

ORDER SELECTION AND INFERENCE WITH LONG MEMORY DEPENDENT DATA

ABHIMANYU GUPTA AND JAVIER HIDALGO

ABSTRACT. In empirical studies selection of the order of a model is routinely invoked. A common example is the order selection of an autoregressive model via Akaike's *AIC*, Schwarz's *BIC* or Hannan and Quinn's *HIC*. The criteria are based on the conditional sum of squares, *CSS*. However the computation of the *CSS* might be difficult for some models such as Bloomfield's exponential model and/or when we allow for long memory dependence. The main aim of the paper is thus to propose an alternative way to compute the criterion by using the decomposition of the variance of the innovation errors in terms of its frequency components. We show its validity to obtain the correct order the model. In addition, as a by-product, we describe a simple (two-step) estimator of the parameters of the model.

1. INTRODUCTION

In empirical studies methods to select the order of a model are often invoked. The methods are based on the minimization of a criterion function which involves the estimation of the one-step-prediction error, that is the conditional sum of squares, *CSS*, plus a penalization function depending on the number of parameters which has been estimated. For instance, when deciding the order of an autoregressive (*AR*) or autoregressive moving average (*ARMA*) model, standard methods are Akaike (1974) *AIC*, Schwarz (1978) *BIC* or Hannan and Quinn's (1979) *HIC* information criterion. The criteria only differ on the penalization function employed. The bulk of the work has been done under the assumption that the data is weakly dependent, see among others Akaike (1974), Hannan and Quinn (1979), Shibata (1976), or Hannan (1980). However, one exception is Beran et al. (1998), who examined the order determination of the autoregressive component of a fractional integrated autoregressive (*FARD*) model with known, but finite, upper limit for the order of the autoregressive polynomial.

However methods based on the computation of the *CSS* require to obtain the innovation sequence, which for some models can be computationally

Date: **April 26, 2019.**

Key words and phrases. Long Memory, spectral decomposition, BIC, HIC.

Abhimanyu Gupta's research was supported by ESRC grant ES/R006032/1.

We thank three referees for comments which led to a much improved version of the paper. Of course all remaining errors are our sole responsibility. Finally, data sharing is not applicable to this article as no new data were created or analyzed in this study.

very demanding, if at all possible. One of these models is the Bloomfield's (1973) exponential model. The reason comes from the observation that they do not have a finite autoregressive representation and more importantly the coefficients of the *AR* representation do not have a close form in terms of the parameters of the model, so that one needs to implement numerical integration to compute the coefficients and so the innovations. The latter computational complication might be exacerbated if one allows for the possibility of long memory dependence.

The main objective of the paper is to introduce and examine the properties of a modified version of the aforementioned criteria. We shall focus on the order selection of the fractional autoregressive, *FARI*, or the fractional integrated Bloomfield's (1973) exponential, *FEXP*, model, see respectively (2.3) and (2.5) for their definitions. The order of the short memory component, p_0 , of the models is assumed to be finite but with an unknown upper bound, which appears to be realistic in empirical applications. Thus, by order of the short memory, p_0 , we mean the order of the *AR* polynomial if the data followed a *FARI* model. So, we relax some of the conditions imposed in the work by Beran et al. (1998), which is the closest to the type of assumption we impose to the dependence structure of the data. Moreover we allow for a larger class of models as we do not focus exclusively on the *FARI* models. To achieve that goal, we analyze a criterion function based on the decomposition of the variance of the innovation/error term of the model in terms of its frequency components by using the canonical decomposition of the spectral density function, see Whittle (1963, p.26), Hannan's (1970, p.147) Theorem 6 or Brillinger's (1981) Theorem 3.8.4 for a definition. We believe that the method has some important features compared to those based on the *CSS*. Apart from the computational aspects already mentioned, there are some additional theoretical considerations which we shall mention when introducing our criterion.

In addition, as a by product, we describe and examine an estimator of the long memory parameter without knowledge of the short memory component which is of $O_p\left(T^{-1/2} \log \log^{1/2} T\right)$. Furthermore, we provide *root-T* consistent estimators of the short memory component for the *AR* or the Bloomfield's models. The motivation to examine these estimators of the parameters comes from the fact that their statistical behaviour is central to investigate the asymptotic properties of the estimation of p_0 . Finally, because the parameters of the model are estimated independently of the dimension (number of parameters) assumed in the model and *root-T* (up to a $\log \log^{1/2} T$ factor) consistent, they can be employed as initial estimates in a two-step iterative algorithm when using the Whittle objective function to estimate the parameters of the model.

The remainder of the paper is organized as follows. In the next section, we present the canonical decomposition of the spectral density function and we discuss how it can be used to estimate the parameters of the model.

Furthermore we examine a simple two-step computational *root-T* consistent and asymptotically normal estimator of the long memory parameter. Section 3 describes the criterion function to estimate the order of a *FARI* and/or a *FEXP* model. We show that the criterion provides a consistent estimator of the order of the model. Section 4 presents a small Monte-Carlo experiment to assess the finite sample performance of the order selection. In Section 5, we give the proofs of our results. Finally in Section 6 we discuss the scenario when the data is nonstationary, that is we allow the long memory parameter to be in the nonstationary region. It appears possible/feasible since the criterion does not need to compute the innovations of the model.

2. ESTIMATION OF THE *FARI* AND *FEXP* MODELS

The main aim of this section is to describe and examine how we can estimate the parameters of the model using the canonical decomposition of the spectral density function. The motivation for its study comes from the observation that to prove the validity of our criterion given in (3.5), one preliminary step is to know the statistical properties of the estimator of the parameters employed in its computation. For that purpose, let $\{x_t\}_{t \in \mathbb{Z}}$ be a zero mean covariance stationary sequence which is assumed to have the autoregressive representation

$$(2.1) \quad A(L)x_t =: \sum_{j=0}^{\infty} a_j x_{t-j} = \varepsilon_t, \quad a_0 = 1,$$

where $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a white noise zero mean sequence with variance σ_ε^2 . We shall assume that (2.1) admits a finite dimensional parameterization in terms of a $(p_0 + 1)$ -dimensional parameter ψ . That is,

$$(2.2) \quad \sum_{j=0}^{\infty} a_j(\psi) x_{t-j} = \varepsilon_t, \quad a_0(\psi) = 1.$$

One possible parameterization of (2.2) is the *FARI* $(p_0; \alpha, 0)$ model

$$(2.3) \quad \Phi(L; \theta) (1 - L)^{\alpha/2} x_t = \varepsilon_t,$$

where $\alpha \in [0, 1)$ and $\Phi(L; \theta) = 1 - \sum_{j=1}^{p_0} \theta_j L^j$ whose roots lie outside the unit circle. For this model $\psi = (\alpha, \theta) \in \Delta$, a compact set in \mathbb{R}^{p_0+1} , and we denote $\theta = (\theta_1, \dots, \theta_{p_0})'$. The parameter α is known as the *long memory* parameter. Model (2.3) has a spectral density function given by

$$(2.4) \quad f_x(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left| 1 - e^{i\lambda} \right|^{-\alpha} \left| \Phi(e^{i\lambda}; \theta) \right|^{-2}, \quad 0 \leq \lambda \leq \pi.$$

A second possible (finite) parameterization of (2.2) is the *FEXP* model, whose spectral density function is defined as

$$(2.5) \quad f_x(\lambda) = \left| 1 - e^{i\lambda} \right|^{-\alpha} \exp \left\{ -\zeta_0 - \sum_{j=1}^{p_0} \zeta_j \cos(j\lambda) \right\}, \quad 0 \leq \lambda \leq \pi.$$

Here $\psi = (\alpha, \zeta')'$, where $\zeta = (\zeta_1, \dots, \zeta_{p_0})'$, after we recall that the parameter ζ_0 is linked to the variance of the innovation sequence ε_t by the transformation $\sigma_\varepsilon^2 = 2\pi \exp(-\zeta_0)$. Observe that the number of parameters governing the *short memory* component is p_0 for both *FEXP* and *FARI* models. As indicated in the introduction, (2.5) does not have a finite *AR* representation even when $\alpha = 0$ and more importantly, see (2.6) below, the coefficient $a_j(\psi)$ in the *AR* representation (2.2) have no close form and they have to be obtained by numerical integration.

Consider the *AR*(∞) representation of x_t in (2.1). Using the canonical factorization of the spectral density function, see Whittle (1963, p.26) or Hannan's (1970, p.147) Theorem 6 or Brillinger's (1981) Theorem 3.8.4, their coefficients a_j can be obtained as

$$(2.6) \quad \begin{aligned} a_j &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp \left\{ - \sum_{\ell=1}^{\infty} \xi_\ell e^{-i\ell\lambda} \right\} e^{ij\lambda} d\lambda, \quad j = 1, 2, \dots \\ \sigma_\varepsilon^2 &= 2\pi e^{-\xi_0}, \end{aligned}$$

where

$$(2.7) \quad \xi_\ell = \frac{1}{\pi} \int_0^\pi \log(f_x(\lambda)) \cos(\ell\lambda) d\lambda, \quad \ell \geq 0,$$

and $f_x(\lambda)$ denoting the spectral density function of the sequence $\{x_t\}_{t \in \mathbb{Z}}$.

Next, suppose that we replace $f_x(\lambda)$ by $g_x(\lambda) =: |1 - e^{i\lambda}|^\alpha f_x(\lambda)$ in (2.7). Then, we would have that for the model (2.5), (2.7) becomes

$$(2.8) \quad \zeta_\ell = \frac{1}{\pi} \int_0^\pi \log(g_x(\lambda)) \cos(\ell\lambda) d\lambda, \quad \ell \geq 0$$

whereas for the model (2.4), (2.6) becomes

$$(2.9) \quad \theta_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp \left\{ - \sum_{\ell=1}^{\infty} \xi_\ell e^{-i\ell\lambda} \right\} e^{ij\lambda} d\lambda, \quad j \geq 1.$$

That is, expressions (2.8) and (2.9) provide respectively the parameters of the short memory component of the *FEXP* and *FARI* models.

However in empirical examples, $g_x(\lambda)$ is unknown, so to compute either (2.8) or (2.9), we need to estimate $g_x(\lambda)$. For that purpose, denote the periodogram of a generic sequence $\{z_t\}_{t=1}^T$ by

$$I_z(\lambda) = |w_z(\lambda)|^2, \quad \lambda \in [0, \pi],$$

where

$$w_z(\lambda) = \frac{1}{T^{1/2}} \sum_{t=1}^T z_t e^{-it\lambda}$$

is the *Discrete Fourier Transform* of the sequence $\{z_t\}_{t=1}^T$. Let $\lambda_j = 2\pi j/T$, $j = 0, \pm 1, \dots, \pm [T/2]$ with $[a]$ denoting the integer part of the number a ,

then a standard estimator of $g_x(\lambda)$ is the weighted periodogram

$$(2.10) \quad \widehat{g}_x(\lambda) = \frac{1}{2m+1} \sum_{j=-m}^m \left| 1 - e^{i(\lambda_j + \lambda)} \right|^{\widehat{\alpha}} I_x(\lambda + \lambda_j),$$

where $\widehat{\alpha}$ is an estimator of the long memory parameter α and $m = m(T)$ a sequence that increases slowly with T , that is $m^{-1} + T/m = o(1)$. For instance, see Geweke and Porter-Hudak (1983) and Robinson (1995a), we can estimate the parameter α by

$$(2.11) \quad \widehat{\alpha} = \sum_{k=1}^m \delta_k(m) \log(I_{x,k}),$$

where, for a generic n , we denote

$$(2.12) \quad \delta_k(n) = \frac{h_k - \bar{h}(n)}{\sum_{k=1}^n (h_k - \bar{h}(n))^2}; \quad h_k = -\log \left| 1 - e^{i\lambda_k} \right|; \quad \bar{h}(n) = n^{-1} \sum_{k=1}^n h_k$$

and for a generic function $\chi(\lambda)$, we abbreviate $\chi(\lambda_k)$ by χ_k .

It is worth giving the intuition behind the estimator in (2.10). Since under suitable regularity conditions, see Robinson (1995a), we have that $\widehat{\alpha} - \alpha = o_p(1)$ and $f_x^{-1}(\lambda + \lambda_j) E(I_x(\lambda + \lambda_j)) \rightarrow 1$, we should expect that $\widehat{g}_x(\lambda)$ will estimate $\sigma_\varepsilon^2 |\Phi(e^{i\lambda}; \theta)|^{-2} / (2\pi)$ or $\exp\left\{-\zeta_0 - \sum_{j=1}^{p_0} \zeta_j \cos(j\lambda)\right\}$, that is the short memory components of the *FARI* or *FEXP* models respectively.

Observe that the estimator given in (2.10) suggests to estimate $f_x(\lambda)$ by

$$\widehat{f}_x(\lambda) = \left| 1 - e^{i\lambda} \right|^{-\widehat{\alpha}} \widehat{g}_x(\lambda)$$

which we can regard as being first prewhitened and then recoloured in the frequency domain, in contrast to the analogue suggested by Press and Tukey (1956) in the time domain when $f_x(\lambda)$ is believed to have sharp peaks, as it might be under our conditions. The estimator $\widehat{f}_x(\lambda)$ was considered in Hidalgo and Yajima (2002) who showed that, for $\lambda > 0$, $\left| (\widehat{f}_x(\lambda) - f_x(\lambda)) / f_x(\lambda) \right| = O_p(m^{-1/2})$ and $\sup_{j \geq 1} \left| (\widehat{f}_{x,j} - f_{x,j}) / f_{x,j} \right| = o_p(1)$. It is interesting to observe that the latter statistical properties do not assume any prior knowledge of p_0 .

So, replacing $g_x(\lambda)$ by $\widehat{g}_x(\lambda)$ in (2.7) and (2.8) and using Riemann's discrete approximation of integrals by sums, we obtain

$$(2.13) \quad \widehat{\zeta}_j = \frac{1}{M} \sum_{\ell=1}^{M-1} (\log \widehat{g}_{x,2m\ell}) \cos(j\lambda_{2m\ell}), \quad j = 0, 1, \dots, M,$$

$$(2.14) \quad \widehat{\Theta}_{2mq} = \overline{\widehat{\Theta}_{-2mq}} = \exp \left\{ - \sum_{j=1}^{M-1} \widehat{\zeta}_j e^{-ij\lambda_{2mq}} \right\}, \quad q = 0, 1, \dots, M,$$

$$(2.15) \quad \widehat{\theta}_\ell = \frac{1}{2M} \sum_{q=-M+1}^M \widehat{\Theta}_{2mq} e^{i\ell\lambda_{2mq}}, \quad \ell = 1, \dots, M, \quad \text{and}$$

$$(2.16) \quad \widehat{\sigma}_\varepsilon^2 = 2\pi e^{-\widehat{\zeta}_0},$$

where \bar{z} denotes the conjugate of the complex number z and $M = \lceil T/2m \rceil$. Observe that (2.13) gives the estimate of the parameters ζ_j , for $j = 0, 1, \dots, p_0$, for the Bloomfield's exponential model in (2.5) whereas $\widehat{\theta}_\ell$ in (2.15) gives those of $\Phi(L; \theta)$ of the *FARI* model in (2.3).

We now introduce our regularity conditions.

Condition C1: $\{x_t\}_{t \in \mathbb{Z}}$ admits a *MA* representation

$$x_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} b_j^2 < \infty, \quad b_0 = 1,$$

where $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is a stochastic process with finite $8 + \delta$ moments for some $\delta > 0$, where $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$, $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = E(\varepsilon_t^2) = \sigma_\varepsilon^2$ and the joint fourth cumulant of ε_t is

$$\text{cum}(\varepsilon_{t_1}, \varepsilon_{t_2}, \varepsilon_{t_3}, \varepsilon_{t_4}) = \begin{cases} \kappa & t_1 = t_2 = t_3 = t_4 \\ 0 & \text{otherwise.} \end{cases}$$

Condition C2: $(\partial/\partial\lambda) |B(\lambda)| = O(|B(\lambda)|/\lambda)$ as $\lambda \rightarrow 0+$, where

$$B(\lambda) = \sum_{j=0}^{\infty} b_j e^{ij\lambda}$$

is twice continuously differentiable in any open set outside the origin. Also, there exist $C \in (0, \infty)$ and $\alpha \in [0, 1)$, such that

$$f_x(\lambda) = C\lambda^{-\alpha} (1 + O(\lambda^2)) \quad \text{as } \lambda \rightarrow 0+.$$

Condition C3: $m^4/T^3 + T^2/m^3 \rightarrow 0$.

Some discussion about our assumptions is in order. Condition *C1* is restrictive in the linearity it imposes but not otherwise. The condition implies that $E|x_t|^{8+\delta} < \infty$ for some $\delta > 0$. Condition *C2* deals with the behaviour of $f_x(\lambda)$. For frequencies $\lambda \rightarrow 0+$, they are the same used elsewhere by, say, Robinson (1995*a, b*) and thus, the same comments apply here, while for frequencies λ in any open set outside the origin, they are standard. It should be

noted that both models (2.3) and (2.5) satisfy these two conditions. Finally Condition *C3* bounds the rate of increase of m to infinity.

We finish the section describing the statistical properties of the estimator of the short memory component of the sequence given in (2.13) and (2.15) as well as those properties for the long memory parameter. First their statistical properties are invariant regardless of the value of the true unknown p_0 . This feature plays a relevant role on the properties of the estimator of p_0 , \hat{p} given in the next section. Furthermore, proceeding as with the proofs of Hidalgo and Yajima's (2002) Theorems 3 and 1 respectively, for any finite collection $\ell_1 < \ell_2 < \dots < \ell_q$,

$$\begin{aligned} T^{1/2} \left(\hat{\theta}_{\ell_1} - \theta_{\ell_1}, \dots, \hat{\theta}_{\ell_q} - \theta_{\ell_q} \right) &\xrightarrow{d} \mathcal{N}(0, V_{FARI}) \\ T^{1/2} \left(\hat{\zeta}_{\ell_1} - \zeta_{\ell_1}, \dots, \hat{\zeta}_{\ell_q} - \zeta_{\ell_q} \right) &\xrightarrow{d} \mathcal{N}(0, V_{FEXP}). \end{aligned}$$

On the other hand, $\hat{g}_x(\lambda; [\log T])$ provides a consistent estimator of $\Phi(e^{i\lambda}; \theta)$ or $\exp\{-\zeta_0 - \sum_{\ell=1}^{p_0} \zeta_\ell \cos(\ell\lambda)\}$, where respectively (2.17)

$$\hat{g}_x(\lambda; p) =: \left| 1 - \hat{\theta}_1 e^{i\lambda} - \dots - \hat{\theta}_p e^{ip\lambda} \right|^{-2} \quad \text{or} \quad =: \exp \left\{ -\hat{\zeta}_0 - \sum_{\ell=1}^p \hat{\zeta}_\ell \cos(\ell\lambda) \right\}.$$

This motivates the following (two-step) estimator of α . Indeed, in view of the previous arguments we may expect that $\hat{g}_x(\lambda; [\log T]) I_x(\lambda)$ will asymptotically behave as $\sigma_\varepsilon^2 |1 - e^{i\lambda}|^{-\alpha} / 2\pi$. So, we can estimate α using the log-periodogram estimator but employing all the Fourier frequencies λ_j , $1 \leq j \leq [T/2]$. More specifically, we estimate α by

$$(2.18) \quad \tilde{\alpha} = \sum_{k=1}^{[T/2]} \delta_k([T/2]) \log(\hat{g}_{x,k}([\log T]) I_{x,k}),$$

where $\delta_k([T/2])$ was defined in (2.12), and recall the notation $\hat{g}_x(\lambda_k; [\log T]) =: \hat{g}_{x,k}([\log T])$

We shall denote $\log_2 T =: \log(\log T)$. We then have the following result.

Theorem 1. *Assuming C1 to C3, $\tilde{\alpha} - \alpha = O_p(T^{-1/2} \log_2 T)$.*

Proof. The proof of this, or any other, theorem is confined in Section 5. \square

We now comment on the result of Theorem 1. Firstly, the rate of convergence is faster than the one achieved by Moulines and Soulier (1999) by a logarithmic factor and very close to the parametric rate except for a $\log_2 T$ factor. Secondly, Hidalgo and Yajima's (2002) Theorems 3 and 1 and Theorem 1 suggest that a step-wise algorithm can be implemented. That is, we can employ $\tilde{\alpha}$ instead of $\hat{\alpha}$ when computing $\hat{\zeta}_j$ and $\hat{\theta}_\ell$ in (2.13) and (2.15) respectively. Of course with the updated new estimates $\hat{\zeta}_j$ and $\hat{\theta}_\ell$, we can update the estimate of α given in (2.18) and so on. This is achieved by

updating our estimator of $g_{x,k}$, i.e. $\hat{g}_{x,k}([\log T])$. However, it is worth noticing that since the initial estimators of the parameters $\theta_\ell, \zeta_\ell, \ell = 1, \dots, p$, are $o_p(T^{-1/4})$ consistent, a two-step estimator will suffice. Finally, we mention that inspecting the proof of Theorem 1, and in particular expression (5.1), we easily observed that if instead of $\tilde{\alpha}$ given in (2.18) we compute

$$\check{\alpha} = \sum_{k=1}^{[T/2]} \delta_k([T/2]) \log(\hat{g}_{x,k}(p_0) I_{x,k}),$$

we then have that $T^{1/2}(\check{\alpha} - \alpha)$ converges in distribution to a normal random variable. The latter follows by Theorem 1, replacing $[\log T]$ by p_0 there.

3. ORDER SELECTION OF THE *FARI* AND *FEXP* MODELS

We now describe and examine the properties of the criterion function to estimate p_0 . The problem of order selection dates back to Akaike's (1974) information criterion *AIC* for the selection of the order p_0 of an *AR*(p_0) model, i.e. model (2.2) with $\alpha = 0$. Akaike suggested to select p_0 as the value \hat{p} which minimizes

$$AIC(p) = \log(\hat{\sigma}_{\varepsilon,p}^2) + \frac{2p}{T},$$

where $\hat{\sigma}_{\varepsilon,p}^2$ is an estimator of the variance of the innovation sequence ε_t in (2.2), after an *AR*(p) model was fitted to the data. That is, $\hat{\sigma}_{\varepsilon,p}^2 = T^{-1} \sum_{t=p+1}^T \hat{\varepsilon}_{t,p}^2$, where $\hat{\varepsilon}_{t,p}, t = p+1, \dots, T$, are the residuals. Alternative criteria have been provided by Mallow's (1973) C_p or Parzen's (1974) *CAT*. However, as was shown by Shibata (1976), these criteria overestimate the true order p_0 with a positive probability even as T increases to infinity.

Since it is always desirable to have consistent estimation procedures, some modifications to the *AIC* were introduced. One of them is Schwarz's (1978) *BIC* criterion, defined as the value \hat{p} which minimizes

$$(3.1) \quad BIC(p) = \log(\hat{\sigma}_{\varepsilon,p}^2) + \frac{p \log T}{T},$$

or Hannan and Quinn's (1979) *HIC*

$$(3.2) \quad HIC(p) = \log(\hat{\sigma}_{\varepsilon,p}^2) + \frac{cp \log_2 T}{T} \quad \text{with } c > 2.$$

We have already pointed out that a possible drawback in the implementation of *BIC*(p) and/or *HIC*(p) is the computation of the *CSS*, i.e. $\hat{\sigma}_{\varepsilon,p}^2$, and in particular the residuals $\hat{\varepsilon}_{t,p}$. Indeed, the residuals are based on

$$(3.3) \quad \hat{\varepsilon}_{t,p} = \sum_{j=0}^{t-1} a_j(\hat{\psi}) x_{t-j},$$

and with long memory dependence data it may not be a good approximation to $\tilde{\varepsilon}_{t,p} =: \sum_{j=0}^{\infty} a_j(\hat{\psi}) x_{t-j}$, or to show that the explicit truncation

used in (3.3) has not consequences is not trivial. See Robinson (2005) for discussion of the latter. Secondly as indicated in the introduction it may be very hazardous to obtain $\widehat{\varepsilon}_{t,p}$, as $a_j(\widehat{\psi})$ can only be computed by integral approximation in models such as Bloomfield's *FEXP*. So, it appears that the method might not be plausible or easy to compute.

So, due to the previous comments, we shall now describe a modification of the above criteria based on the *CSS*. To that end, consider the Whittle's objective function

$$W_T(p; \alpha) =: \frac{1}{[T/2]} \sum_{k=1}^{[T/2]} \frac{|1 - e^{i\lambda_k}|^\alpha I_{x,k}}{g_{x,k}(p)},$$

where similarly to $\widehat{g}_x(\lambda; p)$ given in (2.17), we define

$$g_x(\lambda; p) =: \left| 1 - \theta_1 e^{i\lambda} - \dots - \theta_p e^{ip\lambda} \right|^{-2} \quad \text{or} \quad =: \exp \left\{ -\zeta_0 - \sum_{\ell=1}^p \zeta_\ell \cos(\ell\lambda) \right\}.$$

We then estimate p_0 by

$$(3.4) \quad \widehat{p} = \arg \min_{0 \leq p \leq [\log T]} S_T(p),$$

where

$$(3.5) \quad \begin{aligned} S_T(p) &= : BIC(p; \widehat{\alpha}) = \widehat{W}_T(p; \widehat{\alpha}) + \frac{p \log T}{T} \\ \widehat{W}_T(p; \widehat{\alpha}) &= : \frac{1}{[T/2]} \sum_{k=1}^{[T/2]} \frac{|1 - e^{i\lambda_k}|^{\widehat{\alpha}} I_{x,k}}{\widehat{g}_{x,k}(p)}. \end{aligned}$$

Before we state the properties of \widehat{p} given in (3.4), we give the intuition of the criteria in (3.5). Suppose first that $p < p_0$. Because $E \{ |1 - e^{i\lambda}|^\alpha I_{x,j} - g_{x,j}(p_0) \} = o(1)$, we may expect then that

$$\begin{aligned} BIC(p) - BIC(p_0) &= W_T(p; \alpha) - W_T(p_0; \alpha) + \frac{(p - p_0) \log T}{T} \\ &\simeq \frac{\sigma_\varepsilon^2}{[T/2]} \sum_{k=1}^{[T/2]} \frac{g_{x,k}(p_0)}{g_{x,k}(p)} - \sigma_\varepsilon^2 + \frac{(p - p_0) \log T}{T}. \end{aligned}$$

On the other hand, see Brockwell and Davis (1991, p.377), we know that

$$\frac{\sigma_\varepsilon^2}{[T/2]} \sum_{k=1}^{[T/2]} \frac{g_{x,k}(p_0)}{g_{x,k}(p)} - \sigma_\varepsilon^2 > 0,$$

so that we may conclude that, as $T \rightarrow \infty$,

$$BIC(p) - BIC(p_0) > 0$$

and hence that $\hat{p} \geq p_0$. But when $p > p_0$, we have that

$$\begin{aligned} BIC(p) - BIC(p_0) &\simeq \frac{\sigma_\varepsilon^2}{[T/2]} \sum_{k=1}^{[T/2]} \frac{g_{x,k}(p_0)}{g_{x,k}(p)} - \sigma_\varepsilon^2 + \frac{(p - p_0) [\log T]}{T} \\ &= \frac{(p - p_0) [\log T]}{T} \end{aligned}$$

because in this case $[T/2]^{-1} \sum_{k=1}^{[T/2]} g_{x,k}(p_0) / g_{x,k}(p) = 1$. The latter suggests that $\arg \min BIC(p)$ cannot be achieved for a $p > p_0$ either. Thus, we should expect that $\hat{p} \rightarrow_P p_0$. This is formalized in the next theorem.

Theorem 2. *Assuming Conditions C1 to C3, $\hat{p} \rightarrow_P p_0$ for models (2.3) or (2.5).*

We now comment on the results of Theorem 2. First inspection of the proofs indicate that the theorem holds true if $\log T$ were replaced by $\log^{1+\delta} T$ for any $0 < \delta < 1$. Following Pötscher (1991), one consequence of Theorem 2 is the asymptotic properties of Whittle estimator are the same if p_0 is replaced by \hat{p} . In addition, as in An et al. (1982), we allow p to increase as $\log T$. This generalizes the often scenario where the true value of p_0 is assumed to be less than some (known) positive integer, say P . So, in this regard we relax some of the conditions in Beran et al. (1998) who assumed the upper bound P is known to the practitioner.

Finally, denote the Whittle's estimator as

$$\hat{\psi}(p) = \arg \min_{\psi=(\alpha, \beta)'} W_T(p) =: \frac{1}{[T/2]} \sum_{k=1}^{[T/2]} \frac{|1 - e^{i\lambda_k}|^\alpha I_{x,k}}{g_{x,k}(\beta, p)},$$

where we have adopted the notation β for either the parameters θ or ζ .

Theorem 3. *Assuming Conditions C1 to C3 the statistical properties of the Whittle estimate of the parameters are the same as if p_0 were known. That is, $\hat{\psi}(\hat{p}) - \hat{\psi}(p_0) = o_p(T^{-1/2})$.*

Proof. The proof follows in the line of Pötscher (1991), so it is omitted. \square

4. MONTE CARLO EXPERIMENT

To examine the finite sample performance of \hat{p} , a small Monte Carlo experiment was carried out. In our first set of experiments, we consider *FARI*(1) and *FEXP*(1) models with $\alpha = 0.2, 0.4, 0.6, 0.8$, and θ or $\zeta = -0.8, -0.5, -0.2, 0.2, 0.5, 0.6$, according to whether we consider the *FARI* or *FEXP* model. For each combination, we generated 1000 replications of series of length $T = 128, 256, 512$ using the circulant embedding algorithm of Chandna and Walden (2013) that generates observations from a process with prescribed spectral density.

For the estimation of the parameters θ or ζ (depending on the model under consideration), we choose three different values for $m = T/2M$. These

values are $m = 8, 16, 32$ for $T = 128$, $m = 16, 32, 64$ for $T = 256$ and $m = 32, 64, 128$ for $T = 512$. The choice reflects the theoretical requirement that m increases as a function of T , but maintains three different ratios T/m (16, 8 and 4) for all three sample sizes examined. This makes for a more meaningful comparison across various values of T and m .

The results are tabulated in Tables 1-3, covering $T = 128, 256$ and 512 , respectively. For the *FARI(1)* model, the two largest choices of m lead to perfect AR order selection for all sample sizes, but even the smallest value of m gives less than perfect results only in the $T = 128$ case, which is the smallest sample size. Nevertheless the worst performance only exhibits 2.23% incorrect order selection. Similarly, we only observe a sensitivity of the accuracy of order selection to the long memory parameter α for the smallest values of m , although the proportion of correct orders remains above 84% even in the worst cases. For the largest values of m , correct order selection is perfectly achieved regardless of the value of T, θ or α .

The *FEXP(1)* model is somewhat more sensitive to parameter values and sample sizes. For the smaller values of m and T , the proportion of correct order selection can be quite low. This is especially true when the long memory parameter α is large and is exacerbated by values of ζ with large magnitudes. The problem is mitigated by increasing the sample size T and m , as observed in Table 3 when $(T, m) = (512, 128)$ yields a correct order selection proportion in excess of 99.2% in all cases and perfect order selection in 19 out of 24 cases. Thus our simulations suggest that correct order selection in the *FEXP* model requires larger sample sizes as compared to the *FARI* model, especially in models with larger longer memory or autoregressive parameters. For both the *FARI* and *FEXP* cases, failure to detect the correct order in our simulations exclusively results in overfitting and never underfitting.

Note that for the *FEXP(1)* model, performance sometimes worsens with increasing α for the two smallest bandwidths, but this phenomenon is less common for the largest bandwidth. Since the bandwidths must increase with T , this feature is likely due to the smaller bandwidths not being sufficient for the purpose. In fact perusal of the tables reveals that such a phenomenon is observed even for the *FARI(1)* model, even though the proportion of correct order selection is much higher. Finally, note that even for the largest bandwidths in the *FEXP(1)* model, it is sometimes the case that results worsen as α increases but because the correct order detection rate is already very high these changes are quite negligible. This lends further support to the explanation that this feature arises because of too small a bandwidth, relative to sample size.

In our second set of simulations, we again use the algorithm of Chandna and Walden (2013) to simulate observations from a *FARI(2)* process, choosing the same values for α , T and m as above and experimenting with $\theta^1 = (0.4, 0.4)$, $\theta^2 = (0.3, 0.5)$ and $\theta^3 = (0.5, 0.3)$. The results are reported in Table 4. Order selection is perfect for θ^1 regardless of the values

of α, T or m . On the other hand, for θ^2 and θ^3 the smallest value of m (for each respective value of T), leads to an almost complete failure to detect the true order. The problem is resolved with larger values of m , however, with perfect order selection being achieved. As in the *FARI(1)* case, our order detection criterion never underfits the model.

<i>FARI</i> (1)							
	θ	-0.8	-0.5	-0.2	0.2	0.5	0.6
	α						
$m = 8$	0.2	0.9760	0.9830	0.9880	0.9930	0.9000	0.9920
	0.4	0.9590	0.9850	0.9870	0.9770	0.8780	0.9850
	0.6	0.9530	0.9640	0.9830	0.9490	0.8490	0.9930
	0.8	0.9280	0.9380	0.9630	0.9400	0.8410	0.9770
$m = 16$	0.2	0.9890	0.9870	0.9930	0.9990	0.9680	1.0000
	0.4	0.9810	0.9810	0.9860	0.9950	0.9700	1.0000
	0.6	0.9630	0.9650	0.9850	0.9720	0.9630	1.0000
	0.8	0.9120	0.9440	0.9880	0.9560	0.9790	1.0000
$m = 32$	0.2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
<i>FEXP</i> (1)							
	ζ	-0.8	-0.5	-0.2	0.2	0.5	0.6
	α						
$m = 8$	0.2	0.7150	0.8670	0.9630	0.9910	0.9990	0.9990
	0.4	0.5350	0.6470	0.7780	0.9350	0.9840	0.9960
	0.6	0.4610	0.4320	0.5020	0.7110	0.8610	0.9100
	0.8	0.3950	0.3450	0.3430	0.3630	0.5180	0.5780
$m = 16$	0.2	0.7070	0.8540	0.9740	0.9990	1.0000	1.0000
	0.4	0.6650	0.6640	0.7620	0.9810	0.9980	0.9980
	0.6	0.6770	0.6230	0.6120	0.6750	0.8790	0.9330
	0.8	0.6700	0.6740	0.6110	0.5240	0.5700	0.5990
$m = 32$	0.2	0.9760	0.9630	0.9650	0.9980	1.0000	1.0000
	0.4	1.0000	0.9850	0.9700	0.9860	0.9980	1.0000
	0.6	1.0000	0.9960	0.9960	0.9830	0.9950	0.9910
	0.8	1.0000	0.9990	1.0000	0.9980	0.9940	0.9920

TABLE 1. Empirical proportion of correct order selection with 1000 replications and $T = 128$.

<i>FARI</i> (1)							
	θ	-0.8	-0.5	-0.2	0.2	0.5	0.6
	α						
$m = 16$	0.2	0.9950	0.9960	1.0000	0.9950	0.9490	1.0000
	0.4	0.9830	0.9940	0.9970	0.9900	0.9390	1.0000
	0.6	0.9710	0.9750	0.9930	0.9530	0.9280	1.0000
	0.8	0.9310	0.9730	0.9810	0.9480	0.9410	1.0000
$m = 32$	0.2	0.9990	0.9980	0.9990	1.0000	0.9930	1.0000
	0.4	0.9940	0.9920	0.9980	0.9990	0.9930	1.0000
	0.6	0.9870	0.9830	0.9920	0.9600	0.9980	1.0000
	0.8	0.9420	0.9680	0.9960	0.9430	0.9990	1.0000
$m = 64$	0.2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
<i>FEXP</i> (1)							
	ζ	-0.8	-0.5	-0.2	0.2	0.5	0.6
	α						
$m = 16$	0.2	0.7010	0.8270	0.9590	0.9930	0.9990	1.0000
	0.4	0.5360	0.5500	0.6860	0.9380	0.9870	0.9970
	0.6	0.4560	0.4120	0.4240	0.5410	0.7940	0.8490
	0.8	0.3760	0.3130	0.2810	0.2780	0.2980	0.3240
$m = 32$	0.2	0.7410	0.8060	0.9640	1.0000	1.0000	1.0000
	0.4	0.7380	0.6890	0.7150	0.9390	0.9970	0.9990
	0.6	0.7480	0.6840	0.6470	0.5800	0.7520	0.8480
	0.8	0.7590	0.7170	0.6250	0.4390	0.4280	0.3990
$m = 64$	0.2	1.0000	0.9900	0.9740	1.0000	1.0000	1.0000
	0.4	1.0000	0.9990	0.9950	0.9940	0.9990	1.0000
	0.6	1.0000	1.0000	1.0000	1.0000	0.9990	0.9980
	0.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLE 2. Empirical proportion of correct order selection with 1000 replications and $T = 256$.

<i>FARI</i> (1)							
	θ	-0.8	-0.5	-0.2	0.2	0.5	0.6
	α						
$m = 32$	0.2	1.0000	1.0000	0.9980	1.0000	0.9810	1.0000
	0.4	0.9990	0.9990	0.9980	0.9880	0.9720	1.0000
	0.6	0.9940	0.9880	0.9990	0.9310	0.9760	1.0000
	0.8	0.9710	0.9750	0.9940	0.9570	0.9950	1.0000
$m = 64$	0.2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.4	1.0000	1.0000	0.9990	0.9990	1.0000	1.0000
	0.6	1.0000	0.9990	0.9980	0.9420	1.0000	1.0000
	0.8	0.9860	0.9780	0.9960	0.9700	1.0000	1.0000
$m = 128$	0.2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
<i>FEXP</i> (1)							
	ζ	-0.8	-0.5	-0.2	0.2	0.5	0.6
	α						
$m = 32$	0.2	0.6900	0.7700	0.9470	1.0000	1.0000	1.0000
	0.4	0.5540	0.4890	0.5740	0.8800	0.9950	0.9970
	0.6	0.4470	0.3900	0.3350	0.3500	0.5950	0.7110
	0.8	0.2840	0.2240	0.1880	0.1490	0.1080	0.1110
$m = 64$	0.2	0.8210	0.7950	0.9330	1.0000	1.0000	1.0000
	0.4	0.8200	0.7640	0.7250	0.8560	0.9980	1.0000
	0.6	0.8430	0.7630	0.6490	0.5010	0.5430	0.6350
	0.8	0.8470	0.7900	0.5960	0.4090	0.3020	0.2650
$m = 128$	0.2	1.0000	1.0000	0.9920	1.0000	0.9990	0.9990
	0.4	1.0000	1.0000	0.9990	0.9990	1.0000	1.0000
	0.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLE 3. Empirical proportion of correct order selection with 1000 replications and $T = 512$.

α	$T = 128$			$T = 256$			$T = 512$					
	m	θ^1	θ^2	θ^3	m	θ^1	θ^2	θ^3	m	θ^1	θ^2	θ^3
0.2	8	1.0000	0.0260	0.0070	16	1.0000	0.0110	0.0010	32	1.0000	0.0000	0.0000
0.4		1.0000	0.0440	0.0000		1.0000	0.0380	0.0010		1.0000	0.0030	0.0000
0.6		1.0000	0.0690	0.0020		1.0000	0.0570	0.0010		1.0000	0.0150	0.0000
0.8		1.0000	0.0960	0.0010		1.0000	0.0800	0.0000		1.0000	0.0230	0.0000
0.2	16	1.0000	1.0000	1.0000	32	1.0000	1.0000	1.0000	64	1.0000	1.0000	1.0000
0.4		1.0000	1.0000	1.0000		1.0000	1.0000	1.0000		1.0000	1.0000	1.0000
0.6		1.0000	1.0000	1.0000		1.0000	1.0000	1.0000		1.0000	1.0000	1.0000
0.8		1.0000	1.0000	1.0000		1.0000	1.0000	1.0000		1.0000	1.0000	1.0000
0.2	32	1.0000	1.0000	1.0000	64	1.0000	1.0000	1.0000	128	1.0000	1.0000	1.0000
0.4		1.0000	1.0000	1.0000		1.0000	1.0000	1.0000		1.0000	1.0000	1.0000
0.6		1.0000	1.0000	1.0000		1.0000	1.0000	1.0000		1.0000	1.0000	1.0000
0.8		1.0000	1.0000	1.0000		1.0000	1.0000	1.0000		1.0000	1.0000	1.0000

TABLE 4. Empirical proportion of correct order selection with 1000 replications, $FARI(2)$ model, $\theta^1 = (0.4, 0.4), \theta^2 = (0.5, 0.3), \theta^3 = (0.3, 0.5)$.

5. PROOFS

5.1. Proof of Theorem 1.

From the definition of $\tilde{\alpha}$ in (2.18) and abbreviating $\delta_k([T/2])$ by δ_k , we have that

$$(5.1) \quad \tilde{\alpha} - \alpha = \left\{ \sum_{k=1}^{[T/2]} \delta_k \log(g_{x,k}(p_0) I_{x,k}) - \alpha \right\} + \sum_{k=1}^{[T/2]} \delta_k \log(g_{x,k}^{-1}(p_0) \hat{g}_{x,k}(\hat{p})).$$

The first term on the right of (5.1) is $O_p(T^{-1/2})$ proceeding as in the proof of Robinson's (1995a) Theorem 2, after we observe that it is

$$\sum_{k=1}^{[T/2]} \delta_k \log(I_{x,k}/f_{x,k})$$

and when using Robinson's (1995a) Theorem 2 we notice that $I_{x,k}$ is normalized by the true $f_x(\lambda)$ instead of its approximation given in Condition C2.

So to complete the proof it suffices to show that the second term on the right of (5.1) is $O_p(T^{-1/2} \log_2 T)$. We shall only consider the *FARI* model, the proof for the *FEXP* is similarly handled if it is not easier. To that end and using the convention that $\theta_p = 0$ if $p > p_0$, we have that

$$(5.2) \quad \left| 1 - \hat{\theta}_1 e^{i\lambda_j} - \dots - \hat{\theta}_{[\log T]} e^{i[\log T]\lambda_j} \right|^2 - \left| 1 - \theta_1 e^{i\lambda_j} - \dots - \theta_{p_0} e^{ip_0\lambda_j} \right|^2 \\ = \sum_{p,q=1}^{[\log T]} (\hat{\theta}_p - \theta_p) (\hat{\theta}_q - \theta_q) e^{i(p-q)\lambda_j} \\ + 2 \sum_{p=1}^{[\log T]} (\hat{\theta}_p - \theta_p) \operatorname{Re} \left(\left(1 - \theta_1 e^{i\lambda_j} - \dots - \theta_{p_0} e^{ip_0\lambda_j} \right) e^{p\lambda_j} \right),$$

where $\operatorname{Re}(z)$ denotes the real part of the complex number z . Now, proceeding as in the proof of Hidalgo and Yajima's (2002) Theorem 3 but using a Taylor's expansion up to its third term in their expression (51), we obtain that

$$(5.3) \quad \hat{\theta}_p - \theta_p = \frac{1}{M} \sum_{q=1}^M \varkappa_{n,q} \left(b_{|q-p|_+} + \frac{1}{M} \right) \\ + \frac{1}{M^2} \sum_{q_1, q_2=1}^M \varkappa_{n,q_1} \varkappa_{n,q_2} \left(b_{|p-q_1+q_2|_+} + \frac{1}{M} \right) + O_p(m^{-3/2}),$$

where $E\varkappa_{n,q}^2 < \infty$, after we proceed as in expression (52) in Hidalgo and Yajima (2002). (5.3) implies that $(\hat{\theta}_p - \theta_p) = O_p(T^{-1/2})$ for any $1 \leq p \leq$

$[\log T]$, and hence Taylor's expansion of $\log z$ around $z = 1$ yields that the second term on the right of (5.1) is

$$(5.4) \quad 2 \sum_{p=1}^{[\log T]} \left(\widehat{\theta}_p - \theta_p \right) \sum_{k=1}^{[T/2]} \delta_k \psi_k + O_p \left(T^{-1} \log^2 T \right),$$

where

$$\psi(\lambda) = \frac{\operatorname{Re} \left((1 - \theta_1 e^{i\lambda} - \dots - \theta_{p_0} e^{ip_0 \lambda}) e^{p\lambda} \right)}{|1 - \theta_1 e^{i\lambda} - \dots - \theta_{p_0} e^{ip_0 \lambda}|^2}.$$

Next, by definition of δ_k , we have that (5.4) is, except multiplicative constants,

$$(5.5) \quad \begin{aligned} & \sum_{p=1}^{[\log T]} \left(\widehat{\theta}_p - \theta_p \right) \frac{2}{T} \sum_{k=1}^{[T/2]} \psi_k \log |1 - e^{i\lambda_k}| + O_p \left(T^{-1} \log^2 T \right) \\ &= 2 \sum_{p=1}^{[\log T]} \left(\widehat{\theta}_p - \theta_p \right) \int_0^\pi \psi(\lambda) \log |1 - e^{i\lambda}| d\lambda + O_p \left(T^{-1} \log^2 T \right) \end{aligned}$$

using Brillinger's (1981, p.15) and Robinson's (1995b) Lemma 2 because $\psi(\lambda)$ is a continuous differentiable function. However, because

$$\int_0^\pi \frac{\log |1 - e^{i\lambda}|}{|1 - \theta_1 e^{i\lambda} - \dots - \theta_{p_0} e^{ip_0 \lambda}|^2} \cos(p\lambda) d\lambda = O(p^{-1}),$$

in view of (5.3), the right side of (5.5) is bounded in absolute value by

$$K \sum_{p=1}^{[\log T]} \frac{\widehat{\theta}_p - \theta_p}{p} + O_p \left(T^{-1} \log^2 T \right) = O_p \left(T^{-1/2} \log_2 T \right)$$

for some finite and positive K . This completes the proof of the theorem. ■

5.2. Proof of Theorem 2.

First note that because $\widehat{p} = \arg \min_{0 \leq p \leq [\log T] - p_0} S_T(p)$, then $S_T(\widehat{p}) - S_T(p_0) < 0$. We begin with the proof for model (2.5). We first show that $\Pr \{\widehat{p} > p_0\} \rightarrow 0$. Denote $\widehat{p} = p_0 + j$ for some $1 \leq j \leq [\log T] - p_0$. Using the inequality

$$(5.6) \quad \Pr \{\widehat{p} > p_0\} \leq \Pr \left\{ \sup_{1 \leq j \leq [\log T] - p_0} S_T(p_0) - S_T(p_0 + j) > 0 \right\},$$

it suffices to show that the right side of (5.6) converges to zero.

For any $0 \leq r < s \leq [\log T]$, denote

$$\widehat{h}_q(r, s) = \sum_{\ell=r}^{s-1} \widehat{\zeta}_\ell \cos(\ell \lambda_q) \quad \text{and} \quad h_q(r, s) = \sum_{\ell=r}^{s-1} \zeta_\ell \cos(\ell \lambda_q).$$

Then, writing $p = p_0 + j$, that $S_T(p_0) - S_T(p)$ is

$$(5.7) \quad \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left| 1 - e^{i\lambda_q} \right|^{\tilde{\alpha}} \exp \left\{ \widehat{h}_q(1, p_0) \right\} \left(1 - \exp \left\{ \widehat{h}_q(p_0 + 1, p_0 + j) \right\} \right) - \frac{j \lceil \log T \rceil}{T}.$$

Denoting $1 - \exp \left\{ \widehat{h}_q(p_0 + 1, p_0 + j) \right\} =: b_q(j)$, by standard algebra, the first term of (5.7) is

$$(5.8) \quad \begin{aligned} & \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left| 1 - e^{i\lambda_q} \right|^{\alpha} \exp \left\{ h_q(1, p_0) \right\} \left(\exp \left\{ \widehat{h}_q(1, p_0) - h_q(1, p_0) \right\} - 1 \right) b_q(j) \\ & + \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left\{ \left| 1 - e^{i\lambda_q} \right|^{\tilde{\alpha}} - \left| 1 - e^{i\lambda_q} \right|^{\alpha} \right\} \exp \left\{ h_q(1, p_0) \right\} b_q(j) \\ & + \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left\{ \left| 1 - e^{i\lambda_q} \right|^{\tilde{\alpha}} - \left| 1 - e^{i\lambda_q} \right|^{\alpha} \right\} \left(\exp \left\{ \widehat{h}_q(1, p_0) \right\} - \exp \left\{ h_q(1, p_0) \right\} \right) b_q(j) \\ & + \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left| 1 - e^{i\lambda_q} \right|^{\alpha} \exp \left\{ h_q(1, p_0) \right\} b_q(j). \end{aligned}$$

We begin examining the first term of (5.8). Because by Hidalgo and Yajima (2002), $\widehat{\zeta}_\ell - \zeta_\ell = O_p(T^{-1/2})$, using the convention that $\zeta_\ell = 0$ for $\ell > p_0$, Taylor's expansion of $\log z$ around 1 yields that the the first term of (5.8) is

$$(5.9) \quad \sum_{k=1}^{p_0} \left(\widehat{\zeta}_k - \zeta_k \right) \sum_{\ell=p_0+1}^{p_0+j} \widehat{\zeta}_\ell \frac{1}{T} \sum_{q=1}^{T-1} \frac{I_{x,q} \cos(\ell\lambda_q) \cos(k\lambda_q)}{\left| 1 - e^{i\lambda_q} \right|^{-\alpha} \exp \left\{ -h_q(1, p_0) \right\}} + O_p \left(\frac{\log T}{T^{3/2}} \right).$$

Now,

$$\begin{aligned} \frac{1}{T} \sum_{q=1}^{T-1} \frac{I_{x,q}}{f_{x,q}} \psi_q &= \frac{1}{T} \sum_{q=1}^{T-1} \left(\frac{I_{x,q}}{f_{x,q}} - 2\pi I_{\varepsilon,q} \right) \psi_q + \frac{2\pi}{T} \sum_{q=1}^{T-1} I_{\varepsilon,q} \psi_q \\ &= O_p \left(T^{-2/3} \log^{2/3} T \right) + \frac{2\pi}{T} \sum_{q=1}^{T-1} I_{\varepsilon,q} \psi_q \end{aligned}$$

proceeding as in the proof of (4.8) in Robinson (1995b), and that $I_{x,q}$ is normalized by $f_{x,q}$ instead of its approximation in Condition C2. On the other hand for any continuous differentiable function $\psi(\lambda)$, the second term satisfies that

$$(5.10) \quad \left| \frac{2\pi}{T} \sum_{q=1}^{T-1} I_{\varepsilon,q} \psi_q - \sigma_\varepsilon^2 \int_0^{2\pi} \psi(\lambda) d\lambda \right| = O \left(T^{-1/2} \right)$$

using Brillinger (1980, p.15). So, as $f_{x,q} = |1 - e^{i\lambda q}|^{-\alpha} \exp\{-\zeta_0 - h_q(1, p_0)\}$, it yields that (5.9) is

$$e^{-\zeta_0} \sum_{k=1}^{p_0} (\widehat{\zeta}_k - \zeta_k) \sum_{\ell=p_0+1}^{p_0+j} \widehat{\zeta}_\ell \int_0^{2\pi} \cos(\ell\lambda) \cos(k\lambda) d\lambda + O_p\left(\frac{\log T}{T^{3/2}}\right) = O_p\left(\frac{\log T}{T^{3/2}}\right)$$

because $\int_0^{2\pi} \cos(k\lambda) \cos(\ell\lambda) d\lambda = 0$ if $k \neq \ell$.

Next we examine the second term of (5.8), which proceeding as with the first term and Theorem 1, it is

$$\begin{aligned} & e^{-\zeta_0} (\widetilde{\alpha} - \alpha) \sum_{\ell=p_0+1}^{p_0+j} \widehat{\zeta}_\ell \int_0^{2\pi} \cos(\ell\lambda) \log|1 - e^{i\lambda}| d\lambda + O_p\left(\frac{\log^2 T}{T^{3/2}}\right) \\ &= (\widetilde{\alpha} - \alpha) \sum_{\ell=p_0+1}^{p_0+j} \widehat{\zeta}_\ell \phi_\ell + O_p\left(\frac{\log^2 T}{T^{3/2}}\right), \end{aligned}$$

where $\phi_\ell = O(\ell^{-1})$ is the Fourier coefficient of $e^{-\zeta_0} \log|1 - e^{i\lambda q}|$.

The third term of (5.8) is easily seen to be $O_p(T^{-3/2} \log^2 T)$ using Taylor's expansion as was done for the first two terms. Finally the fourth term of (5.8), using the inequality $|e^z - 1 - z - z^2/2| \leq Kz^3$, is

$$\begin{aligned} & \sum_{\ell=p_0+1}^{p_0+j} \widehat{\zeta}_\ell \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} |1 - e^{i\lambda q}|^\alpha \exp\{h_q(1, p_0)\} \cos(\ell\lambda_q) \\ &+ \frac{1}{2T} \sum_{q=1}^{T-1} I_{x,q} |1 - e^{i\lambda q}|^\alpha \exp\{h_q(1, p_0)\} \left(\sum_{\ell=p_0+1}^{p_0+j} \widehat{\zeta}_\ell \cos(\ell\lambda_q) \right)^2 \\ &+ O_p\left(\frac{\log^2 T}{T^{3/2}}\right). \end{aligned}$$

Now using (5.10), the first term of the last displayed expression is

$$\sum_{\ell=p_0+1}^{p_0+j} \widehat{\zeta}_\ell \int_0^{2\pi} \cos(\ell\lambda) d\lambda + O_p\left(\frac{\log^2 T}{T^{3/2}}\right) = O_p\left(\frac{\log^2 T}{T^{3/2}}\right)$$

because $\int_0^{2\pi} \cos(\ell\lambda) d\lambda = 0$ if $\ell \neq 0$, whereas the same argument yields that the second term is

$$\sum_{\ell=p_0+1}^{p_0+j} \widehat{\zeta}_\ell^2 \int_0^{2\pi} \cos^2(\ell\lambda) d\lambda = K \sum_{\ell=p_0+1}^{p_0+j} \widehat{\zeta}_\ell^2.$$

So, we conclude that the left side of (5.7), that is $S_T(p_0; 1) - S_T(p; 1)$, is

$$K \sum_{\ell=p_0+1}^{p_0+j} \widehat{\zeta}_\ell^2 + (\widetilde{\alpha} - \alpha) \sum_{\ell=p_0+1}^{p_0+j} \widehat{\zeta}_\ell \phi_\ell - \frac{j[\log T]}{T} + O_p\left(\frac{\log^2 T}{T^{3/2}}\right),$$

and hence the right side of (5.6) is bounded by

$$(5.11) \quad \Pr \left\{ \sup_{1 \leq j \leq [\log T] - p_0} K \sum_{\ell=p_0+1}^{p_0+j} \widehat{\zeta}_\ell^2 + o_p(\log T) - j [\log T] > 0 \right\} \\ + \Pr \left\{ \sup_{1 \leq j \leq [\log T] - p_0} \left(T (\tilde{\alpha} - \alpha) \sum_{\ell=p_0+1}^{p_0+j} \widehat{\zeta}_\ell \phi_\ell + o_p(\log T) \right) - j [\log T] > 0 \right\}.$$

So, to complete the proof that $\Pr \{\widehat{p} > p_0\} = o(1)$ it suffices to show that (5.11) converges to zero. Proceeding as with the proof of Hidalgo and Yajima's (2002) Theorem 1 and standard algebra implies that for any $\ell \geq 1$,

$$T^{1/2} \widehat{\zeta}_\ell = \frac{1}{T^{1/2}} \sum_{s=1}^{[T/2]} \left(\frac{2\pi I_{\varepsilon,s}}{\sigma_\varepsilon^2} - 1 \right) \cos(\ell \lambda_s) + O_p \left(\frac{T^{1/2}}{m} + \frac{m^{1/2}}{T^{1/2}} \right) \\ = \frac{1}{T^{1/2}} \sum_{t=2}^T \varepsilon_t \varepsilon_{t-\ell} + O_p \left(\frac{T^{1/2}}{m} + \frac{m^{1/2}}{T^{1/2}} \right),$$

so that the first term of (5.11) is bounded by

$$(5.12) \quad \Pr \left\{ \sup_{1+p_0 \leq \ell \leq [\log T]} \left| \frac{1}{T^{1/2}} \sum_{t=2}^T \varepsilon_t \varepsilon_{t-\ell} \right| > [\log^{1/2} T] \right\}.$$

Now because by Condition $C\mathcal{Q}$, $\varepsilon_t \varepsilon_{t-\ell}$ is a martingale difference, using Markov inequality and that $\sup_\ell |a_\ell| \leq \left(\sum_\ell |a_\ell|^4 \right)^{1/4}$, we obtain that (5.12) is bounded by

$$\frac{1}{[\log T]^2} \sum_{\ell=1}^{[\log T]} E \left| \frac{1}{T^{1/2}} \sum_{t=2}^T \varepsilon_t \varepsilon_{t-\ell} \right|^4 = O([\log T]^{-1}).$$

Next, the second term of (5.11) is bounded by

$$\Pr \left\{ \frac{T^{1/2} (|\tilde{\alpha} - \alpha|)}{[\log T]^{1/2}} > K \right\} + \Pr \left\{ K \sup_{1+p_0 \leq \ell \leq [\log T]} T^{1/2} \widehat{\zeta}_\ell > [\log T]^{1/2} \right\}.$$

The first term of the last displayed expression converges to zero by Theorem 1, whereas the second term is that already examined in (5.12). This concludes the proof that $\Pr \{\widehat{p} > p_0\} = o(1)$.

To complete the proof for model (2.5), it only remains to show that

$$\Pr \left\{ \sup_{0 \leq p < p_0} (S_T(p_0) - S_T(p)) > 0 \right\} \rightarrow 0.$$

Because p_0 is finite, it suffices to show that for any $0 \leq p < p_0$,

$$(5.13) \quad \Pr \{S_T(p_0) - S_T(p) > 0\} \rightarrow 0.$$

To that end, we first notice that

$$\begin{aligned}
S_T(p) &= \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left| 1 - e^{i\lambda_q} \right|^{\tilde{\alpha}} \exp \left\{ \widehat{h}_q(1,p) \right\} + \frac{p \lfloor \log T \rfloor}{T} \\
&= \left\{ \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left| 1 - e^{i\lambda_q} \right|^{\alpha} \exp \{ h_q(1,p) \} + \frac{p \lfloor \log T \rfloor}{T} \right\} \\
&\quad + \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left\{ \left| 1 - e^{i\lambda_q} \right|^{\tilde{\alpha}} - \left| 1 - e^{i\lambda_q} \right|^{\alpha} \right\} \left(\exp \left\{ \widehat{h}_q(1,p) \right\} - h_q(1,p) \right) \\
(5.14) \quad &+ \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left| 1 - e^{i\lambda_q} \right|^{\alpha} \left(\exp \left\{ \widehat{h}_q(1,p) \right\} - h_q(1,p) \right) \\
&\quad + \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left\{ \left| 1 - e^{i\lambda_q} \right|^{\tilde{\alpha}} - \left| 1 - e^{i\lambda_q} \right|^{\alpha} \right\} h_q(1,p).
\end{aligned}$$

The last three terms on the right of (5.14) are $o_p(1)$. Indeed, proceeding as with the second term of (5.8), Taylor's expansion and Theorem 1, we conclude that the last term on the right of (5.14) is

$$\begin{aligned}
&(\tilde{\alpha} - \alpha) \int_0^{2\pi} f_x(\lambda) \left| 1 - e^{i\lambda} \right|^{\alpha} \exp \left\{ \sum_{\ell=1}^p \zeta_\ell \cos(\ell\lambda) \right\} \log \left| 1 - e^{i\lambda} \right| d\lambda + O_p \left(\frac{\log_2 T}{T} \right) \\
&= (\tilde{\alpha} - \alpha) \int_0^{2\pi} \exp \left\{ - \sum_{\ell=p+1}^{p_0} \zeta_\ell \cos(\ell\lambda) \right\} \log \left| 1 - e^{i\lambda} \right| d\lambda + O_p \left(\frac{\log_2 T}{T} \right) \\
&= O_p \left(\frac{\log_2 T}{T^{1/2}} \right).
\end{aligned}$$

Next using Hidalgo and Yajima (2002), cf. (5.3), Taylor's expansion yields that the third term on the right of (5.14) is

$$\begin{aligned}
&\sum_{\ell=1}^p (\widehat{\zeta}_\ell - \zeta_\ell) \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left| 1 - e^{i\lambda_q} \right|^{\alpha} \exp \{ h_q(1,p) \} + O_p(T^{-1}) \\
&= \sum_{\ell=1}^p (\widehat{\zeta}_\ell - \zeta_\ell) \int_0^{2\pi} f_x(\lambda) \left| 1 - e^{i\lambda} \right|^{\alpha} \exp \left\{ \sum_{\ell=1}^p \zeta_\ell \cos(\ell\lambda) \right\} d\lambda + O_p(T^{-1}) \\
&= \sum_{\ell=1}^p (\widehat{\zeta}_\ell - \zeta_\ell) \int_0^{2\pi} \exp \left\{ - \sum_{\ell=p+1}^{p_0} \zeta_\ell \cos(\ell\lambda) \right\} d\lambda + O_p(T^{-1}) \\
&= O_p(T^{-1/2})
\end{aligned}$$

using (5.10). The second term of (5.14) is $o_p(1)$ using the behaviour of the third and fourth terms of (5.14) together with Cauchy-Schwarz inequality.

Finally, the first term on the right of (5.14) is

$$\int_0^{2\pi} f_x(\lambda) \left| 1 - e^{i\lambda} \right|^\alpha \exp \left\{ \sum_{\ell=1}^p \zeta_\ell \cos(\ell\lambda) \right\} d\lambda + \frac{p \lfloor \log T \rfloor}{T} + O_p(T^{-1})$$

using (5.10). So, we conclude that

$$S_T(p) = e^{-\zeta_0} \int_0^{2\pi} \exp \left\{ \sum_{\ell=p+1}^{p_0} \zeta_\ell \cos(\ell\lambda) \right\} d\lambda + \frac{p \lfloor \log T \rfloor}{T} + o_p(1)$$

and hence standard algebra yields that

$$S_T(p_0) - S_T(p) = 2\pi e^{-\zeta_0} \left(1 - \frac{1}{2\pi} \int_0^{2\pi} \exp \left\{ \sum_{\ell=p+1}^{p_0} \zeta_\ell \cos(\ell\lambda) \right\} d\lambda \right) + o_p(1).$$

But the first term on the right of the last displayed expression is negative since

$$\frac{1}{2\pi} \int_0^{2\pi} \exp \left\{ \sum_{\ell=p+1}^{p_0} \zeta_\ell \cos(\ell\lambda) \right\} d\lambda > 1$$

proceeding as in Brockwell and Davis (1991, p.377). From here we conclude the proof of (5.13) and hence part **(a)**.

Now we examine model **(2.3)**. Proceeding as with the proof of model **(2.5)**, we first show that $\Pr\{\hat{p} > p_0\} \rightarrow 0$. Denote $\hat{p} = p_0 + j$ for some $1 \leq j \leq \lfloor \log T \rfloor - p_0$. Using the inequality (5.6), it suffices to show that

$$(5.15) \quad \Pr \left\{ \sup_{1 \leq j \leq \lfloor \log T \rfloor - p_0} S_T(p_0) - S_T(p_0 + j) > 0 \right\} \rightarrow 0.$$

To that end, we first notice that

$$(5.16) \quad \begin{aligned} & S_T(p_0) - S_T(p_0 + j) \\ &= \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left| 1 - e^{i\lambda_q} \right|^{\tilde{\alpha}} \left(\left| \Phi \left(e^{i\lambda_q}; \hat{\theta}(p_0) \right) \right|^2 - \left| \Phi \left(e^{i\lambda_q}; \hat{\theta}(p_0 + j) \right) \right|^2 \right) \\ & \quad - \frac{j \lfloor \log T \rfloor}{T}, \end{aligned}$$

where $\hat{\theta}(p) = (\hat{\theta}_1, \dots, \hat{\theta}_p)'$ and $\theta(p) = (\theta_1, \dots, \theta_p)'$. We shall first examine the behaviour of

$$(5.17) \quad \tilde{S}_T(p_0 + j) = \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left| 1 - e^{i\lambda_q} \right|^{\tilde{\alpha}} \left(\left| \Phi \left(e^{i\lambda_q}; \theta(p_0) \right) \right|^2 - \left| \Phi \left(e^{i\lambda_q}; \hat{\theta}(p_0 + j) \right) \right|^2 \right).$$

Using (5.2), $\tilde{S}_T(p_0 + j)$ is

$$\begin{aligned}
& \sum_{p_1, p_2=1}^{p_0+j} (\hat{\theta}_{p_1} - \theta_{p_1}) (\hat{\theta}_{p_2} - \theta_{p_2}) \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left(\left| 1 - e^{i\lambda_q} \right|^{\tilde{\alpha}} - \left| 1 - e^{i\lambda_q} \right|^{\alpha} \right) e^{i(p_1-p_2)\lambda_q} \\
(5.18) \quad & + \sum_{p_1, p_2=1}^{p_0+j} (\hat{\theta}_{p_1} - \theta_{p_1}) (\hat{\theta}_{p_2} - \theta_{p_2}) \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left| 1 - e^{i\lambda_q} \right|^{\alpha} e^{i(p_1-p_2)\lambda_q} \\
& + 2 \sum_{p=1}^{p_0+j} (\hat{\theta}_p - \theta_p) \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left(\left| 1 - e^{i\lambda_q} \right|^{\tilde{\alpha}} - \left| 1 - e^{i\lambda_q} \right|^{\alpha} \right) \operatorname{Re} \left(\Phi \left(e^{i\lambda_q}; \theta(p_0) \right) e^{ip\lambda_q} \right) \\
& + 2 \sum_{p=1}^{p_0+j} (\hat{\theta}_p - \theta_p) \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left| 1 - e^{i\lambda_q} \right|^{\alpha} \operatorname{Re} \left(\Phi \left(e^{i\lambda_q}; \theta(p_0) \right) e^{ip\lambda_q} \right).
\end{aligned}$$

Because $\hat{\theta}_p - \theta_p = O_p(T^{-1/2})$ by Hidalgo and Yajima (2002), proceeding as with the second term of (5.8), the first term of (5.18) is $O_p(T^{-3/2} \log^2 T)$. The second term of (5.18) is

$$\begin{aligned}
& \sum_{p_1, p_2=1}^{p_0+j} (\hat{\theta}_{p_1} - \theta_{p_1}) (\hat{\theta}_{p_2} - \theta_{p_2}) \int_0^{2\pi} \left| \Phi \left(e^{i\lambda}; \theta(p_0) \right) \right|^{-2} e^{i(p_1-p_2)\lambda} d\lambda + O_p \left(\frac{\log^2 T}{T^{3/2}} \right) \\
= & \sum_{p_1, p_2=1}^{p_0+j} (\hat{\theta}_{p_1} - \theta_{p_1}) (\hat{\theta}_{p_2} - \theta_{p_2}) \gamma_{p_1-p_2} + O_p \left(\frac{\log^2 T}{T^{3/2}} \right)
\end{aligned}$$

using (5.10) and where γ_p denotes the autocovariance function of an $AR(p_0)$ process with spectral density function $|\Phi(e^{i\lambda}; \theta(p_0))|^{-2}/2\pi$.

Next, the third term of (5.18) is

$$\begin{aligned}
& 2(\tilde{\alpha} - \alpha) \sum_{p=1}^{p_0+j} (\hat{\theta}_p - \theta_p) \sum_{\ell=0}^{p_0} \theta_\ell \frac{1}{T} \sum_{q=1}^{T-1} I_{x,q} \left| 1 - e^{i\lambda_q} \right|^{\alpha} \log \left| 1 - e^{i\lambda_q} \right| \operatorname{Re} \left(e^{i(p+\ell)\lambda_q} \right) \\
& + O_p \left(\frac{\log^2 T}{T^{3/2}} \right) \\
= & 2(\tilde{\alpha} - \alpha) \sum_{p=1}^{p_0+j} (\hat{\theta}_p - \theta_p) \sum_{\ell=0}^{p_0} \theta_\ell \int_0^{2\pi} \left| \Phi \left(e^{i\lambda}; \theta(p_0) \right) \right|^{-2} \log \left| 1 - e^{i\lambda} \right| \operatorname{Re} \left(e^{i(p+\ell)\lambda} \right) d\lambda \\
& + O_p \left(\frac{\log^2 T}{T^{3/2}} \right) \\
& + 2(\tilde{\alpha} - \alpha) \sum_{p=1}^{p_0+j} (\hat{\theta}_p - \theta_p) \sum_{\ell=0}^{p_0} \theta_\ell \phi_\ell + O_p \left(\frac{\log^2 T}{T^{3/2}} \right),
\end{aligned}$$

where ϕ_ℓ is the ℓ th Fourier coefficient of $|\Phi(e^{i\lambda}; \theta(p_0))|^{-2} \log |1 - e^{i\lambda}| \operatorname{Re}(e^{i(p+\ell)\lambda})$ and using (5.10) for the first equality and Theorem 1 in the second equality and that $\widehat{\theta}_p - \theta_p = O_p(T^{-1/2})$. Observe that $\phi_\ell = O(\ell^{-1})$.

Finally the last term of (5.18), which is

$$(5.19) \quad 2 \sum_{p=1}^{p_0+j} (\widehat{\theta}_p - \theta_p) \sum_{\ell=0}^{p_0} \theta_\ell \frac{1}{T} \sum_{q=1}^{T-1} I_{\varepsilon, q} \left| \Phi(e^{i\lambda_q}; \theta(p_0)) \right|^{-2} \operatorname{Re}(e^{i(p+\ell)\lambda_q}) + O_p\left(\frac{\log^2 T}{T^{4/3}}\right).$$

Because $\gamma_\ell =: \int_0^{2\pi} |\Phi(e^{i\lambda}; \theta(p_0))|^{-2} \operatorname{Re}(e^{i\ell\lambda}) d\lambda$ satisfies the equation in difference $\sum_{\ell=0}^{p_0} \gamma_\ell \theta_\ell = 0$ and Brillinger (1991, p.15) implies that

$$\frac{1}{T} \sum_{q=1}^{T-1} \left| \Phi(e^{i\lambda_q}; \theta(p_0)) \right|^{-2} \operatorname{Re}(e^{ip\lambda_q}) - \int_0^{2\pi} |\Phi(e^{i\lambda}; \theta(p_0))|^{-2} \operatorname{Re}(e^{ip\lambda}) d\lambda = O\left(\frac{1}{T}\right),$$

we have that for any s ,

$$\varsigma_T(s) =: \frac{1}{T} \sum_{q=1}^{T-1} \left| \Phi(e^{i\lambda_q}; \theta(p_0)) \right|^{-2} \operatorname{Re}(e^{ip\lambda_q}) e^{is\lambda_q} = O\left(\frac{1}{T}\right),$$

and hence (5.19) is $O_p(T^{-4/3} \log^2 T)$. Indeed,

$$\begin{aligned} & \sum_{\ell=0}^{p_0} \theta_\ell \frac{1}{T} \sum_{q=1}^{T-1} I_{\varepsilon, q} \left| \Phi(e^{i\lambda_q}; \theta(p_0)) \right|^{-2} \operatorname{Re}(e^{i(p+\ell)\lambda_q}) \\ &= \sum_{\ell=0}^{p_0} \theta_\ell \left\{ \frac{1}{T} \sum_{t=1}^T \varepsilon_t^2 \varsigma_T(0) + \frac{2}{T} \sum_{t=2}^T \varepsilon_t \sum_{s=1}^{t-1} \varepsilon_s \varsigma_T(t-s) \right\} \\ &= O_p(T^{-1}) \end{aligned}$$

by definition of $\varsigma_T(p, s)$, that $T^{-1} \sum_{t=1}^T \varepsilon_t^2 - \sigma_\varepsilon^2 = o_p(1)$ and $\varepsilon_t \sum_{s=1}^{t-1} \varepsilon_s \xi_T(p_0, t-s)$ is a martingale difference sequence triangular array. Observe that $|T\varsigma_T(p, s)| < K$. So together with the fact that $\widehat{\theta}_p - \theta_p = O_p(T^{-1/2})$, we conclude that the first term of (5.19) is $O_p(T^{-3/2} \log^2 T)$ and then the last term of (5.18) is $O_p(T^{-4/3} \log^2 T)$. The latter implies that (5.17) can be written as

$$\begin{aligned} \widetilde{S}_T(p_0 + j) &= \sum_{p_1, p_2=1}^{p_0+j} \gamma_{p_1-p_2} (\widehat{\theta}_{p_1} - \theta_{p_1}) (\widehat{\theta}_{p_2} - \theta_{p_2}) \\ &+ (\widetilde{\alpha} - \alpha) \sum_{p=1}^{p_0+j} (\widehat{\theta}_p - \theta_p) \sum_{\ell=0}^{p_0} \phi_\ell \theta_\ell + O_p\left(\frac{\log^2 T}{T^{4/3}}\right). \end{aligned}$$

Thus, the left side of (5.15) is bounded by

$$\Pr \left\{ \sup_{1 \leq j \leq \lfloor \log T \rfloor - p_0} \sum_{p_1, p_2=1}^{p_0+j} \gamma_{p_1-p_2} T \left(\widehat{\theta}_{p_1} - \theta_{p_1} \right) \left(\widehat{\theta}_{p_2} - \theta_{p_2} \right) - j \log T > 0 \right\} \\ + \Pr \left\{ T^{1/2} (\widetilde{\alpha} - \alpha) \sup_{1 \leq j \leq \lfloor \log T \rfloor - p_0} \sum_{p=1}^{p_0+j} T^{1/2} \left(\widehat{\theta}_p - \theta_p \right) \sum_{\ell=0}^{p_0} \phi_\ell \theta_\ell - j \log T > 0 \right\}$$

converges to 0 proceeding similarly as with the proof of (5.11). This completes the proof of the theorem. \blacksquare

6. CONCLUSIONS AND EXTENSIONS

In this paper we have introduced and examined a criterion for the order selection of the fractional integrated autoregressive polynomial or Bloomfield's exponential model based on the spectral decomposition of the variance of the innovations or error term of the model. The criterion has some similarities with statistics frequently employed for goodness-of-fit tests. We have shown its consistency and examined its finite sample performance in a Monte-Carlo experiment. A close inspection of the proofs suggests that it is possible to show the validity using other criterion such as that of Hannan and Quinn's (1979) *HIC*, given by

$$\arg \min_{p \leq \lfloor \log_2 T \rfloor} HIC(p; \alpha) = W_T(p; \alpha) + \frac{cp \log_2 T}{T} \quad \text{with } c > 2$$

proceeding as in Hidalgo (2002). Observe that in the last displayed expression the upper bound, although increases with the sample size, has changed from $\lfloor \log T \rfloor$ to $\lfloor \log_2 T \rfloor$.

We now discuss how our criterion can be used when the data is non-stationary and/or strongly negative dependent in particular when the true value of α can be greater than or equal to 1 or less than 0. That is, we assume that $\alpha \in [-1, 2]$. For that purpose, we define the taper periodogram of a generic sequence $\{z_t\}_{t=1}^T$ by

$$\mathring{I}_z(\lambda) = |\mathring{w}_z(\lambda)|^2, \quad \lambda \in [0, \pi],$$

where

$$\mathring{w}_z(\lambda) = \frac{1}{\left(\sum_{t=1}^T h^2(t/T) \right)^{1/2}} \sum_{t=1}^T h(t/T) z_t e^{-it\lambda}$$

is the (taper) *Discrete Fourier Transform*, where $h(z)$ is the taper function. Then a standard estimator is the weighted periodogram

$$\check{g}_x(\lambda) = \frac{1}{2m+1} \sum_{j=-m}^m \left| 1 - e^{i(\lambda_j + \lambda)} \right|^{\widetilde{\alpha}} \mathring{I}_x(\lambda + \lambda_j),$$

where $\check{\alpha}$ is an estimator of the long memory parameter α

$$(6.1) \quad \check{\alpha} = \sum_{k=1}^m \delta_k(m) \log \left(\hat{I}_{x,k} \right),$$

see Velasco (1999). From here we proceed as in Section 2 to compute the estimators of the short memory parameter but replacing $\hat{\alpha}$ by $\check{\alpha}$ there. That is,

$$\begin{aligned} \hat{\zeta}_j &= \frac{1}{M} \sum_{\ell=1}^{M-1} (\log \check{g}_{x,2m\ell}) \cos(j\lambda_{2m\ell}), \quad j = 0, 1, \dots, M, \\ \hat{\Theta}_{2mq} &= \overline{\hat{\Theta}_{-2mq}} = \exp \left\{ - \sum_{j=1}^{M-1} \hat{\zeta}_j e^{-ij\lambda_{2mq}} \right\}, \quad q = 0, 1, \dots, M, \\ \hat{\theta}_\ell &= \frac{1}{2M} \sum_{q=-M+1}^M \hat{\Theta}_{2mq} e^{i\ell\lambda_{2mq}}, \quad \ell = 1, \dots, M, \quad \text{and} \\ \hat{\sigma}_\varepsilon^2 &= 2\pi e^{-\hat{\zeta}_0}, \end{aligned}$$

where

$$\check{g}_x(\lambda) = \frac{1}{2m+1} \sum_{j=-m}^m \left| 1 - e^{i(\lambda_j + \lambda)} \right|^{\check{\alpha}} \hat{I}_x(\lambda + \lambda_j).$$

Hence we can estimate p_0 as

$$\begin{aligned} \hat{p} &= : \arg \min_{p \leq [\log T]} BIC(p) = \hat{W}_T(p; \check{\alpha}) + \frac{p \log T}{T} \\ \hat{p} &= : \arg \min_{p \leq [\log_2 T]} HIC(p) = \hat{W}_T(p; \check{\alpha}) + \frac{cp \log_2 T}{T} \quad \text{with } c > 2 \end{aligned}$$

with

$$\hat{W}_T(p; \alpha) =: \frac{1}{[T/2]} \sum_{k=1}^{[T/2]} \frac{|1 - e^{i\lambda_k}|^\alpha \hat{I}_{x,k}}{g_{x,k}(p)},$$

which in view of results in Velasco (1999) and/or Velasco and Robinson (2000), we envisage that \hat{p} should converge in probability to the true value p_0 , proceeding as we did in the proof of Theorem 2.

REFERENCES

- [1] AKAIKE, H. (1974), A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, 716-723.
- [2] AN, H.Z., CHEN, Z.G. AND HANNAN, E.J. (1982), Autocorrelation, autoregression and autoregressive approximation. *Annals of Statistics*, **10**, 926-936.
- [3] 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.
- [4] BLOOMFIELD, P. (1973), An exponential model for the spectrum of a scalar time series. *Biometrika*, **60**, 217-226.
- [5] BRILLINGER, D.R. (1981), Time Series, Data Analysis and Theory. San Francisco: Holden-Day.
- [6] CHANDNA, S. AND WALDEN, A.T. (2013), Simulation methodology for inference on physical parameters of complex vector-valued signals. *IEEE Transactions on Signal Processing*, **61**, 5260-5269.
- [7] GEWEKE, J. AND PORTER-HUDAK, S. (1983), The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221-238.
- [8] HANNAN, E.J. (1963), Regression for time series. In M. Rosenblatt (ed.), *Time Series Analysis*, 17-37. New York: Wiley.
- [9] HANNAN, E.J. (1970), Multiple Time Series. John Wiley: New York.
- [10] HANNAN, E.J. (1980), The estimation of the order of an ARMA process. *Annals of Statistics*, **8**, 1071-1081.
- [11] HANNAN, E.J. AND QUINN, B.G. (1979), The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Ser. B*, **41**, 190-195.
- [12] HIDALGO, J. (2002), Consistent order selection with strongly dependent data and its application to efficient estimation. *Journal of Econometrics*, **110**, 213-239.
- [13] HIDALGO, J. AND YAJIMA, Y. (2002), Prediction of strongly dependent processes in the frequency domain with application to signal extraction. *Econometric Theory*, **18**, 584-624.
- [14] MALLOWS, C.L. (1973), Some comments on C_p . *Technometrics*, **15**, 661-675.
- [15] MOULINES, E. AND SOULIER, P. (1999), Broadband log-periodogram regression of time series with long-range dependence. *Annals of Statistics*, **27**, 1415-143.
- [16] PARZEN, E. (1974), Some recent advances in time series analysis. *IEEE Transactions on Automatic Control*, AC-19, 723-730.
- [17] PÖTSCHER, B.M. (1991), Effects of model selection on inference. *Econometric Theory*, **7**, 163-185.
- [18] PRESS, H. AND TUKEY, J.W. (1956), Power spectral methods of analysis and their application to problems in airplane dynamics. flight test manual, NATO, Advisory Group for Aeronautical Research and Development, vol. IV-C, 1-41. Reprinted as Bell System Monograph 2606.
- [19] ROBINSON, P.M. (1979), Distributed lag approximation to linear time-invariant systems. *Annals of Statistics*, **7**, 507-515.
- [20] ROBINSON, P.M. (1995a), Log-periodogram regression for time series with long range dependence. *Annals of Statistics*, **23**, 1048-1072.
- [21] ROBINSON, P.M. (1995b), Gaussian semiparametric estimation of long-range dependence. *Annals of Statistics*, **23**, 1630-1661.
- [22] ROBINSON, P.M. (2005), Efficiency improvements in inference on stationary and nonstationary fractional time series. *Annals of Statistics*, **33**, 1800-1842.
- [23] SCHWARZ, G. (1978), Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.
- [24] SERFLING, R.J. (1980), Approximation theorems of mathematical statistics. John Wiley and Sons.

- [25] SHIBATA, R. (1976), Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, **63**, 117-126.
- [26] SHIBATA, R. (1980), Asymptotic efficiency selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics*, **8**, 147-164.
- [27] SHIBATA, R. (1981), An optimal selection of regression variables. *Biometrika*, **68**, 45-54.
- [28] VELASCO, C. (1999), Non-stationary log-periodogram regression. *Journal of Econometrics*, **91**, 325-371.
- [29] VELASCO, C. AND ROBINSON, P.M. (2000), Whittle Pseudo-Maximum Likelihood Estimates of Non-Stationary Time Series. *Journal of the American Statistical Association*, **95**, 1229-1243.

DEPARTMENT OF ECONOMICS, UNIVERSITY OF ESSEX, WIVENHOE PARK, COLCHESTER, CO4 3SQ, U.K.

E-mail address: a.gupta@essex.ac.uk

LONDON SCHOOL OF ECONOMICS, HOUGHTON STREET, LONDON WC2A 2AE, U.K.

E-mail address: f.j.hidalgo@lse.ac.uk