Submitted to the Annals of Statistics
arXiv: math.PR/0000000

GAUSSIAN PSEUDO-MAXIMUM LIKELIHOOD ESTIMATION OF FRACTIONAL TIME SERIES MODELS

BY JAVIER HUALDE, PETER M. ROBINSON

Universidad Pública de Navarra and London School of Economics

We consider the estimation of parametric fractional time series models in which not only is the memory parameter unknown, but one may not know whether it lies in the stationary/invertible region or the nonstationary or noninvertible regions. In these circumstances a proof of consistency (which is a prerequisite for proving asymptotic normality) can be difficult owing to non-uniform convergence of the objective function over a large admissible parameter space. In particular, this is the case for the conditional sum of squares estimate, which can be expected to be asymptotically efficient under Gaussianity. Without the latter assumption, we establish consistency and asymptotic normality for this estimate in case of a quite general univariate model. For a multivariate model we establish asymptotic normality of a one-step estimate based on an initial $\sqrt{n}$-consistent estimate.

1. Introduction. Autoregressive moving average (ARMA) models have featured prominently in the analysis of time series. The versions initially stressed in the theoretical literature (e.g. [11], [26]) are stationary and invertible. Following [6], unit root nonstationarity has frequently been incorporated, while "overdifferenced" non-invertible processes have also featured. Stationary ARMA processes automatically have short memory with "memory parameter", denoted δ_0, taking the value zero, implying a huge behavioural gap relative to unit root versions, where $\delta_0 = 1$. This has been bridged by "fractionally-differenced", or long memory, models, a leading class being the fractional autoregressive integrated ARMA (FARIMA). A FARIMA (p_1, δ_0, p_2) process x_t is given by

$$x_t = \Delta^{-\delta_0}\{u_t \mathbb{1}(t>0)\},\ t = 0, \pm 1, ..., \tag{1.1}$$

$$\alpha(L)u_t = \beta(L)\varepsilon_t,\ t = 0, \pm 1, ..., \tag{1.2}$$

where $\{x_t\}$ is the observable series; L is the lag operator; $\Delta = 1 - L$;

$$(1-L)^{-\zeta} = \sum_{j=0}^{\infty} a_j(\zeta)L^j,\quad a_j(\zeta) = \frac{\Gamma(j+\zeta)}{\Gamma(\zeta)\Gamma(j+1)},$$

AMS 2000 subject classifications: 62M10; 62F12

Keywords and phrases: Fractional processes, nonstationarity, noninvertibility, Gaussian estimation, consistency, asymptotic normality, multiple time series

with $\Gamma(\zeta) = \infty$ for $\zeta = 0, -1, ...$, and by convention $\Gamma(0)/\Gamma(0) = 1$; $\mathbb{1}(\cdot)$ is the indicator function; $\alpha(L)$ and $\beta(L)$ are real polynomials of degrees p_1 and p_2, which share no common zeros, and all of their zeros are outside the unit circle in the complex plane; and the ε_t are serially uncorrelated and homoscedastic with zero mean. The reason (1.1) features the truncated process $u_t\mathbb{1}(t > 0)$ rather than simply u_t is to simultaneously cover δ_0 falling in both the stationary region $\left(\delta_0 < \frac{1}{2}\right)$ and the nonstationary region ($\delta_0 \geq \frac{1}{2}$, where otherwise the process would "blow up"). In the former case the truncation implies that x_t is only "asymptotically stationary". In recent years fractional modelling has found many applications in the sciences and social sciences, for example with respect to environmental and financial data.

Early work on asymptotic statistical theory for fractional models assumed $\delta_0 < \frac{1}{2}$ (and replaced $u_t\mathbb{1}(t > 0)$ by u_t in (1.1)). Assuming $\delta_0 \in (0, \frac{1}{2})$, [8], [9], [10] and [12] showed consistency and asymptotic normality of Whittle estimates (of δ_0 and other parameters, such as the coefficients of α and β), thereby achieving analogous results to those of [11], [26] for stationary ARMA processes (i.e. (1.2) with $u_t = x_t$) and other short memory models. More recently, [16] considered empirical maximum likelihood inference covering this setting. Note that [8], [9], [10] and [12], and much other work, not only excluded $\delta_0 \geq \frac{1}{2}$ but also the short-memory case $\delta_0 = 0$, as well as negatively dependent processes where $\delta_0 < 0$. To some degree other δ_0 can be covered, for example for $\delta_0 \in (1, 3/2)$ one can first-difference the data, apply the methods and theory of [8], [9], [10] and [12], and then add 1 to the memory parameter estimate, but this still requires prior knowledge that δ_0 lies in an interval of length no more than $\frac{1}{2}$.

On the other hand, [3] argued that the same desirable properties should hold without so restricting δ_0, in case of a conditional-sum-of-squares estimate, and this would be consistent with the classical asymptotic properties established by [18] for score tests for a unit root and other hypotheses against fractional alternatives, by comparison with the nonstandard behaviour of unit root tests against stationary autoregressive alternatives. However, the proof of asymptotic normality in [3] appears to assume that the estimate lies in a small neighbourhood of δ_0, without first proving consistency (see also [24]). Due to a lack of uniform convergence, consistency of this implicitly-defined estimate is especially difficult to establish when the set of admissible values of δ is large. In particular this is the case when δ_0 is known only to lie in an interval of length greater than $\frac{1}{2}$. In the present paper, we establish consistency and asymptotic normality when the interval is arbitrarily large, including (simultaneously) stationary, nonstationary, invertible and non-invertible values of δ_0. Thus prior knowledge of which of these phenomena obtains is unnecessary, and this seems especially practically desirable given, for example, that estimates near the $\delta_0 = \frac{1}{2}$ or $\delta_0 = 1$ boundaries frequently occur in practice, while empirical interest in autoregressive models with two unit roots suggests allowance for values in the region of $\delta_0 = 2$ also, and (following [1]) antipersistence and the possibility of overdifferencing imply the possibility that $\delta_0 < 0$.

We in fact consider a more general model than (1.1), (1.2), retaining (1.1) but generalizing (1.2) to

$$u_t = \theta(L; \boldsymbol{\varphi}_0)\varepsilon_t, \; t = 0, \pm 1, ..., \tag{1.3}$$

where ε_t is a zero-mean unobservable white noise sequence, $\boldsymbol{\varphi}_0$ is an unknown $p \times 1$ vector, $\theta(s; \boldsymbol{\varphi}) = \sum_{j=0}^{\infty} \theta_j(\boldsymbol{\varphi}) s^j$, where for all $\boldsymbol{\varphi}$, $\theta_0(\boldsymbol{\varphi}) = 1$, $\theta(s; \boldsymbol{\varphi})$: $\mathbb{C} \times \mathbb{R}^p$ is continuous in s and $|\theta(s; \boldsymbol{\varphi})| \neq 0$, $|s| \leq 1$. More detailed conditions will be imposed below. The role of θ in (1.3), like α and β in (1.2), is to permit parametric short memory autocorrelation. We allow for the simplest case FARIMA$(0, \delta_0, 0)$ by taking $\boldsymbol{\varphi}_0$ to be empty. Another model covered by (1.3) is the exponential-spectrum one of [5] (which in conjunction with fractional differencing leads to a relatively neat covariance matrix formula [18]). Semiparametric models (where u_t has nonparametric autocovariance structure, see e.g. [19], [23]) afford still greater flexibility than (1.3), but also require larger samples in order for comparable precision to be achieved. In more moderate-sized samples, investment in a parametric model can prove worthwhile, even the simple FARIMA$(1, \delta_0, 0)$ employed in the Monte Carlo simulations reported in the supplementary material [14], while model choice procedures can be employed to choose p_1 and p_2 in the FARIMA(p_1, δ_0, p_2), as illustrated in the empirical examples included in the supplementary material [14].

We wish to estimate $\boldsymbol{\tau}_0 = (\delta_0, \boldsymbol{\varphi}_0')'$ from observations x_t, $t = 1, ..., n$. For any admissible $\boldsymbol{\tau} = (\delta, \boldsymbol{\varphi}')'$, define

$$\varepsilon_t(\boldsymbol{\tau}) = \Delta^{\delta}\theta^{-1}(L; \boldsymbol{\varphi})x_t, \; t \geq 1, \tag{1.4}$$

noting that (1.1) implies $x_t = 0$, $t \leq 0$. For a given user-chosen optimizing set $\mathcal{T}$, define as an estimate of $\boldsymbol{\tau}_0$

$$\widehat{\boldsymbol{\tau}} = \arg\min_{\boldsymbol{\tau} \in \mathcal{T}} R_n(\boldsymbol{\tau}), \tag{1.5}$$

where

$$R_n(\boldsymbol{\tau}) = \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^2(\boldsymbol{\tau}), \tag{1.6}$$

and $\mathcal{T} = \mathcal{I} \times \Psi$, where $\mathcal{I} = \{\delta : \triangledown_1 \leq \delta \leq \triangledown_2\}$ for given $\triangledown_1$, $\triangledown_2$ such that $\triangledown_1 < \triangledown_2$, Ψ is a compact subset of $\mathbb{R}^p$, and $\boldsymbol{\tau}_0 \in \mathcal{T}$.

The estimate $\widehat{\boldsymbol{\tau}}$ is sometimes termed "conditional sum of squares" (though "truncated sum of squares" might be more suitable). It has the anticipated advantage of having the same limit distribution as the maximum likelihood estimate of $\boldsymbol{\tau}_0$ under Gaussianity, in which case it is asymptotically efficient (though here we do not assume Gaussianity). It was employed by [6] in estimation of non-fractional ARMA models (when δ_0 is a given integer), by [15], [21] in stationary FARIMA models, where $0 < \delta_0 < 1/2$, and by [3], [24] in nonstationary FARIMA models, allowing $\delta_0 \geq 1/2$.

The following section sets down detailed regularity conditions, a formal statement of asymptotic properties and the main proof details. Section 3 provides asymptotically normal estimates in a multivariate extension of (1.1), (1.3). Joint modelling of related processes is important both for reasons of parsimony and interpretation, and multivariate fractional processes are currently relatively untreated, even in the stationary case. Further possible extensions are discussed in Section 4. Useful lemmas are stated in Section 5. Due to space restrictions, the proofs of these lemmas, along with an analysis of finite-sample performance of the procedure and an empirical application, are included in the supplementary material [14].

2. Consistency and asymptotic normality.

2.1. *Consistency of* $\widehat{\boldsymbol{\tau}}$. Our first two assumptions will suffice for consistency.

A1. **(i)**

$$\left|\theta\left(s;\boldsymbol{\varphi}\right)\right| \neq \left|\theta\left(s;\boldsymbol{\varphi}_0\right)\right|,$$

for all $\boldsymbol{\varphi} \neq \boldsymbol{\varphi}_0$, $\boldsymbol{\varphi} \in \Psi$, on a set $S \subset \{s : |s| = 1\}$ of positive Lebesgue measure;

(ii) for all $\boldsymbol{\varphi}$, $\theta\left(e^{i\lambda};\boldsymbol{\varphi}\right)$ is differentiable in λ with derivative in $Lip\left(\varsigma\right)$, $\varsigma > 1/2$;

(iii) for all λ, $\theta\left(e^{i\lambda};\boldsymbol{\varphi}\right)$ is continuous in $\boldsymbol{\varphi}$;

(iv) for all $\boldsymbol{\varphi} \in \Psi$, $\left|\theta\left(s;\boldsymbol{\varphi}\right)\right| \neq 0$, $|s| \leq 1$.

Condition (i) provides identification while (ii) and (iv) ensure that u_t is an invertible short-memory process (with spectrum that is bounded and bounded away from zero at all frequencies). Further, by (ii) the derivative of $\theta(e^{i\lambda};\boldsymbol{\varphi})$ has Fourier coefficients $j\theta_j\left(\boldsymbol{\varphi}\right) = O\left(j^{-\varsigma}\right)$ as $j \to \infty$, for all $\boldsymbol{\varphi}$, from p.46 of [27], so that, by compactness of Ψ and continuity of $\theta_j(\boldsymbol{\varphi})$ in $\boldsymbol{\varphi}$ for all j,

$$\sup_{\boldsymbol{\varphi}\in\Psi} |\theta_j(\boldsymbol{\varphi})| = O\left(j^{-(1+\varsigma)}\right) \text{ as } j \to \infty. \tag{2.1}$$

Also, writing $\theta^{-1}(s;\boldsymbol{\varphi}) = \phi(s;\boldsymbol{\varphi}) = \sum_{j=0}^{\infty} \phi_j(\boldsymbol{\varphi}) s^j$, we have $\phi_0(\boldsymbol{\varphi}) = 1$ for all $\boldsymbol{\varphi}$, and (ii), (iii) and (iv) imply that

$$\sup_{\boldsymbol{\varphi}\in\Psi} \left|\phi_j(\boldsymbol{\varphi})\right| = O\left(j^{-(1+\varsigma)}\right) \text{ as } j \to \infty. \tag{2.2}$$

Finally, (ii) also implies that

$$\inf_{\substack{|s|=1\\ \boldsymbol{\varphi}\in\Psi}} |\phi(s;\boldsymbol{\varphi})| > 0. \tag{2.3}$$

A1 is easily satisfied by standard parameterizations of stationary and invertible ARMA processes (1.2) in which autoregressive and moving average orders are not both over-specified. More generally, A1 is similar to conditions employed in asymptotic theory for the estimate $\widehat{\boldsymbol{\tau}}$ and other forms of Whittle estimate that restrict to stationarity (see e.g. [8], [9], [10], [12], [21]) and not only is it readily verifiable because θ is a known parametric function, but in practice θ satisfying A1 are invariably employed by practitioners.

A2. The ε_t in (1.3) are stationary and ergodic with finite fourth moment, and

$$E\left(\varepsilon_t | \mathcal{F}_{t-1}\right) = 0, \quad E\left(\varepsilon_t^2 \middle| \mathcal{F}_{t-1}\right) = \sigma_0^2 \tag{2.4}$$

almost surely, where $\mathcal{F}_t$ is the σ-field of events generated by ε_s, $s \le t$, and conditional (on $\mathcal{F}_{t-1}$) third and fourth moments of ε_t equal the corresponding unconditional moments.

A2 avoids requiring independence or identity of distribution of ε_t, but rules out conditional heteroskedasticity. It has become fairly standard in the time series asymptotics literature since [11].

THEOREM 2.1. *Let (1.1), (1.3) and A1, A2 hold. Then as $n \to \infty$*

$$\widehat{\boldsymbol{\tau}} \to_p \boldsymbol{\tau}_0. \tag{2.5}$$

PROOF. We give the proof for the most general case where $\triangledown_1 < \delta_0 - \frac{1}{2}$, but our proof trivially covers the $\triangledown_1 \ge \delta_0 - \frac{1}{2}$ situation, for which some of the steps described below are superfluous. The proof begins standardly. For $\varepsilon > 0$ define $N_\varepsilon = \{\boldsymbol{\tau} : \|\boldsymbol{\tau} - \boldsymbol{\tau}_0\| < \varepsilon\}$, $\overline{N}_\varepsilon = \{\boldsymbol{\tau} : \boldsymbol{\tau} \notin N_\varepsilon, \boldsymbol{\tau} \in \mathcal{T}\}$. For small enough ε,

$$\Pr\left(\widehat{\boldsymbol{\tau}} \in \overline{N}_\varepsilon\right) \leq \Pr\left(\inf_{\boldsymbol{\tau}\in\overline{N}_\varepsilon} S_n(\boldsymbol{\tau}) \leq 0\right), \tag{2.6}$$

where $S_n(\boldsymbol{\tau}) = R_n(\boldsymbol{\tau}) - R_n(\boldsymbol{\tau}_0)$. The remainder of the proof reflects the fact that $R_n(\boldsymbol{\tau})$, and thus $S_n(\boldsymbol{\tau})$, converges in probability to a well-behaved function when $\delta > \delta_0 - \frac{1}{2}$, and diverges when $\delta < \delta_0 - \frac{1}{2}$, while the need to establish uniform convergence, especially in a neighbourhood of $\delta = \delta_0 - \frac{1}{2}$, requires additional special treatment. Consequently, for arbitrarily small $\eta > 0$, such that $\eta < \delta_0 - \frac{1}{2} - \bigtriangledown_1$, we define the non-intersecting sets $\mathcal{I}_1 = \left\{\delta : \bigtriangledown_1 \leq \delta \leq \delta_0 - \frac{1}{2} - \eta\right\}$, $\mathcal{I}_2 = \left\{\delta : \delta_0 - \frac{1}{2} - \eta < \delta < \delta_0 - \frac{1}{2}\right\}$, $\mathcal{I}_3 = \left\{\delta : \delta_0 - \frac{1}{2} \leq \delta \leq \delta_0 - \frac{1}{2} + \eta\right\}$, $\mathcal{I}_4 = \left\{\delta : \delta_0 - \frac{1}{2} + \eta < \delta \leq \bigtriangledown_2\right\}$. Correspondingly, define $\mathcal{T}_i = \mathcal{I}_i \times \Psi$, $i = 1, ..., 4$, so $\mathcal{T} = \cup_{i=1}^4 \mathcal{T}_i$. Thus from (2.6) it remains to prove

$$\Pr\left(\inf_{\boldsymbol{\tau}\in\overline{N}_\varepsilon\cap\mathcal{T}_i} S_n(\boldsymbol{\tau}) \leq 0\right) \to 0, \text{ as } n \to \infty, \ i = 1, ..., 4. \tag{2.7}$$

Each of the four proofs differs, and we describe them in reverse order.
Proof of (2.7) for $i = 4$. By a familiar argument, the result follows if for $\boldsymbol{\tau} \in \mathcal{T}_4$ there is a deterministic function $U(\boldsymbol{\tau})$ (not depending on n), such that

$$S_n(\boldsymbol{\tau}) = U(\boldsymbol{\tau}) - T_n(\boldsymbol{\tau}),$$

where

$$\inf_{\overline{N}_\varepsilon\cap\mathcal{T}_4} U(\boldsymbol{\tau}) > \epsilon, \tag{2.8}$$

ϵ throughout denoting a generic arbitrarily small positive constant, and

$$\sup_{\mathcal{T}_4} |T_n(\boldsymbol{\tau})| = o_p(1). \tag{2.9}$$

Since $x_t = 0$, $t \leq 0$, for $\boldsymbol{\tau} \in \mathcal{T}_4$ we set (cf. (1.4)), $\zeta_t(\boldsymbol{\tau}) = \Delta^{\delta-\delta_0}\phi(L;\boldsymbol{\varphi})\, u_t$, $U(\boldsymbol{\tau}) = E\zeta_t^2(\boldsymbol{\tau}) - \sigma_0^2$, and $T_n(\boldsymbol{\tau}) = R_n(\boldsymbol{\tau}_0) - \sigma_0^2 - \left\{R_n(\boldsymbol{\tau}) - E\zeta_t^2(\boldsymbol{\tau})\right\}$. We may write

$$U(\boldsymbol{\tau}) = \sigma_0^2\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{g(\lambda)}{g_0(\lambda)}d\lambda - 1\right),$$

where

$$g(\lambda) = \left|1 - e^{i\lambda}\right|^{2(\delta - \delta_0)} \left|\phi\left(e^{i\lambda}; \boldsymbol{\varphi}\right)\right|^2, \; g_0(\lambda) = g(\lambda)|_{\boldsymbol{\tau} = \boldsymbol{\tau}_0}.$$

For all $\boldsymbol{\tau}$ $(2\pi)^{-1} \int_{-\pi}^{\pi} \log\left(g(\lambda)/g_0(\lambda)\right) d\lambda = 0$, so by Jensen's inequality

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{g(\lambda)}{g_0(\lambda)} d\lambda \geq 1. \tag{2.10}$$

Under A1(i), we have strict inequality in (2.10) for all $\boldsymbol{\tau} \neq \boldsymbol{\tau}_0$, so that by continuity in $\boldsymbol{\tau}$ of the left side of (2.10), (2.8) holds. Next, write

$$\varepsilon_t(\boldsymbol{\tau}) = \sum_{j=0}^{t-1} c_j(\boldsymbol{\tau}) u_{t-j}, \quad \zeta_t(\boldsymbol{\tau}) = \sum_{j=0}^{\infty} c_j(\boldsymbol{\tau}) u_{t-j},$$

where $c_j(\boldsymbol{\tau}) = \sum_{k=0}^{j} \phi_k(\boldsymbol{\varphi}) a_{j-k}(\delta_0 - \delta)$. Because, given A2, the $\varepsilon_t^2 - \sigma_0^2$ are stationary martingale differences,

$$R_n(\boldsymbol{\tau}_0) - \sigma_0^2 = \frac{1}{n} \sum_{t=1}^{n} \left(\varepsilon_t^2 - \sigma_0^2\right) \to_p 0, \text{ as } n \to \infty. \tag{2.11}$$

Then defining $\gamma_k = E(u_t u_{t-k})$, and henceforth writing $c_j = c_j(\boldsymbol{\tau})$, (2.9) would hold on showing that

$$\sup_{\mathcal{T}_4} \left| \frac{1}{n} \sum_{t=1}^{n} \left[\left(\sum_{j=0}^{t-1} c_j u_{t-j} \right)^2 - E\left(\sum_{j=0}^{t-1} c_j u_{t-j} \right)^2 \right] \right| = o_p(1), \tag{2.12}$$

$$\sup_{\mathcal{T}_4} \left| \frac{1}{n} \sum_{t=1}^{n} \sum_{j=0}^{t-1} \sum_{k=t}^{\infty} c_j c_k \gamma_{j-k} \right| = o_p(1), \tag{2.13}$$

$$\sup_{\mathcal{T}_4} \left| \frac{1}{n} \sum_{t=1}^{n} \sum_{j=t}^{\infty} \sum_{k=t}^{\infty} c_j c_k \gamma_{j-k} \right| = o_p(1). \tag{2.14}$$

We first deal with (2.12). The term whose modulus is taken is

$$\frac{1}{n} \sum_{j=0}^{n-1} c_j^2 \sum_{l=1}^{n-j} \left(u_l^2 - \gamma_0\right) + \frac{2}{n} \sum_{j=0}^{n-2} \sum_{k=j+1}^{n-1} c_j c_k \sum_{l=k-j+1}^{n-j} \left\{u_l u_{l-(k-j)} - \gamma_{j-k}\right\}$$

$$= (a) + (b). \tag{2.15}$$

First,

$$E\sup_{\mathcal{T}_4}|(a)| \le \frac{1}{n}\sum_{j=0}^{n-1}\sup_{\mathcal{T}_4} c_j^2 E\left|\sum_{l=1}^{n-j}\left(u_l^2-\gamma_0\right)\right|.$$

It can be readily shown that, uniformly in j, $Var\left(\sum_{l=1}^{n-j}u_l^2\right)=O(n)$, so

$$\sup_{\mathcal{T}_4}|(a)| = O_p\left(n^{-\frac{1}{2}}\sum_{j=1}^{\infty}j^{-2\eta-1}\right)=O_p\left(n^{-\frac{1}{2}}\right),$$

by Lemma 1. Next, by summation by parts, (b) is equal to

$$\begin{aligned}
& \frac{2c_{n-1}}{n}\sum_{j=0}^{n-2}c_j\sum_{k=j+1}^{n-1}\sum_{l=k-j+1}^{n-j}\left\{u_l u_{l-(k-j)}-\gamma_{j-k}\right\} \\
& -\frac{2}{n}\sum_{j=0}^{n-2}c_j\sum_{k=j+1}^{n-2}\left(c_{k+1}-c_k\right)\sum_{r=j+1}^{k}\sum_{l=r-j+1}^{n-j}\left\{u_l u_{l-(r-j)}-\gamma_{j-r}\right\} \\
= & \ (b_1)+(b_2).
\end{aligned}$$

It can be easily shown that, uniformly in j,

$$Var\left(\sum_{k=j+1}^{n-1}\sum_{l=k-j+1}^{n-j}u_l u_{l-(k-j)}\right)=O\left(n^2\right),$$

so we have

$$E\sup_{\mathcal{T}_4}|(b_1)| \le Kn^{-\eta-\frac{3}{2}}\sum_{j=1}^{n}j^{-\eta-\frac{1}{2}}\left\{Var\left(\sum_{k=j+1}^{n-1}\sum_{l=k-j+1}^{n-j}u_l u_{l-(k-j)}\right)\right\}^{\frac{1}{2}} \le Kn^{-2\eta},$$

by Lemma 1, where K throughout denotes a generic finite but arbitrarily large positive constant. Similarly,

$$E\sup_{\mathcal{T}_4}|(b_2)| \le Kn^{-1}\sum_{j=1}^{n}j^{-\eta-\frac{1}{2}}\sum_{k=j+1}^{n}k^{\max\left(-\eta-\frac{3}{2},-(1+\varsigma)\right)}\left\{Var\left(\sum_{r=j+1}^{k}\sum_{l=r-j+1}^{n-j}u_l u_{l-(r-j)}\right)\right\}^{\frac{1}{2}},$$

by Lemma 1, where ς was introduced in A1 (ii). It can be readily shown that

$$Var\left(\sum_{r=j+1}^{k}\sum_{l=r-j+1}^{n-j}u_l u_{l-(r-j)}\right) \le K(k-j)(n-j).$$

Take η such that $\eta + \frac{3}{2} < 1 + \varsigma$. Then

$$\begin{aligned} E \sup_{\mathcal{T}_4} |(b_2)| &\leq Kn^{-\frac{1}{2}} \sum_{j=1}^{n} j^{-\eta-\frac{1}{2}} \sum_{k=j+1}^{n} k^{-\eta-\frac{3}{2}} (k-j)^{\frac{1}{2}} \\ &\leq Kn^{-\frac{1}{2}} \sum_{j=1}^{n} j^{-\eta-\frac{1}{2}} \sum_{k=1}^{n} (k+j)^{-\eta-\frac{3}{2}} k^{\frac{1}{2}}. \end{aligned}$$

This is bounded by

$$Kn^{-\frac{1}{2}} \sum_{j=1}^{n} j^{-3\eta-\frac{1}{2}} \sum_{k=1}^{n} k^{\eta-1}, \tag{2.16}$$

because $(k+j)^{-\eta-\frac{3}{2}} \leq j^{-2\eta} k^{\eta-\frac{3}{2}}$. For small enough η, (2.16) is bounded by $Kn^{-2\eta}$, to complete the proof of (2.12). Next, the term whose modulus is taken in (2.13) is

$$\frac{1}{n} \sum_{t=1}^{n} \int_{-\pi}^{\pi} f(\lambda) \sum_{j=0}^{t-1} \sum_{k=t}^{\infty} c_j c_k e^{i(j-k)\lambda} d\lambda, \tag{2.17}$$

where $f(\lambda)$ denotes the spectral density of u_t. By boundedness of f (implied by Assumption A1) and the Cauchy inequality, (2.17) is bounded by

$$Kn^{-1} \sum_{t=1}^{n} \left\{ \int_{-\pi}^{\pi} \left| \sum_{j=0}^{t-1} c_j e^{ij\lambda} \right|^2 d\lambda \int_{-\pi}^{\pi} \left| \sum_{k=t}^{\infty} c_k e^{-ik\lambda} \right|^2 d\lambda \right\}^{\frac{1}{2}} \leq Kn^{-1} \sum_{t=1}^{n} \left\{ \sum_{j=0}^{t-1} c_j^2 \sum_{k=t}^{\infty} c_k^2 \right\}^{\frac{1}{2}},$$

so the left side of (2.13) is bounded by

$$Kn^{-1} \sum_{t=1}^{n} \left\{ \sum_{j=1}^{t} j^{-2\eta-1} \sum_{k=t}^{\infty} k^{-2\eta-1} \right\}^{\frac{1}{2}} \leq Kn^{-1} \sum_{t=1}^{n} t^{-\eta} \leq Kn^{-\eta} = o(1),$$

by Lemma 1, to establish (2.13). Finally, by a similar reasoning, the term whose modulus is taken in (2.14) is bounded by

$$Kn^{-1} \sum_{t=1}^{n} \left\{ \int_{-\pi}^{\pi} \left| \sum_{j=t}^{\infty} c_j e^{ij\lambda} \right|^2 d\lambda \right\}^{\frac{1}{2}} \leq Kn^{-1} \sum_{t=1}^{n} t^{-2\eta} \leq Kn^{-2\eta},$$

to conclude the proof of (2.14), and thence of (2.9). Thus (2.7) is proved for $i = 4$. With respect to (2.7) for $i = 1, 2, 3$, note from $\mathcal{T}_i \cap \overline{N}_\varepsilon \equiv \mathcal{T}_i$ for such i, and (2.11), that these results follow if

$$\Pr\left(\inf_{\mathcal{T}_i} R_n(\boldsymbol{\tau}) \le K\right) \to 0 \text{ as } n \to \infty,\ i = 1, 2, 3. \tag{2.18}$$

Proof of (2.7) for $i = 3$. Denote, for any sequence ζ_t, $w_\zeta(\lambda) = n^{-\frac{1}{2}} \sum_{t=1}^{n} \zeta_t e^{it\lambda}$, $I_\zeta(\lambda) = |w_\zeta(\lambda)|^2$, the discrete Fourier transform and periodogram respectively, and $\lambda_j = 2\pi j/n$. For $V_n(\boldsymbol{\tau})$ satisfying Lemma 3, setting $\boldsymbol{\tau}^* = (\delta, \boldsymbol{\varphi}_0')'$,

$$R_n(\boldsymbol{\tau}) = \frac{1}{n}\sum_{j=1}^{n} I_{\varepsilon(\boldsymbol{\tau})}(\lambda_j) = \frac{1}{n}\sum_{j=1}^{n} \left|\xi\left(e^{i\lambda_j}; \boldsymbol{\varphi}\right)\right|^2 I_{\varepsilon(\boldsymbol{\tau}^*)}(\lambda_j) + \frac{1}{n} V_n(\boldsymbol{\tau}),$$

where $\xi(s; \boldsymbol{\varphi}) = \theta(s; \boldsymbol{\varphi}_0)/\theta(s; \boldsymbol{\varphi}) = \sum_{j=0}^{\infty} \xi_j(\boldsymbol{\varphi}) s^j$. Then

$$\inf_{\mathcal{T}_3} R_n(\boldsymbol{\tau}) \ge \inf_{\substack{\lambda \in [-\pi,\pi] \\ \boldsymbol{\varphi} \in \Psi}} \left|\xi\left(e^{i\lambda}; \boldsymbol{\varphi}\right)\right|^2 \inf_{\delta \in \mathcal{I}_3} R_n(\boldsymbol{\tau}^*) - \sup_{\mathcal{T}_3} \frac{1}{n} |V_n(\boldsymbol{\tau})|. \tag{2.19}$$

A1 implies (see (2.3))

$$\inf_{\substack{\lambda \in [-\pi,\pi] \\ \boldsymbol{\varphi} \in \Psi}} \left|\xi\left(e^{i\lambda}; \boldsymbol{\varphi}\right)\right|^2 > \epsilon.$$

Thus

$$\begin{aligned} \inf_{\mathcal{T}_3} R_n(\boldsymbol{\tau}) \ &\ge\ \epsilon \inf_{\mathcal{I}_3} \frac{1}{n} \sum_{t=1}^{n} \left(\sum_{j=0}^{t-1} a_j \varepsilon_{t-j}\right)^2 \\ &\quad - \sup_{\mathcal{T}_3} \frac{1}{n} |V_n(\boldsymbol{\tau})| - \sup_{\mathcal{I}_3} \frac{1}{n} |W_n(\delta)|, \end{aligned} \tag{2.20}$$

where $a_j = a_j(\delta_0 - \delta)$, and by Lemma 2

$$W_n(\delta) = \epsilon \sum_{t=1}^{n} v_t^2(\delta) + 2\epsilon \sum_{t=1}^{n} v_t(\delta) \sum_{j=0}^{t-1} a_j \varepsilon_{t-j}.$$

By Lemma 2 and (0.6) in the proof of Lemma 3 in the supplementary material [14] (taking $\kappa = 1/2$ there in both cases)

$$\sup_{\mathcal{I}_3} \frac{1}{n} |W_n(\delta)| = O_p\left(n^{-1} + \frac{\log n}{n^{\frac{1}{2}}}\right) = o_p(1), \tag{2.21}$$

and also by Lemma 3 (with $\kappa = 1/2$ there)

$$\sup_{\mathcal{I}_3} \frac{1}{n} \left| V_n(\boldsymbol{\tau}) \right| = O_p\left(\frac{\log^2 n}{n} \right) = o_p(1). \tag{2.22}$$

Next, note that for $\delta \in \mathcal{I}_3$

$$\frac{\partial a_j^2}{\partial \delta} = -2\left(\psi(j + \delta_0 - \delta) - \psi(\delta_0 - \delta) \right) a_j^2 < 0, \tag{2.23}$$

where we introduce the digamma function $\psi(x) = (d/dx)\log\Gamma(x)$. From (2.23) and the fact that $\psi(x)$ is strictly increasing in $x > 0$,

$$\begin{aligned} \inf_{\mathcal{I}_3} n^{-1} \sum_{t=1}^{n} \left(\sum_{j=0}^{t-1} a_j \varepsilon_{t-j} \right)^2 &\geq n^{-1} \sum_{t=1}^{n} \sum_{j=0}^{t-1} a_j^2 \left(\frac{1}{2} - \eta \right) \varepsilon_{t-j}^2 \\ &\quad - \sup_{\mathcal{I}_3} \left| \frac{1}{n} \sum_{t=1}^{n} \sum\sum_{j \neq k}^{t-1} a_j a_k \varepsilon_{t-j} \varepsilon_{t-k} \right|. \end{aligned} \tag{2.24}$$

By a very similar analysis to that of (b) in (2.15), the second term on the right of (2.24) is bounded by

$$\begin{aligned} &\frac{2}{n} \sup_{\mathcal{I}_3} \left| \sum_{j=0}^{n-2} \sum_{k=j+1}^{n-1} a_j a_k \sum_{l=k-j+1}^{n-j} \varepsilon_l \varepsilon_{l-(k-j)} \right| \leq \frac{2}{n} \sup_{\mathcal{I}_3} \left| \sum_{j=0}^{n-2} a_j \sum_{k=j+1}^{n-1} \sum_{l=k-j+1}^{n-j} \varepsilon_l \varepsilon_{l-(k-j)} \right| \\ &+ \frac{2}{n} \sup_{\mathcal{I}_3} \left| \sum_{j=0}^{n-2} a_j \sum_{k=j+1}^{n-2} (a_{k+1} - a_k) \sum_{r=j+1}^{k} \sum_{l=r-j+1}^{n-j} \varepsilon_l \varepsilon_{l-(k-j)} \right|, \end{aligned}$$

which has expectation bounded by

$$\begin{aligned} &\frac{K}{n^{\frac{1}{2}}} \sum_{j=1}^{n} j^{-\frac{1}{2}} + \frac{K}{n^{\frac{1}{2}}} \sum_{j=1}^{n} j^{-\frac{1}{2}} \sum_{k=1}^{n} (k+j)^{-\frac{3}{2}} k^{\frac{1}{2}} \\ \leq\ & K \left(1 + \frac{1}{n^{\frac{1}{2}}} \sum_{j=1}^{n} j^{-\frac{1}{2}-a} \sum_{k=1}^{n} k^{-1+a} \right) \leq K, \end{aligned} \tag{2.25}$$

for any $0 < a < 1/2$. Therefore, there exists a large enough K such that

$$\Pr\left(\sup_{\mathcal{I}_3} \left| n^{-1} \sum_{t=1}^{n} \sum\sum_{j \neq k}^{t-1} a_j a_k \varepsilon_{t-j} \varepsilon_{t-k} \right| > K \right) \to 0, \tag{2.26}$$

as $n \to \infty$. Then, noting (2.20), (2.21), (2.22), (2.26), we deduce (2.18) for $i = 3$ if

$$\Pr\left(\frac{1}{n}\sum_{t=1}^{n}\sum_{j=0}^{t-1} a_j^2\left(\frac{1}{2}-\eta\right)\varepsilon_{t-j}^2 \le K\right) \to 0 \text{ as } n \to \infty. \tag{2.27}$$

Now

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\sum_{j=0}^{t-1} a_j^2\left(\frac{1}{2}-\eta\right)\varepsilon_{t-j}^2 &= \sigma_0^2\frac{\Gamma(2\eta)}{\Gamma^2\left(\frac{1}{2}+\eta\right)} + \frac{1}{n}\sum_{t=1}^{n}\sum_{j=0}^{t-1} a_j^2\left(\frac{1}{2}-\eta\right)\left(\varepsilon_{t-j}^2-\sigma_0^2\right) \\ &\quad -\frac{\sigma_0^2}{n}\sum_{t=1}^{n}\sum_{j=t}^{\infty} a_j^2\left(\frac{1}{2}-\eta\right).\end{aligned}$$

The third term on the right is clearly $O\left(n^{-2\eta}\right)$, whereas, as in the treatment of (a) in (2.15), the second is $O_p\left(n^{-1/2}\right)$, so that (2.27) holds as $\Gamma(2\eta)/\Gamma^2\left(\frac{1}{2}+\eta\right)$ can be made arbitrarily large for small enough η. This proves (2.18), and thus (2.7), for $i = 3$.

Proof of (2.7) for $i = 2$. Take $\eta < 1/4$ and note that $\mathcal{I}_2 \subset [\delta_0 - \kappa, \delta_0 - \frac{1}{2} + \eta)$ for $\kappa = \eta + \frac{1}{2}$. It follows from Lemma 2 and (0.6) in the proof of Lemma 3 (see supplementary material [14]) that

$$\begin{aligned}\sup_{\mathcal{I}_2}\frac{1}{n}\left|W_n(\delta)\right| &= O_p\left(\frac{1}{n}\sum_{t=1}^{n} t^{2\eta-1} + \frac{1}{n}\sum_{t=1}^{n} t^{\eta-\frac{1}{2}}t^{\eta}\right) \\ &= O_p\left(n^{2\eta-\frac{1}{2}}\right) = o_p(1).\end{aligned} \tag{2.28}$$

It follows from Lemma 3 that

$$\sup_{\mathcal{I}_2}\frac{1}{n}\left|V_n(\boldsymbol{\tau})\right| = O_p\left(n^{2\eta-1}\right) = o_p(1). \tag{2.29}$$

Denote $f_n(\delta) = n^{-1}\sum_{t=1}^{n}\left(\sum_{j=0}^{t-1} a_j\varepsilon_{t-j}\right)^2$. By (2.28), (2.29), it follows that (2.18) for $i = 2$ holds if for arbitrarily large K

$$\Pr\left(\inf_{\mathcal{I}_2} f_n(\delta) > K\right) \to 1, \tag{2.30}$$

as $n \to \infty$. Clearly,

$$\inf_{\mathcal{I}_2} f_n(\delta) \ge \inf_{\mathcal{I}_2}\frac{n^{2(\delta_0-\delta)}}{n}\inf_{\mathcal{I}_2}\frac{1}{n^{2(\delta_0-\delta)}}\sum_{t=1}^{n}\left(\sum_{j=0}^{t-1} a_j\varepsilon_{t-j}\right)^2. \tag{2.31}$$

Defining $b_{j,n}(d) = a_j(d)/n^{d-1}$, $b_{j,n} = b_{j,n}(\delta_0 - \delta)$, the right side of (2.31) is bounded below by

$$\inf_{\mathcal{I}_2} \frac{1}{n^2} \sum_{j=0}^{n-1} b_{j,n}^2 \sum_{l=1}^{n-j} \varepsilon_l^2 - \sup_{\mathcal{I}_2} \frac{2}{n^2} \left| \sum_{j=0}^{n-2} \sum_{k=j+1}^{n-1} b_{j,n} b_{k,n} \sum_{l=k-j+1}^{n-j} \varepsilon_l \varepsilon_{l-(k-j)} \right|. \tag{2.32}$$

For $1 \le j \le n$,

$$\inf_{\mathcal{I}_2} b_{j,n} \ge \inf_{\mathcal{I}_2} \frac{\epsilon}{\Gamma(\delta_0 - \delta)} \inf_{\mathcal{I}_2} \left(\frac{j}{n}\right)^{\delta_0 - \delta - 1} \ge \frac{\epsilon}{\Gamma\left(\frac{1}{2} + \eta\right)} \left(\frac{j}{n}\right)^{\eta - \frac{1}{2}},$$

$$\sup_{\mathcal{I}_2} b_{j,n} \le \sup_{\mathcal{I}_2} \frac{K}{\Gamma(\delta_0 - \delta)} \sup_{\mathcal{I}_2} \left(\frac{j}{n}\right)^{\delta_0 - \delta - 1} \le \frac{K}{\sqrt{\pi}} \left(\frac{j}{n}\right)^{-\frac{1}{2}}. \tag{2.33}$$

Then by (2.33), using summation by parts as in the analysis of (b) in (2.15), the expectation of the second term in (2.32) is bounded by

$$\frac{K}{n} \sum_{j=1}^{n} \left(\frac{j}{n}\right)^{-\frac{1}{2}} + \frac{K}{n^{\frac{1}{2}}} \sum_{j=1}^{n} j^{-\frac{1}{2}} \sum_{k=1}^{n} k^{\frac{1}{2}} (k+j)^{-\frac{3}{2}},$$

which, noting (2.25), is $O(1)$. Next, the first term in (2.32) is bounded below by

$$\frac{\sigma_0^2}{n^2} \sum_{j=0}^{n-1} (n-j) b_{j,n}^2 (1/2 + \eta) - \frac{1}{n^2} \sum_{j=0}^{n-1} b_{j,n}^2 (1/2) \left| \sum_{l=1}^{n-j} \left(\varepsilon_l^2 - \sigma_0^2\right) \right|. \tag{2.34}$$

Using (2.33) it can be easily shown that the second term in (2.34) is $O_p\left(n^{-\frac{3}{2}} \sum_{j=1}^{n} n/j\right) = O_p(n^{-\frac{1}{2}} \log n)$, whereas the first term is bounded below by

$$\begin{aligned} & \frac{\epsilon}{n} \sum_{j=1}^{n} \left\{ \left(\frac{j}{n}\right)^{2\eta - 1} - \left(\frac{j}{n}\right)^{2\eta} \right\} \\ \ge\ & \frac{\epsilon}{2} \int_{1/n}^{1} \left\{ x^{2\eta - 1} - x^{2\eta} \right\} dx = \frac{\epsilon}{2} \left[\frac{x^{2\eta}}{2\eta} - \frac{x^{2\eta+1}}{2\eta + 1} \right]_{1/n}^{1} \\ =\ & \frac{\epsilon}{4\eta(2\eta + 1)} - O_p(n^{-2\eta}). \end{aligned} \tag{2.35}$$

Then (2.30) holds because the right side of (2.35) can be made arbitrarily large on setting η arbitrarily close to zero. This proves (2.18), and thus (2.7), for $i = 2$.

Proof of (2.7) for $i = 1$. Noting that $R_n(\boldsymbol{\tau}) \geq n^{-2}\left(\sum_{t=1}^{n} \varepsilon_t(\boldsymbol{\tau})\right)^2$,

$$\Pr\left(\inf_{\mathcal{T}_1} R_n(\boldsymbol{\tau}) > K\right) \geq \Pr\left(n^{2\eta} \inf_{\mathcal{T}_1}\left(\frac{1}{n^{\delta_0-\delta+\frac{1}{2}}}\sum_{t=1}^{n}\varepsilon_t(\boldsymbol{\tau})\right)^2 > K\right), \tag{2.36}$$

because $\delta_0 - \delta \geq 1/2 + \eta$. Clearly $\sum_{t=1}^{n}\varepsilon_t(\boldsymbol{\tau}) = \sum_{j=0}^{n-1} d_j(\boldsymbol{\tau}) u_{n-j}$, where

$$d_j(\boldsymbol{\tau}) = \sum_{k=0}^{j} c_k(\boldsymbol{\tau}) = \sum_{k=0}^{j}\phi_k(\boldsymbol{\varphi})\sum_{l=0}^{j-k} a_l(\delta_0-\delta) = \sum_{k=0}^{j}\phi_k(\boldsymbol{\varphi}) a_{j-k}(\delta_0-\delta+1).$$

For arbitrarily small $\epsilon > 0$, the right side of (2.36) is bounded from below by

$$\Pr\left(\inf_{\mathcal{T}_1}\left(\frac{1}{n^{\delta_0-\delta+\frac{1}{2}}}\sum_{t=1}^{n}\varepsilon_t(\boldsymbol{\tau})\right)^2 > \epsilon\right), \tag{2.37}$$

for n large enough, so it suffices to show (2.37)$\rightarrow 1$ as $n \to \infty$. First

$$\frac{1}{n^{\delta_0-\delta+\frac{1}{2}}}\sum_{t=1}^{n}\varepsilon_t(\boldsymbol{\tau}) = \phi(1;\boldsymbol{\varphi})\,\theta(1;\boldsymbol{\varphi}_0)\,h_n(\delta) + r_n(\boldsymbol{\tau}),$$

where $h_n(\delta) = n^{-1/2}\sum_{j=0}^{n-1} b_{j,n}(\delta_0-\delta+1)\,\varepsilon_{n-j}$, $b_{j,n}(\cdot)$ was defined below (2.31), and

$$\begin{aligned} r_n(\boldsymbol{\tau}) &= -\frac{1}{n^{\frac{1}{2}}}\sum_{j=0}^{n-1} b_{j,n}(\delta_0-\delta+1)\sum_{k=j+1}^{\infty}\phi_k(\boldsymbol{\varphi})\,u_{n-j} - \frac{1}{n^{\frac{1}{2}}}\sum_{j=1}^{n-1} s_{j,n}(\boldsymbol{\tau})\,u_{n-j} \\ &\quad + \frac{\phi(1;\boldsymbol{\varphi})}{n^{\frac{1}{2}}}\sum_{j=0}^{n-1} b_{j,n}(\delta_0-\delta+1)\left(u_{n-j} - \theta(1;\boldsymbol{\varphi}_0)\,\varepsilon_{n-j}\right), \end{aligned} \tag{2.38}$$

for

$$s_{j,n}(\boldsymbol{\tau}) = \sum_{k=0}^{j-1}\left(b_{k+1,n}(\delta_0-\delta+1) - b_{k,n}(\delta_0-\delta+1)\right)\sum_{l=0}^{k}\phi_{j-l}(\boldsymbol{\varphi}),$$

where (2.38) is routinely derived, noting that by summation by parts

$$d_j(\boldsymbol{\tau}) = a_j(\delta_0-\delta+1)\sum_{k=0}^{j}\phi_k(\boldsymbol{\varphi}) - \sum_{k=0}^{j-1}\left(a_{k+1}(\delta_0-\delta+1) - a_k(\delta_0-\delta+1)\right)\sum_{l=0}^{k}\phi_{j-l}(\boldsymbol{\varphi}).$$

Now

$$\begin{aligned}\inf_{\mathcal{I}_1}\left(\frac{1}{n^{\delta_0-\delta+\frac{1}{2}}}\sum_{t=1}^{n}\varepsilon_t(\boldsymbol{\tau})\right)^2 &\geq \theta^2(1;\boldsymbol{\varphi}_0)\inf_{\boldsymbol{\Psi}}\phi^2(1;\boldsymbol{\varphi})\inf_{\mathcal{I}_1}h_n^2(\delta)\\ &\quad -K\sup_{\boldsymbol{\Psi}}|\phi(1;\boldsymbol{\varphi})|\sup_{\mathcal{I}_1}|h_n(\delta)|\sup_{\mathcal{I}_1}|r_n(\boldsymbol{\tau})|.\end{aligned}$$

Noting (2.3) and that under A1, $\sup_{\boldsymbol{\Psi}}|\phi(1;\boldsymbol{\varphi})|<\infty$, the required result follows on showing that

$$\sup_{\mathcal{I}_1}|r_n(\boldsymbol{\tau})| = o_p(1), \tag{2.39}$$

$$\sup_{\mathcal{I}_1}|h_n(\delta)| = O_p(1), \tag{2.40}$$

$$\Pr\left(\inf_{\mathcal{I}_1}h_n^2(\delta)>\epsilon\right) \to 1 \tag{2.41}$$

as $n\to\infty$.

The proof of (2.40) is omitted as it is similar to and much easier than the proof of (2.39), which we now give. Let $r_n(\boldsymbol{\tau})=\sum_{i=1}^{3}r_{in}(\boldsymbol{\tau})$. By the Cauchy inequality

$$\sup_{\mathcal{I}_1}|r_{1n}(\boldsymbol{\tau})|\leq\frac{1}{n^{\frac{1}{2}}}\left(\sum_{j=0}^{n-1}\sup_{\mathcal{I}_1}b_{j,n}^2(\delta_0-\delta+1)\left(\sup_{\Psi}\sum_{k=j+1}^{\infty}|\phi_k(\boldsymbol{\varphi})|\right)^2\sum_{j=1}^{n}u_j^2\right)^{\frac{1}{2}},$$

so that by (2.2), noting that $E\left(\sum_{j=1}^{n}u_j^2\right)^{1/2}\leq Kn^{1/2}$,

$$\begin{aligned}E\sup_{\mathcal{I}_1}|r_{1n}(\boldsymbol{\tau})| &\leq K\left(\sum_{j=1}^{n}\sup_{\mathcal{I}_1}\left(\frac{j}{n}\right)^{2(\delta_0-\delta)}\left(\sum_{k=j+1}^{\infty}k^{-1-\varsigma}\right)^2\right)^{\frac{1}{2}}\\ &\leq K\left(\sum_{j=1}^{n}\left(\frac{j}{n}\right)^{1+2\eta}j^{-2\varsigma}\right)^{\frac{1}{2}}\leq Kn^{\frac{1}{2}-\varsigma}=o(1),\end{aligned}$$

because $\varsigma>1/2$ by A1(ii). Next, by summation by parts

$$r_{2n}(\boldsymbol{\tau})=-\frac{s_{n-1,n}(\boldsymbol{\tau})}{n^{\frac{1}{2}}}\sum_{j=1}^{n-1}u_{n-j}+\frac{1}{n^{\frac{1}{2}}}\sum_{j=1}^{n-2}(s_{j+1,n}(\boldsymbol{\tau})-s_{j,n}(\boldsymbol{\tau}))\sum_{k=1}^{j}u_{n-k},$$

so

$$\sup_{\mathcal{I}_1} |r_{2n}(\boldsymbol{\tau})| \le \frac{\sup_{\mathcal{I}_1} |s_{n-1,n}(\boldsymbol{\tau})|}{n^{\frac{1}{2}}} \left| \sum_{j=1}^{n-1} u_{n-j} \right| + \frac{1}{n^{\frac{1}{2}}} \sum_{j=1}^{n-2} \sup_{\mathcal{I}_1} |s_{j+1,n}(\boldsymbol{\tau}) - s_{j,n}(\boldsymbol{\tau})| \left| \sum_{k=1}^{j} u_{n-k} \right|. \tag{2.42}$$

Given that $a_{k+1}(\delta_0 - \delta + 1) - a_k(\delta_0 - \delta + 1) = a_{k+1}(\delta_0 - \delta)$,

$$s_{j,n}(\boldsymbol{\tau}) = \frac{1}{n^{\delta_0 - \delta}} \sum_{k=0}^{j-1} a_{k+1}(\delta_0 - \delta) \sum_{l=0}^{k} \phi_{j-l}(\boldsymbol{\varphi}),$$

so as $E\left|\sum_{j=1}^{n-1} u_j\right| \le K n^{1/2}$, noting (2.2) and Stirling's approximation, the expectation of the first term on the right side of (2.42) is bounded by

$$K \sum_{k=1}^{n} \sup_{\mathcal{I}_1} \left(\frac{k}{n}\right)^{\delta_0 - \delta} k^{-1} \sum_{l=1}^{k} (n-l)^{-1-\varsigma} \le \frac{K}{n^{\frac{1}{2}+\eta}} \sum_{k=1}^{n} k^{-\frac{1}{2}+\eta} (n-k)^{-\frac{1}{2}} \le \frac{K}{n^{\frac{1}{2}}} \frac{1}{n} \sum_{k=1}^{n} \left(\frac{k}{n}\right)^{-\frac{1}{2}+\eta} \left(1 - \frac{k}{n}\right)^{-\frac{1}{2}} \le K n^{-\frac{1}{2}}.$$

Next, noting that $a_{j+1}(\delta_0 - \delta) - a_j(\delta_0 - \delta) = a_{j+1}(\delta_0 - \delta - 1)$, it can be shown that

$$s_{j+1,n}(\boldsymbol{\tau}) - s_{j,n}(\boldsymbol{\tau}) = \frac{1}{n^{\delta_0 - \delta}} \sum_{k=1}^{j} \phi_k(\boldsymbol{\varphi}) \sum_{l=j-k+2}^{j+1} a_l(\delta_0 - \delta - 1) + \frac{\phi_{j+1}(\boldsymbol{\varphi})}{n^{\delta_0 - \delta}} \sum_{l=1}^{j+1} a_l(\delta_0 - \delta). \tag{2.43}$$

Thus, noting that, uniformly in j, $E\left|\sum_{k=1}^{j} u_{n-k}\right| \le K j^{1/2}$, by previous arguments the contribution of the last term on the right side of (2.43) to the expectation of the second term on the right side of (2.42) is bounded by

$$\frac{K}{n^{\frac{1}{2}}} \sum_{j=1}^{n} j^{\frac{1}{2}} j^{-1-\varsigma} \sup_{\mathcal{I}_1} \left(\frac{j}{n}\right)^{\delta_0 - \delta} \le \frac{K}{n^{\frac{1}{2}}} \sum_{j=1}^{n} j^{-\frac{1}{2}-\varsigma} \left(\frac{j}{n}\right)^{\frac{1}{2}+\eta} \le K n^{-\varsigma}.$$

By identical arguments, the contribution of the first term on the right side of (2.43) to the expectation of the last term on the right side of (2.42) is bounded by

$$\frac{K}{n^{\frac{1}{2}}}\sum_{j=1}^{n} j^{\frac{1}{2}}\sum_{k=1}^{j-1} k^{-1-\varsigma}\sum_{l=j-k}^{j}\sup_{\mathcal{I}_1}\left(\frac{l}{n}\right)^{\delta_0-\delta} l^{-2}$$

$$\leq \frac{K}{n^{1+\eta}}\sum_{j=1}^{n} j^{\frac{1}{2}}\sum_{k=1}^{j-1} k^{-1-\varsigma}\sum_{l=j-k}^{j} l^{-\frac{3}{2}+\eta}. \tag{2.44}$$

Given that $\sum_{l=j-k}^{j} l^{-\frac{3}{2}+\eta} \leq K(j-k)^{-\frac{3}{2}+\eta} k$, the right side of (2.44) is bounded by

$$\frac{K}{n^{1+\eta}}\sum_{j=1}^{n} j^{\frac{1}{2}}\sum_{k=1}^{j-1} k^{-\varsigma}(j-k)^{-\frac{3}{2}+\eta}$$

$$\leq \frac{K}{n^{1+\eta}}\sum_{j=1}^{n} j^{\frac{1}{2}}\sum_{k=1}^{[j/2]} k^{-\varsigma}(j-k)^{-\frac{3}{2}+\eta}$$

$$+\frac{K}{n^{1+\eta}}\sum_{j=1}^{n} j^{\frac{1}{2}}\sum_{k=[j/2]+1}^{j-1} k^{-\varsigma}(j-k)^{-\frac{3}{2}+\eta}, \tag{2.45}$$

where $[\cdot]$ denotes integer part. Clearly, the right side of (2.45) is bounded by

$$\frac{K}{n^{1+\eta}}\sum_{j=1}^{n} j^{\frac{1}{2}}\left(j^{-\frac{3}{2}+\eta} j^{1-\varsigma} + j^{-\varsigma}\sum_{k=1}^{\infty} k^{-\frac{3}{2}+\eta}\right) \leq K\left(n^{-\varsigma} + n^{\frac{1}{2}-\varsigma-\eta}\right),$$

so $\sup_{\mathcal{T}_1}|r_{2n}(\boldsymbol{\tau})| = o_p(1)$ because $\varsigma > 1/2$. Next, writing $u_t = \theta(1;\boldsymbol{\varphi}_0)\varepsilon_t + \widetilde{\varepsilon}_{t-1} - \widetilde{\varepsilon}_t$, for $\widetilde{\varepsilon}_t = \sum_{j=0}^{\infty}\widetilde{\theta}_j(\boldsymbol{\varphi}_0)\varepsilon_{t-j}$, $\widetilde{\theta}_j(\boldsymbol{\varphi}_0) = \sum_{k=j+1}^{\infty}\theta_k(\boldsymbol{\varphi}_0)$, where, by A1, A2, $\widetilde{\varepsilon}_t$ is well defined in the mean square sense, we have

$$r_{3n}(\boldsymbol{\tau}) = -\frac{\phi(1;\boldsymbol{\varphi})}{n^{\delta_0-\delta+\frac{1}{2}}}\left(\sum_{j=0}^{n-1} a_j(\delta_0-\delta)\widetilde{\varepsilon}_{n-k} - a_{n-1}(\delta_0-\delta+1)\widetilde{\varepsilon}_0\right).$$

In view of previous arguments, it is straightforward to show that $\sup_{\mathcal{T}_1}|r_{3n}(\boldsymbol{\tau})| = o_p(1)$, to conclude the proof of (2.39).

Finally we prove (2.41). Considering $h_n(\delta)$ as a process indexed by δ, we show first that

$$h_n(\delta) \Rightarrow \int_0^1 \frac{(1-s)^{\delta_0-\delta}}{\Gamma(\delta_0-\delta+1)} dB(s), \tag{2.46}$$

where $B(s)$ is a scalar Brownian motion with variance σ_0^2 and $\Rightarrow$ means weak convergence in the space of continuous functions on $\mathcal{I}_1$. We give this space the uniform topology. Convergence of the finite dimensional distributions follows by Theorem 1 of [13], noting that A2 implies conditions A(i), A(ii) and A(iii) in [13] (in particular A2 implies that the fourth-order cumulant spectral density function of ε_t is bounded). Next, by Theorem 12.3 of [4], if for all fixed $\delta \in \mathcal{I}_1$ $h_n(\delta)$ is a tight sequence, and if for all $\delta_1, \delta_2 \in \mathcal{I}_1$ and for K not depending on δ_1, δ_2, n

$$E\left(h_n(\delta_1) - h_n(\delta_2)\right)^2 \leq K\left(\delta_1 - \delta_2\right)^2, \tag{2.47}$$

then the process $h_n(\delta)$ is tight, and (2.46) would follow. First, for fixed δ, it is straightforward to show that $\sup_n E\left(h_n^2(\delta)\right) < \infty$, so $h_n(\delta)$ is uniformly integrable and therefore tight. Next

$$E\left(h_n(\delta_1) - h_n(\delta_2)\right)^2 = \frac{\sigma_0^2}{n}\sum_{j=0}^{n-1}\left(b_{j,n}(\delta_0 - \delta_1 + 1) - b_{j,n}(\delta_0 - \delta_2 + 1)\right)^2$$

$$= \frac{\sigma_0^2(\delta_1 - \delta_2)^2}{n}\sum_{j=0}^{n-1}\frac{\left(a_j'\left(\delta_0 - \overline{\delta} + 1\right) - a_j\left(\delta_0 - \overline{\delta} + 1\right)\log n\right)^2}{n^{2\left(\delta_0 - \overline{\delta}\right)}},$$

by the mean value theorem, where $\overline{\delta} = \overline{\delta}_n$ is an intermediate point between δ_1 and δ_2. As in Lemma D.1 of [22],

$$\begin{aligned} & a_j'\left(\delta_0 - \overline{\delta} + 1\right) - a_j\left(\delta_0 - \overline{\delta} + 1\right)\log n \\ = & \left(\psi\left(j + \delta_0 - \overline{\delta} + 1\right) - \psi\left(\delta_0 - \overline{\delta} + 1\right) - \log n\right) a_j\left(\delta_0 - \overline{\delta} + 1\right). \end{aligned}$$

Now (2.47) holds on showing that, for $\overline{\delta} \in \mathcal{I}_1$,

$$\frac{\psi^2\left(\delta_0 - \overline{\delta} + 1\right)}{n}\sum_{j=0}^{n-1} b_{j,n}^2\left(\delta_0 - \overline{\delta} + 1\right) \leq K, \tag{2.48}$$

$$\frac{1}{n}\sum_{j=0}^{n-1}\left(\psi\left(j + \delta_0 - \overline{\delta} + 1\right) - \log n\right)^2 b_{j,n}^2\left(\delta_0 - \overline{\delta} + 1\right) \leq K. \tag{2.49}$$

By Stirling's approximation, the left side of (2.48) is bounded by

$$K\frac{\psi^2(\delta_0 - \nabla_1 + 1)}{n}\sum_{j=1}^{n}\sup_{\mathcal{I}_1}\left(\frac{j}{n}\right)^{2(\delta_0 - \delta)} \leq K\frac{\psi^2(\delta_0 - \nabla_1 + 1)}{n}\sum_{j=1}^{n}\sup_{\mathcal{I}_1}\left(\frac{j}{n}\right)^{1+2\eta} \leq K.$$

Regarding (2.49), it can be shown that uniformly in $\mathcal{I}_1$, $\psi\left(j+\delta_0-\overline{\delta}+1\right)=\log j+O\left(j^{-1}\right)$ (see, e.g. [2], p.259). Thus, apart from a remainder term of smaller order, the left side of (2.49) is bounded by

$$K\frac{1}{n}\sum_{j=1}^{n}\left(\log\frac{j}{n}\right)^2 b_{j,n}^2\left(\delta_0-\overline{\delta}+1\right)\leq K\frac{1}{n}\sum_{j=1}^{n}\left(\log\frac{j}{n}\right)^2\left(\frac{j}{n}\right)^{1+2\eta}, \tag{2.50}$$

uniformly in $\mathcal{I}_1$, the right side of (2.50) being bounded by $K\int_0^1(\log x)^2\,dx=2K$, to conclude the proof of tightness. Then by the continuous mapping theorem

$$\inf_{\mathcal{I}_1}h_n^2\left(\delta\right)\to_d\inf_{\mathcal{I}_1}\left(\int_0^1\frac{(1-s)^{\delta_0-\delta}}{\Gamma\left(\delta_0-\delta+1\right)}dB\left(s\right)\right)^2.$$

This is a.s. positive because the quantity whose infimum is taken is a χ_1^2 random variable times $\sigma_0^2/\left[\left\{2\left(\delta_0-\delta\right)+1\right\}\Gamma\left(\delta_0-\delta+1\right)^2\right]$, which is bounded away from zero on $\mathcal{I}_1$. Thus as $n\to\infty$

$$\Pr\left(\inf_{\mathcal{I}_1}h_n^2\left(\delta\right)>\epsilon\right)\to\Pr\left(\inf_{\mathcal{I}_1}\left(\int_0^1\frac{(1-s)^{\delta_0-\delta}}{\Gamma\left(\delta_0-\delta+1\right)}dB\left(s\right)\right)^2>\epsilon\right),$$

and (2.41) follows as ϵ is arbitrarily small. Then we conclude (2.18), and thus (2.7), for $i=1$. □

2.2. *Asymptotic normality of* $\widehat{\boldsymbol{\tau}}$. This requires an additional regularity condition.

A3. (i)

$$\boldsymbol{\tau}_0\in int\mathcal{T};$$

- **(ii)** for all λ, $\theta\left(e^{i\lambda};\boldsymbol{\varphi}\right)$ is twice continuously differentiable in $\boldsymbol{\varphi}$ on a closed neighbourhood $\mathcal{N}_\epsilon\left(\boldsymbol{\varphi}_0\right)$ of radius $0<\epsilon<1/2$ about $\boldsymbol{\varphi}_0$;
- **(iii)** the matrix

$$\mathbf{A}=\begin{pmatrix}\pi^2/6 & -\sum_{j=1}^{\infty}\mathbf{b}_j'\left(\boldsymbol{\varphi}_0\right)/j\\ -\sum_{j=1}^{\infty}\mathbf{b}_j\left(\boldsymbol{\varphi}_0\right)/j & \sum_{j=1}^{\infty}\mathbf{b}_j\left(\boldsymbol{\varphi}_0\right)\mathbf{b}_j'\left(\boldsymbol{\varphi}_0\right)\end{pmatrix}$$

is non-singular, where $\mathbf{b}_j\left(\boldsymbol{\varphi}_0\right)=\sum_{k=0}^{j-1}\theta_k\left(\boldsymbol{\varphi}_0\right)\partial\phi_{j-k}\left(\boldsymbol{\varphi}_0\right)/\partial\boldsymbol{\varphi}$.

By compactness of $\mathcal{N}_\epsilon(\boldsymbol{\varphi}_0)$ and continuity of $\partial\phi_j(\boldsymbol{\varphi})/\partial\varphi_i$, $\partial^2\phi_j(\boldsymbol{\varphi})/\partial\varphi_i\partial\varphi_l$, for all j, with $i,l=1,...,p$, where φ_i is the i-th element of $\boldsymbol{\varphi}$, A1(ii), A1(iv) and A3(ii) imply that, as $j\to\infty$

$$\sup_{\boldsymbol{\varphi}\in\mathcal{N}_\epsilon(\boldsymbol{\varphi}_0)}\left|\frac{\partial\phi_j(\boldsymbol{\varphi})}{\partial\varphi_i}\right| = O\left(j^{-(1+\varsigma)}\right), \qquad \sup_{\boldsymbol{\varphi}\in\mathcal{N}_\epsilon(\boldsymbol{\varphi}_0)}\left|\frac{\partial^2\phi_j(\boldsymbol{\varphi})}{\partial\varphi_i\partial\varphi_l}\right| = O\left(j^{-(1+\varsigma)}\right),$$

which again is satisfied in the ARMA case. As with A1, A3 is similar to conditions employed under stationarity, and can readily be checked in general.

THEOREM 2.2. *Let (1.1), (1.3) and A1-A3 hold. Then as $n\to\infty$*

$$n^{\frac{1}{2}}\left(\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0\right)\to_d N\left(0,\mathbf{A}^{-1}\right). \tag{2.51}$$

PROOF. The proof standardly involves use of the mean value theorem, approximation of a score function by a martingale so as to apply a martingale convergence theorem, and convergence in probability of a Hessian in a neighbourhood of $\boldsymbol{\tau}_0$. From the mean value theorem, (2.51) follows if we prove that

$$\frac{\sqrt{n}}{2}\frac{\partial R_n(\boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}}\to_d N\left(0,\sigma_0^4\mathbf{A}\right), \tag{2.52}$$

$$\frac{1}{2}\frac{\partial^2 R_n(\overline{\boldsymbol{\tau}})}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}\to_p \sigma_0^2\mathbf{A}, \tag{2.53}$$

where $\|\overline{\boldsymbol{\tau}}-\boldsymbol{\tau}_0\|\le\|\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0\|$.
Proof of (2.52). It suffices to prove

$$\frac{\sqrt{n}}{2}\frac{\partial R_n(\boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}}-\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\varepsilon_t\sum_{j=1}^{\infty}\mathbf{m}_j(\boldsymbol{\varphi}_0)\varepsilon_{t-j}=o_p(1) \tag{2.54}$$

and

$$\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\varepsilon_t\sum_{j=1}^{\infty}\mathbf{m}_j(\boldsymbol{\varphi}_0)\varepsilon_{t-j}\to_d N\left(0,\sigma_0^4\mathbf{A}\right), \tag{2.55}$$

where $\mathbf{m}_j(\boldsymbol{\varphi}_0)=\left(-j^{-1},\mathbf{b}_j'(\boldsymbol{\varphi}_0)\right)'$. By Lemma 2, the left side of (2.54) is the $(p+1)\times 1$ vector $\left(r_1+r_2+r_3,(\mathbf{s}_1+\mathbf{s}_2)'\right)'$, where

$$\begin{aligned}
r_1 &= \frac{1}{\sqrt{n}}\sum_{t=2}^{n}\varepsilon_t\sum_{j=t}^{\infty}\frac{1}{j}\varepsilon_{t-j}, \quad r_2=\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\varepsilon_t\sum_{j=1}^{t-1}\frac{1}{j}\sum_{k=t-j}^{\infty}\phi_k\left(\boldsymbol{\varphi}_0\right)u_{t-j-k},\\
r_3 &= -\frac{1}{\sqrt{n}}\sum_{t=2}^{n}v_t\left(\delta_0\right)\sum_{j=1}^{t-1}\frac{1}{j}\sum_{k=0}^{t-j-1}\phi_k\left(\boldsymbol{\varphi}_0\right)u_{t-j-k}, \quad \mathbf{s}_1=\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\varepsilon_t\sum_{j=t}^{\infty}\frac{\partial\phi_j\left(\boldsymbol{\varphi}_0\right)}{\partial\boldsymbol{\varphi}}u_{t-j},\\
\mathbf{s}_2 &= \frac{1}{\sqrt{n}}\sum_{t=2}^{n}v_t\left(\delta_0\right)\sum_{j=1}^{t-1}\frac{\partial\phi_j\left(\boldsymbol{\varphi}_0\right)}{\partial\boldsymbol{\varphi}}u_{t-j}.
\end{aligned}$$

Clearly, $E\left(r_1\right)=0$, and

$$Var\left(r_1\right)=\frac{1}{n}\sum_{t=2}^{n}\sum_{j=t}^{\infty}\sum_{s=2}^{n}\sum_{k=s}^{\infty}\frac{1}{jk}E\left(\varepsilon_t\varepsilon_s\varepsilon_{t-j}\varepsilon_{s-k}\right)=\frac{\sigma_0^4}{n}\sum_{t=2}^{n}\sum_{j=t}^{\infty}\frac{1}{j^2}=O\left(\frac{\log n}{n}\right),$$

noting that, by A2, the ε_t and $\varepsilon_t^2-\sigma_0^2$ are martingale difference sequences. Thus, $r_1=O_p\left(n^{-1/2}\log^{1/2}n\right)$. Next, $E\left(r_2\right)=0$, and $Var\left(r_2\right)$ equals

$$\frac{1}{n}\sum_{t=2}^{n}\sum_{j=1}^{t-1}\sum_{k=t-j}^{\infty}\sum_{s=2}^{n}\sum_{l=1}^{s-1}\sum_{m=s-l}^{\infty}\frac{\phi_k\left(\boldsymbol{\varphi}_0\right)\phi_m\left(\boldsymbol{\varphi}_0\right)}{jl}E\left(\varepsilon_t\varepsilon_s u_{t-j-k}u_{s-l-m}\right). \tag{2.56}$$

From (1.3) and A2, the expectation is $\sigma_0^2\gamma_{j+k-\ell-m}$ for $s=t$, and zero otherwise. By A1, u_t has bounded spectral density. Thus, (2.56) is bounded by

$$\begin{aligned}
K\frac{1}{n}\sum_{t=2}^{n}\int_{-\pi}^{\pi}\left|\sum_{j=1}^{t-1}\sum_{k=t-j}^{\infty}\frac{\phi_k\left(\boldsymbol{\varphi}_0\right)}{j}e^{i(j+k)\mu}\right|^2 d\mu &\leq \frac{K}{n}\sum_{t=2}^{n}\sum_{j=1}^{t-1}\sum_{k=t-j}^{\infty}\sum_{l=1}^{t-1}\frac{\phi_k(\boldsymbol{\varphi}_0)\phi_{j+k-l}(\boldsymbol{\varphi}_0)}{jl}\\
&\leq \frac{K}{n}\sum_{t=2}^{n}\sum_{j=1}^{t-1}\sum_{k=t-j}^{\infty}\sum_{l=1}^{t-1}\frac{k^{-1-\varsigma}(j+k-l)^{-1-\varsigma}}{jl}\\
&\leq \frac{K}{n}\sum_{t=2}^{n}\sum_{l=1}^{t-1}\frac{(t-l)^{-1-\varsigma}}{l}\sum_{j=1}^{t-1}\frac{(t-j)^{-\varsigma}}{j}.
\end{aligned}$$

Now

$$\sum_{l=1}^{t-1}\frac{(t-l)^{-1-\varsigma}}{l}=\sum_{l=1}^{[t/2]}\frac{(t-l)^{-1-\varsigma}}{l}+\sum_{l=[t/2]+1}^{t-1}\frac{(t-l)^{-1-\varsigma}}{l}\leq K\left(t^{-1-\varsigma}\log t+t^{-1}\right)\leq\frac{K}{t}.$$

Then $Var(r_2) = O\left(n^{-1}\sum_{t=2}^{n} t^{-1}\sum_{j=1}^{t-1} j^{-1}\right) = O\left(n^{-1}\log^2 n\right)$, so $r_2 = O_p\left(\log n/n^{1/2}\right)$. Next, by Lemma 2

$$r_3 = O_p\left(n^{-\frac{1}{2}}\sum_{t=2}^{n} t^{-\frac{1}{2}-\varsigma}\log t\right) = O_p\left(n^{-\frac{1}{2}}\right).$$

Also, $E(\mathbf{s}_1) = 0$ and

$$\begin{aligned} Var(\mathbf{s}_1) &= O\left(\left\|\frac{1}{n}\sum_{t=2}^{n}\sum_{j=t}^{\infty}\sum_{k=t}^{\infty}\frac{\partial\phi_j(\boldsymbol{\varphi}_0)}{\partial\boldsymbol{\varphi}}\frac{\partial\phi_k(\boldsymbol{\varphi}_0)}{\partial\boldsymbol{\varphi}'}E(u_{t-j}u_{t-k})\right\|\right) \\ &= O\left(\frac{1}{n}\sum_{t=2}^{n}\int_{-\pi}^{\pi}\left\|\sum_{j=t}^{\infty}\frac{\partial\phi_j(\boldsymbol{\varphi}_0)}{\partial\boldsymbol{\varphi}}e^{ij\lambda}\right\|^2 d\lambda\right) \\ &= O\left(\frac{1}{n}\sum_{t=2}^{n}\sum_{j=t}^{\infty}\left\|\frac{\partial\phi_j(\boldsymbol{\varphi}_0)}{\partial\boldsymbol{\varphi}}\right\|^2\right) = O\left(\frac{1}{n}\sum_{t=2}^{n} t^{-1-2\varsigma}\right) = O\left(n^{-1}\right), \end{aligned}$$

since $\varsigma > 1/2$, $\|\cdot\|$ denoting Euclidean norm. Finally, by Lemmas 2 and 4

$$\mathbf{s}_2 = O_p\left(n^{-\frac{1}{2}}\sum_{t=1}^{n} t^{-\frac{1}{2}-\varsigma}\right) = O_p(n^{-\frac{1}{2}}),$$

to conclude the proof of (2.54).

Next, (2.55) holds by the Cramer-Wold device and, for example, Theorem 1 of [7] on showing that

$$E\left(\varepsilon_t\sum_{j=1}^{\infty}\mathbf{m}_j(\boldsymbol{\varphi}_0)\varepsilon_{t-j}\middle|\mathcal{F}_{t-1}\right) = 0 \qquad \text{a.s.,} \tag{2.57}$$

and

$$\frac{1}{n}\sum_{t=2}^{n}E\left(\varepsilon_t^2\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\mathbf{m}_j(\boldsymbol{\varphi}_0)\mathbf{m}_k'(\boldsymbol{\varphi}_0)\varepsilon_{t-j}\varepsilon_{t-k}\middle|\mathcal{F}_{t-1}\right)$$

$$-\frac{1}{n}\sum_{t=2}^{n}E\left(\varepsilon_t^2\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\mathbf{m}_j(\boldsymbol{\varphi}_0)\mathbf{m}_k'(\boldsymbol{\varphi}_0)\varepsilon_{t-j}\varepsilon_{t-k}\right) \to_p 0, \tag{2.58}$$

because $E\left(\varepsilon_t^2\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\mathbf{m}_j(\boldsymbol{\varphi}_0)\mathbf{m}_k'(\boldsymbol{\varphi}_0)\varepsilon_{t-j}\varepsilon_{t-k}\middle|\mathcal{F}_{t-1}\right)$ has expectation $\sigma_0^2\mathbf{A}$, noting that the Lindeberg condition is satisfied as $\varepsilon_t\sum_{j=1}^{\infty}\mathbf{m}_j(\boldsymbol{\varphi}_0)\varepsilon_{t-j}$

is stationary with finite variance. Now (2.57) follows as ε_{t-j}, $j \geq 1$ is $\mathcal{F}_{t-1}$-measurable, whereas the left side of (2.58) is

$$\frac{\sigma_0^2}{n}\sum_{t=2}^{n}\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\mathbf{m}_j\left(\boldsymbol{\varphi}_0\right)\mathbf{m}_k'\left(\boldsymbol{\varphi}_0\right)\left(\varepsilon_{t-j}\varepsilon_{t-k}-E\left(\varepsilon_{t-j}\varepsilon_{t-k}\right)\right)\to_p 0,$$

because $\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\mathbf{m}_j\left(\boldsymbol{\varphi}_0\right)\mathbf{m}_k'\left(\boldsymbol{\varphi}_0\right)\left(\varepsilon_{t-j}\varepsilon_{t-k}-E\left(\varepsilon_{t-j}\varepsilon_{t-k}\right)\right)$ is stationary ergodic with mean zero. This completes the proof of (2.55), and thus (2.52). Proof of (2.53). Denote by N_ϵ an open neighbourhood of radius $\epsilon < 1/2$ about $\boldsymbol{\tau}_0$, and

$$\text{(2.59)} \qquad \mathbf{A}_n\left(\boldsymbol{\tau}\right) = \frac{1}{n}\sum_{t=2}^{n}\sum_{j=0}^{t-1}\sum_{k=1}^{t-1}\left(c_j\frac{\partial^2 c_k}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}+\frac{\partial c_j}{\partial\boldsymbol{\tau}}\frac{\partial c_k}{\partial\boldsymbol{\tau}'}\right)\gamma_{k-j},$$

$$\text{(2.60)} \qquad \mathbf{A}\left(\boldsymbol{\tau}\right) = \sum_{j=0}^{\infty}\sum_{k=1}^{\infty}\left(c_j\frac{\partial^2 c_k}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}+\frac{\partial c_j}{\partial\boldsymbol{\tau}}\frac{\partial c_k}{\partial\boldsymbol{\tau}'}\right)\gamma_{k-j}.$$

Trivially,

$$\frac{1}{2}\frac{\partial^2 R_n\left(\overline{\boldsymbol{\tau}}\right)}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}=\frac{1}{2}\frac{\partial^2 R_n\left(\overline{\boldsymbol{\tau}}\right)}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}-\mathbf{A}_n\left(\overline{\boldsymbol{\tau}}\right)+\mathbf{A}_n\left(\overline{\boldsymbol{\tau}}\right)-\mathbf{A}\left(\overline{\boldsymbol{\tau}}\right)+\mathbf{A}\left(\overline{\boldsymbol{\tau}}\right)-\mathbf{A}\left(\boldsymbol{\tau}_0\right)+\mathbf{A}\left(\boldsymbol{\tau}_0\right).$$

Because $c_j(\boldsymbol{\tau}_0)=\phi_j(\boldsymbol{\tau}_0)$, it follows that $\sum_{j=0}^{\infty}c_j(\boldsymbol{\tau}_0)u_{t-j}=\varepsilon_t$, so the first term in $\mathbf{A}\left(\boldsymbol{\tau}_0\right)$ is identically zero. Also, as in the proof of (2.55), the second term of $\mathbf{A}\left(\boldsymbol{\tau}_0\right)$ is identically $\sigma_0^2\mathbf{A}$. Thus, given that by Slutzky's theorem and continuity of $\mathbf{A}\left(\boldsymbol{\tau}\right)$ at $\boldsymbol{\tau}_0$, $\mathbf{A}\left(\overline{\boldsymbol{\tau}}\right)-\mathbf{A}\left(\boldsymbol{\tau}_0\right)=o_p\left(1\right)$, (2.53) holds on showing

$$\text{(2.61)} \qquad \sup_{\boldsymbol{\tau}\in N_\epsilon}\left\|\frac{1}{2}\frac{\partial^2 R_n\left(\boldsymbol{\tau}\right)}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}-\mathbf{A}_n\left(\boldsymbol{\tau}\right)\right\| = o_p\left(1\right),$$

$$\text{(2.62)} \qquad \sup_{\boldsymbol{\tau}\in N_\epsilon}\left\|\mathbf{A}_n\left(\boldsymbol{\tau}\right)-\mathbf{A}\left(\boldsymbol{\tau}\right)\right\| = o_p\left(1\right),$$

for some $\epsilon > 0$, as $n\to\infty$. As $\epsilon < 1/2$, the proof for (2.61) is almost identical to that for (2.12), noting the orders in Lemma 4. To prove (2.62), we show that

$$\text{(2.63)} \qquad \sup_{\boldsymbol{\tau}\in N_\epsilon}\left\|\frac{1}{n}\sum_{t=2}^{n}\sum_{j=0}^{t-1}\sum_{k=1}^{t-1}c_j\frac{\partial^2 c_k}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}\gamma_{k-j}-\sum_{j=0}^{\infty}\sum_{k=1}^{\infty}c_j\frac{\partial^2 c_k}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}'}\gamma_{k-j}\right\|$$

is $o_p\left(1\right)$, the proof for the corresponding result concerning the difference between the second terms in (2.59), (2.60) being almost identical. By Lemma

4, (2.63) is bounded by

$$\frac{K}{n}\sum_{t=1}^{n}\sum_{j=1}^{t}\sum_{k=t+1}^{\infty} j^{\epsilon-1}k^{\epsilon-1}(k-j)^{-1-\varsigma}\log^2 k+\frac{K}{n}\sum_{t=1}^{n}\sum_{j=t}^{\infty} j^{2\epsilon-2}\log^2 j$$
$$+\frac{K}{n}\sum_{t=1}^{n}\sum_{j=t}^{\infty}\sum_{k=j+1}^{\infty} j^{\epsilon-1}k^{\epsilon-1}(k-j)^{-1-\varsigma}\log^2 k, \tag{2.64}$$

noting that (2.1) implies that $\gamma_j = O\left(j^{-1-\varsigma}\right)$. The first term in (2.64) is bounded by

$$\frac{K}{n}\sum_{t=1}^{n} t^{\epsilon}\sum_{k=t+1}^{\infty} k^{\epsilon+a-1}(k-t)^{-1-\varsigma} \leq \frac{K}{n}\sum_{t=1}^{n} t^{\epsilon}\sum_{k=1}^{\infty}(k+t)^{\epsilon+a-1}k^{-1-\varsigma}, \tag{2.65}$$

for any $a > 0$. Choosing a such that $2\epsilon + a < 1$, (2.65) is bounded by

$$\frac{K}{n}\sum_{t=1}^{n} t^{2\epsilon+a-1}\sum_{k=1}^{\infty} k^{-1-\varsigma} = O\left(n^{2\epsilon+a-1}\right) = o(1).$$

Similarly, the second term in (2.64) can be easily shown to be $o(1)$, whereas the third term is bounded by

$$\frac{K}{n}\sum_{t=1}^{n}\sum_{j=t}^{\infty} j^{2\epsilon+a-2}\sum_{k=j+1}^{\infty}(k-j)^{-1-\varsigma}, \tag{2.66}$$

for any $a > 0$, so choosing again a such that $2\epsilon+a < 1$, (2.66) is $O\left(n^{2\epsilon+a-1}\right) = o(1)$, to conclude the proof of (2.53), and thus of the theorem.

□

3. Multivariate extension. When observations on several related time series are available joint modelling can achieve efficiency gains. We consider a vector $\mathbf{x}_t = (x_{1t}, ..., x_{rt})'$ given by

$$\mathbf{x}_t = \mathbf{\Lambda}_0^{-1}\left\{\mathbf{u}_t \mathbb{1}(t>0)\right\},\ t = 0, \pm 1, ..., \tag{3.1}$$

where $\mathbf{u}_t = (u_{1t}, ..., u_{rt})'$,

$$\mathbf{u}_t = \mathbf{\Theta}(L; \boldsymbol{\varphi}_0)\,\boldsymbol{\varepsilon}_t,\ t = 0, \pm 1, ..., \tag{3.2}$$

in which $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, ..., \varepsilon_{rt})'$, $\boldsymbol{\varphi}_0$ is (as in the univariate case) a $p \times 1$ vector of short-memory parameters, $\mathbf{\Theta}(s;\boldsymbol{\varphi}) = \sum_{j=0}^{\infty}\mathbf{\Theta}_j(\boldsymbol{\varphi})s^j$, $\mathbf{\Theta}_0(\boldsymbol{\varphi}) = I_r$ for all $\boldsymbol{\varphi}$, and $\mathbf{\Lambda}_0 = diag\left(\Delta^{\delta_{01}}, ..., \Delta^{\delta_{0r}}\right)$, where the memory parameters δ_{0i}

are unknown real numbers. In general, they can all be distinct but for the sake of parsimony we allow for the possibility that they are known to lie in a set of dimension $q < r$. For example, perhaps as a consequence of pretesting, we might believe some or all the δ_{0i} are equal, and imposing this restriction in the estimation could further improve efficiency. We introduce known functions $\delta_i = \delta_i(\boldsymbol{\delta})$, $i = 1, ..., r$, of $q \times 1$ vector $\boldsymbol{\delta}$, such that for some $\boldsymbol{\delta}_0$ we have $\delta_{0i} = \delta_i(\boldsymbol{\delta}_0)$, $i = 1, ..., r$. We denote $\boldsymbol{\tau} = (\boldsymbol{\delta}', \boldsymbol{\varphi}')'$ and define (cf. (1.4))

$$\boldsymbol{\varepsilon}_t(\boldsymbol{\tau}) = \boldsymbol{\Theta}^{-1}(L; \boldsymbol{\varphi})\boldsymbol{\Lambda}(\boldsymbol{\delta})\,\mathbf{x}_t, \ t \geq 1,$$

where $\boldsymbol{\Lambda}(\boldsymbol{\delta}) = diag\left(\Delta^{\delta_1}, ..., \Delta^{\delta_r}\right)$. Gaussian likelihood considerations suggest the multivariate analogue to (1.6)

$$R_n^*(\boldsymbol{\tau}) = \det\{\boldsymbol{\Sigma}_n(\boldsymbol{\tau})\}, \tag{3.3}$$

where $\boldsymbol{\Sigma}_n(\boldsymbol{\tau}) = n^{-1}\sum_{t=1}^{n} \boldsymbol{\varepsilon}_t(\boldsymbol{\tau})\boldsymbol{\varepsilon}_t'(\boldsymbol{\tau})$, assuming that no prior restrictions link $\boldsymbol{\tau}_0$ with the covariance matrix of $\boldsymbol{\varepsilon}_t$. Unfortunately our consistency proof for the univariate case does not straightforwardly extend to an estimate minimizing (3.3) if $q > 1$. Also (3.3) is liable to pose a more severe computational challenge than (1.6) since p is liable to be larger in the multivariate case and q may exceed 1; it may be difficult to locate an approximate minimum of (3.3) as a preliminary to iteration. We avoid both these problems by taking a single Newton step from an initial $\sqrt{n}$-consistent estimate $\widetilde{\boldsymbol{\tau}}$. Defining

$$\begin{aligned} \mathbf{H}_n(\boldsymbol{\tau}) &= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial \boldsymbol{\varepsilon}_t(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}'}\right)' \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\tau})\frac{\partial \boldsymbol{\varepsilon}_t(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}'}, \\ \mathbf{h}_n(\boldsymbol{\tau}) &= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial \boldsymbol{\varepsilon}_t(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}'}\right)' \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\tau})\,\boldsymbol{\varepsilon}_t(\boldsymbol{\tau}), \end{aligned}$$

we consider the estimate

$$\widehat{\boldsymbol{\tau}} = \widetilde{\boldsymbol{\tau}} - \mathbf{H}_n^{-1}(\widetilde{\boldsymbol{\tau}})\mathbf{h}_n(\widetilde{\boldsymbol{\tau}}). \tag{3.4}$$

We collect together all the requirements for asymptotic normality of $\widehat{\boldsymbol{\tau}}$ in:

A4. (i) For all $\boldsymbol{\varphi}$, $\Theta\left(e^{i\lambda}; \boldsymbol{\varphi}\right)$ is differentiable in λ with derivative in $Lip(\varsigma)$, $\varsigma > 1/2$;

(ii) for all $\boldsymbol{\varphi}$, $\det\{\boldsymbol{\Theta}(s; \boldsymbol{\varphi})\} \neq 0$, $|s| \leq 1$;

(iii) the $\boldsymbol{\varepsilon}_t$ in (3.2) are stationary and ergodic with finite fourth moment, $E\left(\boldsymbol{\varepsilon}_t \middle| \mathcal{F}_{t-1}\right)=0$, $E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' \middle| \mathcal{F}_{t-1}\right)=\boldsymbol{\Sigma}_0$ almost surely, where $\boldsymbol{\Sigma}_0$ is positive definite, $\mathcal{F}_t$ is the σ-field of events generated by $\boldsymbol{\varepsilon}_s$, $s\leq t$, and conditional (on $\mathcal{F}_{t-1}$) third and fourth moments and cross-moments of elements of $\boldsymbol{\varepsilon}_t$ equal the corresponding unconditional moments;

(iv) for all λ, $\Theta\left(e^{i\lambda};\boldsymbol{\varphi}\right)$ is twice continuously differentiable in $\boldsymbol{\varphi}$ on a closed neighbourhood $\mathcal{N}_\epsilon\left(\boldsymbol{\varphi}_0\right)$ of radius $0<\epsilon<1/2$ about $\boldsymbol{\varphi}_0$;

(v) the matrix B having (i,j)th element

$$\sum_{k=1}^{\infty} tr\left\{\left(\mathbf{d}_k^{(i)}\left(\boldsymbol{\varphi}_0\right)\right)'\boldsymbol{\Sigma}_0^{-1}\mathbf{d}_k^{(j)}\left(\boldsymbol{\varphi}_0\right)\boldsymbol{\Sigma}_0\right\}$$

is non-singular, where

$$\begin{aligned}\mathbf{d}_k^{(i)}\left(\boldsymbol{\varphi}_0\right) &= -\frac{\partial\delta_i\left(\boldsymbol{\delta}_0\right)}{\partial\delta_i}\sum_{l=1}^{k}\frac{1}{l}\sum_{m=0}^{k-l}\boldsymbol{\Phi}_m^{(i)}\left(\boldsymbol{\varphi}_0\right)\boldsymbol{\Theta}_{k-l-m}\left(\boldsymbol{\varphi}_0\right), \quad 1\leq i\leq r,\\ &= \sum_{l=1}^{k}\frac{\partial\boldsymbol{\Phi}_l\left(\boldsymbol{\varphi}_0\right)}{\partial\varphi_i}\boldsymbol{\Theta}_{k-l}\left(\boldsymbol{\varphi}_0\right), \quad r+1\leq i\leq r+p,\end{aligned}$$

the $\Phi_j\left(\boldsymbol{\varphi}\right)$ being coefficients in the expansion $\boldsymbol{\Theta}^{-1}\left(s;\boldsymbol{\varphi}\right)=\boldsymbol{\Phi}\left(s,\boldsymbol{\varphi}\right)=\sum_{j=0}^{\infty}\boldsymbol{\Phi}_j\left(\boldsymbol{\varphi}\right)s^j$, where $\Phi_m^{(i)}\left(\boldsymbol{\varphi}_0\right)$ is an $r\times r$ matrix whose i-th column is the i-th column of $\Phi_i\left(\boldsymbol{\varphi}_0\right)$ and whose other elements are all zero;

(vi) $\delta_i\left(\boldsymbol{\delta}\right)$ is twice continuously differentiable in $\boldsymbol{\delta}$, for $i=1,...,r$;

(vii) $\widetilde{\boldsymbol{\tau}}$ is a $\sqrt{n}$-consistent estimate of $\boldsymbol{\tau}_0$.

The components of A4 are mostly natural extensions of ones in A1, A2 and A3, are equally checkable, and require no additional discussion. The important exception is (vii). When $\boldsymbol{\Theta}(s;\boldsymbol{\varphi})$ is a diagonal matrix (as in the simplest case $\boldsymbol{\Theta}(s;\boldsymbol{\varphi})\equiv\mathbf{I}_r$, when x_{it} is a FARIMA$(0,\delta_{0i},0)$ for $i=1,...,r$) then $\widetilde{\boldsymbol{\tau}}$ can be obtained by first carrying out r univariate fits following the approach of Section 2, and then if necessary reducing the dimensionality in a common-sense way: for example if some of the δ_{0i} are *a priori* equal then the common memory parameter might be estimated by the arithmetic mean of estimates from the relevant univariate fits. Notice that in the diagonal$-\boldsymbol{\Theta}$ case with no cross-equation parameter restrictions the efficiency improvement afforded by $\widehat{\boldsymbol{\tau}}$ is due solely to cross-correlation in $\boldsymbol{\varepsilon}_t$, i.e. non-diagonality of $\boldsymbol{\Sigma}_0$.

When $\boldsymbol{\Theta}(s;\boldsymbol{\varphi})$ is not diagonal it is less clear how to use the $\sqrt{n}$-consistent outcome of Theorem 2 to form $\widetilde{\boldsymbol{\tau}}$. We can infer that $\mathbf{u}_t$ has spectral density matrix $(2\pi)^{-1}\mathbf{\Theta}(e^{i\lambda};\boldsymbol{\varphi}_0)\mathbf{\Sigma}_0\mathbf{\Theta}(e^{-i\lambda};\boldsymbol{\varphi}_0)'$. From the i-th diagonal element of this (the power spectrum of u_{it}) we can deduce a form for the Wold representation of u_{it}, corresponding to (1.3). However, starting from innovations $\boldsymbol{\varepsilon}_t$ in (3.2) satisfying (iii) of A4, it does not follow in general that the innovations in the Wold representation of u_{it} will satisfy a condition analogous to (2.4) of A2, indeed it does not help if we simply strengthen A4 such that the $\boldsymbol{\varepsilon}_t$ are independent and identically distributed. However, (2.4) certainly holds if $\boldsymbol{\varepsilon}_t$ is Gaussian, which motivates our estimation approach from an efficiency perspective. Notice that if $\mathbf{u}_t$ is a vector ARMA process with non-diagonal $\mathbf{\Theta}$, in general all r univariate AR operators are identical, and of possibly high degree; the formation of $\widetilde{\boldsymbol{\tau}}$ is liable to be affected by a lack of parsimony, or some ambiguity.

An alternative approach could involve first estimating the δ_{0i} by some semiparametric approach, using these estimates to form differenced $\mathbf{x}_t$ and then estimating $\boldsymbol{\varphi}_0$ from these proxies for $\mathbf{u}_t$. This initial estimate will be less-than-$\sqrt{n}$-consistent, but its rate can be calculated given a rate for the bandwidth used in the semiparametric estimation. One can then calculate the (finite) number of iterations of form (3.4) needed to produce an estimate satisfying (2.51), following Theorem 5 and the discussion on p.539 of [17].

THEOREM 3.1. *Let (3.1), (3.2) and A4 hold. Then as* $n\to\infty$

$$n^{\frac{1}{2}}\left(\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0\right)\to_d N\left(\mathbf{0},\mathbf{B}^{-1}\right). \tag{3.5}$$

PROOF. Because $\widehat{\boldsymbol{\tau}}$ is explicitly defined in (3.4), we start, standardly, by approximating $h_n\left(\widetilde{\boldsymbol{\tau}}\right)$ by the mean value theorem. Then in view of A4 (vii), (3.5) follows on showing

$$\sqrt{n}\mathbf{h}_n(\boldsymbol{\tau}_0)\to_d N\left(\mathbf{0},\mathbf{B}\right), \tag{3.6}$$

$$\mathbf{H}_n(\boldsymbol{\tau}_0)\to_p \mathbf{B}, \tag{3.7}$$

$$\mathbf{H}_n\left(\overline{\boldsymbol{\tau}}\right)-\mathbf{H}_n\left(\boldsymbol{\tau}_0\right)\to_p 0, \tag{3.8}$$

for $\|\overline{\boldsymbol{\tau}}-\boldsymbol{\tau}_0\|\le\|\widetilde{\boldsymbol{\tau}}-\boldsymbol{\tau}_0\|$. We only show (3.6), as (3.7), (3.8) follow from similar arguments to those given in the proof of (2.53). Noting that $\partial\boldsymbol{\varepsilon}_1(\boldsymbol{\tau}_0)/\partial\boldsymbol{\tau}'=0$, whereas for $t\ge 2$, $\partial\boldsymbol{\varepsilon}_t(\boldsymbol{\tau}_0)/\partial\boldsymbol{\tau}'$ equals

$$\sum_{j=1}^{t-1}\left(-\mathbf{\Phi}_j^{(1)}\left(\boldsymbol{\varphi}_0\right)\sum_{k=1}^{t-j-1}\frac{\mathbf{u}_{t-j-k}}{k},...,-\mathbf{\Phi}_j^{(r)}\left(\boldsymbol{\varphi}_0\right)\sum_{k=1}^{t-j-1}\frac{\mathbf{u}_{t-j-k}}{k},\right.$$
$$\left.\frac{\partial\mathbf{\Phi}_j\left(\boldsymbol{\varphi}_0\right)}{\partial\varphi_1}\mathbf{u}_{t-j},...,\frac{\partial\mathbf{\Phi}_j\left(\boldsymbol{\varphi}_0\right)}{\partial\varphi_p}\mathbf{u}_{t-j}\right),$$

by similar arguments to those in the proof of Theorem 2, it can be shown that the left side of (3.6) equals

$$\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\left(\begin{array}{ccc}\sum_{j=1}^{\infty}\mathbf{d}_j^{(1)}(\boldsymbol{\varphi}_0)\,\varepsilon_{t-j} & \cdots & \sum_{j=1}^{\infty}\mathbf{d}_j^{(r+p)}(\boldsymbol{\varphi}_0)\,\varepsilon_{t-j}\end{array}\right)'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\varepsilon}_t+o_p(1).$$

Then by the Cramer-Wold device, (3.6) holds if for any $(r+p)$-dimensional vector $\boldsymbol{\vartheta}$ (with i-th component ϑ_i)

$$\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\sum_{j=1}^{\infty}\boldsymbol{\varepsilon}_{t-j}'\mathbf{M}_j'(\boldsymbol{\varphi}_0)\,\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\varepsilon}_t\to_d N\left(0,\boldsymbol{\vartheta}'\mathbf{B}\boldsymbol{\vartheta}\right), \tag{3.9}$$

where $\mathbf{M}_j(\boldsymbol{\varphi}_0)=\sum_{k=1}^{r+p}\vartheta_k\mathbf{d}_j^{(k)}(\boldsymbol{\varphi}_0)$. As in the proof of (2.55), (3.9) holds by Theorem 1 of [7], for example, noting that

$$\begin{aligned}
& E\left(\sum_{j=1}^{\infty}\boldsymbol{\varepsilon}_{t-j}'\mathbf{M}_j'(\boldsymbol{\varphi}_0)\,\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\varepsilon}_t\right)^2 \\
= \; & E\left(\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\boldsymbol{\varepsilon}_{t-j}'\mathbf{M}_j'(\boldsymbol{\varphi}_0)\,\boldsymbol{\Sigma}_0^{-1}E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right|\mathcal{F}_{t-1})\,\boldsymbol{\Sigma}_0^{-1}\mathbf{M}_k(\boldsymbol{\varphi}_0)\,\boldsymbol{\varepsilon}_{t-k}\right) \\
= \; & E\left(\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}tr\left\{\boldsymbol{\varepsilon}_{t-j}'\mathbf{M}_j'(\boldsymbol{\varphi}_0)\,\boldsymbol{\Sigma}_0^{-1}\mathbf{M}_k(\boldsymbol{\varphi}_0)\,\boldsymbol{\varepsilon}_{t-k}\right\}\right) \\
= \; & \sum_{j=1}^{\infty}tr\left\{\mathbf{M}_j'(\boldsymbol{\varphi}_0)\,\boldsymbol{\Sigma}_0^{-1}\mathbf{M}_j(\boldsymbol{\varphi}_0)\,\boldsymbol{\Sigma}_0\right\}=\boldsymbol{\vartheta}'\mathbf{B}\boldsymbol{\vartheta},
\end{aligned}$$

to conclude the proof. □

4. Further comments and extensions.

1. Our univariate and multivariate structures cover a wide range of parametric models for stationary and nonstationary time series, with memory parameters allowed to lie in a set that can be arbitrarily large. Unit root series are a special case, but unlike in the bulk of the large unit root literature, we do not have to assume knowledge that memory parameters are 1. Indeed, in Monte Carlo [14] our method out-performs one which correctly assumes the unit interval in which δ_0 lies, while in empirical examples our findings conflict with previous, unit root, ones.
2. As the nondiagonal structure of $\mathbf{A}$ and $\mathbf{B}$ suggests, there is efficiency loss in estimating $\boldsymbol{\varphi}_0$ if memory parameters are unknown, but on the other hand if these are misspecified, $\boldsymbol{\varphi}_0$ will in general be inconsistently

estimated. Our limit distribution theory can be used to test hypotheses on the memory and other parameters, after straightforwardly forming consistent estimates of $\mathbf{A}$ or $\mathbf{B}$.

3. Our multivariate system (3.1), (3.2) does not cover fractionally cointegrated systems because $\mathbf{\Sigma}_0$ is required to be positive definite. On the other hand our theory for univariate estimation should cover estimation of individual memory parameters, so long as Assumption A2, in particular, can be reconciled with the full system specification. Moreover, again on an individual basis, it should be possible to derive analogous properties of estimates of memory parameters of cointegrating errors based on residuals that use simple estimates of cointegrating vectors, such as least squares.
4. In a more standard regression setting, for example with deterministic regressors such as polynomial functions of time, it should be possible to extend our theory for univariate and multivariate models to residual-based estimates of memory parameters of errors.
5. Adaptive estimates, which have greater efficiency at distributions of unknown, non-Gaussian form, can be obtained by taking one Newton step from our estimates (as in [20]).
6. Our methods of proof should be extendable to cover seasonally and cyclically fractionally differenced processes.
7. Nonstationary fractional series can be defined in many ways. Our definition ((1.1) and (3.1)) is a leading one in the literature, and has been termed "Type II". Another popular one ("Type I") was used by [25] for an alternate type of estimate. That estimate assumes invertibility and is generally less efficient than $\widehat{\boldsymbol{\tau}}$ due to the tapering required to handle nonstationarity. It seems likely that the asymptotic theory derived in this paper for $\widehat{\boldsymbol{\tau}}$ can also be established in a "Type I" setting.

5. Technical Lemmas. The proofs of the following lemmas appear in [14].

LEMMA 1. *Under A1*

$$\varepsilon_t(\boldsymbol{\tau}) = \sum_{j=0}^{t-1} c_j(\boldsymbol{\tau})\, u_{t-j}, \tag{5.1}$$

with $c_0(\boldsymbol{\tau}) = 1$ *where for any* $\delta \in \mathcal{I}$, *as* $j \to \infty$,

$$\sup_{\boldsymbol{\varphi}\in\Psi} |c_j(\boldsymbol{\tau})| = O\left(j^{\max(\delta_0-\delta-1,-1-\varsigma)}\right),$$

$$\sup_{\boldsymbol{\varphi}\in\Psi} |c_{j+1}(\boldsymbol{\tau}) - c_j(\boldsymbol{\tau})| = O\left(j^{\max(\delta_0-\delta-2,-1-\varsigma)}\right). \tag{5.2}$$

Lemma 2. *Under A1, A2*

$$\varepsilon_t(\boldsymbol{\tau}^*) = \sum_{j=0}^{t-1} a_j \varepsilon_{t-j} + v_t(\delta),$$

$\boldsymbol{\tau}^* = (\delta, \boldsymbol{\varphi}_0)$ *and for any* $\kappa \geq 1/2$

$$\sup_{\delta_0 - \kappa \leq \delta < \delta_0 - \frac{1}{2} + \eta} |v_t(\delta)| = O_p\left(t^{\kappa-1}\right),$$

and $v_t(\delta_0) = O_p\left(t^{-1/2-\varsigma}\right)$.

Lemma 3. *Under A1, A2*

$$\sum_{j=1}^{n} I_{\varepsilon(\boldsymbol{\tau})}(\lambda_j) = \sum_{j=1}^{n} \left| \frac{\theta\left(e^{i\lambda_j}; \boldsymbol{\varphi}_0\right)}{\theta\left(e^{i\lambda_j}; \boldsymbol{\varphi}\right)} \right|^2 I_{\varepsilon(\boldsymbol{\tau}^*)}(\lambda_j) + V_n(\boldsymbol{\tau}), \tag{5.3}$$

where for any real number $\kappa \geq 1/2$

(5.4)
$$\sup_{\substack{\delta_0 - \kappa \leq \delta < \delta_0 - \frac{1}{2} + \eta \\ \boldsymbol{\varphi} \in \Psi}} |V_n(\boldsymbol{\tau})| = O_p\left(\log^2 n \mathbb{1}(\kappa = 1/2) + n^{2\kappa - 1} \mathbb{1}(\kappa > 1/2)\right).$$

Lemma 4. *Under A3, given an open neighbourhood* N_ϵ *of radius* $\epsilon < 1/2$ *about* $\boldsymbol{\tau}_0$, *as* $j \to \infty$,

$$\begin{aligned}
\sup_{\boldsymbol{\tau} \in N_\epsilon} |c_j(\boldsymbol{\tau})| &= O\left(j^{\epsilon-1}\right), \ \sup_{\boldsymbol{\tau} \in N_\epsilon} \left| \frac{\partial c_j(\boldsymbol{\tau})}{\partial \delta} \right| = O\left(j^{\epsilon-1} \log j\right), \\
\sup_{\boldsymbol{\tau} \in N_\epsilon} |c_{j+1}(\boldsymbol{\tau}) - c_j(\boldsymbol{\tau})| &= O\left(j^{\max(\epsilon-2, -1-\varsigma)}\right), \\
\sup_{\boldsymbol{\tau} \in N_\epsilon} \left| \frac{\partial}{\partial \delta} (c_{j+1}(\boldsymbol{\tau}) - c_j(\boldsymbol{\tau})) \right| &= O\left(j^{-1-\varsigma} + j^{\epsilon-2} \log j\right), \\
\sup_{\boldsymbol{\tau} \in N_\epsilon} \left| \frac{\partial^2 c_j(\boldsymbol{\tau})}{\partial \delta^2} \right| &= O\left(j^{\epsilon-1} \log^2 j\right), \ \sup_{\boldsymbol{\tau} \in N_\epsilon} \left\| \frac{\partial c_j(\boldsymbol{\tau})}{\partial \boldsymbol{\varphi}} \right\| = O\left(j^{\epsilon-1}\right), \\
\sup_{\boldsymbol{\tau} \in N_\epsilon} \left| \frac{\partial^2}{\partial \delta^2} (c_{j+1}(\boldsymbol{\tau}) - c_j(\boldsymbol{\tau})) \right| &= O\left(j^{-1-\varsigma} + j^{\epsilon-2} \log^2 j\right), \\
\sup_{\boldsymbol{\tau} \in N_\epsilon} \left\| \frac{\partial}{\partial \boldsymbol{\varphi}} (c_{j+1}(\boldsymbol{\tau}) - c_j(\boldsymbol{\tau})) \right\| &= O\left(j^{\max(\epsilon-2, -1-\varsigma)}\right), \\
\sup_{\boldsymbol{\tau} \in N_\epsilon} \left\| \frac{\partial^2 c_j(\boldsymbol{\tau})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'} \right\| &= O\left(j^{\epsilon-1}\right), \ \sup_{\boldsymbol{\tau} \in N_\epsilon} \left\| \frac{\partial^2 c_j(\boldsymbol{\tau})}{\partial \boldsymbol{\varphi} \partial \delta} \right\| = O\left(j^{\epsilon-1} \log j\right),
\end{aligned}$$

$$\sup_{\boldsymbol{\tau}\in N_\epsilon}\left\|\frac{\partial^2}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'}\left(c_{j+1}\left(\boldsymbol{\tau}\right)-c_j\left(\boldsymbol{\tau}\right)\right)\right\| = O\left(j^{\max(\epsilon-2,-1-\varsigma)}\right),$$
$$\sup_{\boldsymbol{\tau}\in N_\epsilon}\left\|\frac{\partial^2}{\partial\boldsymbol{\varphi}\partial\delta}\left(c_{j+1}\left(\boldsymbol{\tau}\right)-c_j\left(\boldsymbol{\tau}\right)\right)\right\| = O\left(j^{-1-\varsigma}+j^{\epsilon-2}\log j\right).$$

Acknowledgements. We thank the Associate Editor and two referees for constructive comments that have improved the presentation. We also thank Søren Johansen and Morten ∅. Nielsen for helpful comments. The first author's research was supported by the Spanish Ministerio de Ciencia e Innovación through a Ramón y Cajal contract and ref. ECO2008-02641. The second author's research was supported by ESRC Grant RES-062-23-0036 and Spanish Plan Nacional deI+D+I Grant SEJ2007-62908/ECON, and some work was carried out while visiting Universidad Carlos III, Madrid, holding a Cátedra de Excelencia.

SUPPLEMENTARY MATERIAL

Supplement to "Gaussian Pseudo-Maximum Likelihood Estimation of Fractional Time Series Models." The supplementary material contains a Monte Carlo experiment of finite sample performance of the proposed procedure, an empirical application to US income and consumption data, and the proofs of the lemmas given in Section 5 of the present paper.

References.

[1] Adensted, R.K. (1974). On large-sample estimation for the mean of a stationary random sequence. *Annals of Statistics* **2**, 1095–1107.
[2] Abramowitz, M. and Stegun, I. (1970). *Handbook of mathematical functions.* Dover, New York.
[3] Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society, Ser. B.* **57**, 659–672.
[4] Billingsley, P. (1968). *Convergence of probability measures.* Wiley, New York.
[5] Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60**, 217-226.
[6] Box, G.E.P. and Jenkins, G.M. (1971). *Time series analysis. Forecasting and control.* Holden-Day, San Francisco.
[7] Brown, B.M. (1971). Martingale central limit theorems. *Annals of Mathematical Statistics* **42**, 59–66.
[8] Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics* **17**, 1749–1766.
[9] Fox, R. and Taqqu, M.S. (1986). Large-sample properties of parameter estimates for strongly dependent processes. *Annals of Statistics* **14**, 517–532.
[10] Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate. *Probability Theory and Related Fields* **86**, 87–104.

[11] Hannan, E.J. (1973) The asymptotic theory of linear time-series models *Journal of Applied Probability* **10**, 130-145.

[12] Hosoya, Y. (1996). The quasi-likelihood approach to statistical inference on multiple time series with long-range dependence. *Journal of Econometrics* **73**, 217–236.

[13] Hosoya, Y. (2005). Fractional invariance principle. *Journal of Time Series Analysis* **26**, 463–486.

[14] Hualde, J. and Robinson P.M. (2011). Supplement to "Gaussian pseudo-maximum likelihood estimation of fractional time series models."

[15] Li, W.K. and McLeod, A.I. (1986). Fractional time series modelling. *Biometrika* **73**, 217–221.

[16] Nordman, D. and Lahiri, S.N. (2006). A frequency domain empirical likelihood for short- and long-range dependence, *Annals of Statistics* **34**, 3019-3050.

[17] Robinson, P.M. (1988). The stochastic difference between econometric statistics. *Econometrica* **56**, 531–548.

[18] Robinson, P.M. (1994). Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* **89**, 1420-1437.

[19] Robinson, P.M. (1995). Gaussian semiparametric estimation of long-range dependence. *Annals of Statistics* **23**, 1630-1661.

[20] Robinson, P.M. (2005). Efficiency improvements in inference on stationary and nonstationary fractional time series. *Annals of Statistics* **33**, 1800–1842.

[21] Robinson, P.M. (2006). Conditional-sum-of-squares estimation of models for stationary time series with long memory. In *Time Series and Related Topics: In Memory of Ching-Zong Wei*. (H.-C. Ho, C.-K. Ing and T.L. Lai, eds.). IMS Lecture Notes - Monograph Series **52**, 130-137.

[22] Robinson, P.M. and Hualde, J. (2003). Cointegration in fractional systems with unknown integration orders. *Econometrica* **71**, 1727-1766.

[23] Shimotsu, K. and Phillips, P.C.B. (2005). Exact local Whittle estimation of fractional integration. *Annals of Statistics* **33**, 1890-1933.

[24] Tanaka, K. (1999). The nonstationary fractional unit root. *Econometric Theory* **15**, 549–582.

[25] Velasco, C. and Robinson, P.M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association* **95**, 1229–1243.

[26] Walker, A.M. (1964) Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *Journal of the Australian Mathematical Society* **4**, 363-384.

[27] Zygmund, A. (1977). *Trigonometric series*. Cambridge University Press, Cambridge.

J. Hualde
Departamento de Economía
Universidad Pública de Navarra
Campus Arrosadía
31006 Pamplona
Spain
E-mail: javier.hualde@unavarra.es

P.M. Robinson
Department of Economics
London School of Economics
Houghton Street
London WC2A 2AE
United Kingdom
E-mail: P.M.Robinson@lse.ac.uk