	0.91	0.08	1.22	1.38	0.01	0.91	0.08	1.22	1.39

Table 6
ARFIMA(1, d, 1) empirical results.

Inflation (Feb. 1958–Jul. 2011), Exch. Rate (Dec. 1, 1986–June 30, 1999)							
	h	*Crit.*					
		Japan's inflation		UK's inflation		DM–Yen	
		MIC/BIC	MIC/HQ	MIC/BIC	MIC/HQ	MIC/BIC	MIC/HQ
LW	1	0.966	0.966	0.988	0.989	0.963	0.954
	3	0.988	0.988	1.008	1.009	0.950	0.943
	12	0.942	0.942	0.958	0.961	0.992	0.992
LPW, $r = 1$	1	0.978	0.978	0.982	0.992	0.963	0.954
	3	0.998	0.998	1.037	1.031	0.950	0.943
	12	0.926	0.926	1.010	1.011	0.992	0.992
ELW	1	0.955	0.941	0.966	0.972	0.971	0.965
	3	0.967	0.948	0.986	0.987	0.960	0.954
	12	0.988	0.994	0.991	0.991	0.997	0.999
FELW	1	0.975	0.973	1.011	1.011	0.988	0.982
	3	1.020	1.019	1.061	1.061	0.989	0.987
	12	0.936	0.937	1.010	1.010	1.001	1.002

This section also provides strong evidence that the MIC should be preferred over the standard criteria and that it generally performs well.

5. Empirical applications

Two examples of the above methodology for model selection are given in this section: first we analyze monthly inflation series for Japan and the UK, and second we investigate the order selection in the daily realized volatility series of DM–Yen exchange rate returns. Starting with the inflation example, the sample period spans from February 1958 to July 2011. The series of data was used until July 2006 for model selection purposes and the last 5 years (i.e. 60 months) were retained for assessing the out of sample forecasting ability of the various selected models. The observed realized volatility of DM–Yen is calculated at daily level using the high frequency squared returns aggregated over the day and were computed for 3045 days from December 1, 1986 to June 30, 1999. Similarly, to the inflation series, the last 3 trading months (i.e. 60 days) observations were retained for out of sample forecast comparison purposes.

Table 6 presents the relative Root Mean Squared Forecast Error, RMSFE, results from the estimated ARFIMA models for the above data. Again, the same choice for the bandwidth is fixed at $\alpha = 0.4$. It should be noted that the inflation series in these countries is quite likely to exhibit nonstationary long memory, so that the ELW and FELW are ideal estimators. To this extent the MIC based on these estimators is superior. The forecast horizons are one, three and twelve steps ahead, and correspond to one month, one quarter and one year respectively for the inflation series. The benchmark criteria used here are BIC and HQ.

The left panel of Table 6 provides the results for Japan's inflation. It can be seen that the MIC provides better forecasts compared to the other criteria using LW and LPW. It is clear that BIC and HQ select the same model orders in both estimations. In the LW case, the relative RMSFEs of MIC compared to both benchmarks are 0.966, 0.988 and 0.942 for the first, third and twelfth steps ahead respectively. It slightly weakens when LPW is used with 0.978, 0.998 for the first and third steps and 0.926 for the twelfth step. Using ELW, it is seen that MIC again outperforms the benchmarks with 0.955, 0.967 and 0.988 relative RMSFEs compared to BIC and 0.941, 0.948 and 0.994 compared to HQ. In the case of the FELW, the MIC fails in the third step with relative RMSFE above unity compared to both benchmarks. However, in the first and twelfth steps ahead MIC provides more accurate forecasts. For example for the twelfth step ahead the relative RMSFEs of MIC are 0.936 and 0.937 compared to BIC and HQ respectively.

The mid panel of Table 6 provides information on the methodology applied to the UK's inflation. In the short run, i.e. one month ahead, MIC provides better forecasts using LW, LPW and ELW compared to the BIC and HQ. However, for longer forecast horizons the MIC is better only using LW and ELW.

The right panel of Table 6 presents the results using the realized volatility series of the DM–Yen exchange rate returns. MIC dominates in terms of RMSFE compared to all other criteria in all steps ahead and most estimation methods apart from the twelfth step using FELW. In the first and third steps ahead the relative RMSFE of MIC compared to the other criteria fluctuates from 0.943 to 0.971. For the twelfth step ahead, the MIC's RMSFE is very close but still less than unity taking values between 0.992 and 0.997.

In summary, these empirical results in inflation and realized volatility series indicate that forecasts derived from models chosen from the MIC are generally preferable to those derived from the BIC and/or HQ procedures. However, it should be noted that some of the procedures involve estimators that are not valid in the nonstationary regions of d. When a researcher has no a priori knowledge concerning the degree of fractional integration of the series, a safe strategy is to use either the FELW or ELW estimators.

6. Conclusions

This paper has addressed the problem of model selection of the short memory component of a univariate time series with long memory. While there is a lot of evidence of the existence of long memory in economic time series, the most appropriate methods for selecting models have previously been unresolved. The application of the standard information criteria such as AIC, BIC and HQ, are not consistent when the long memory exponent is semi parametrically estimated. We establish a Modified Information Criterion (MIC) that overcomes these difficulties and provide proofs showing its asymptotic validity. The semi parametric estimators of the long memory parameter converge at a slower, and non standard, rate to the true value compared with MLE and this affects the theoretical properties and functional form of the MIC. The results here are general and cover any short memory model as defined in Sin and White (1996) and could, for example, be further applied to ESTAR, or some smooth nonlinear parameterization in the conditional mean, or even ARFIMA models with GARCH errors. Simulation evidence based on the workhorse ARFIMA(p, d, q) model and empirical evidence in monthly inflation and daily realized volatility generally indicate the desirability of using the new modified criterion.

7. Proofs

7.1. Proof of Theorem 1

It is first necessary to show the following,

$$\hat{\theta}_{\hat{d},i} - \hat{\theta}_{d^0,i} = O_p\left(T^{-\delta}\right), \quad i = 1, 2, \tag{23}$$

$$\sum_{t=1}^{T} q_{it,\hat{d}}\left(\hat{\theta}_{\hat{d},i}\right) - \sum_{t=1}^{T} q_{it,d^0}\left(\hat{\theta}_{d^0,i}\right) = O_p\left(T^{1-\delta}\right), \quad i = 1, 2. \tag{24}$$

Then,

$$\sum_{t=1}^{T} q_{it,\hat{d}}\left(\hat{\theta}_{\hat{d},i}\right) - \sum_{t=1}^{T} q_{it,d^0}\left(\hat{\theta}_{d^0,i}\right) = \sum_{t=1}^{T} q_{it,\hat{d}}\left(\hat{\theta}_{\hat{d},i}\right) - \sum_{t=1}^{T} q_{it,d^0}\left(\hat{\theta}_{\hat{d},i}\right) + \sum_{t=1}^{T} q_{it,d^0}\left(\hat{\theta}_{\hat{d},i}\right) - \sum_{t=1}^{T} q_{it,d^0}\left(\hat{\theta}_{d^0,i}\right).$$

Then, Eq. (24) follows if,

$$\sum_{t=1}^{T} q_{it,\hat{d}}\left(\hat{\theta}_{\hat{d},i}\right) - \sum_{t=1}^{T} q_{it,d^0}\left(\hat{\theta}_{\hat{d},i}\right) = O_p\left(T^{1-\delta}\right), \quad i = 1, 2, \tag{25}$$

and,

$$\sum_{t=1}^{T} q_{it,d^0}\left(\hat{\theta}_{\hat{d},i}\right) - \sum_{t=1}^{T} q_{it,d^0}\left(\hat{\theta}_{d^0,i}\right) = O_p\left(T^{1-\delta}\right), \quad i = 1, 2. \tag{26}$$

For Eqs. (23) and (25), it is sufficient to show that,

$$\sup_{\theta_i}\left(\sum_{t=1}^{T} q_{it,\hat{d}}\left(\theta_i\right) - \sum_{t=1}^{T} q_{it,d^0}\left(\theta_i\right)\right) = O_p\left(T^{1-\delta}\right), \quad i = 1, 2.$$

Applying the mean value theorem we have that,

$$q_{it,\hat{d}}\left(\theta_i\right) = q_{it,d^0}\left(\theta_i\right) + \left.\frac{\partial q_{it,d}\left(\theta_i\right)}{\partial d}\right|_{d=\bar{d}}\left(\hat{d} - d^0\right).$$

So,

$$\begin{aligned}\sup_{\theta_i}\left(\sum_{t=1}^{T} q_{it,\hat{d}}\left(\theta_i\right) - \sum_{t=1}^{T} q_{it,d^0}\left(\theta_i\right)\right) &= \sup_{\theta_i}\left(\sum_{t=1}^{T} \left.\frac{\partial q_{it,d}\left(\theta_i\right)}{\partial d}\right|_{d=\bar{d}}\left(\hat{d} - d^0\right)\right) \\ &\leq \left(\lim_{\varepsilon\to 0}\sup_{\theta_i}\sup_{\bar{d}\in B\left(d^0,\varepsilon\right)}\sup_{t}\left.\frac{\partial q_{it,d}\left(\theta_i\right)}{\partial d}\right|_{d=\bar{d}}\right)\left(\hat{d} - d^0\right),\end{aligned}$$

where $B\left(d^0, \varepsilon\right)$ denotes the ball of radius ε centered on d^0. So as long as,

$$\lim_{\varepsilon\to 0}\sup_{\theta_i}\sup_{\bar{d}\in B\left(d^0,\varepsilon\right)}\sup_{t}\left.\frac{\partial q_{it,d}\left(\theta_i\right)}{\partial d}\right|_{d=\bar{d}} = O_p\left(1\right),$$

Eqs. (23) and (25) hold. For Eq. (26), we have,

$$q_{it,d^0}\left(\hat{\theta}_{\hat{d},i}\right) - q_{it,d^0}\left(\hat{\theta}_{d^0,i}\right) = \left.\frac{\partial q_{it,d^0}\left(\theta_i\right)}{\partial \theta_i}\right|_{\theta_i=\bar{\theta}} \left(\hat{\theta}_{\hat{d},i} - \hat{\theta}_{d^0,i}\right),$$

where $\bar{\theta}$ lies between $\hat{\theta}_{\hat{d},i}$ and $\hat{\theta}_{d^0,i}$. So, given that $\hat{\theta}_{d^0,i} \to_p \theta_i^*$ and as long as,

$$\lim_{\varepsilon\to 0} \sup_{\bar{\theta}\in B(\theta_i^*,\varepsilon)} \sup_t \left.\frac{\partial q_{it,d^0}\left(\theta_i\right)}{\partial \theta_i}\right|_{\theta_i=\bar{\theta}} = O_p(1).$$

Eq. (26) holds. Having shown Eqs. (23) and (24), we use these parts to establish the theorem. First, show Eq. (16) by using Eq. (24), then

$$Q_{iT,\hat{d}}\left(\hat{\theta}_{\hat{d},i}\right) = Q_{iT,d^0}\left(\hat{\theta}_{d^0,i}\right) + O_p\left(T^{1-\delta}\right). \tag{27}$$

Further,

$$Q_{iT,d^0}\left(\hat{\theta}_{d^0,i}\right) = Q_{iT,d^0}\left(\theta_i^*\right) + O_p(1). \tag{28}$$

This follows by a two term mean value expansion,

$$Q_{iT,d^0}\left(\theta_i^*\right) = Q_{iT,d^0}\left(\hat{\theta}_{d^0,i}\right) + \left.\frac{\partial Q_{iT,d^0}\left(\theta_i\right)}{\partial \theta_i'}\right|_{\theta_i=\hat{\theta}_{d^0,i}} \left(\hat{\theta}_{d^0,i} - \theta_i^*\right) + 1/2\left(\hat{\theta}_{d^0,i} - \theta_i^*\right)' \left.\frac{\partial^2 Q_{iT,d^0}\left(\theta_i\right)}{\partial \theta_i \partial \theta_i'}\right|_{\theta_i=\bar{\theta}} \left(\hat{\theta}_{d^0,i} - \theta_i^*\right).$$

By $\left.\frac{\partial Q_{iT,d^0}(\theta_i)}{\partial \theta_i'}\right|_{\theta_i=\hat{\theta}_{d^0,i}} = 0$, and Eq. (15), $\left(\hat{\theta}_{d^0,i} - \theta_i^*\right)' \left.\frac{\partial^2 Q_{iT,d^0}(\theta_i)}{\partial \theta_i \partial \theta_i'}\right|_{\theta_i=\bar{\theta}} \left(\hat{\theta}_{d^0,i} - \theta_i^*\right) = O_p(1)$. So,

$$Q_{iT,\hat{d}}\left(\hat{\theta}_{\hat{d},i}\right) = Q_{iT,d^0}\left(\theta_i^*\right) + O_p\left(T^{1-\delta}\right) = Q_{iT,d^0}\left(\theta_i^*\right) + o_p(T).$$

Then,

$$\begin{aligned} \mathrm{IC}\left(\hat{d}\right) &= Q_{1T,\hat{d}}\left(\hat{\theta}_{\hat{d},1}\right) - Q_{2T,\hat{d}}\left(\hat{\theta}_{\hat{d},2}\right) - c_T = E\left(Q_{1T,d^0}\left(\theta_1^*\right)\right) - E\left(Q_{2T,d^0}\left(\theta_2^*\right)\right) + o_p\left(T\right) - c_T \\ &= E\left(Q_{1T,d^0}\left(\theta_1^*\right)\right) - E\left(Q_{2T,d^0}\left(\theta_2^*\right)\right) + o_p(T) > 0, \end{aligned}$$

with probability approaching one, since $E\left(Q_{1T,d^0}\left(\theta_1^*\right)\right) - E\left(Q_{2T,d^0}\left(\theta_2^*\right)\right) = O(T)$, proving Eq. (16). This finally shows Eq. (17). Again, by Eqs. (27) and (28),

$$\mathrm{IC}\left(\hat{d}\right) = Q_{1T,\hat{d}}\left(\hat{\theta}_{\hat{d},1}\right) - Q_{2T,\hat{d}}\left(\hat{\theta}_{\hat{d},2}\right) - c_T = Q_{1T,d^0}\left(\theta_1^*\right) - Q_{2T,d^0}\left(\theta_2^*\right) + O_p\left(T^{1-\delta}\right) - c_T \le 0,$$

with probability approaching 1 using the facts that $Q_{1T,d^0}\left(\theta_1^*\right) - Q_{2T,d^0}\left(\theta_2^*\right) = O_p(1)$ and $T^{1-\delta}c_T^{-1} = o_p(1)$, proving Eq. (17) and the theorem.

7.2. Proof of Theorem 2

To analyze the special case where,

$$q_{it}\left(z_d, \theta_{d,i}\right) = q_{it}\left(z_{d,t}^{(p)}, \theta_{d,i}\right),$$

and $z_{d,t}^{(p)} = \left(z_{d,t}, \ldots, z_{d,t-p}\right)'$ for some finite constant p. This simplifies Eq. (13) since,

$$\begin{aligned} &\sup_{\theta_i}\left(\sum_{t=1}^{T} q_{it}\left(z_{\hat{d},t}^{(p)}, \theta_i\right) - \sum_{t=1}^{T} q_{it}\left(z_{d^0,t}^{(p)}, \theta_i\right)\right) \\ &= \sup_{\theta_i}\left(\sum_{t=1}^{T}\left(\frac{\partial q_{it}\left(z_{d^0,t}^{(p)}, \theta_i\right)}{\partial z_{d^0,t}^{(p)}}\right)'\left(z_{\hat{d},t}^{(p)} - z_{d^0,t}^{(p)}\right)\right) + O_p\left(\sum_{t=1}^{T}\left(z_{\hat{d},t}^{(p)} - z_{d^0,t}^{(p)}\right)^2\right) \\ &\le \left(\sup_{\theta_i}\sup_z\sup_t \frac{\partial q_{it}\left(z, \theta_i\right)}{\partial z}\right)'\left(\sum_{t=1}^{T} z_{\hat{d},t}^{(p)} - z_{d^0,t}^{(p)}\right) + O_p\left(\sum_{t=1}^{T}\left(z_{\hat{d},t}^{(p)} - z_{d^0,t}^{(p)}\right)^2\right). \end{aligned}$$

By Wright (1995),

$$\sum_{t=1}^{T}\left(z_{\hat{d},t}-z_{d^0,t}\right)=O_p\left(T^{1-\delta}\right),$$

and,

$$\sum_{t=1}^{T}\left(z_{\hat{d},t}^{(p)}-z_{d^0,t}^{(p)}\right)^2=o_p\left(T^{1-\delta}\right),$$

So as long as,

$$\sup_{\theta_i}\sup_{z}\sup_{t}\frac{\partial q_{it}\left(z,\theta_i\right)}{\partial z}<\infty,\quad i=1,2,$$

Eq. (13) holds.

7.3. *Proof of Corollary 1*

The next step is to consider a leading case where Eq. (18) does not hold. This is the case where $z_{d^0,t}$ is an MA process. For simplicity an MA(1) process is assumed so that,

$$z_{d^0,t}=\epsilon_t-\theta\epsilon_{t-1},$$

where $\epsilon_t\sim$ i.i.d.(0, 1) and $|\theta|<1$. This implies a log-likelihood function of the form,

$$L\left(d,\theta\right)=-\frac{T}{2}\log\left(2\pi\right)+\sum_{t=1}^{T}q_{t,d}\left(\theta\right)=-\frac{T}{2}\log\left(2\pi\right)+\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\theta^j z_{d,t-j}\right).$$

It is necessary to examine Eqs. (13) and (14) since Eq. (15) can be proven similarly to Eq. (14). Then,

$$\sup_{t}\sup_{\theta}\lim_{\varepsilon\to 0}\sup_{\bar{d}\in B\left(d^0,\varepsilon\right)}\left.\frac{\partial q_{it,d}\left(\theta\right)}{\partial d}\right|_{d=\bar{d}}=\sup_{t}\sup_{|\theta|<1}\lim_{\varepsilon\to 0}\sup_{\bar{d}\in B\left(d^0,\varepsilon\right)}\sum_{j=0}^{t-1}\theta^j\left.\frac{\partial z_{d,t-j}}{\partial d}\right|_{d=d^0}. \tag{29}$$

We note that, by Eqs. (4.11)–(4.17) of Wright (1995) and assuming that $y_t=0,\ t=\cdots-1,0$,

$$\lim_{\varepsilon\to 0}\sup_{\bar{d}\in B\left(d^0,\varepsilon\right)}\left.\frac{\partial z_{d,t-j}}{\partial d}\right|_{d=d^0}=\lim_{\varepsilon\to 0}\sup_{\bar{d}\in B\left(d^0,\varepsilon\right)}\sum_{j=1}^{t-1}a_{1j}\left(\bar{d}-d^0\right)\left(\epsilon_t-\theta\epsilon_{t-1}\right)=\sum_{j=1}^{t-1}a_{1j}(0)\left(\epsilon_t-\theta\epsilon_{t-1}\right),$$

where $a_{1j}(x)=\frac{da_j(x)}{dx}$, $a_j(x)=\frac{\Gamma(j-x)}{\Gamma(-x)\Gamma(j+1)}$, and therefore, $a_{1j}(0)=j^{-1}$. By Lemma 1 of Wright (1995),

$$\sup_{t}\sum_{j=1}^{t-1}a_{1j}(0)\left(\epsilon_t-\theta\epsilon_{t-1}\right)=O_p(1),$$

and so,

$$\sup_{t}\sup_{|\theta|<1}\sum_{j=0}^{t-1}\theta^j\left.\frac{\partial z_{d,t-j}}{\partial d}\right|_{d=d^0}=O_p(1),$$

proving Eq. (13). For Eq. (14) it is convenient to use,

$$\left.\frac{\partial q_{it,d^0}\left(\theta_i\right)}{\partial\theta_i}\right|_{\theta_i=\bar{\theta}}=j\sum_{j=0}^{t-1}\theta^{j-1}z_{d,t-j}. \tag{30}$$

Then, it is obvious that Eq. (14) holds.

Acknowledgments

The authors are grateful to the editor, the associate editor and three anonymous referees for their helpful comments. Any remaining errors are the authors' sole responsibility.

References

Abadir, K.M., Distaso, W., Giraitis, L., 2007. Non-stationarity extended Local Whittle estimation. Journal of Econometrics 141, 1353–1384.
Abadir, K.M., Distaso, W., Giraitis, L., 2011. An I(d) model with trend and cycles. Journal of Econometrics 163, 186–199.
Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.
Andrews, D.W.K., Sun, Y., 2004. Adaptive local polynomial Whittle estimation of long-range dependence. Econometrica 72 (2), 569–614.
Baillie, R.T., Kapetanios, G., 2007. Testing for neglected nonlinearity in long memory models. Journal of Business and Economic Statistics 25, 447–461.
Baillie, R.T., Kapetanios, G., 2008. Nonlinear models for strongly dependent processes with financial applications. Journal of Econometrics 147, 60–71.
Baillie, R.T., Kapetanios, G., 2013. Estimation and inference for impulse response functions from univariate strongly persistent processes. Econometrics Journal (forthcoming).
Beran, J., Bhansali, R.J., Ocker, D., 1998. On unified model selection for stationary and nonstationary short and long memory autoregressive processes. Biometrika 85, 921–934.
Bisaglia, L., Guegan, M., 1998. A comparison of techniques of estimation in long-memory processes. Computational Statistics & Data Analysis 27, 61–81.
Crato, N., Ray, B.K., 1996. Model selection and forecasting for long range dependent processes. Journal of Forecasting 15, 107–125.
Dalla, V., Giraitis, L., Hidalgo, J., 2005. Consistent estimation of the long memory parameter for nonlinear time series. Journal of Time Series Analysis 27, 211–251.
Doornik, J.A., Ooms, M., 2003. Computational aspects of maximum likelihood estimation of autoregressive, fractionally integrated, moving average models. Computational Statistics and Data Analysis 42, 333–348.
Fox, C., Taqqu, M.S., 1986. Large sample properties of parameter estimates for strongly dependent stationary Gaussian processes. Annals of Statistics 14, 517–532.
Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–238.
Granger, C.W.J., 1980. Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14, 227–238.
Granger, C.W.J., Joyeux, R., 1980. An introduction to long memory time series models and fractional differencing. Journal of Time Series Analysis 1, 15–39.
Haldrup, N., Nielsen, M.O., 2007. Estimation of fractional integration in the presence of data noise. Computational Statistics and Data Analysis 51, 3100–3114.
Hannan, E.J., 1980. The estimation of the order of an ARMA process. Annals of Statistics 8, 1071–1081.
Hannan, E.J., Quinn, B.G., 1979. The determination of the order of an autoregression. Journal of the Royal Statistical Society: Series B 41, 190–195.
Henry, M., 2001. Robust automatic bandwidth for long memory. Journal of Time Series Analysis 22 (3), 293–316.
Hidalgo, J., 2002. Consistent order selection with strongly dependent data and its application to efficient estimation. Journal of Econometrics 110, 213–239.
Hosking, J.R.M., 1981. Fractional differencing. Biometrika 65, 165–176.
Hosoya, Y., 1997. A limit theory for long range dependence and statistical inference on related models. Annals of Statistics 25, 105–137.
Kapetanios, G., 2001. Model selection in threshold models. Journal of Time Series Analysis 22, 733–754.
Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Annals of Mathematical Statistics 22, 79–86.
Mayoral, L., 2007. Minimum distance estimation of stationary and non-stationary ARFIMA processes. Econometrics Journal 10, 124–148.
Phillips, P.C.B., 2007. Unit root periodogram. Journal of Econometrics 138, 104–124.
Poskitt, D., 2007. Autoregressive approximation in nonstandard situations: the non-invertible and fractionally integrated case. Annals of the Institute of Statistical Mathematics 59, 697–725.
Robinson, P.M., 1995. Gaussian semiparametric estimation of long range dependence. Annals of Statistics 23, 1630–1661.
Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.
Shibata, R., 1976. Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika 63, 117–126.
Shibata, R., 1980. Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. Annals of Statistics 8, 147–164.
Shimotsu, K., Phillips, P.C.B., 2005. Exact local Whittle estimation of fractional integration. Annals of Statistics 33, 1890–1933.
Shimotsu, K., Phillips, P.C.B., 2006. Local whittle estimation of fractional integration and some of its variants. Journal of Econometrics 130, 209–233.
Sin, C.Y., White, H., 1996. Information criteria for selecting possibly misspecified parametric models. Journal of Econometrics 71, 207–225.
Sowell, F.B., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Journal of Econometrics 53, 165–188.
Tieslau, M.A., Schmidt, P., Baillie, R.T., 1996. A minimum distance estimator for long-memory processes. Journal of Econometrics 71, 249–264.
Wright, J.H., 1995. Stochastic orders of magnitude associated with two-stage estimators of fractional ARMA systems. Journal of Time Series Analysis 16, 119–125.
Zevallos, M., Palma, W., 2013. Minimum distance estimation of *ARFIMA* processes. Computational Statistics and Data Analysis 58, 242–256.