Efficiency Improvements in Inference on Stationary and Nonstationary Fractional Time Series

by

P M Robinson*
London School of Economics and Political Science

Contents:

Discussion Paper
No.EM/04/480
November 2004

The Suntory Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020 7955 6679

---

* Research supported by a Leverhulme Trust Personal Research Professorship and ESRC grant R000239936.

Abstract

We consider a time series model involving a fractional stochastic component, whose integration order can lie in the stationary/invertible or nonstationary regions and be unknown, and additive deterministic component consisting of a generalised polynomial. The model can thus incorporate competing descriptions of trending behaviour. The stationary input to the stochastic component has parametric autocorrelation, but innovation with distribution of unknown form. The model is thus semiparametric, and we develop estimates of the parametric component which are asymptotically normal and achieve an M-estimation efficiency bound, equal to that found in work using an adaptive LAM/LAN approach. A major technical feature which we treat is the effect of truncating the autoregressive representation in order to form innovation proxies. This is relevant also when the innovation density is parameterised, and we provide a result for that case also. Our semiparametric estimates employ nonparametric series estimation, which avoids some complications and conditions in kernel approaches featured in much work on adaptive estimation of time series models; our work thus also contributes to methods and theory for non-fractional time series models, such as autoregressive moving averages. A Monte Carlo study of finite sample performance of the semiparametric estimates is included.

Keywords: Fractional processes, efficient semiparametric estimation, adaptive estimation, nonstationary processes, series estimation, M-estimation.

JEL No.: C22.

AMS 200 subject classifications: Primary 62M10, secondary 62F11, 62G10, 62J05.

Contact details:
P M Robinson
Department of Economics
London School of Economics
tel: +44 20 7955 7516
email: p.m.robinson@lse.ac.uk

1. INTRODUCTION

This paper obtains efficient estimates in stationary or nonstationary, possibly fractional, time series. Consider a regression model given by

$$y_t = \mu^T z_t + x_t, \quad t \in \mathbb{Z}, \tag{1.1}$$

where $\mathbb{Z} = \{t : t = 0, \pm 1, ...\}$, z_t is a deterministic $q \times 1$ vector sequence, μ is an unknown $q \times 1$ vector, T denotes transposition, x_t is a zero-mean stochastic process, and y_t is an observable sequence. Any nonstationarity in the mean of y_t would be due to z_t, nonstationarity in variance to x_t, but cases when $\mu^T z_t$ is *a priori* constant and x_t is stationary are also of interest.

To describe x_t, denote by B the back-shift operator, so $Bx_t = x_{t-1}$, and by $\Delta = 1 - B$ the difference operator; formally, for all real d

$$\Delta^{-d} = \sum_{j=0}^{\infty} \Delta_j(d) B^j, \qquad \Delta_j(d) = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)},$$

with Γ denoting the gamma function such that $\Gamma(d) = \infty$ for $d = 0, -1, -2, ...$, and $\Gamma(0)/\Gamma(0) = 1$. Assume the sequence x_t is given by

$$x_t = \Delta^{-m_0} v_t^{\#}, \qquad t \in \mathbb{Z}, \tag{1.2}$$

where m_0 is a non-negative integer,

$$v_t^{\#} = v_t 1(t \geq 1), \qquad t \in \mathbb{Z}, \tag{1.3}$$

for $1(.)$ the indicator function, and

$$v_t = \Delta^{-\zeta_0} u_t, \qquad t \in \mathbb{Z}, \tag{1.4}$$

for $|\zeta_0| < \frac{1}{2}$, with u_t a zero-mean covariance stationary process with absolutely continuous spectral distribution function and spectral density $f(\lambda)$ that is at least positive and finite for all λ.

The process v_t is then also covariance stationary, having "long memory" for $\zeta_0 > 0$, "short memory" for $\zeta_0 = 0$ and "negative memory" for $\zeta_0 < 0$. When $m_0 = 0$, we have $x_t = v_t^{\#} = v_t$ for $t \geq 1$. When $m_0 \geq 1$, x_t "integrates" $v_t^{\#}$, and the truncation in (1.2) implies that x_t has variance that is finite, albeit evolving with t. With $\xi_0 = m_0 + \zeta_0$, x_t is well-defined for

$$\xi_0 \in S \subset \{\xi : -\frac{1}{2} < \xi < \infty, \quad \xi \neq \frac{1}{2}, \frac{3}{2}, ...\}. \tag{1.5}$$

The requirement $\xi_0 > -\frac{1}{2}$ excludes non-invertible processes, and the final qualification in (1.5) excludes ξ_0 that cannot be reduced to the stationary/invertible region $(-\frac{1}{2}, \frac{1}{2})$ by integer differencing. Alternative definitions of nonstationary fractional x_t are available, e.g. $\Delta^{-\xi_0} u_t^{\#}$.

Suppose ξ_0 is unknown; m_0 may also be unknown. Suppose u_t is assumed to have parametric autocorrelation:

$$f(\lambda) = \frac{\sigma_0^2}{2\pi} \left|\beta(e^{i\lambda}; \nu_0)\right|^2, \quad \lambda \in (-\pi, \pi], \tag{1.6}$$

such that $cov(u_0, u_j) = \int_{-\pi}^{\pi} f(\lambda) \cos(j\lambda) d\lambda$, $j \in \mathbb{Z}$, $\beta(s; \nu)$ is a smooth, given function of complex-valued s and column-vector $\nu \in V \subset \mathbb{R}^{p_1 - 1}$, $p_1 \geq 1$, satisfying

$$\beta_0(\nu) = 1, \quad \beta(s; \nu) \neq 0, \quad |s| \leq 1, \quad \nu \in V, \tag{1.7}$$

where $\beta_j(\nu) = \int_{-\pi}^{\pi} \beta(e^{i\lambda}; \nu) \cos(j\lambda) d\lambda$, and $\nu_0 \in V$ and $\sigma_0^2 > 0$ are unknown. Then u_t is the variance of the best linear predictor for u_t. For example, u_t can be a standardly-parameterized autoregressive moving average (ARMA) process of autoregressive (AR) order p_{11} and moving average (MA) order p_{12}, such that $p_1 - 1 \leq p_{11} + p_{12} < \infty$; when ν_0 consists precisely of the AR and MA coefficients we have $p_{11} + p_{12} = p_1 - 1$, otherwise the coefficients obey prior restrictions. We call v_t a FARIMA$(p_{11}, \zeta_0, p_{12})$, and x_t a FARIMA(p_{11}, ξ_0, p_{12}). Whereas ν_t is stationary, due to the truncation (1.2) x_t is nonstationary even when $\xi_0 < \frac{1}{2}$ (it could be called "asymptotically stationary" then). The case when $x_t = \nu_t$ for all $t \in \mathbb{Z}$, so x_t is stationary, can be dealt with similarly but we impose the truncation in (1.2) for all $m_0 \geq 0$ for the sake of a unified presentation. The set V is contained in the "stationary and invertible region". The case $p_1 = 1$ means ν_0 is empty, and if $\beta \equiv 1$, x_t is a FARIMA$(0, \xi_0, 0)$. An alternative model for u_t is due to Bloomfield (1972).

The main focus of the paper is estimation of $\theta_{01} = \left(\xi_0, \nu_0^T\right)^T$, and we restrict to a specialized form of z_t in (1.1)

$$z_t = \left(t^{\tau_1}, ..., t^{\tau_q}\right)^T 1(t \geq 1), \quad \tau_1 < \tau_2 < ... < \tau_q, \tag{1.8}$$

where the τ_j are real-valued. Debate has centred on the origin - deterministic or stochastic - of nonstationarity in time series. A notable feature is competition at low frequencies, and given the fractional model for x_t this is most neatly expressed by (1.8). Some components of z_t may have negligible effect on fractionally differenced y_t. Denote by μ_j the j-th element of μ and $\mathcal{T}_1 = \left\{j : \tau_j < \xi_0 - \frac{1}{2}\right\}$, $\mathcal{T}_2 = \left\{j : \tau_j = \xi_0\right\}$, $\mathcal{T}_3 = \left\{j : \xi_0 - \frac{1}{2} \leq \tau_j < \xi_0; \tau_j > \xi_0\right\}$, where any of these sets can be empty. We cannot estimate μ_j for $j \in \mathcal{T}_1$, and do not discuss estimation of μ_j for $j \in \mathcal{T}_2$. Write $s_t = \Sigma_{j \in \mathcal{T}_1} \mu_j t^{\tau_j}$ and for $p_2 = \#\mathcal{T}_3 \leq q$ introduce the $p_2 \times 1$ vectors z_{2t} and θ_{02}, whose j-th elements are the elements of z_t and μ whose index is the j-th largest element of $\mathcal{T}_3$. It will be convenient to write $z_{2t} = (t^{\chi_1}, ..., t^{\chi_{p_2}})^T$, where the χ_j are appropriate τ_j, and satisfy $\frac{1}{2} \leq \chi_1 < ... < \chi_{p_2}$. We can write (1.1) as

$$y_t = s_t + \mu^* t^{\xi_0} + \theta_{02}^T z_{2t} + x_t, \tag{1.9}$$

where $\mu^* = 0$ if $\tau_j \neq \xi_0$ for all j.

We discuss estimation of θ_{02}, along with θ_{01}. For this we require that the τ_j, $j \in \mathcal{T}_3$, are known. The boundary case of $\mathcal{T}_3$, $\tau_j = \xi_0 - \frac{1}{2}$, thus strictly implies ξ_0 is known, but this provision is instead designed to cover a situation in which $\tau_j < \xi_0 - \frac{1}{2}$ for all $j \in \mathcal{T}_1$ is anticipated, with ξ_0 unknown, but in fact $\tau_j = \xi_0 - \frac{1}{2}$ for some j. For

$\theta_1 = (\xi, \nu^T)^T \in S \times V$, introduce the function $\alpha(s;\theta_1)$. $\mathbb{R} \times \mathbb{R}^{p_1} \to \mathbb{R}$, and consider $\alpha(s;\theta_1^{(-)})$, where $\theta_1^{(-)} = (0, \nu^T)^T$, such that

$$\alpha(s;\theta_1) = (1-s)^{\xi} \alpha(s;\theta_1^{(-)}). \tag{1.10}$$

Take $\alpha(s;\theta_1^{(-)}) = \beta(s;\nu)^{-1}$ for $|s| \leq 1$, $\nu \in V$, and note that $\int_{-\pi}^{\pi} \alpha(e^{i\lambda};\theta_1^{(-)})d\lambda = 1$, $\nu \in V$. From (1.6) and (1.7), u_t has one-sided AR representation

$$\alpha(B;\theta_{01}^{(-)})u_t = \sigma_0 \varepsilon_t, \quad t \in \mathbb{Z}, \tag{1.11}$$

where $\theta_{01}^{(-)} = (0, \nu_0^{\tau})^T$, and the ε_t are uncorrelated with zero mean and unit variance. Introduce square-summable coefficients $\alpha_j(\theta_1)$ in the expansion

$$\alpha(s;\theta_1) = \sum_{j=0}^{\infty} \alpha_j(\theta_1) s^j, \quad |s| \leq 1, \quad \xi \in S, \quad \nu \in V, \tag{1.12}$$

so $\alpha_0(\theta_1) \equiv 1$. For given $\theta = \left(\theta_1^T, \theta_2^T\right)^T$, define the computable

$$e_t(\theta) = \sum_{j=0}^{t-1} \alpha_j(\theta_1)\left(y_{t-j} - \theta_2^T z_{2,t-j}\right), \; E_t(\theta) = e_t(\theta) - \frac{1}{n}\sum_{t=1}^{n} e_t(\theta), \qquad t \geq 1, \tag{1.13}$$

the latter being proxies for $\sigma_0 \varepsilon_t$, with s_t ignored in $e_t(\theta)$ because it is anticipated to have negligible effect, and $\mu^* t^{\xi_0}$ ignored in view of the mean-correction in $E_t(\theta)$.

Given observations y_t, $t = 1, ..., n$, define

$$Q_\rho(\theta, \theta_3) = \frac{1}{n}\sum_{t=1}^{n} \rho\left(E_t(\theta)/\tilde{\sigma}; \theta_3\right), \tag{1.14}$$

for an $n^{\frac{1}{2}}$-consistent estimate $\tilde{\sigma}$ of σ_0, a given non-negative function $\rho : \mathbb{R} \times \mathbb{R}^{p_3} \Rightarrow \mathbb{R}$, and any admissible value θ_3 of an unknown $p_3 \times 1$ parameter vector θ_{03}; θ_3 may be empty, as when $\rho(s;\theta_3) = s^2$. Consider the estimate $\left(\bar{\theta}_\rho^T, \bar{\theta}_{3\rho}^T\right) = \arg\min_{\Theta \times \Theta_3} Q_p(\theta, \theta_3)$, for compact sets $\Theta \in \mathbb{R}^p$, $\Theta_3 \in \mathbb{R}^{p_3}$. One anticipates (see e.g. Martin's (1982) discussion of M-estimates of ARMA models) that under suitable conditions $\bar{\theta}_\rho, \bar{\theta}_{3\rho}$ are asymptotically independent and the asymptotic variance matrix of $\bar{\theta}_\rho$ depends on ρ only through the scalar factor $\mathcal{H} = \int \rho'(s)^2 g(s) ds / \left\{\int \rho''(s) g(s) ds\right\}^2$, where the prime indicates differentiation, double-prime indicates twice differentiation, and reference to θ_{03} is suppressed. If integration-by-parts can be conducted, this and the Schwarz inequality indicate that $\mathcal{H} \geq \mathcal{J}^{-1}$, defining the information

$$\mathcal{J} = \int \psi(s)^2 g(s) ds \tag{1.15}$$

and the score function

$$\psi(s) = -g'(s)/g(s). \tag{1.16}$$

The lower bound is attained by $\bar{\theta}_{\log g}$, and the paper obtains estimates that are efficient in the sense of having the same asymptotic variance as $\bar{\theta}_{\log \rho}$. In Theorem 2 of Section 3 we justify such an estimate on the basis of known $g(s;\theta_3)$. If g is misspecified not only will the estimate not be efficient but it may even be inconsistent. Our main result is Theorem 1 of Section 3, which justifies efficient semiparametric estimates, in which the density of ε_t is nonparametric. These estimates are adaptive in the sense of Stone (1975) and are described in the following section. Section 4 describes a Monte Carlo study of finite sample behaviour of the semiparametric estimates. Section 5 attempts to place the work in perspective, relative to the literature. Section 6 presents the main proof details, which use a series of lemmas that make up Section 7. Some of these, such as Lemmas 1, 2, 7, 8, 13, 15 and 16, may be useful in other work. A principal technical feature is our handling of the approximation of the $\sigma_0\varepsilon_t$ in (1.11) by the $e_t(\theta_0)$ defined by (1.13), a delicate matter in fractional models.

2. SEMIPARAMETRIC ESTIMATES

As in much adaptive estimation literature we take an approximate Newton step from an initial consistent estimate $\tilde{\theta}$ of θ_0, with the same rate of convergence as $\bar{\theta}_{\log g}$. This requires estimating $\psi(s)$. We employ an approach developed by Beran (1976), Newey (1988). Beran (1976) proposed a series estimate of $\psi(s)$ (with respect to innovations in an AR(p) model) that employs integration-by-parts. His estimate of $\psi(s)$ was actually not a smoothed nonparametric one because he fixed the number of terms, L, in the series. Newey (1988) allowed L to increase slowly with n, in adapting to error distribution of unknown form in cross-sectional regression.

Let $\phi_\ell(s)$, $\ell = 1, 2, ...,$ be a sequence of given, continuously differentiable functions. For $L \geq 1$, scalar h_t, $t = 1, ..., n$, and $h = (h_1, ..., h_n)^T$, define $\phi^{(L)}(h_t) = (\phi_1(h_t), ..., \phi_L(h_t))^T$, $\Phi^{(L)}(h_t) = \phi^{(L)}(h_t) - n^{-1}\sum_{s=1}^n \phi^{(L)}(h_s)$, $\phi^{'(L)}(h_t) = (\phi_1'(h_t), ..., \phi_L'(h_t))^T$ and

$$W^{(L)}(h) = n^{-1}\sum_{t=1}^n \Phi^{(L)}(h_t)\Phi^{(L)}(h_t)^T, w^{(L)}(h) = n^{-1}\sum_{t=1}^n \phi^{'(L)}(h_t),$$
$$\widehat{a}^{(L)}(h) = W^{(L)}(h)^{-1}w^{(L)}(h),\ \psi^{(L)}(h_t; \widehat{a}^{(L)}(h)) = \widehat{a}^{(L)}(h)^T\Phi^{(L)}(h_t).$$

With $E(\theta) = (E_1(\theta), ..., E_n(\theta))^T$ define

$$\widetilde{\psi}_t^{(L)}(\theta, \sigma) = \psi^{(L)}(E_t(\theta)/\sigma; \widehat{a}^{(L)}(E(\theta)/\sigma)),$$

where it will follow from our conditions that in a neighbourhood of θ_0, σ_0, $W^{(L)}(E(\theta)/\sigma)$ is nonsingular with probability approaching 1 as $n \to \infty$. We then compute the $\tilde{\psi}_t^{(L)}(\tilde{\theta}, \tilde{\sigma})$. Following Beran (1976), Newey (1988) we have approximated $\psi(\varepsilon_t)$ by $\Sigma_{\ell=1}^L a_\ell \{\phi_\ell(\varepsilon_t) - E\phi_\ell(\varepsilon_t)\}$ (imposing the restriction $E\psi(\varepsilon_t) = 0$), noted that (under conditions to be given) integration-by-parts implies $E\left\{\phi^{(L)}(\varepsilon_t)\psi(\varepsilon_0)\right\} = E\left\{\phi^{(L)}(\varepsilon_t)\right\}$, estimated $(a_1, ..., a_L)^T$ by $a^{(L)}(E(\tilde{\theta})/\tilde{\sigma})$, and then $\psi(\varepsilon_t)$ by $\tilde{\psi}_t^{(L)}(\tilde{\theta}, \tilde{\sigma})$.

Define (see (1.10)-(1.13))

$$e_t'(\theta) = (\partial/\partial\theta)e_t(\theta) = \left(e_{t1}'(\theta)^T, e_{t2}'(\theta)^T\right)^T,$$

where

$$e_{t1}'(\theta) = \alpha'(B; \theta_1)\left(y_t - \theta_2^T z_{2t}\right), \qquad e_{t2}'(\theta) = -\alpha(B; \theta_1)z_{2t},$$

with

$$\begin{aligned}\alpha'(s;\theta_1) &= (\partial/\partial\theta_1)\alpha(s;\theta_1) = (1-s)^{\xi}\alpha(s;\theta_1^{(-)})\gamma(s;\nu),\\ \gamma(s;\nu) &= \left[\log(1-s), \{(\partial/\partial\nu)^T\alpha(s;\theta_1^{(-)})\}/\alpha(s;\theta_1^{(-)})\right]^T. \end{aligned} \tag{2.1}$$

Define

$$E'_{ti}(\theta) = e'_{ti}(\theta) - n^{-1}\sum_{s=1}^{n} e'_{si}(\theta), \quad i = 1, 2,$$

$$\begin{aligned} r_{Li}(\theta,\sigma) &= \sum_{t=1}^{n} \widetilde{\psi}_t^{(L)}(\theta,\sigma)E'_{ti}(\theta), \quad R_i(\theta) = \sum_{t=1}^{n} E'_{ti}(\theta)E'_{ti}(\theta)^T, i = 1, 2,\\ \mathcal{J}_L(\theta,\sigma) &= n^{-1}\sum_{t=1}^{n} \widetilde{\psi}_t^{(L)}(\theta,\sigma)^2. \end{aligned}$$

Estimate θ_{01}, θ_{02} by

$$\widehat{\theta}_i = \widetilde{\theta}_i + \left\{R_i(\widetilde{\theta})\mathcal{J}_L(\widetilde{\theta},\widetilde{\sigma})\right\}^{-1} r_{Li}(\widetilde{\theta},\widetilde{\sigma}), \quad i = 1, 2, \tag{2.2}$$

respectively, for $\tilde{\theta} = (\tilde{\theta}_1^T, \tilde{\theta}_2^T)^T$.

As in Newey (1988) we restrict to $\phi_\ell(s)$ satisfying

$$\phi_\ell(s) = \phi(s)^\ell, \tag{2.3}$$

for a smooth function $\phi(s)$. Examples are

$$\phi(s) = s \tag{2.4}$$

$$\phi(s) = s(1+s)^{-\frac{1}{2}}. \tag{2.5}$$

Our conditions require L to increase very slowly with n, and allow the increase to be arbitrarily slow; in practice, for moderate n, (2.2) might be computed for a few small integers L, starting with $L = 1$. Recursive formulae are available, using partitioned regression, such that the elements of $W^{(L)}(E(\tilde{\theta})/\tilde{\sigma})$, $w^{(L)}(E(\tilde{\theta})/\tilde{\sigma})$ can be used in computing $\tilde{\psi}_t^{(L+1)}(\tilde{\theta},\tilde{\sigma})$.

3. MAIN RESULTS

We introduce the following regularity conditions.

Assumption A1 *The sequence y_t is generated by (1.1) with x_t generated by (1.2)-(1.4) and (1.11), where the ε_t are independent and identically distributed (iid) with zero mean and variance 1, and z_t is given by (1.8).*

Assumption A2 *Either:*

(a) $E\varepsilon_0^4 < \infty$; *or*

(b) *for some* $\omega > 0$ *the moment generating function* $E\left(e^{t|\varepsilon_0|^\omega}\right)$ *exists for some* $t > 0$; *or*

(c) ε_0 *is almost surely bounded.*

Assumption A3 ε_0 *has density,* $g(s)$*, that is differentiable, and*

$$0 < \mathcal{J} < \infty,$$

where $\mathcal{J}$ *is defined in (1.15).*

Assumption A4 *The sentence including (1.6) and (1.7) is true,* ν_0 *is an interior point of* V *and in a neighbourhood* $\mathcal{N}$ *of* ν_0, $\alpha(s;\theta_1^{(-)}) = \beta(s;\nu)^{-1}$ *is thrice continuously differentiable in* ν *for* $|s| = 1$ *and*

$$\sum_{j=1}^{\infty} j^3 \left\{ \left|\beta_j(\nu_0)\right| + \sup_{\mathcal{N}} \left|\alpha_j(\theta_1^{(-)})\right| \right.$$

$$\left. + \sup_{\mathcal{N}} \left|\alpha_j^{(k)}(\theta_1^{(-)})\right| + \sup_{\mathcal{N}} \left|\alpha_j^{(k,\ell)}(\theta_1^{(-)})\right| + \sup_{\mathcal{N}} \left|\alpha_j^{(k,\ell,m)}(\theta_1^{(-)})\right| \right\} < \infty,$$

for all $k, \ell, m = 1, ..., p_1 - 1$, *where* $\alpha_j(\theta_1^{(-)})$ *is defined by (1.10), (1.12) and* $\alpha_j^{(k)}(\theta_1^{(-)}) = (\partial/\partial\nu_k)\alpha_j(\theta_1^{(-)})$, $\alpha_j^{(k,\ell)}(\theta_1^{(-)}) = (\partial/\partial\nu_\ell)\alpha_j^{(k)}(\theta_1^{(-)})$, $\alpha_j^{(k,\ell,m)}(\theta_1^{(-)}) = (\partial/\partial\nu_m)\alpha_j^{(k,\ell)}(\theta_1^{(-)})$, ν_k *being the* k*-th element of* ν.

Assumption A5 *For all* $(p_1-1)\times 1$ *non-null vectors* λ, $\lambda^T \left\{(\partial/\partial\nu)\alpha(e^{i\lambda};\theta_{01}^{(-)})\right\}\beta(e^{i\lambda};\nu_0) \neq 0$ *on a subset of* $(-\pi,\pi]$ *of positive measure.*

Assumption A6

$$0 < \sigma_0^2 < \infty.$$

Assumption A7

$$n^{\frac{1}{2}}\left(\tilde{\theta}_1 - \theta_{01}\right) = O_p(1), \quad D_n(\tilde{\theta}_2 - \theta_{02}) = O_p(1), \quad n^{\frac{1}{2}}(\tilde{\sigma}^2 - \sigma_0^2) = O_p(1),$$

where

$$\begin{aligned} D_n &= diag\left\{n^{\chi_1-\xi_0+\frac{1}{2}}1(\chi_1-\xi_0 > -\frac{1}{2}) + (\log n)^{\frac{1}{2}}1(\chi_1-\xi_0 = -\frac{1}{2}),\right. \\ &\quad \left. n^{\chi_2-\xi_0+\frac{1}{2}}, ..., n^{\chi_{p_2}-\xi_0+\frac{1}{2}}\right\}. \end{aligned}$$

Assumption A8 $\phi_\ell(s)$ *satisfies (2.3), where* $\phi(s)$ *is strictly increasing and thrice continuously differentiable and is such that, for some* $\kappa \geq 0$, $K < \infty$,

$$|\phi(s)| \leq 1(|s| \leq 1) + |s|^\kappa 1(|s| > 1), \tag{3.1}$$

$$|\phi'(s)| + |\phi''(s)| + |\phi'''(s)| \leq C(1 + |\phi(s)|^K). \tag{3.2}$$

Assumption A9

$$L \to \infty \text{ as } n \to \infty \tag{3.3}$$

and either:

(a)

$$\underline{\lim}_{n\to\infty}\left(\frac{\log n}{L}\right) > 8\left\{\log\eta + \max(\log\varphi, 0)\right\} \simeq 7.05 + 8\max(\log\varphi, 0); \tag{3.4}$$

or

(b)

$$\varliminf_{n\to\infty}\left(\frac{\log n}{L\log L}\right) > \max\left(\frac{8\kappa}{\omega}, \frac{4\kappa(\omega+1)}{\omega}\right); \tag{3.5}$$

or

(c)

$$\varliminf_{n\to\infty}\left(\frac{\log n}{L\log L}\right) > 4\kappa, \tag{3.6}$$

where

$$\eta = 1 + 2^{\frac{1}{2}} \simeq 2.414$$

and

$$\varphi = \frac{1+|\phi(s_1)|}{\phi(s_2)-\phi(s_1)},$$

$[s_1, s_2]$ *being an interval on which* $g(s)$ *is bounded away from zero.*

Remark 1 Parts (a), (b) and (c) of A2 increase in strength and entail trade-offs with A8 and A9. When $\kappa = 0$ in A8, so $\phi(s)$ is bounded, (a) of A2 and (a) of A9 suffice; a finite fourth moment seems hard to avoid in dealing with the deviation $e_t(\theta_0) - \sigma_0\varepsilon_t$. Part (b) of A2 holds with $\omega = 1$ for Laplace ε_t and with $\omega = 2$ for Gaussian ε_t. We require (b) of A2 when $\kappa > 0$ in A8, so $\phi(s)$ can be unbounded, and also (b) of A9. If (c) of A2 holds, then a fortiori we can have $\kappa > 0$ in A8, and can relax (b) of A9 to (c).

Remark 2 Assumption A3 is virtually necessary.

Remark 3 Assumption A4 is stronger than necessary, but is chosen for brevity of presentation and because it is readily checked for short memory and invertible AR (α) and MA (β) filters arising in models of most practical interest, such as ARMA and Bloomfield (1972) models, and in any case conditions on the short memory component are of only secondary interest here. A property useful in several places (see in particular Lemma 13 of Section 7) that is ensured by A4 is as follows. A (possibly vector) sequence α_j, $j \geq 0$, has property $P_r(d)$, $r \geq 0$, if

$$\|\alpha_j\| \leq C\left\{\log(j+2)\right\}^r (j+1)^{d-1}, \quad \|\alpha_j - \alpha_{j+1}\| \leq C\left\{\log(j+2)\right\}^r (j+1)^{d-2}, \quad j \geq 0,$$

where $\|.\|$ denotes Euclidean norm. For $|s| \leq 1$ and $\theta_1^{(+)} = (\zeta, \nu^T)^T$, define square-summable $\pi_j(\theta_1^{(+)})$ such that

$$\pi(s;\theta_1^{(+)}) = (1-s)^{-\zeta}\beta(s;\nu) = \sum_{j=0}^{\infty} \pi_j(\theta_1^{(+)})s^j, \quad |\zeta| < \frac{1}{2}, \quad \nu \in V.$$

Then, with $\theta_{01}^{+} = (\zeta_0, \nu_0^T)^T$, $\pi_j(\theta_{01}^{(+)})$ has property $P_0(\zeta_0)$, $\alpha_j(\theta_{01}^{(+)})$ has property $P_0(-\zeta_0)$ and $(\partial/\partial/\theta_1^{(+)T})\alpha_j(\theta_{01}^{(+)})$ has property $P_1(-\zeta_0)$. This follows from Lemmas 11 and 12 of Section 7 on noting that, for $\alpha(s) = \sum_{j=0}^{\infty} \alpha_j s^j$, $\beta(s) = \sum_{j=0}^{\infty} \beta_j s^j$, the coefficient of s^j in $\alpha(s)\beta(s)$ is $\sum_{k=0}^{j} \alpha_k \beta_{j-k}$, that the coefficients of s^j in $(1-s)^{-d}$ and $-\log(1-s)$ are $\Delta_j(d)$ and j^{-1}, that $\pi(1;\theta_{01}^{(+)}) = 0$ for $\zeta_0 < 0$, and that $\alpha(1;\theta_{01}^{(+)}) = 0$, $(\partial/\partial/\theta^{(+)T})\alpha(1;\theta_{01}^{(+)}) = 0$ for $\zeta_0 > 0$.

Remark 4 A5 is an identifiability condition, violated if, for example, u_t is specified as an ARMA with both AR and MA orders over-stated. A5, with A4, implies that

$$\begin{aligned} \Omega_1 &= \frac{1}{2\pi}\int_{-\pi} \gamma\left(e^{i\lambda};\nu_0\right)\gamma\left(e^{-i\lambda};\nu_0\right)^T d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \begin{bmatrix} \log\left|1-e^{i\lambda}\right|^2 \\ 2\frac{\partial}{\partial\nu}\log\left|\beta(e^{i\lambda};\nu_0)\right| \end{bmatrix} \begin{bmatrix} \log\left|1-e^{i\lambda}\right|^2 \\ 2\frac{\partial}{\partial\nu}\log\left|\beta(e^{i\lambda};\nu_0)\right| \end{bmatrix}^T d\lambda \end{aligned} \tag{3.7}$$

is positive definite, with γ given by (2.1). Ω_1 is proportional to the inverse of the limiting covariance matrix of $\hat{\theta}_1$. We define also the corresponding matrix with respect to $\hat{\theta}_2$,

$$\Omega_2 = \frac{\sigma_0^2}{2\pi}\beta(1;\nu_0)^2 \left(\frac{\left\{2(\chi_i-\xi_0)+1\right\}^{\frac{1}{2}}\left\{2(\chi_j-\xi_0)+1\right\}^{\frac{1}{2}}(\chi_i-\xi_0)(\chi_j-\xi_0)}{(\chi_i+\chi_j-2\xi_0+1)(\chi_i-\xi_0+1)(\chi_j-\xi_0+1)}\right), \tag{3.8}$$

when $\chi_1 - \xi_0 > -\frac{1}{2}$, where the (i,j)-th element of the matrix is displayed; because $\left((\chi_i + \chi_j - 2\xi_0 + 1)^{-1} \right)$ is a Cauchy matrix (see Knuth, 1968, p.30), and the inequalities in (1.8) hold, Ω_2 is positive definite. The same is true when $\tau_j - \xi_0 = -\frac{1}{2}$ for some j, Ω_2 being defined by replacing the $(1,1)$-th element of the matrix in (3.8) by 1, and the other elements in the first row and column by zero.

Remark 5 The middle part of A7 is likely to be satisfied by the least squares estimate of θ_{02}, under similar conditions to ours. A substantial literature justifies $\tilde{\theta}_1$ satisfying A7; typically $\theta_{02}' z_{2t}$ is assumed constant *a priori* but the results should go through more generally with x_t replaced by least squares residuals. Various estimates of θ_{01} (which we collectively call Whittle estimates) have been shown to be $n^{\frac{1}{2}}$-consistent and asymptotically $N(0, \Omega_1^{-1})$ when $0 \leq \xi_0 < \frac{1}{2}$ under Gaussianity of x_t (when they achieve the efficiency bound of Section 1 and are as good as maximum likelihood estimates), and under more general conditions (see e.g. Hannan (1973), Fox and Taqqu (1986), Dahlhaus (1989), Giraitis and Surgailis (1990)). The estimate minimizing (1.14) with $\rho(s) = s^2$ (usually with $E_t(\theta)$ replaced by $e_t(\theta)$) falls within this class. This estimate (used by Li and McLeod,1986, for fractional models and Box and Jenkins, 1971, for ARMA ones) is sometimes called a conditional sum of squares (CSS) estimate (though it is based on formulae for the truncated AR representation rather than for the conditional expectation given the finite past record). Beran (1995) argued that it has the same desirable asymptotic properties for $\xi_0 > \frac{1}{2}$, tying in with Robinson's (1994) derivation of standard asymptotics for score tests, based on the same objective function, for unit root and more general nonstationary hypotheses against fractional alternatives. These authors employed a different definition of fractional nonstationarity from ours, but for our definition Velasco and Robinson (2000) established the same properties for a Whittle estimate when $-\frac{1}{2} < \xi_0 < \frac{3}{4}$, and for a tapered version of this for $-\frac{1}{2} < \xi_0 < \infty$, though the tapering inflates asymptotic variance. They established consistency of their implicitly-defined optimizer despite lack of uniform convergence over an admissible parameter set that includes a wide range of nonstationary values of ξ. The definition of improved estimate as a Newton step from a previously established $n^{\frac{1}{2}}$-consistent estimate avoids a similar difficulty. Velasco and Robinson's (2000) estimate of σ_0^2 should satisfy the final part of A7 (with (e) sufficient within A2).

Remark 6 When $\kappa = 0$ in A8, then $|\phi(s)| \leq 1$ for all s, under (3.1); there would be no gain in generality by specifying ϕ to satisfy a larger finite bound. For $\kappa > 0$ we might take $\phi(s) = s^\kappa$, cf (2.4). The reason for imposing different bounds on $\phi(s)$ over $|s| \leq 1$ and $|s| > 1$ is to allow possibly different rates of approach to zero and infinity. A8 is stronger than the corresponding assumption of Newey (1988), and is driven by the presence of $e_t(\theta_0)$ for small t, when it does not approximate $\sigma_0 \varepsilon_t$; we prefer this to trimming out small t, which introduces further ambiguity. It is hard to think of reasons for choosing ϕ that do not satisfy (3.1), (3.2), which imply power-law bounds on $\phi'(s)$, $\phi''(s)$ and $\phi'''(s)$ as $s \to \infty$.

Remark 7 The weakest of the conditions in A9, (a), can only apply when $\kappa = 0$ in A8, in which case $\log \varphi > 0$. Subject to this, the hope is that s_1 and s_2 exist such that φ is arbitrarily close to 1, as when $g(s) > 0$ for all s; then the strict inequality in (3.4) applies with $\log \varphi = 0$. The mysterious constant η is due to approximating $W^{(L)}$ in the proof in terms of the Cauchy matrix with (i,j)-th element $\int_{-1}^{1} u^{i+j-2} du$ (see Lemma 7 of Section 7). Since ϕ is defined for negative and positive arguments this seems more natural than Newey's (1988) use of the Hilbert matrix $\left(\int_0^1 u^{i+j-2} du\right)$ and affords some slight improvement over it due to the many zero elements in this Cauchy matrix (following a similar proof for the Hilbert matrix to Lemma 7's, η would be replaced by $\eta^2 \simeq 5.828$). In fact a constant such as η does not arise in Newey's work because he is content with a slightly stronger condition than any in

A9, $L \log L / \log n \to 0$, irrespective of whether or not ϕ is bounded, and without considering the impact of bounded ε_t. This is because he accepts a bound of form L^{CL} at several points of his proof. Our slightly sharper bounds suggest that when ϕ is bounded it is effectively the denominator of $\psi^{(L)}$ (i.e. the inverse of $W^{(L)}$) that dominates, while when ϕ is unbounded the numerator dominates. In the former case, the slow L corresponds to the notorious ill-conditioning of Cauchy/Hilbert matrices. One disadvantage of a bounded ϕ is that a larger L might be needed to approximate a ψ of infinite range, though our slightly milder condition on L in A9(a) might help to justify this. Another is that it excludes (2.4), which "nests" the Gaussian case, though it would be possible to modify our theory to allow inclusion of $\phi_1(s) = s$, say, followed by polynomial ϕ_ℓ (2.3) using bounded ϕ such as (2.5). Though the partly known nature of the bounds in A9 is interesting, and their reflection of other assumptions is intuitively reasonable in a relative sense, not only is the improvement over Newey's rate slight but even after guessing ω and φ no practical choices of L in finite samples can be concluded, indeed the same asymptotic bounds result if any fixed integer is added to or subtracted from L. As in much other semiparametric work, no information towards an optimal choice of L emerges, indeed as in Newey (1988) there is no lower bound on L, besides that it must increase with n.

Theorem 1 *Let Assumptions A1-A9 hold, such that when $\kappa = 0$ A2(a) holds with A9(a), or when $\kappa > 0$ either A2(b) holds with A9(b) or A2(c) holds with A9(c). Then as $n \to \infty$, $n^{\frac{1}{2}}(\hat{\theta}_1 - \theta_{01})$ and $D_n(\hat{\theta}_2 - \theta_{02})$ converge to independent $N(0, \mathcal{J}^{-1}\Omega_1^{-1})$, $N(0, \mathcal{J}^{-1}\Omega_2^{-1})$ vectors, respectively, where the limiting covariance matrices are consistently estimated by $\left\{\mathcal{J}_L(\tilde{\theta}, \tilde{\theta}) R_1(\tilde{\theta})/n\right\}^{-1}$, $\left\{\mathcal{J}_L(\tilde{\theta}, \tilde{\theta}) D_n^{-1} R_2(\tilde{\theta}) D_n^{-1}\right\}^{-1}$, respectively.*

To place Theorem 1 in perspective and to further balance the focus on Whittle estimation in the long memory literature, we also consider the fully parametric case, where $g(s; \theta_3)$ is a prescribed parametric form, as described after (1.14), on the basis of which define $\hat{\theta}_3 = \arg\min_{\Theta_3} Q_{\log g}(\tilde{\theta}; \theta_3)$, and, with $\psi(s;\theta_3) = -(\partial/\partial s)g(s;\theta_3)/g(s;\theta_3)$,

$$\begin{aligned} \mathcal{J}_n(\theta,\sigma,\theta_3) &= n^{-1}\sum_{t=1}^{n}\psi\left(E_t(\theta)/\sigma;\theta_3\right)^2, \\ r_i(\theta,\sigma,\theta_3) &= \sum_{t=1}^{n}\psi\left(E_t(\theta)/\sigma;\theta_3\right)E'_{ti}(\theta), \quad i=1,2, \end{aligned}$$

and redefine $\hat{\theta}_i$, $i = 1, 2$ of (2.2) as

$$\hat{\theta}_i = \tilde{\theta}_i + \left\{R_i(\tilde{\theta})\mathcal{J}_n(\tilde{\theta},\tilde{\sigma},\hat{\theta}_3)\right\}^{-1} r_i(\tilde{\theta},\tilde{\sigma},\hat{\theta}_3), \qquad i=1,2.$$

We introduce the following additional assumptions.

Assumption A10 *Θ_3 is compact and θ_{03} is an interior point of Θ_3.*

Assumption A11 *For all $\theta_3 \in \Theta - \{\theta_{03}\}$, $g(s;\theta_3) \neq g(s;\theta_{03})$ on a set of positive measure.*

Assumption A12 *In a neighbourhood $\mathcal{N}$ of θ_{03}, $\log g(s;\theta_3)$ is thrice continuously differentiable in θ_3 for all s and*

$$\int_{-\infty}^{\infty}\left\{\sup_{\mathcal{N}}\left|g^{(k)}(s;\theta_3)\right| + \sup_{\mathcal{N}}\left|g^{(k,\ell)}(s;\theta_3)\right| + \sup_{\mathcal{N}}\left|g^{(k,\ell,m)}(s;\theta_3)\right|\right\}ds < \infty,$$

where $g^{(k)}$, $g^{(k,\ell)}$, $g^{(k,\ell,m)}$ represent partial derivatives of g with respect to the k-th, the k-th and ℓ-th, and the k-th, ℓ-th and m-th elements of θ_3, respectively.

Assumption A13 *$\Omega_3 = E\{(\partial/\partial\theta_3)\log g(\varepsilon_t;\theta_{03})(\partial/\partial\theta_3^T)\log g(\varepsilon_0;\theta_{03})\}$ is positive definite.*

Theorem 2 *Let Assumptions A1, A2(a), A3-A7, A10-A13 hold. Then as $n \to \infty$, $n^{\frac{1}{2}}(\hat{\theta}_1 - \theta_{01})$, $D_n^{\frac{1}{2}}(\hat{\theta}_2 - \theta_{02})$ and $n^{\frac{1}{2}}(\hat{\theta}_3 - \theta_{03})$ converge to independent $N(0, \mathcal{J}^{-1}\Omega_1^{-1})$, $N(0, \mathcal{J}^{-1}\Omega_2^{-1})$ and $N(0, \Omega_3^{-1})$ vectors respectively, where the limiting covariance matrices are consistently estimated by $\left\{\mathcal{J}_n(\tilde{\theta},\tilde{\sigma},\hat{\theta}_3)R_1(\tilde{\theta})/n\right\}^{-1}$, $\left\{\mathcal{J}_n(\tilde{\theta},\tilde{\sigma},\hat{\theta}_3)D_n^{-1}R_2(\tilde{\theta})D_n^{-1}\right\}^{-1}$*

and

$$\left\{ n^{-1} \sum_{t=1}^{n} \left[(\partial/\partial\theta_3) \log g \left(E_t(\tilde{\theta})/\tilde{\sigma}; \hat{\theta}_3 \right) \right] \left[(\partial/\partial\theta_3^T) \log g \left(E_t(\tilde{\theta})/\tilde{\sigma}; \hat{\theta}_3 \right) \right] \right\}^{-1},$$

respectively.

The proof (which entails an initial consistency proof for the implicitly-defined extremum estimate $\hat{\theta}_3$) is omitted because it combines relatively standard arguments with elements of the proof of Theorem 1, notably concerning the $e_t(\theta_0) - \sigma_0\varepsilon_t$ issue. Our treatment of this would also lead to a theorem for M-estimates of θ_0 minimizing (1.14) in which $\rho(s)$ is a completely specified function, not necessarily $\log g(s)$, but we omit this to conserve on space, and because the efficiency improvement of the paper's title would in general not be achieved.

Theorems 1 and 2 suggest locally more powerful (Wald-type) tests on θ_{01} than those implied by CLTs for Whittle estimates. For example, the hypothesis of short memory, $\xi_0 = 0$, can be efficiently tested, as can, say, the significance of AR coefficients in a FARIMA$(p_{11}, \xi_0, 0)$, for any unknown $\xi_0 > -\frac{1}{2}$. We can also efficiently investigate the question of relative success of deterministic and stochastic components in describing trending time series. For example, we can apply the theorems to test $\theta_{02} = 0$, or, with $p_2 = 1$, $p_2 = t^\tau$, test $\xi_0 = \tau + \frac{1}{2}$ against the one-sided alternative $\xi_0 > \tau + \frac{1}{2}$ (see the discussion after (1.9)); in the first case rejection implies a significant deterministic trend, in the latter, a dominant stochastic one. Tests based on $\hat{\theta}_2$ are in general more powerful than those based on least squares (see Yajima, 1988) or generalized least squares (see Dahlhaus, 1995).

4. FINITE SAMPLE PERFORMANCE

A small Monte Carlo study was carried out to investigate the success of our semi-parametric estimates in small and moderate samples. Along with the value of n, major influential features seem likely to be the form of $g(s)$, the value of ξ_0 and the choice of ϕ and L.

We focussed on the simple FARIMA$(0, \xi_0, 0)$ model for y_t (knowing $\mu^T z_0 \equiv 0$) for

(i) $\xi_0 = -0.25$ ("antipersistent")
(ii) $\xi_0 = 0.25$ ("stationary with long memory")
(iii) $\xi_0 = 0.75$ ("nonstationary but mean-reverting")
(iv) $\xi_0 = 1.25$ ("nonstationary, non-mean-reverting").

For ε_t we considered the following distributions (the scalings referred to producing $var(\varepsilon_t) = 1$):

(a) $N(0, 1)$

(b) $0.5N(-3, 1) + 0.5N(3, 1)$

(c) (scaled) $0.05N(0, 25) + 0.95N(0, 1)$

(d) (scaled) Laplace

(e) (scaled) t_5.

These were mostly chosen for the sake of consistency with other Monte Carlo studies of adaptive estimates. The benchmark case (a), and the two (symmetric and asymmetric) mixed normal distributions (b) and (c), were used by Kreiss (1987) in a stationary AR model, with kernel estimates of ψ, and by Newey (1988) (in a cross-sectional regression model), Ling (2003) also using (b) in a FARIMA$(0, \xi_0, 0)$ model with kernel estimates of ψ. Kreiss (1987) also used (d). The point of (e) is that it only just satisfies the minimal fourth moment condition on ε_0, A2(a). Kernel approaches, from Stone (1975), Bickel (1982) for location and regression models for independent observations, through Kreiss (1987, for example), Drost *et al.* (1997), Koul and Schick (1997) for short memory time series models, and Hallin, Taniguchi, Serroukh and Choy (1999), Hallin and Serroukh (1999), Ling (2003) for long memory ones, have been popular in the adaptive estimation literature. Besides requiring choice of a kernel and bandwidth (analogous to our ϕ and L), they typically involve one or more forms of trimming, in part due to the presence of a kernel density estimate in the denominator of the estimate of $\psi(s)$, and sometimes sample splitting and discretization of the initial estimate. Theorem 1 of course implies semiparametric efficient estimates using series estimation for short memory models. For ϕ we used both (2.4) and (2.5), and tried $L = 1, 2, 3, 4$, with $n = 64$ and 128. For $\tilde{\xi} = \tilde{\theta}$ and $\tilde{\sigma}^2$ Velasco and Robinson's (2000) estimates were employed, with a cosine bell taper; this is sufficient to satisfy A7 for all ξ_0 considered, albeit unnecessary when $\xi_0 = \pm 0.25$.

5. FINAL COMMENTS

In various stationary, short-memory time series models, Kreiss (1987, for example) Drost *et al.* (1997), Koul and Schick (1997), and others, developed local asympotitic normality (LAN) and local asymptotic minimaxity (LAM) theory of Le Cam (1960), Hajek (1972) to establish $\sqrt{n}$-consistent, asymptotically normal and asymptotically efficient estimates, and, further, adaptive estimates that achieve the same properties in the presence of nonparametric g. A similar approach was followed by Hallin, Taniguchi, Serroukh and Choy (1999), Hallin and Serroukh (1999) and Ling (2003) in case of stationary and nonstationary fractional models. LAN theory commences from a log likelihood ratio, but in view of the difficulty in constructing likelihoods in a general non-Gaussian setting, the latter authors commenced not from the likelihood for $y_1, ..., y_n$ but from a "likelihood" for $y_1, ..., y_n$ and the infinite set of unobservable variables ε_t, $t \leq 0$, in terms of the density g of ε_t, or a "conditional likelihood" for $y_1, ..., y_n$ given the ε_t, $t \leq 0$, or the y_t, $t \leq 0$. We do not employ such constructions and do not establish local optimality properties. However our M-estimate efficiency bound is of course the same as the asymptotic variance resulting from a LAM/LAN approach.

Another motivation for our more elementary efficiency criterion is to allow space to focus on the main technical difficulty distinguishing asymptotic distribution theory for fractional models from that for short-memory ones. This is due to the need to approximate the truncated AR transforms $e_t = e_t(\theta_0)$ (see (1.13)) by scaled innovations $\sigma_0\varepsilon_t$. Consider a simplified version of the problem in which $y_t = x_t$ *a priori*, so $\theta = \theta_1$, and define $\delta_t = e_t - \sigma_0\varepsilon_t$. In the following section (relying heavily on Lemmas 13 and 14 of Section 7) we find that $E\left|\delta_t\right|^r \leq Ct^{-r/2}$, $r \geq 2$, given a sufficient moment condition on ε_t. This property is useful in our proof that e_t can be replaced by $\sigma_0\varepsilon_t$ in a $\widehat{a}^{(L)}\left(E(\theta_0)/\sigma_0\right)$ (see Lemma 19). In some cases it is possible to show that the upper bound provides a sharp rate. Consider the stationary $FARIMA(0, \xi_0, 0)$ (cf. Hallin and Serroukh, 1999), where $0 < \xi_0 = \zeta_0 < \frac{1}{2}$ and $x_t = v_t$, $t \in \mathbb{Z}$. Noting that $cov(x_0, x_j) \geq j^{2\xi_0 - 1}/C$, $\alpha_j(\xi_0) \geq j^{-\xi_0-1}/C$ for $j > 0$, where C denotes a finite but arbitrarily large generic constant,

$$\begin{aligned}
E(\delta_t^2) &= \sum_{j=t}^{\infty}\sum_{k=t}^{\infty} \alpha_j(\xi_0)\alpha_k(\xi_0)cov(x_j, x_k) \\
&\geq C^{-1} \sum_{\substack{j=t \\ 1\leq|j-k|\leq t}}^{\infty}\sum_{k=t}^{\infty} j^{-\xi_0-1}k^{-\xi_0-1}\left|j-k\right|^{2\xi_0-1} \\
&\geq C^{-1}t^{2\xi_0-1}\sum_{j=t}^{\infty}\sum_{k=t+1}^{t+j}(jk)^{-\xi_0-1} \\
&\geq C^{-1}t^{2\xi_0-1}\sum_{j=t}^{2t} j^{-\xi_0}(t+j)^{-\xi_0-1} \\
&\geq (Ct)^{-1}.
\end{aligned}$$

This contrasts with the exponential rate occurring with ARMA models. In this stationary $FARIMA(0, \xi_0, 0)$,

$$\delta_t = \sum_{j=0}^{t-1} \alpha_j(\xi_0) x_{t-j} - \sigma_0 \varepsilon_t = \sum_{j=0}^{t-1} \alpha_j(\xi_0) \nu_{t-j} - \sigma_0 \varepsilon_t = -\sum_{j=t}^{\infty} \alpha_{t+j}(\xi_0) \nu_{t-j}. \tag{5.1}$$

In our "asymptotically stationary" version of the $FARIMA(0, \xi_0, 0)$, also with $0 < \xi_0 < \frac{1}{2}$, we have $x_t = x_t^{\#}$ but again (5.1) results, from (1.4), (1.10), (1.11) and Lemma 5 of Section 7. In this connection, note that for general ξ_0, Ling (2003) took $x_t = \Delta^{-m_0} v_t^{\#} + v_t 1(t \leq 0)$ in place of our (1.2), but this different prescription of x_t for $t \leq 0$ makes no difference to e_t, which depends on x_s for $s \geq 1$ only.

The above upper bound for $E\left|\delta_t\right|^r$, combined with the Schwarz inequality, is insufficient to deal completely with the replacement of e_t by $\sigma_0 \varepsilon_t$,even when ψ is smooth. Staying with the case $y_t = x_t$ *a priori*, the proof of Theorems 1 and 2 entails establishing asymptotic normality of a quantity of form $c_{1n} = n^{-\frac{1}{2}} \sum_{t=1}^n \psi(e_t) h_t$, where h_t is $\{\varepsilon_s, s \leq t-1\}$-measurable and has finite variance; c_{1n} is called a "central sequence" by Hallin, Taniguchi, Serroukh and Choy (1999) (see their (2.15), (3.11)) and Hallin and Serroukh (1999) (see their (2.4)). Asymptotic normality of $c_{2n} = n^{-\frac{1}{2}} \sum_{t=1}^n \psi(\varepsilon_t) h_t$ follows straightforwardly from a martingale CLT. This leaves the relatively difficult task of showing that $c_{1n} - c_{2n} = o_p(1)$. In fact our proof does not directly consider $c_{1n} - c_{2n}$ because we do not assume ψ is smooth; we instead approximate the e_t by the $\sigma_0 \varepsilon_t$ within the smooth estimate of ψ and then appeal to mean square approximation of $\psi(\varepsilon_t)$ by its least squares projection on the $\phi(\varepsilon_t)^\ell$, $\ell = 1, ..., L$, as $L \to \infty$, as in Newey (1988). However for this, $S_n = n^{-\frac{1}{2}} \sum_t \delta_t h_t$ (i.e. $c_{1n} - c_{2n}$ with $\psi(x)$ replaced by x) is relevant, and the sharper a bound we obtain for it the weaker some other conditions can be; we obtain $S_n = O_p\left((\log n)^{3/2} n^{-\frac{1}{2}}\right)$.

The same kind of issue arises in theory for Whittle estimation. For short-memory stationary processes, with $\xi_0 = 0$, Hannan (1973) established the CLT for various Whittle estimates. His proof does not work under stationary long memory,

$0 < \xi_0 < \frac{1}{2}$, due to the bad behaviour of the periodogram and spectral density at low frequencies. However, in this case Fox and Taqqu (1986), Dahlhaus (1989) and Giraitis and Surgailis (1990) delicately exploited a kind of balance between these quantities in order to establish CLTs. The CSS estimate minimizing $\sum_{t=1}^{n} e_t^2(\theta)$ (see Remark 5 in Section 3 concerning (1.4)) is not one of those considered by these authors, but its CLT requires showing $S_n = o_p(1)$, which entails similar challenge to results they established for the somewhat different quadratic forms arising from their parameter estimates. Our results for replacing e_t by $\sigma_0 \varepsilon_t$ can be employed to provide a proof of asymptotic normality of the CSS version of Whittle estimate. Whittle and adaptive estimation are both areas in which asymptotic results under short- and long-memory are qualitatively the same, but sufficient methods of proof significantly differ.

6. PROOF OF THEOREM 1

The consistency of the covariance matrix estimates is implied by the proof of the CLT. By far the most significant features of this are accomplished in the lemmas in the following section. Their application is mostly relatively straightforward, and is thus described here in abbreviated form. For notational convenience we now write $\theta_3 = \sigma$ and expand θ as $\theta = \left(\theta_1^T, \theta_2^T, \theta_3\right)^T$. We also abbreviate $\sum_{t=1}^{n}$ to $\sum_t$, and $E_t(\theta_0)$, $E(\theta_0)$, $E_{ti}(\theta_0)$ to E_t, E, E_{ti} respectively, $i = 1, 2$. By the mean value theorem, for $i = 1, 2$,

$$\hat{\theta}_i - \theta_{0i} = \left\{ I_{p_i} + \frac{R_i(\tilde{\theta})^{-1}}{\mathcal{J}_L(\tilde{\theta})} \bar{S}_{Lii} \right\} \left(\tilde{\theta}_i - \theta_{0i}\right) + \frac{R_i(\tilde{\theta})^{-1}}{\mathcal{J}_L(\tilde{\theta})} \left\{ \sum_{\substack{j=1 \\ j \neq i}}^{3} \bar{S}_{Lij}(\tilde{\theta}_j - \theta_{0j}) + r_{Li}(\theta_0) \right\},$$

where, with $[S_{Li1}(\theta), S_{Li2}(\theta), S_{Li3}(\theta)] = (\partial/\partial\theta^T) r_{Li}(\theta)$, each row of $\bar{S}_{Lij}$ is formed from the corresponding row of $S_{Lij}(\theta)$ by replacing θ by $\bar{\theta}$ such that $\|\bar{\theta}-\theta_0\| \le \left\|\tilde{\theta}-\theta_0\right\|$ where $\|A\| = \{tr(A^T A)\}^{\frac{1}{2}}$. Write $D_{1n} = D_{3n} = n^{\frac{1}{2}}$, $D_{2n} = D_n$ and define $\mathcal{N} = \{\theta : \|D_{in}(\theta_i - \theta_{0i})\| \le 1,\ i = 1,2,3\}$. The result follows if

$$\sup_{\mathcal{N}} \left\|D_{in}^{-1}\{R_i(\theta) - R_i(\theta_0)\}D_{in}^{-1}\right\| \ \to_p \ 0, \quad i \ = \ 1,2, \tag{6.1}$$

$$\sup_{\mathcal{N}} \left\|D_{in}^{-1}\{S_{Lij}(\theta) - S_{Lij}(\theta_0)\}D_{jn}^{-1}\right\| \ \to_p \ 0, \quad i \ = \ 1,2, \quad j = 1,2,3, \tag{6.2}$$

$$\sup_{\mathcal{N}} |\mathcal{J}_L(\theta) - \mathcal{J}_L(\theta_0)| \ \to_p \ 0, \tag{6.3}$$

$$D_{in}^{-1} R_i(\theta_0) D_{in}^{-1} \ \to_p \ \Omega_i, \quad i = 1,2, \tag{6.4}$$

$$\{R_i(\theta_0)\mathcal{J}_L(\theta_0)\}^{-1} S_{Lij}(\theta_0) \ \to_p \ -I_{p_i} 1(i=j), \quad i = 1,2, \ j = 1,2,3, \tag{6.5}$$

$$\mathcal{J}_L(\theta_0) \ \to_p \ \mathcal{J}, \tag{6.6}$$

$$\begin{bmatrix} n^{-\frac{1}{2}} r_1 \\ D_n^{-1} r_2 \end{bmatrix} \ \to_d \ N\left(0, \begin{bmatrix} \mathcal{J}\Omega_1 & 0 \\ 0 & \mathcal{J}\Omega_2 \end{bmatrix}\right), \tag{6.7}$$

$$D_{in}^{-1}\{r_{Li}(\theta_0) - r_i\} \ \to_p \ 0, \quad i = 1,2, \tag{6.8}$$

where

$$r_1 = \sum_t \psi(\varepsilon_t)\varepsilon'_{t1}, \quad r_2 = \sum_t \psi(\varepsilon_t) E'_{t2},$$

with $\varepsilon'_{t1} = \left(\partial/\partial\theta_1^{(+)T}\right) \alpha\left(B;\theta_1^{(+)}\right)/\sigma_0 = \gamma(B;\nu_0)\,\varepsilon_t$.

The most difficult and distinctive problems occur in (6.8) for $i = 1$, which faces the $e_t - \sigma_0\varepsilon_t$ problem, as well as the increasing L, in the presence of normalization only by D_{1n}^{-1}. The first of these aspects is also in (6.1), (6.4) and both are in (6.2), (6.3), (6.5) and (6.6) but the normalizations make (6.4)-(6.6) much easier to deal with and the proof details are otherwise relatively standard, albeit lengthy.

The same may also be said for (6.1)-(6.3), except for the approximation of the fractional difference Δ^{ξ_0} by Δ^{ξ} for $|\xi - \xi_0| \leq n^{-\frac{1}{2}}$, bearing in mind that "nonstationary" values of ξ, ξ_0 are permitted. The basic steps in proving (6.1) - (6.3) are illustrated by the least complicated case (6.1). By elementary inequalities it suffices to show that $\sup_{\mathcal{N}} \sum_t \left\| D_{in}^{-1}(e'_{ti}(\theta) - e'_{ti}(\theta_0)) \right\|^2 \rightarrow_p 0$, $i = 1, 2$. Write $\alpha = \alpha(B; \theta^{(-)})$, $\alpha' = \alpha'(B; \theta^{(-)})$ with α_0, α'_0 denoting these quantities at $\nu = \nu_0$. For $i = 2$, it suffices to apply Lemmas 1, 2, 3 and (with $m = \xi_0$) 4, the j-th elements of $\alpha_0(\Delta^{\xi} - \Delta^{\xi_0})z_{2t}$, and $(\alpha - \alpha_0)\Delta^{\xi_0} z_{2t}$ being respectively $O\left(n^{-\frac{1}{2}}(\log t) t^{\chi_j - \xi_0}\right)$ and $O\left(n^{-\frac{1}{2}} t^{\chi_j - \xi_0}\right)$ uniformly in $\mathcal{N}$, noting that $\xi_0 > -\frac{1}{2}$ and $\chi_j \geq \xi_0 - \frac{1}{2}$ implies $\chi_j > -1$ and $\xi_0 < \chi_j + 1$. For $i = 1$, the terms in z_{2t} are dealt with similarly, while Lemmas 1-4 give, for example, $\alpha'_0 \left(\Delta^{\xi} - \Delta^{\xi_0}\right)(s_t + \mu^* t^{\xi_0}) = O\left(n^{-\frac{1}{2}}(\log t)^2\right)$ and $(\alpha' - \alpha'_0)\Delta^{\xi_0}(s_t + \mu^* t^{\xi_0}) = O(n^{-\frac{1}{2}})$ uniformly in $\mathcal{N}$. In the above we apply first Lemma 3, then Lemma 1 and then Lemma 2, noting that in case (ii) of Lemma 1 must be used (either for a leading term or remainder) the coefficient of s^j in the expression of $-\log(1-s)$, and thus of $(-\log(1-s))^r$, is positive for all $j \geq 1$, so far nonnegative sequences g_t, h_0, such that $g_t \leq h_t$, we have $|(-\log \Delta)^r g_t| \leq |(-\log \Delta)^r x_t|$. So far as contributions from x_t are concerned, from Lemma 5

$$\sup_{\mathcal{N}} \left\| (\alpha' - \alpha'_0)\Delta^{\xi_0} x_t \right\| \leq \sum_{j=0}^{t-1} \left\{ \sup_{\mathcal{N}} \left\| \alpha'_j - \alpha'_{0j} \right\| \right\} \left\{ \left| \Delta^{\zeta_0} v^{\#}_{t-j} \right| + \left| (\log \Delta) \Delta^{\zeta_0} v^{\#}_{t-j} \right| \right\},$$

where α'_j, α'_{0j} are the j-th Fourier coefficients of α', α'_0. By the mean value theorem and Lemma 6 this has second moment $O(n^{-1})$. The same result holds for $\alpha'_0 \left(\Delta^{\xi} - \Delta^{\xi_0}\right) x_t$ after taking $m = m_0$ in Lemma 4, noting that its supremum over $\mathcal{N}$ is bounded by

$$Cn^{-\frac{1}{2}} \left\| \alpha'_0 \Delta^{\xi_0} x_t \right\| + Cn^{-\frac{1}{2}} \left\| (\log \Delta) \alpha'_0 \Delta^{\xi_0} x_t \right\| + Cn^{-1} \left(\sum_{j=1}^{t} v^2_{t-j} \right)^{\frac{1}{2}}$$

and applying Lemmas 5 and 6. The proof of (6.1) is readily completed.

Before coming to (6.8) we briefly discuss (6.7). Consider variates $U = \left(n^{-\frac{1}{2}} r_1^T, (D_n^{-1} r_2)^T\right)^T$, $V = \lambda^T (EUU^T)^{-\frac{1}{2}} U$ for a $(p_1 + p_2) \times 1$ vector λ such that $\lambda^T \lambda = 1$. We have $EV = 0$, $EV^2 = 1$, since $E\psi(\varepsilon_0) = 0$ and ε'_{t1} is independent of ε_t, so (6.7) follows from Theorem 2 of Scott (1973) if

$$\sum_t \begin{bmatrix} n^{-\frac{1}{2}} \varepsilon'_{t1} \\ D_n^{-1} E'_{t2} \end{bmatrix} \begin{bmatrix} n^{-\frac{1}{2}} \varepsilon'_{t1} \\ D_n^{-1} E'_{t2} \end{bmatrix}^T \to_p \begin{bmatrix} \Omega_1 & 0 \\ 0 & \Omega_2 \end{bmatrix}, \tag{6.9}$$

$$\sum_t \psi(\varepsilon_t^2) \left\{ n^{-1} \left\| \varepsilon'_{t1} \right\|^2 1 \left(\left\| \psi(\varepsilon_t) \varepsilon'_{t1} \right\| \geq \delta n^{\frac{1}{2}} \right) + \left\| D_n^{-1} E'_{t2} \right\|^2 1 \left(\left\| \psi(\varepsilon_t) D_n^{-1} E'_{t2} \right\| \geq \delta \right) \right\} \to_p 0 \tag{6.10}$$

for any $\delta > 0$. The proof of (6.9) follows from Lemmas 1 and 3 and approximating sums by integrals, while that of (6.10) follows from stationarity and finite variance of $\psi(\varepsilon_t)$ and ε'_{t1} and the slowly changing character of z_{2t}.

We prove (6.8) only for $i = 1$, the case $i = 2$ involving some of the same steps but being much easier. Define $\Xi^{(L)}(s) = \phi^{(L)}(s) - E\phi^{(L)}(\varepsilon_t)$, $W^{(L)} = E\left\{\Xi^{(L)}(\varepsilon_t)\Xi^{(L)}(\varepsilon_t)^T\right\}$. It follows from Lemma 8 that $W^{(L)}$ is non-singular, and thence we define $a^{(L)} = W^{(L)-1} w^{(L)}$ where $w^{(L)} = E\left\{\phi'^{(L)}(\varepsilon_t)\right\} = E\left\{\phi^{(L)}(\varepsilon_t)\psi(\varepsilon_t)\right\}$, by integration-by-parts, as in Beran (1976) and as justified under our conditions by Lemma 2.2 of Newey (1988). Defining also $\bar{\psi}^{(L)}(\varepsilon_t; a^{(L)}) = a^{(L)T} \Xi^{(L)}(\varepsilon_t)$ we have

$$n^{-\frac{1}{2}} \left\{ r_{L1}(\theta_0) - r_1 \right\} = \sum_{i=1}^{4} \sum_{j=1}^{2} A_{ij} - A_{11},$$

where $A_{ij} = n^{-\frac{1}{2}} \sum_t B_{it} C_{jt}$ and $B_{1t} = \psi(\varepsilon_t)$, $B_{2t} = \bar{\psi}^{(L)}(\varepsilon_t; a^{(L)}) - \psi(\varepsilon_t)$, $B_{3t} = \psi^{(L)}(\varepsilon_t; \hat{a}^{(L)}(\varepsilon)) - \bar{\psi}^{(L)}(\varepsilon_t; a^{(L)})$, $B_{4t} = \tilde{\psi}^{(L)}(\theta_0, \sigma_0) - \psi^{(L)}(\varepsilon_t; \hat{a}^{(L)}(\varepsilon))$, $C_{1t} = \sigma_0 \varepsilon'_{t1}$, $C_{2t} = E'_{t1} - \sigma_0 \varepsilon'_{t1}$.

Since ε'_{t1} is $\{\varepsilon_s, s < t\}$-measurable and $E\left\|\varepsilon'_{01}\right\|^2 \leq C\left\|\Omega_1\right\| < \infty$, while B_{2t} has zero mean, $E\left\|A_{21}\right\|^2 \leq CEB_{20}^2 \to 0$, as $L \to \infty$ from Freud (1971, pp.74-77), Newey (1988, Lemma 2.2) since the moments of $\phi(\varepsilon_0)$ characterize its distribution under A2, A8.

Before discussing other A_{ij} define

$$\mu_a = 1 + E\left\{|\varepsilon_0|^a 1(|\varepsilon_t| > 1)\right\},$$

for $a > 0$, and the following sequences:

$$\begin{aligned}\rho_{aL} &= CL, \quad \text{if } a = 0, \\ &= (CL)^{aL/\omega}, \quad \text{if } a > 0 \text{ and A2(b) holds}, \\ &= C^L, \quad \text{if } a > 0 \text{ and A2(c) holds},\end{aligned}$$

suppressing reference in ρ_{aL} to the arbitrarily large constant C; and also

$$\pi_L = (\log L)\eta^{2L}1(\varphi < 1) + (L\log L)\eta^{2L}1(\varphi = 1) + (\log L)(\eta\varphi)^{2L}1(\varphi > 1),$$

for $L > 1$.

Write $A_{31} = (b_{1n} - b_{2n}b_{3n})\left\{\hat{a}^{(L)}(\varepsilon) - a^{(L)}\right\} - b_{2n}b_{3n}a^{(L)}$, where $b_{1n} = n^{-\frac{1}{2}}\sigma_0 \sum_t \varepsilon'_{t1}\Xi^{(L)}(\varepsilon_t)^T$, $b_{2n} = n^{-1}\sum_t \varepsilon'_{t1}$, $b_{3n} = n^{-\frac{1}{2}}\sigma_0 \sum_t \Xi^{(L)}(\varepsilon_t)^T$. We have $E\left|\phi(\varepsilon_0)\right|^r \le \mu_{\kappa r}$ and thus from Lemma 9

$$E\left\|b_{1n}\right\|^2 + E\left\|b_{3n}\right\|^2 \le C\sum_{\ell=1}^{L}\left(E\left\|\varepsilon'_{01}\right\|^2 + 1\right)E\phi^{2\ell}(\varepsilon_0) \le \rho_{2\kappa L}.$$

Since $b_{2n} = O_p(n^{-\frac{1}{2}}\log n)$ from Lemma 17, we deduce from Lemma 10 that

$$A_{31} = O\left(\frac{L\rho_{2\kappa L}\pi_L}{n^{\frac{1}{2}}}\left(\log n + L^{\frac{1}{2}}\rho_{4\kappa L}^{\frac{1}{2}}\pi_L\right)\right). \tag{6.10}$$

Before imposing A9, we estimate A_{41}, which can be written

$$n^{-\frac{1}{2}}\sigma_0\left[\sum_t \varepsilon'_{t1}\left\{\Phi^{(L)}\left(E_t/\sigma_0\right) - \Phi^{(L)}(\varepsilon_t)\right\}\right]\hat{a}^{(L)}(E/\sigma_0) \tag{6.11}$$

$$+n^{-\frac{1}{2}}\sigma_0\sum_t \varepsilon'_{t1}\Phi^{(L)}(\varepsilon_t)^T\left\{\hat{a}^{(L)}(E/\sigma_0) - \hat{a}^{(L)}(\varepsilon)\right\}. \tag{6.12}$$

The square-bracketed quantity in (6.11) has norm bounded by

$$\left(\sum_{\ell=1}^{L}\left\|\sum_t \varepsilon'_{ti}\delta_{\ell t}\right\|^2\right)^{\frac{1}{2}} + n^{-1}\left\|\sum_t \varepsilon'_{ti}\right\|\left\{\sum_{\ell=1}^{L}\left(\sum_t \delta_{\ell t}\right)^2\right\}^{\frac{1}{2}}, \tag{6.13}$$

where $\delta_{\ell t} = \phi_\ell(E_t/\sigma_0) - \phi_\ell(\varepsilon_t)$. We have

$$\delta_{\ell t} = \phi_\ell'(\varepsilon_t)d_t + \frac{1}{2}\phi_\ell''(\bar{\varepsilon}_t)d_t^2, \tag{6.14}$$

where $|\bar{\varepsilon}_t - \varepsilon_t| \leq |d_t|$, $d_t = E_t/\sigma_0 - \varepsilon_t$. Now $e_t = \alpha\left(B;\theta_{01})(s_t + \mu^* t^{\xi_0} + x_t\right)$, and from Lemma 5 (see also (1.13))

$$\alpha\left(B;\theta_{01}\right)x_t = \alpha(B;\theta_{01}^{(+)})v_t^{\#} = \sigma_0\varepsilon_t - \sum_{j=0}^{\infty}\alpha_{t+j}(\theta_{01}^{(+)})v_{-j} = \sigma_0\varepsilon_t + d_{1t},$$

where

$$d_{1t} = -\sum_{j=1}^{\infty}\lambda_{jt}\varepsilon_{t-j}, \qquad \lambda_{jt} = \sum_{k=0}^{j}\alpha_{k+t}(\theta_{01}^{(+)})\beta_{j-k}(\theta_{01}^{(+)}).$$

Since $\alpha(B;\theta_0)s_t = o(t^{-\frac{1}{2}})$ and $\alpha(B;\theta_{01})t^{\xi_0} = \alpha(1;\theta_0^{(-)})\Gamma(\xi_0+1)+O(t^{-1})$ from Lemma 1, it follows that

$$d_t = d_{1t} + d_2 + d_3 + o(t^{-\frac{1}{2}}), \tag{6.15}$$

where $d_2 = n^{-1}\sum_{j=0}^{\infty}(\sum_t \lambda_{jt})\varepsilon_{-j}$, $d_3 = n^{-1}\sum_t \varepsilon_t$. From Lemmas 13, 14 and 18, for $2 \leq r \leq 4$ under A2(a) and $r > 4$ under A2(b) and A2(c),

$$E\left|d_{1t}\right|^r \;\leq\; (Cr)^{2r}t^{-r/2}\mu_{r_+}^{r/r_+}, \tag{6.16}$$

$$E\left|d_2\right|^r + E\left|d_3\right|^r \;\leq\; (Cr)^{2r}n^{-r/2}\mu_{r_+}^{r/r_+}, \tag{6.17}$$

where r_+ is the smallest even integer such that $r \leq r_+$. Returning to (6.13), we have

$$\left\|\sum_t \varepsilon_{t1}'\delta_{\ell t}\right\| \;\leq\; \left\|\sum_t \varepsilon_{t1}'\left\{\phi_\ell'(\varepsilon_t) - E\phi_\ell'(\varepsilon_0)\right\}d_{1t}\right\| \tag{6.18}$$

$$+\left\|\sum_t \varepsilon_{t1}'\left\{\phi_\ell'(\varepsilon_t) - E\phi_\ell'(\varepsilon_0)\right\}\right\|(|d_2| + |d_3|) \tag{6.19}$$

$$+\left|E\phi_\ell'(\varepsilon_0)\right|\left\|\sum_t \varepsilon_{t1}'d_{1t}\right\| \tag{6.20}$$

$$+\left|E\phi_\ell'(\varepsilon_0)\right|\left\|\sum_t \varepsilon_{t1}'\right\|(|d_2| + |d_3|) \tag{6.21}$$

$$+\left\|\sum_t \varepsilon_{t1}'\phi_\ell''(\bar{\varepsilon}_t)d_t^2\right\|. \tag{6.22}$$

Now

$$
\begin{aligned}
\left|\phi_\ell'(s)\right| &= \ell\left|\phi'(s)\phi^{\ell-1}(s)\right| \\
&\leq C\ell(1+|\phi(s)|^K\left\{1\left(|s|\leq 1\right)+|s|^{\kappa(\ell-1)}1(|s|>1)\right\} \\
&\leq C\ell\left\{1\,|s|\leq 1+|s|^{\kappa(\ell-1+K)}1(|s|>1\right\}, \qquad (6.23)
\end{aligned}
$$

and since ε_t is independent of $\varepsilon_{ti}'d_{1t}$, the right side of (6.18) is

$$
O_p\left(\left\{E\phi_\ell'(\varepsilon_0)^2\right\}^{\frac{1}{2}}\sum_t\left(E\left\|\varepsilon_{ti}'\right\|^4Ed_{1t}^4\right)^{\frac{1}{2}}\right)=O_p\left(\ell\mu_{2\kappa(\ell+K)}^{\frac{1}{2}}\log n\right),
$$

using (6.16). The same bound applies to (6.19)-(6.21), proceeding similarly and using respectively (6.17), Lemma 16, and (6.17) with Lemma 17; note that it is the second factor in (6.20) which leads to the main work in handling the quantity S_n discussed in Section 5. So far as (6.22) is concerned, note that as in (6.23)

$$
\left|\phi_\ell''(s)\right|\leq C\ell^2\left\{1\left(|s|\leq 1\right)+|s|^{\kappa(\ell-1+2K)}1(|s|>1)\right\},
$$

so by the c_r-inequality (Loeve, 1977, p.157) (6.22) is bounded by

$$
\begin{aligned}
&C^{\kappa\ell+1}\ell^2\sum_t\left\|\varepsilon_{t1}'\right\|\left\{d_{1t}^2+d_{1t}^2\left|\varepsilon_t\right|^{\kappa(\ell+K)}+\left|d_{1t}\right|^{\kappa(\ell+K)+2}\right\} \qquad (6.24)\\
&+C^{\kappa\ell+1}\ell^2\sum_t\left\|\varepsilon_{t1}'\right\|\left\{(d_t-d_{1t})^2(1+\left|\varepsilon_t\right|^{\kappa(\ell+K)})+\left|d_t-d_{1t}\right|^{\kappa(\ell+K)+2}\right\}. \qquad (6.25)
\end{aligned}
$$

By (6.16) and Hölder's and Jensen's inequalities, (6.24) has expectation bounded by

$$
C^{\kappa\ell+1}\ell^2\left\{\mu_{\kappa(\ell+K)}\log n+\sum_t\left(E\left|d_{1t}\right|^{2\kappa(\ell+K)+4}\right)^{\frac{1}{2}}\right\}\leq C(C\ell)^{2\kappa\ell}\ell^2\mu_{r_\ell}^{\frac{1}{2}}\log n,
$$

r_ℓ being the smallest integer such that $r_\ell\geq 2\kappa(\ell+K)+4$. From (6.14) and (6.17), (6.25) is of smaller order in probability. It follows from Lemma 9 that

$$
\left(\sum_{\ell=1}^L\left\|\sum_t\varepsilon_{t1}'\delta_{\ell t}\right\|^2\right)^{\frac{1}{2}}=O_p\left((CL)^{2\kappa L+2}\rho_{2\kappa L}^{\frac{1}{2}}\log n\right).
$$

By a similar but easier proof, the second term in (6.13) has the same bound, and by Lemmas 10 and 19

$$(6.11) \;= O_p\left((CL)^{2\kappa L+3}\rho_{2\kappa L}\pi_L n^{-\frac{1}{2}}\log n\right).$$

Next, from similar but simpler arguments to those above

$$n^{-\frac{1}{2}}\left\|\sum_t \varepsilon'_{t1}\Phi^{(L)}(\varepsilon_t)^T\right\| = O_p\left(\rho_{2\kappa L}^{\frac{1}{2}}\log n\right).$$

Application of Lemma 9 indicates that (6.12) is

$$O_p\left(\rho_{2\kappa L}^2\pi_L^2\left(L^2 n^{-\frac{1}{2}}\log n + (CL)^{4\kappa L+3}n^{-1}(\log n)^2\right)\right).$$

Thus

$$A_{41} = O_p\left(\rho_{2\kappa L}\pi_L\left(\rho_{2\kappa L}\pi_L L^2 + (CL)^{2\kappa L+3} + \rho_{2\kappa L}\pi_L(CL)^{4\kappa L+3}n^{-\frac{1}{2}}\right)n^{-\frac{1}{2}}\log n\right). \tag{6.26}$$

Comparison of (6.10) and (6.26) indicates that A_{31} is dominated by A_{41}, whose behaviour under A9 we thus now consider. Take $\kappa = 0$. From Lemma 9, under A9(i)

$$\begin{aligned} A_{41} &= O_p\left(L^4\pi_L^2 n^{-\frac{1}{2}}\log n\right) \\ &= O_p\left(\exp\left[\log n\left\{\left(4\log L + \log\log n + 2\log\pi_L\right)/\log n - \frac{1}{2}\right\}\right]\right) \end{aligned}$$

which is $o_p(1)$ if $\overline{\lim}\log\pi_L/\log n < \frac{1}{4}$, as is clearly implied by (3.4). Now take $\kappa > 0$ under A2(b). From Lemma 9, under A9(ii)

$$A_{41} = O_p\left(\left(L^{4\kappa L/\omega+2} + L^{2\kappa L(1+1/\omega)+3}\right)n^{-\frac{1}{2}}\log n\right) = o_p(1),$$

on proceeding as before. Under A2(c), Lemma 9 and A9(iii) give

$$A_{41} = O_p\left((CL)^{2\kappa L}n^{-\frac{1}{2}}\log n\right) = o_p(1).$$

To consider A_{12}, we can proceed as earlier to write

$$E'_{t1} - \varepsilon'_{t1} = D_{1t} + D_2 + D_3 + \left(t^{-\frac{1}{2}}\log t\right),$$

where

$$D_{1t} = -\sum_{j=1}^{\infty} \tilde{\lambda}_t \varepsilon_{t-j}, \quad D_2 = n^{-1} \sum_{j=0}^{\infty} \left(\sum_t \tilde{\lambda}_{jt} \right) \varepsilon_{-j}, \quad D_3 = n^{-1} \sum_t \varepsilon'_{t1}$$

and $\tilde{\lambda}_{jt} = \sum_{k=0}^{j} (\partial/\partial\theta_1^{(+)T}) \alpha_{k+t}(\theta_{01}^{(+)}) \beta_{j-k}(-\theta_{01}^{(+)})$. Using (7.23) and (7.24) of Lemma 13, we deduce that $\left|\tilde{\lambda}_{jt}\right| \leq C(\log t) j^{\zeta_0} t^{-\zeta_0 - 1}$, $j \leq t$, and $\left|\tilde{\lambda}_{jt}\right| \leq C(\log t) j^{\zeta_0 - 1} \max(j^{-\zeta_0}, t^{-\zeta_0})$, $j > t$, and then proceeding as in Lemma 14, that $\sum_{j=0}^{\infty} \tilde{\lambda}_{jt}^2 \leq C t^{-1} \log^2 t$, $\sum_{j=0}^{\infty} \left(\sum_{t=1}^{n} \tilde{\lambda}_{jt} \right)^2 \leq Cn \log^2 n$. Noting that $E\left(\sum_t \psi(\varepsilon_t) D_{1t}\right)^2 \leq C \sum_t E D_{1t}^2$, using also Lemma 17 and proceeding as in the proof for (6.11) it follows that $A_{12} = O_p\left(n^{-\frac{1}{2}} \log^{3/2} n\right)$.

The remainder of the proof of (6.8) with $i = 1$ deals in similar if easier ways with quantities already introduced and is thus omitted. □

7. TECHNICAL LEMMAS

To simplify lemma statements, we take it for granted that, where needed, Assumptions A1-A9 hold.

Part (ii) of the following lemma is only needed to show that s_t in (1.9) contributes negligibly, in particular when it includes $\tau_1 \leq \xi_0 - 1$.

Lemma 1 *(i) For $w_t = t^\gamma$ with $\gamma > -1$ and $\xi \in (-\frac{1}{2}, \gamma + 1)$,*

$$\Delta^\xi w_t^\# = \frac{\Gamma(\gamma+1)}{\Gamma(\gamma-\xi+1)} t^{\gamma-\xi} + O\left(t^{\gamma-\xi-1} + t^{\gamma-m-1} 1(\xi > 0)\right),$$

as $t \to \infty$, where m is the integer such that $\xi - 1 < m \leq \xi$.

(ii) For $w_t = (\log t)^r t^\gamma$, $r \geq 0$, $\xi > -\frac{1}{2}$,

$$\Delta^\xi w_t^\# = O\left(t^{\max(\gamma,-1)-\xi+d}\right), \quad \text{as } t \to \infty,$$

for any $\delta > 0$.

Proof: (i) The proof when ξ is a nonnegative integer is straightforward, so we assume this is not the case. We have

$$\sum_{j=0}^{\infty} j^k \Delta_j(-\xi) = 0, \qquad j = 0, ..., m, \tag{7.1}$$

when $m \geq 0$ and $\xi > 0$, $(1-s)^{\xi}$ and its first m derivatives in s being zero at $s = 1$. With $a_k = \Delta_k(-\gamma)$,

$$\begin{aligned}
\Delta^{\xi} w_t^{\#} &= \sum_{j=0}^{t-1} \Delta_j(-\xi)(t-j)^{\gamma} \\
&= t^{\gamma} \sum_{j=0}^{t-1} \Delta_j(-\xi) \sum_{k=0}^{\infty} a_k \,(j/t)^k \\
&= -t^{\gamma} \textstyle\sum_k' (t-k)^{-k} a_k \displaystyle\sum_{j=t}^{\infty} j^k \Delta_j(-\xi) 1(m \geq 0) + t^{\gamma} \textstyle\sum_k'' (t-k)^{-k} a_k \displaystyle\sum_{j=0}^{t-1} j^k \Delta_j(-\xi),
\end{aligned} \tag{7.2}$$

where $\sum_k' = \sum_{k=0}^{m}$, $\sum_k'' = \sum_{k=\max(m+1,0)}^{\infty}$ and we apply (7.1). By Stirling's approximation

$$\left| \Delta_j(-\xi) - \frac{j^{-\xi-1}}{\Gamma(-\xi)} \right| \leq C j^{-\xi-2}, \qquad j \geq 1, \tag{7.3}$$

so (7.2) differs from

$$\frac{t^{\gamma}}{\Gamma(-\xi)} \left\{ -\textstyle\sum_k' (t-k)^{-k} a_k \displaystyle\sum_{j=t}^{\infty} j^{k-\xi-1} 1(m \geq 0) + \textstyle\sum_k'' (t-k)^{-k} a_k \displaystyle\sum_{j=0}^{t-1} j^{k-\xi-1} \right\} \tag{7.4}$$

by

$$O\left(t^{\gamma} \textstyle\sum_k' t^{-k} \left| a_k \right| \displaystyle\sum_{j=t}^{\infty} j^{k-\xi-2} 1(m \geq 0) + t^{\gamma} \textstyle\sum_k'' t^{-k} \left| a_k \right| \displaystyle\sum_{j=0}^{t-1} j^{k-\xi-2} \right). \tag{7.5}$$

Now

$$\begin{aligned}
\sum_{j=0}^{t-1} j^{-\alpha} &= t^{1-\alpha}/(1-\alpha) + O(t^{-\alpha}), \qquad \alpha < 1, \\
\sum_{j=t}^{\infty} j^{-\alpha} &= t^{1-\alpha}/(\alpha-1) + O(t^{-\alpha}), \qquad \alpha > 1.
\end{aligned} \tag{7.6}$$

Thus (7.5) is

$$O\left(t^{\gamma-\xi-1}\sum_k' \frac{|a_k|}{\xi+1-k}1(m\geq 0)+t^{\gamma-m-1}\frac{|a_{m+1}|}{\xi-m}1(m\geq -1)\right.$$
$$\left.+t^{\gamma-\xi-1}\sum_k''' \frac{|a_k|}{k-\xi-1}\right)$$
$$= O\left(t^{\gamma-\xi-1}\left\{\sum_k' |a_k|\, 1(m\geq 0)+\sum_k''(k+1)^{-1}|a_k|+Ct^{\gamma-m-1}|a_{m+1}|\, 1(m\geq -1)\right\}\right)$$

where $\sum_k''' = \sum_{k=\max(m+2,0)}^{\infty}$. The first sum in braces is finite because m and the a_k are, while the second sum is finite because $|a_k| \leq Ck^{-\gamma-1}$. Thus since $\gamma > -1$, (7.5) is $O\left(t^{\gamma-m-1}\right)$ for $\xi > 0$ and $O\left(t^{\gamma-\xi}\right)$ for $\xi < 0$. Applying (7.6) again, (7.4) is

$$\frac{t^{\gamma-\xi}}{\Gamma(-\xi)}\sum_{k=0}^{\infty}\frac{a_k}{k-\xi}+O\left(t^{\gamma-\xi-1}\right),$$

and the leading term is $\{\Gamma(\gamma+1)/\Gamma(\gamma-\xi+1)\}t^{\gamma-\xi}$, from Whittaker and Watson (1940, p.260).

(ii) We have

$$\Delta^{\xi}w_t^{\#} = \sum_{j=0}^{t-1}\Delta_j(-\xi)\left\{\log(t0j)\right\}^r(t-j)^{\gamma}.$$

Noting that $\Delta_j(-\xi) = O(j^{-\xi-1})$ and (7.1) holds with $k = 0$ for $\xi > 0$,

$$\sum_{j=0}^{s}\Delta_j(-\xi)\{\log(t-j)\}^r(t-j)^{\gamma} \sim (\log t)^r t^{\gamma}\sum_{j=0}^{s}\Delta_j(-\xi) = O\left(t^{\gamma+\delta_1}s^{-\xi}\right)$$

for $s = o(t)$, $\delta_1 > 0$. On the other hand

$$\left|\sum_{j=s+1}^{t-1}\Delta_j(-\xi)\{\log(t0j)\}^r(t-j)^{\gamma}\right| \leq Cs^{-\xi-1}(\log t)^r\sum_{j=1}^{t}j^{\gamma}.$$

The sum on the right is $O(t^{1+\gamma})$ for $\gamma > -1$, $O\left((\log t)\right)$ for $\gamma = -1$ and $O(1)$ for $\gamma < -1$. Thus choosing $s = t^{1-\delta_2/(\xi+1)}$, $\delta_2 > 0$, produces the result.

Lemma 2 *For $w_t = t^{\gamma}$ and any integer $r > 0$, as $t \to \infty$*

$$(-\log\Delta)^r w_t^{\#} \sim (\log t)^r t^{\gamma} \quad \textit{for } \gamma > -1, \qquad (7.7)$$
$$= O(t^{-1}(\log t)^{r-1}\{1(\gamma < -1)+(\log t)1(\gamma = -1)\}), \quad \textit{for } \gamma \leq -1. \qquad (7.8)$$

Proof: Suppose (7.7) is true for a given r. Then as $t \to \infty$

$$(-\log \Delta)^{r+1} w_t^{\#} \sim (-\log \Delta)(\log t)^r w_t^{\#} = \sum_{j=1}^{t-1} j^{-1}\{\log(t-j)\}^r (t-j)^\gamma. \qquad (7.9)$$

The difference between this and

$$(\log t)^r \sum_{j=1}^{t-1} j^{-1}(t-j)^\gamma \qquad (7.10)$$

is bounded by $C(\log t)^{r-1}$ times

$$\sum_{j=1}^{t-1} j^{-1}\{\log t - \log(t-j)\}(t-j)^\gamma \le \sum_{j=1}^{t-1} j^{-1} \left|\log(1-j/t)\right| (t-j)^\gamma.$$

Splitting this into sums over $j \in [1, [t/2]]$, and $t \in [[t/2]+1, t-1]$, it is seen that the first of these is bounded by

$$t^{-1} \sum_{j=1}^{t-1} (t-j)^\gamma \le C t^\gamma$$

since $|\log(1-x)| \le x$ for $x \in (0, \frac{1}{2})$, while the second is bounded by

$$C t^{-1} \sum_{j=1}^{t-1} \left|\log(j/t)\right| j^\gamma \le C t^\gamma \log t.$$

The difference between (7.10) and

$$(\log t)^r t^\gamma \sum_{j=1}^{t-1} j^{-1} \qquad (7.11)$$

is bounded by

$$C(\log t)^r t^\gamma \sum_{j=1}^{t-1} j^{-1} \left|(1-j/t)^\gamma - 1\right| \le C(\log t)^r t^\gamma.$$

Then (7.11) $\sim (\log t)^{r+1} t^\gamma$ as $t \to \infty$. For $\gamma \le -1$, we can write

$$(-\log \Delta)^r w_t^{\#} = \sum_{j=1}^{t-1} a_j^{(r)} (t-j)^\gamma$$

where $a_j^{(r)} = O\left(\{\log(j+1)\}^{r-1} j^{-1}\right)$. Splitting the sum as before, the first one is $O\left((\log t)^r t^\gamma\right)$ and the second is $O\left((\log t)^{r-1} t^{-1}\right)$ for $\gamma < -1$ and $O\left((\log t)^r t^{-1}\right)$ for $\gamma = -1$. □

In the following four lemmas $b(e^{i\lambda})$ is taken to be a function with absolutely convergent Fourier series, and $b_j = (2\pi)^{-1}\int_{-\pi}^{\pi} b(e^{i\lambda})e^{ij\lambda}d\lambda$.

Lemma 3 *For* $w_t = t^\gamma$,

$$b(B)w_t^{\#} \sim b(1)t^\gamma, \qquad \textit{as } t \to \infty.$$

Proof: The left side equals $t^\gamma \sum_{j=0}^{t-1} b_j + \sum_{j=0}^{t-1} b_j\{(t-j)^\gamma - t^\gamma\}$. The first term differs by $o(t^\gamma)$ from $b(1)t^\gamma$, and the second is bounded by

$$Ct^\gamma \sum_{j=0}^{t-1} |b_j| \left|1 - \left(1 - \frac{j}{t}\right)^\gamma\right| \le Ct^{\gamma-1}\sum_{j=0}^{t} j\,|b_j| = o(t^\gamma),$$

from the Toeplitz lemma. □

Lemma 4 *For a sequence* w_t *such that* $w_t = 0,\ t \le 0$, *and any integer* r, *as* $\xi \to \xi_0$

$$\begin{aligned}(\log\Delta)^r(\Delta^\xi - \Delta^{\xi_0})b(B)w_t \;&\equiv\; (\log\Delta)^{r+1}\Delta^{\xi_0}b(B)w_t(\xi-\xi_0)\\ &\quad + O\left(\left\{\sum_{j=1}^{t}(\Delta^m w_{t-j})^2\right\}^{\frac{1}{2}}(\xi-\xi_0)^2\right) \qquad (7.12)\end{aligned}$$

for $m \in (\xi_0 - \frac{1}{2}, \xi_0 + \frac{1}{2})$.

Proof: By the mean value theorem the left hand side of (7.12) is

$$(\log\Delta)^{r+1}\Delta^{\xi_0}b(B)w_t(\xi-\xi_0) + \frac{1}{2}(\log\Delta)^{r+2}b(B)\Delta^{\bar{\xi}}w_t(\xi-\xi_0)^2,$$

for $\left|\bar{\xi}-\xi_0\right| \le |\xi-\xi_0|$. The last term can be written $\frac{1}{2}\sum_{j=1}^{t-1} c_j\Delta^m w_{t-j}(\xi-\xi_0)^2$, where c_j is the coefficient of s^j in the Taylor expansion of $\{\log(1-s)\}^{r+2}(1-s)^{\bar{\xi}-m}$. From Stirling's approximation, $c_j \sim (\log j)^{r+2}j^{m-\bar{\xi}-1}$ as $j \to \infty$. Now $m - \bar{\xi} \le m - \xi_0 + |\xi-\xi_0|$. The right side of this is less than $\frac{1}{2}$ if $|\xi-\xi_0| < \frac{1}{2} - m + \xi_0$, where the right side of the latter inequality is positive. Thus for $|\xi - \xi_0|$ small enough, $m - \bar{\xi} - 1 < -\frac{1}{2}$. Then $\sum_{j=1}^{\infty} c_j^2 < \infty$ for all r, so the proof is completed by the Cauchy inequality. □

Lemma 5 *For real ξ, and m_0 defined by (1.2),*

$$\Delta^{\xi} b(B) x_t = \Delta^{\xi - m_0} b(B) v_t^{\#}, \qquad t \in \mathbb{Z}. \tag{7.13}$$

Proof: The left hand side of (7.13) is

$$\Delta^{\xi} b(B) \Delta^{-m_0} v_t^{\#} = \Delta^{\xi - m_0} b(B) v_t^{\#}, \qquad t \in \mathbb{Z}. \qquad \square$$

The next Lemma gives a uniform bound for the variance of a process that is only "asymptotically stationary".

Lemma 6 *For all $r \geq 0$, and ζ_0 defined by (1.4),*

$$E\left\{(-\log \Delta)^r \Delta^{\zeta_0} b(L) v_t^{\#}\right\}^2 \leq C < \infty. \tag{7.14}$$

Proof: The left side of (7.14) is

$$\int_{-\pi}^{\pi} \left| \sum_{0}^{t-1} c_j e^{ij\lambda} \right|^2 \left|1 - e^{i\lambda}\right|^{-2\zeta_0} f(\lambda) d\lambda \leq C \left(\sum_{0}^{\infty} |c_j| \right)^2, \tag{7.15}$$

for $\zeta_0 > 0$ since $\left|1 - e^{i\lambda}\right|^{-2\zeta_0} f(\lambda)$ is integrable, c_j being the j-th Fourier coefficient of $[\{-\log(1 - e^{i\lambda})\}^r (1 - e^{i\lambda})^{\zeta_0}] b(e^{i\lambda})$. The j-th Fourier coefficient of the factor in braces is $O\left((\log j)^r j^{-\zeta_0 - 1}\right)$, so since the b_j are summable so are the c_j. For $\zeta_0 \leq 0$ $\left|1 - e^{i\lambda}\right|^{-2\zeta_0} f(\lambda)$ is bounded so the left side of (7.15) is bounded by $\sum_0^{\infty} c_j^2 < \infty$. □

Lemma 7 *Let S_m be the $m \times m$ matrix with (j,k)-th element $(j,k \geq 1)$,*

$$\int_{-1}^{1} u^{j+k-2} du = 2(j+k-1)^{-1} 1(j+k \text{ even}).$$

Then for m sufficiently large

$$tr(S_m^{-1}) < (2\pi)^{-2} \left[\frac{8}{3} + \frac{1}{2} \log \left\{ (2m-3) \left(\frac{2m}{3} - 1 \right) \right\} \right] \eta^{2m}.$$

Proof: It is clear that, like S_m, S_m^{-1} must have (j,k)-th element that is zero for all odd $j+k$. This immediately ensures the necessary property that even rows (columns) of S_m are orthogonal to odd rows (columns) of S_m^{-1}. It then suffices to study the two square matrices $S_{1,m}$ and $S_{2,m}$ formed from, respectively, the odd and even rows and columns of S_m. These exclude only and all zero elements of S_m, and S_m^{-1} is the $m \times m$ matrix whose $(2j-1, 2k-1)$-th element is the $(j+k)$-th element of $S_{1,m}^{-1}$, whose $(2j, 2k)$-th element is the (j,k)-th element of $S_{2,m}^{-1}$, and whose other elements are all zero. Thus it suffices to consider $S_{1,m}^{-1}$ and $S_{2,m}^{-1}$, and indeed $tr(S_m^{-1}) = tr(S_{1,m}^{-1}) + tr(S_{2,m}^{-1})$. We take m to be even; details for m odd are only slightly different and since we want a result only for large m this outcome will clearly be unaffected.

$S_{1,m}$ and $S_{2,m}$ are both Cauchy matrices (see e.g. Knuth 1968, p.36), having (j,k)-th element of form $(a_j + a_k)^{-1}$, in particular, $(j+k-\frac{3}{2})^{-1}$, $(j+k-\frac{1}{2})^{-1}$, respectively. From Knuth (1968, p.36) the j-th diagonal elements of $S_{1,m}^{-1}$, $S_{2,m}^{-1}$ are respectively $2U_1^2(j)/(4j-3)$, $2U_2^2(j)/(4j-1)$, where we define, for real s,

$$U_1(s) = \frac{\prod_{1 \leq i \leq m/2} (i+s-\frac{3}{2})^2}{\prod_{\substack{1 \leq i \leq m/2 \\ i \neq s}} (i-s)},$$

$$U_2(s) = \frac{\prod_{1 \leq i \leq m/2} (i+s-\frac{1}{2})^2}{\prod_{\substack{1 \leq i \leq m/2 \\ i \neq j}} (i-s)}.$$

Thus

$$\begin{aligned} tr(S_m^{-1}) &= 2\sum_{j=1}^{m/2} \left\{(4j-3)^{-1}U_1^2(j) + (4j-1)^{-1}U_2^2(j)\right\} \\ &\leq \left\{2+\frac{1}{2}\log(2m-3)\right\} \max_{1\leq j\leq \frac{m}{2}} U_1^2(j) \\ &\quad + \left\{\frac{2}{3}+\frac{1}{2}\log\left(\frac{2m}{3}-\frac{1}{3}\right)\right\} \max_{1\leq j\leq \frac{m}{2}} U_2^2(j). \end{aligned}$$

For $s \in (0, m/2-1)$

$$U_1(s) - U_1(s+1) = U_1(s)\left\{1 - \frac{(s+\frac{m}{2}-\frac{1}{2})(\frac{m}{2}-s)}{(s-\frac{1}{2})s}\right\}.$$

The factor in braces is $2-m(m-1)/\{2s(2s-1)\}$, which is negative for $s < s(m)$ and positive for $s > s(m)$, where $s(m) = \frac{1}{4} + \{2m(m-1)+1\}^{\frac{1}{2}}/4 \sim m/\sqrt{8}$ as $m \to \infty$. Thus, as $m \to \infty$

$$\max_{1\leq j\leq \frac{m}{2}} U_1(j) \sim \frac{\Gamma\left((\frac{1}{2}+1/\sqrt{8})m-\frac{1}{2}\right)}{\Gamma\left(m/\sqrt{8}-\frac{1}{2}\right)\Gamma(m/\sqrt{8})\Gamma\left((\frac{1}{2}-1/\sqrt{8})m+1\right)}. \qquad (7.16)$$

Applying Stirling's approximation, that is $\Gamma(am+b) \sim (2\pi)^{\frac{1}{2}}e^{-am}(am)^{am+b-\frac{1}{2}}$ as $m \to \infty$, and noting that

$$\left\{\frac{(1+2^{-\frac{1}{2}})^{1+2^{-\frac{1}{2}}}2^{2^{-\frac{1}{2}}}}{(1-2^{-\frac{1}{2}})^{1-2^{-\frac{1}{2}}}}\right\}^{\frac{1}{2}} = 1+2^{\frac{1}{2}},$$

(7.16) is $(2\pi)^{-1}\eta^m(1+o(1))$. In the same way it can be seen that $U_2(s)$ is maximized at $\{2m(m+1)+1\}^{\frac{1}{2}}/4-\frac{1}{4} \sim m/\sqrt{8}$, whence $\max_{1\leq j\leq m/2} U_2(j) \sim (2\pi)^{-1}\eta^m(1+o(1))$ also. The proof is then routinely completed. □

Denote by $\underline{\lambda}(A)$ the smallest eigenvalue of the matrix A.

Lemma 8 *As* $L \to \infty$,

$$\underline{\lambda}\left(W^{(L)}\right)^{-1} = O(\pi_L).$$

Proof: The method of proof, given Lemma 7, is similar to one in Newey (1988), but we obtain a refinement. Define $\phi_+^{(L)}(s) = \left(1, \phi^{(L)}(s)^T\right)^T$, $W_+^{(L)} = E\left\{\phi_+^{(L)}(\varepsilon_t)\phi_+^{(L)}(\varepsilon_t)^T\right\}$, so $W^{(L)} = PW_+^{(L)}P^T$, where the $L \times (L+1)$ matrix P consists of the last L rows of the $(L+1)$-rowed identity matrix. Then $\underline{\lambda}\left(W^{(L)}\right) \geq \underline{\lambda}\left(W_+^{(L)}\right)\underline{\lambda}(PP^T) = \underline{\lambda}\left(W_+^{(L)}\right)$. If $(-1,1) \subset (\phi(s_1), \phi(s_2))$ (which implies $\varphi \leq 1$) then (since $\phi'(s)$ is bounded on (s_1, s_2)) $\underline{\lambda}\left(W_+^{(L)}\right) \geq \underline{\lambda}(S_{L+1})/C \geq tr(S_{L+1}^{-1})^{-1}/C$, where we use S_m defined as in Lemma 7, which can then be applied. Otherwise, $W_+^{(L)}$ exceeds, by a non-negative definite matrix,

$$C^{-1}\int_{\phi(s_1)}^{\phi(s_2)} u^{(L)}u^{(L)T}du = \left\{\frac{\phi(s_2)-\phi(s_1)}{C}\right\} A \int_{-1}^{1} u^{(L)}u^{(L)T}duA^T, \tag{7.17}$$

where $u^{(L)} = (1, u, ..., u^L)^T$ and A is the lower-triangular matrix with (i,j)-th element $\begin{pmatrix} i-1 \\ j-i \end{pmatrix} \phi(s_1)^{i-j}\left\{\phi(s_2)-\phi(s_1)\right\}^{j-1}$, $j \leq i$. The smallest eigenvalue of (7.17) is no less than $C^{-1}\{\phi(s_2)-\phi(s_1)\}\underline{\lambda}(AA^T)\underline{\lambda}(S_{L+1})$. Now $\underline{\lambda}(AA^T) \geq \left\|A^{-1}\right\|^{-2}$, where by recursive calculation A^{-1} is seen to be lower-triangular with (i,j)-th element $a^{ij} = {}^{i-1}C_{i-j}\left\{-\phi(s_1)\right\}^{i-j}\left\{\phi(s_2)-\phi(s_1)\right\}^{1-i}$, $j \leq i$. Thus

$$\left\|A^{-1}\right\|^2 = \sum_{i=1}^{L+1}\left(\sum_{j=1}^{i} a^{ij2}\right) \leq \sum_{i=1}^{L+1}\left(\sum_{j=1}^{i}\left|a^{ij}\right|\right)^2 \leq \sum_{i=1}^{L+1}\varphi^{2(i-1)}.$$

This is bounded by $(1-\varphi^2)^{-1}$ for $\varphi < 1$, by $L+1$ for $\varphi = 1$, and by $(\varphi^2-1)^{-1}\varphi^{2(L+1)}$ for $\varphi > 1$. □

Lemma 9 *For $a \geq 0, b \geq 0$,*

$$\sum_{\ell=1}^{L} \mu_{a\ell+b} \leq \rho_{aL}. \tag{7.18}$$

Proof: In case $a = 0$, or $a > 0$ but A2(c) holds, this is trivial. For $a > 0$ under A2(b), monotonic non-decrease of μ_a in real a implies that the left side of (7.18) is bounded by

$$C\sum_{\ell=1}^{[aL+b]}\mu_{\kappa\ell} \leq \left(\frac{CL}{t}\right)^{(a/\kappa)L} E\left(e^{t|\varepsilon_0|^{\kappa}}\right)$$

for any $t \in (0,1)$, and by A2(b) there exists such t that this is bounded by ρ_{aL}. $\square$

Lemma 10 *As $n \to \infty$,*

$$\begin{aligned} \left\| a^{(L)} \right\| &= O\left(L \rho_{2\kappa L}^{\frac{1}{2}} \pi_L \right), \\ \left\| \widehat{a}^{(L)}(\varepsilon) - a^{(L)} \right\| &= O\left(\frac{L}{n^{\frac{1}{2}}} \rho_{2\kappa L}^{\frac{1}{2}} \pi_L \left(1 + L^{\frac{1}{2}} \rho_{4\kappa L}^{\frac{1}{2}} \pi_L \right) \right). \end{aligned}$$

Proof: Write

$$\hat{a}^{(L)}(\varepsilon) - a^{(L)} = \left\{ W^{(L)}(\varepsilon)^{-1} - W^{(L)-1} \right\} w^{(L)}(\varepsilon) + W^{(L)-1} \left\{ w^{(L)}(\varepsilon) - w^{(L)} \right\}.$$

From (6.23), the Schwarz inequality and Lemma 9

$$\left\| w^{(L)} \right\|^2 = \sum_{\ell=1}^{L} \ell^2 \left\{ E \left\{ \phi'(\varepsilon_0) \phi^{\ell-1}(\varepsilon_0) \right\} \right\}^2 \leq C L^2 \sum_{\ell=1}^{L} \mu_{2\kappa(\ell+K)} \leq L^2 \rho_{2\kappa L}.$$

Similarly, and from independence of the ε_t,

$$E \left\| w^{(L)}(\varepsilon) - w^{(L)} \right\|^2 \leq n^{-1} \sum_{\ell=1}^{L} \ell^2 E \left\{ \phi'(\varepsilon_0) \phi^{\ell-1}(\varepsilon_0) \right\}^2 \leq (L^2/n) \rho_{2\kappa L},$$

$$E \left\| W^{(L)}(\varepsilon) - W^{(L)} \right\|^2 \leq n^{-1} \sum\sum_{k,\ell=1}^{L} E \left\{ \phi(\varepsilon_0)^{2(k+\ell)} \right\} \leq (L/n) \rho_{4\kappa L}.$$

Now apply Lemma 8. $\square$

Lemma 11 *For $j \geq 0$ let $\alpha_j = \Delta_j(d)$ for $d \leq 1$ and $\left|\beta_j\right| \leq C(j+1)^{-3}$. Then the sequence $\sum_{k=0}^{j} \alpha_{j-k} \beta_k$, $j \geq 0$, has property $P_0(d)$.*

Proof: By Stirling's approximation α_j has property $P_0(d)$, whence the proof is completed by splitting sums around $j/2$ and elementary bounding of each. $\square$

Lemma 12 *For $j \geq 0$ let the sequence α_j, $j \geq 0$, have property $P_0(-d)$ and for $d > 0$ let $\sum_{j=0}^{\infty} \alpha_j = 0$. Then for $|d| < 1$ the sequence*

$$\gamma_j = \sum_{k=0}^{j} (j+1-k)^{-1} \alpha_k, \qquad j \geq 0,$$

has property $P_1(-d)$.

Proof: We give the proof only of $\left|\gamma_j - \gamma_{j+1}\right| \leq C \{\log(j+1)\} j^{-d-2}$, the proof of $\left|\gamma_j\right| \leq C \{\log(j+1)\} j^{-d-1}$ being similar and simpler. We have

$$\begin{aligned} \gamma_j - \gamma_{j+1} &= \sum_{k=0}^{\tilde{j}} \left\{ (j+1-k)^{-1} - (j+2-k)^{-1} \right\} \alpha_k - (j+1-\tilde{j})^{-1} \alpha_{\tilde{j}+1} \\ &\quad + \sum_{k=\tilde{j}+1}^{j} (j+1-k)^{-1} (\alpha_k - \alpha_{k+1}), \end{aligned}$$

where $\tilde{j} = [j/2]$. The second term is bounded by Cj^{-d-2} and the third by $C(\log j) j^{-d-2}$. For $d < 0$ the first term is bounded by Cj^{-d-2} and for $d = 0$ by $C(\log j) j^{-d-2}$. For $d > 0$ we apply summation-by-parts to this first term and $\sum_{j=0}^{\infty} \alpha_j = 0$ to obtain the bound Cj^{-d-2} again. □

Lemma 13 *Let the sequence α_j, $j \geq 0$, have property $P_0(-d)$ and the sequence β_j, $j \geq 0$, have property $P_0(e)$, and let*

$$\sum_{j=0}^{\infty} |\alpha_j| < \infty, \qquad \text{if } d = 0, \tag{7.20}$$

$$\sum_{j=0}^{\infty} \left|\beta_j\right| < \infty, \qquad \text{if } e = 0,$$

$$\sum_{j=0}^{\infty} \beta_j = 0, \qquad \text{if } e < 0.$$

Then for $|d| < 1$, $|e| < 1$ it follows that for all $j > 0$, $t > 0$

$$\left| \sum_{k=0}^{j} \alpha_{k+t} \beta_{j-k} \right| \leq C j^e t^{-d-1}, \qquad j \leq t, \tag{7.21}$$

$$\leq C j^{e-1} \max(j^{-d}, t^{-d}), \qquad j > t. \tag{7.22}$$

If instead α_j has property $P_1(-d)$ and (7.20) is not imposed,

$$\left|\sum_{k=0}^{j}\alpha_{k+t}\beta_{j-k}\right| \leq C(\log^{r+1} t)j^e t^{-d-1}, \quad j \leq t, \tag{7.23}$$

$$\leq C(\log^{r+1} j)j^{e-1}\max(j^{-d}, t^{-d}), \quad j > t. \tag{7.24}$$

Proof: We prove only (7.21) and (7.22), the proof of (7.23) and (7.24) being very similar but notationally slightly more complex and less elegant. Write $S_{ab} = \sum_{k=a}^{b}\alpha_{t+k}\beta_{j-k}$. We have

$$|S_{0j}| \leq t^{-d-1}\sum_{k=0}^{j}|\beta_k| \leq Cj^e t^{-d-1}, \quad e \geq 0.$$

This proves (7.21) for $e \geq 0$ and all d. On the other hand, with $\tilde{j} = [j/2]$, summation-by-parts gives

$$\begin{aligned}\left|S_{0\tilde{j}}\right| &\leq \sum_{k=0}^{\tilde{j}-1}\left|\beta_{j-k}-\beta_{j-k-1}\right|\sum_{i=0}^{k}|\alpha_{t+i}| + \left|\beta_{j-\tilde{j}}\right|\sum_{k=0}^{\tilde{j}}|\alpha_{t+k}| \\ &\leq Ct^{-d}\left\{\sum_{k=0}^{\tilde{j}}(j-k)^{e-2}+j^{e-1}\right\} \\ &\leq Cj^{e-1}t^{-d}, \quad d \geq 0, \quad \text{all } e,\end{aligned} \tag{7.25}$$

while

$$\left|S_{\tilde{j}+1,j}\right| \leq C(t+\tilde{j})^{-d-1}j^e \leq Cj^{e-d-1}, \quad \text{all } d, \quad e \geq 0. \tag{7.26}$$

This proves (7.22) for $d \geq 0$, $e \geq 0$ since $j^{e-d-1} \leq j^{e-1}t^{-d}$, $j > t$. For $e < 0$

$$S_{0j} = -\sum_{k=0}^{j-1}\{\alpha_{j-k+t}-\alpha_{j-k-1+t}\}\sum_{i=k+1}^{\infty}\beta_i - \alpha_t\sum_{k=j+1}^{\infty}\beta_i$$

since $\sum_{j=0}^{\infty}\beta_j = 0$. This is bounded by $C\left\{t^{-d-2}j^{e+1}+t^{-d-1}j^e\right\} \leq Cj^e t^{-d-1}$ for $j \leq t$, to prove (7.21) for $e < 0$ and all d. For $e < 0$ and all d

$$\begin{aligned}S_{\tilde{j}+1,j} &= \sum_{k=0}^{j-\tilde{j}-1}\alpha_{j+t-k}\beta_k \\ &= -\sum_{k=0}^{j-\tilde{j}-2}(\alpha_{j+t-k}-\alpha_{j+t-k-1})\sum_{i=k+1}^{\infty}\beta_i - \alpha_{t+\tilde{j}-1}\sum_{k=j-\tilde{j}}^{\infty}\beta_k\end{aligned}$$

and this is bounded by $C\left\{(t+\tilde{j})^{-d-2}j^{e+1}+(t+\tilde{j})^{-d-1}j^{e}\right\}\leq Cj^{e-1}t^{-d}$, which with (7.25) proves (7.22) for $d\geq 0$, $e<0$. Finally, for $d<0$ and all e

$$\left|S_{0\tilde{j}}\right|=\left|\sum_{k=j-\tilde{j}}^{j}\alpha_{j+t-k}\beta_{k}\right|\leq Cj^{e-d-1}$$

which with (7.26) completes the proof of (7.22). □

Lemma 14 *For* $|\zeta_0|<\frac{1}{2}$,

$$\sum_{j=0}^{\infty}\lambda_{jt}^{2} \leq Ct^{-1} \tag{7.27}$$

$$\sum_{j=0}^{\infty}\left(\sum_{t=1}^{n}\lambda_{jt}\right)^{2} \leq Cn. \tag{7.28}$$

Proof: In this and subsequent proofs we drop the zero subscript from ζ_0.We omit the proof for $\zeta=0$ as is it simple. From Lemma 13

$$\sum_{j=1}^{\infty}\lambda_{jt}^{2}\leq Ct^{-2\zeta-2}\sum_{j=1}^{t}j^{2\zeta}+C\sum_{j=t}^{\infty}j^{2\zeta-2}\max\left(j^{-2\zeta},t^{-2\zeta}\right).$$

The first sum is bounded by $Ct^{2\zeta+1}$ and the second by $Ct^{-2\zeta}\sum_{j=t}^{\infty}j^{2\zeta-2}\leq Ct^{-1}$ when $\zeta>0$ and by $C\sum_{j=t}^{\infty}j^{-2}\leq Ct^{-1}$ when $\zeta<0$, to prove (7.27). For $j<n$ and $\zeta\neq 0$

$$\begin{aligned}\left|\sum_{t}\lambda_{jt}\right| &\leq Cj^{\zeta-1}\sum_{t=1}^{j}\max\left(j^{-\zeta},t^{-\zeta}\right)+Cj^{\zeta}\sum_{t=j+1}^{n}t^{-\zeta-1}\\ &\leq C\max(1,(j/n)^{\zeta}).\end{aligned}$$

For $j\geq n$

$$\left|\sum_{t}\lambda_{jt}\right|\leq Cj^{\zeta-1}\sum_{t=1}^{n}\max(j^{-\zeta},t^{-\zeta})\leq C\max(n/j,(n/j)^{1-\zeta}).$$

Thus

$$\begin{aligned}
\sum_{j=0}^{\infty}\left(\sum_t \lambda_{jt}\right)^2 &= \sum_{j=0}^{n}\left(\sum_t \lambda_{jt}\right)^2 + \sum_{j=n+1}^{\infty}\left(\sum_t \lambda_{jt}\right)^2 \\
&\le Cn + Cn^{2-2\zeta}\sum_{j=n}^{\infty} j^{2\zeta-2} \le Cn, \qquad \zeta > 0, \\
&\le Cn^{-2\zeta}\sum_{j=1}^{n} j^{2\zeta} + n^2 \sum_{j=n}^{\infty} j^{-2} \le Cn, \qquad \zeta < 0,
\end{aligned}$$

to prove (7.28). □

Define

$$h_{jk} = \sum_t (t+j)^{-1}\left|\lambda_{kt}\right| \qquad j,k \ge 1.$$

Lemma 15 *For* $0 < \zeta_0 < \frac{1}{2}$ *and* $j \ge 1$

$$\begin{aligned}
h_{jk} &\le Cj^{-\frac{1}{2}}\min\left(j^{-\frac{1}{2}}, k^{-\frac{1}{2}}\right), \quad 1 \le k \le n && (7.29)\\
&\le Cj^{-1}k^{\zeta_0-1}n^{\frac{1}{2}-\zeta_0}\min(j^{\frac{1}{2}}, n^{\frac{1}{2}}), \quad k \ge n. && (7.30)
\end{aligned}$$

For $-\frac{1}{2} < \zeta_0 \le 0$ *and* $j \ge 1$

$$\begin{aligned}
h_{jk} &\le C\min\left(j^{-\frac{1}{2}-\varepsilon}k^{-\frac{1}{2}+\varepsilon}, k^{-1}\log k\right), \quad O < \varepsilon < \frac{1}{2}+\zeta_0, \quad 1 \le k < n, && (7.31)\\
&\le Ck^{-1}\min(n/j, \log n), \quad k \ge n. && (7.32)
\end{aligned}$$

Proof: It follows from Lemma 13 that for $1 \le k \le n$

$$h_{jk} \le Ck^{\zeta-1}\sum_{t=1}^{k}(t+j)^{-1}\max(k^{-\zeta}, t^{-\zeta}) + Ck^{\zeta}\sum_{t=k}^{n}(t+j)^{-1}t^{-\zeta-1}. \qquad (7.33)$$

Suppose $\zeta > 0$. The first term on the right is bounded by

$$\begin{aligned}
Cj^{-1}k^{\zeta-1}\sum_{t=1}^{k} t^{-\zeta} &\le Cj^{-1}, \quad j \ge n, \\
Cj^{-\frac{1}{2}}k^{\zeta-1}\sum_{t=1}^{k} t^{-\zeta-\frac{1}{2}} &\le C(jk)^{-\frac{1}{2}}, \quad j \le k.
\end{aligned}$$

The second term on the right of (7.33) is bounded by

$$
\begin{aligned}
Cj^{-1}k^{\zeta}\sum_{t=k}^{n}t^{-\zeta-1} &\leq Cj^{-1}, \quad j \geq k, \\
Cj^{-\frac{1}{2}}k^{\zeta}\sum_{t=k}^{n}t^{-\zeta-3/2} &\leq C(jk)^{-\frac{1}{2}}, \quad j \leq k.
\end{aligned}
$$

This proves (7.29). Let $\zeta \leq 0$. The first term on the right of (7.33) is bounded by

$$
Ck^{-1}\sum_{t=1}^{k}(t+j)^{-1} \leq C\min(j^{-1}, k^{-1}\log k)
$$

and the second by

$$
\begin{aligned}
Ck^{\zeta}j^{-\frac{1}{2}-\varepsilon}\sum_{t=k}^{\infty}t^{-\zeta-3/2+\varepsilon} &\leq Cj^{-\frac{1}{2}-\varepsilon}k^{-\frac{1}{2}+\varepsilon}, \quad j \geq k, \\
Ck^{\zeta}\sum_{t=k}^{n}t^{-\zeta-2} &\leq Ck^{-1}, \quad j \leq k.
\end{aligned}
$$

This proves (7.31). For $k \geq n$ (7.30) and (7.32) are readily deduced from

$$
h_{jk} \leq Ck^{\zeta-1}\sum_{t}(t+j)^{-1}t^{-\zeta}1(\zeta > 0) + Ck^{-1}\sum_{t}(t+j)^{-1}1(\zeta \leq 0).
$$

□

Lemma 16 *For* $|\zeta_0| < \frac{1}{2}$,

$$
E\left\|\sum_{t}\varepsilon_{t1}'\sum_{j=0}^{\infty}\lambda_{jt}\varepsilon_{-j}\right\|^2 \leq C(\log n)^3.
$$

Proof: Writing $\gamma(s;\nu_0) = \sum_{j=0}^{\infty}\gamma_j s^j$, the expression within the norm is

$$
\sum_{t}\sum_{j=1-t}^{-1}\gamma_{t+j}\varepsilon_{-j}\sum_{k=0}^{\infty}\lambda_{kt}\varepsilon_{-k} + \sum_{j,k=0}^{\infty}\sum H_{jk}\varepsilon_{-j}\varepsilon_{-k}. \tag{7.34}
$$

where $H_{jk} = \sum_t \gamma_{j+t}\lambda_{kt}$. The squared norm of the first term has expectation bounded by

$$
\sum_{s}\sum_{t}\left(\sum_{j=\max(1-s,1-t)}^{-1}\left\|\gamma_{s+j}\right\|\left\|\gamma_{t+j}\right\|\right)\left(\sum_{k=0}^{\infty}\lambda_{sk}\lambda_{tk}\right).
$$

For $s \leq t$ the first bracketed factor is $O\left((t-s+1)^{-1}\log n\right)$ because $\left\|\gamma_j\right\| \leq C(j+1)^{-1}$, while the second one is bounded by

$$\begin{aligned}
& Ct^{-\zeta-1}s^{-\zeta-1}\sum_{j=1}^{s} j^{2\zeta} + Ct^{-\zeta-1}\sum_{j=s+1}^{t} j^{2\zeta-1}\max(j^{-\zeta}, s^{-\zeta}) \\
& +C\sum_{j=t+1}^{\infty} j^{2\zeta-2}\max(j^{-\zeta}, s^{-\zeta})\max(j^{-\zeta}, t^{-\zeta}) \\
\leq\ & C\left\{s^{-\zeta}t^{\zeta-1}1(\zeta>0) + s^{\zeta}t^{-\zeta-1}1(\zeta<0) + (st)^{-\frac{1}{2}}1(\zeta=0)\right\} \leq C(st)^{-\frac{1}{2}}.
\end{aligned}$$

We have

$$\begin{aligned}
\sum_{s=1}^{t}(t-s+1)^{-1}s^{-\frac{1}{2}} &\leq \sum_{s=1}^{[t/2]}(t-s+1)^{-1}s^{-\frac{1}{2}} + \sum_{s=[t/2]}^{t}(t-s+1)^{-1}s^{-\frac{1}{2}} \\
&\leq C(\log t)t^{-\frac{1}{2}}, \\
C(\log n)\sum_{t}(\log t)t^{-1} &\leq C(\log n)^3.
\end{aligned}$$

Next, since $|H_{jk}| \leq Ch_{jk}$, the squared norm of the second term on the right of (7.34) has expectation bounded by

$$C\sum\sum_{j,k=0}^{\infty}\left(h_{jk}^2 + h_{jj}h_{kk} + h_{jk}h_{kj}\right).$$

We apply Lemma 15 to complete the proof. For $\zeta > 0$

$$\begin{aligned}
\sum\sum_{j,k=0}^{\infty} h_{jk}^2 &\leq C\sum_{k=1}^{n}\sum_{j=1}^{k}(jk)^{-1} + C\sum_{k=1}^{n}\sum_{j=k}^{\infty} j^{-2} \\
&\quad +Cn^{1-2\zeta}\sum_{k=n}^{\infty}\sum_{j=1}^{n} j^{-1}k^{2\zeta-2} + Cn^{2-2\zeta}\sum_{k=n}^{\infty}\sum_{j=n}^{\infty} j^{-2}k^{2\zeta-2} \\
&\leq C(\log n)^2 \\
\sum_{j=0}^{\infty} h_{jj} &\leq C\sum_{j=1}^{n} j^{-1} + n^{1-\zeta}\sum_{j=n}^{\infty} j^{\zeta-2} \leq C\log n \\
\sum\sum_{j,k=0}^{\infty} h_{jk}h_{kj} &\leq C\sum_{k=1}^{n}\sum_{j=1}^{k}(jk)^{-1} + Cn^{\frac{1}{2}-\zeta}\sum_{k=1}^{n}\sum_{j=k}^{\infty} j^{\zeta-2}k^{-\frac{1}{2}} \\
&\quad +Cn^{2-2\zeta}\sum\sum_{j,k=n}^{\infty}(jk)^{\zeta-2} \\
&\leq C(\log n)^2.
\end{aligned}$$

For $\zeta \leq 0$

$$
\begin{aligned}
\sum\sum_{j,k=0}^{\infty} h_{jk}^2 &\leq \sum_{k=1}^{n}\sum_{j=1}^{k}(k^{-1}\log k)^2 + C\sum_{k=1}^{n}\sum_{j=k}^{\infty} j^{-1-2\varepsilon}k^{-1+2\varepsilon} \\
&\quad + C(\log n)^2 \sum_{k=n}^{\infty}\sum_{j=1}^{n} k^{-2} + Cn^2 \sum\sum_{j,k=n}^{\infty}(jk)^{-2} \\
&\leq C(\log n)^3, \\
\sum_{j=0}^{\infty} h_{jj} &\leq C\sum_{j=1}^{n} j^{-1} + Cn\sum_{j=n}^{\infty} j^{-2} \leq C\log n, \\
\sum\sum_{j,k=0}^{\infty} h_{jk}h_{kj} &\leq C\sum_{k=1}^{n}\sum_{j=1}^{k} j^{-\frac{1}{2}+\varepsilon}k^{-3/2-\varepsilon}\log k \\
&\quad + C\log n \sum_{k=1}^{n}\sum_{j=n}^{\infty} j^{-\frac{1}{2}-\varepsilon}k^{-\frac{1}{2}+\varepsilon}j^{-1} + Cn^2\sum\sum_{j,k=n}^{\infty}(jk)^{-2} \\
&\leq C(\log n)^2.
\end{aligned}
$$

□

Lemma 17

$$
E\left\|\sum_t \varepsilon'_{t1}\right\|^4 \leq C(\log n)^4 n^2.
$$

Proof: We have,

$$
\sum_t \varepsilon'_{t1} = \sum_{j=1}^{n-1}\left(\sum_{i=1}^{n-j}\gamma_i\right)\varepsilon_j + \sum_{j=0}^{\infty}\left(\sum_{i=j+1}^{j+n}\gamma_i\right)\varepsilon_{-j}.
$$

Thus

$$
E\left\|\sum_t \varepsilon'_{t1}\right\|^4 \leq C\left(\sum_{j=1}^{n-1}\left\|\sum_{i=1}^{n-j}\gamma_i\right\|^2\right)^2 + C\left(\sum_{j=0}^{\infty}\left\|\sum_{i=j+1}^{j+n}\gamma_i\right\|^2\right)^2.
$$

Since

$$
\begin{aligned}
\left\|\sum_{i=1}^{n-j}\gamma_i\right\| &\leq \sum_{i=1}^{n-j}\|\gamma_i\| \leq C\sum_{i=1}^{n} i^{-1} \leq C\log n, \quad 1 \leq j < n, \\
\left\|\sum_{i=j+1}^{j+n}\gamma_i\right\| &\leq C\sum_{i=j+1}^{j+n} i^{-1} \leq C\log n, \quad 1 \leq j \leq n, \\
&\leq Cn/j, \quad j \geq n,
\end{aligned}
$$

the proof is readily completed. □

Lemma 18 *For any sequence c_j, $j \geq 0$, and any $r \geq 1$, if $\mu_{r+} < \infty$,*

$$E\left|\sum_{j=0}^{\infty} c_j \varepsilon_{-j}\right|^r \leq (Cr)^{2r}\left(\sum_{j=0}^{\infty} c_j^2\right)^{r/2} \mu_{r_+}^{r/r_+},$$

where r_+ is the smallest even integer such that $r_+ \geq r$.

Proof: For $r \leq 2$ the proof follows by Jensen's inequality and direct calculation. For $r > 2$ the Marcinkiewicz-Zygmund inequality indicates that

$$E\left|\sum_0^{\infty} c_j \varepsilon_{-j}\right|^r \leq C_r E\left(\sum_0^{\infty} c_j^2 \varepsilon_{-j}^2\right)^{r/2} \tag{7.35}$$

where $C_r = \{18r^{3/2}(r-1)^{-\frac{1}{2}}\}^r$ (see Hall and Heyde, 1980, p.23). By the c_r-inequality (7.35) is bounded by

$$\begin{aligned} & C_r 2^{r/2-1}\left\{E\left|\sum_0^{\infty} c_j^2(\varepsilon_{-j}^2 - 1)\right|^{r/2} + \left(\sum_0^{\infty} c_j^2\right)^{r/2}\right\} \\ \leq\ & C_r 2^{r/2-1}\left\{C_{r/2} E\left|\sum_0^{\infty} c_j^4(\varepsilon_{-j}^2 - 1)^2\right|^{r/4} + \left(\sum_0^{\infty} c_j^2\right)^{r/2}\right\}. \end{aligned}$$

For $2 < r \leq 4$ the first expectation in the last line is bounded by

$$\left\{E\sum_0^{\infty} c_j^4(\varepsilon_{-j}^2 - 1)^2\right\}^{r/4} \leq \left(\sum_0^{\infty} c_j^4 E\varepsilon_0^4\right)^{r/4} \leq \left(\sum_0^{\infty} c_j^2\right)^{r/2} \mu_4^{r/4}.$$

For $r > 4$ we instead apply the c_r-inequality to that expectation, and then the Marcinkiewicz-Zygmund inequality again, and so on, eventually bounding (7.35) by

$$C_r C_{r/2} C_{r/4} ... C_2 . 2^{r/2} . 2^{r/4} . 2^{r/8} ... 1 \left(\sum_0^{\infty} c_j^2\right)^{r/2} \mu_{r_+}^{r/r_+}.$$

The result follows on noting that $r.r^{\frac{1}{2}}.r^{\frac{1}{4}}...r^{1/r} < r^2$, $\quad 2^{\frac{1}{2}}.2^{\frac{1}{4}}...1 < 2$, $2^{\frac{1}{2}}.4^{\frac{1}{4}}...r^{1/r} > 1$ and $j/(j-1) \leq 2$ for all $j \geq 2$. □

Lemma 19 *As $n \to \infty$*

$$\left\|\widehat{a}^{(L)}\left(E/\sigma_0\right) - \widehat{a}^{(L)}(\varepsilon)\right\| = O_p\left(\rho_{2\kappa L}^{3/2}\pi_L^2\left(L^2 n^{-\frac{1}{2}} + (CL)^{4\kappa L+3}n^{-1}\log n\right)\right).$$

Proof: Because the proof is similar to details in Section 3 we sketch it. It turns out that $\left\{W^{(L)}(E/\sigma_0)^{-1} - W^{(L)}(\varepsilon)^{-1}\right\} w^{(L)}(E/\sigma_0)$ dominates $W^{(L)}(\varepsilon)^{-1}\left\{w^{(L)}(E/\sigma_0) - w^{(L)}(\varepsilon)\right\}$, so we look only at the former. $\left\|W^{(L)}(E/\sigma_0) - W^{(L)}(\varepsilon)\right\|$ is bounded by

$$Cn^{-1}\left[\sum_{k,\ell=1}^{L}\sum\left\{\left(\sum_t \delta_{kt}\delta_{\ell t}\right)^2 + \left(\sum_t \phi_k(\varepsilon_t)\delta_{\ell t}\right)^2\right\}\right]^{\frac{1}{2}} \tag{7.36}$$

(incorporating a term due to the mean-correction, which is of smaller order). Using (6.14),

$$\sum_t \phi_k(\varepsilon_t)\delta_{\ell t} = \sum_t \phi_k(\varepsilon_t)\phi_\ell'(\varepsilon_t)d_t + \frac{1}{2}\sum_t \phi_k(\varepsilon_t)\phi_\ell''(\bar{\varepsilon}_t)d_t^2. \tag{7.37}$$

We have

$$\begin{aligned} E\left\|\sum_t\left\{\phi_k(\varepsilon_t)\phi_\ell'(\varepsilon_t) - E\phi_k(\varepsilon_0)\phi_\ell'(\varepsilon_0)\right\}d_{1t}\right\|^2 &\leq CE\left\{\phi_k(\varepsilon_0)\phi_\ell'(\varepsilon_0)\right\}^2\sum_t Ed_{1t}^2 \\ &\leq C\ell^2\mu_{2\kappa(k+\ell+K)}\log n. \end{aligned}$$

Replacing d_{1t} by $d_t - d_{1t}$ gives no greater bound, by virtue of (6.15) and (6.17). On the other hand,

$$\left\{E\phi_k(\varepsilon_0)\phi_\ell'(\varepsilon_0)\right\}\sum_t d_t = O_p\left(\ell\mu_{2\kappa k}^{\frac{1}{2}}\mu_{2\kappa(\ell+K)}^{\frac{1}{2}}n^{\frac{1}{2}}\right)$$

because $\sum_t E_t = 0$ implies $\sum_t d_t = \sum_t \varepsilon_t$. Next

$$\left|\sum_t \phi_k(\varepsilon_t)\phi_\ell''(\bar{\varepsilon}_t)d_t^2\right| \leq C^{\kappa\ell+1}\ell^2\sum_t\left|\phi_k(\varepsilon_t)\right|\left(1 + |\varepsilon_t|^{\kappa(\ell+K)} + |d_t|^{\kappa(\ell+K)}\right)d_t^2.$$

Proceeding as in Section 6, this is $O_p\left((C\ell)^{2\kappa\ell+2}\mu_{\kappa k}\mu_{r_\ell}\log n\right)$, where r_ℓ is the smallest even integer exceeding $\kappa(\ell+K)+2$. It follows that

$$\sum_{k,\ell=1}^{L}\sum\left(\sum_t \phi_k(\varepsilon_t)\delta_{\ell t}\right)^2 = O_p\left(\rho_{2\kappa L}^2\left(L^2 n + (CL)^{4\kappa L+4}(\log n)^2\right)\right).$$

Also

$$\left\{\sum_{k,\ell=1}^{L}\sum\left(\sum_{t}\delta_{k\ell}\delta_{\ell t}\right)^{2}\right\}^{\frac{1}{2}} \leq \sum_{\ell=1}^{L}\sum_{t}\delta_{\ell t}^{2} \leq \sum_{\ell=1}^{L}\sum_{t}\phi_{\ell}'(\bar{\varepsilon}_{t})^{2}d_{t}^{2},$$

and by proceeding as before this is $O_p\left((CL)^{4\kappa L+2}\rho_{2\kappa L}\log n\right)$. The proof is completed by application of Lemmas 8 and 10. $\square$

ACKNOWLEDGEMENTS

I thank Fabrizio Iacone for carrying out the computations reported in Section 4, and two referees for numerous comments which have led to an improved presentation.

Table 1: $\varepsilon_t \sim N(0,1)$

Monte Carlo MSE($\hat{\xi}$)/MSE($\tilde{\xi}$) with $n = 64$ and 1000 replications

		$\phi(s) = s$				$\phi(s) = s(1+s^2)^{-\frac{1}{2}}$			
	L	1	2	3	4	1	2	3	4
ξ_0	-0.25	.62	.62	.62	.62	.66	.67	.63	.65
	0.25	.47	.48	.51	.61	.49	.52	.53	.60
	0.75	.46	.49	.53	.62	.50	.54	.55	.60
	1.25	.47	.50	.52	.61	.52	.53	.52	.56

Table 2: $\varepsilon_t \sim 0.5N(-3,1) + 0.5N(3,1)$

Monte Carlo MSE($\hat{\xi}$)/MSE($\tilde{\xi}$) with $n = 64$ and 1000 replications

		$\phi(s) = s$				$\phi(s) = s(1+s^2)^{-\frac{1}{2}}$			
	L	1	2	3	4	1	2	3	4
ξ_0	-0.25	.92	.92	.83	.90	.94	.93	.82	.83
	0.25	.90	.91	.89	.93	.91	.91	.88	.89
	0.75	.90	.91	.89	.94	.90	.92	.89	.89
	1.25	.88	.89	.88	.92	.89	.89	.87	.87

Table 3: $\varepsilon_t \sim$ (scaled) $0.5N(0,26) + 0.95N(0,1)$

Monte Carlo MSE($\hat{\xi}$)/MSE($\tilde{\xi}$) with $n = 64$ and 1000 replications

		$\phi(s) = s$				$\phi(s) = s(1+s^2)^{-\frac{1}{2}}$			
	L	1	2	3	4	1	2	3	4
ξ_0	-0.25	.71	.71	.62	.77	.81	.76	.63	.70
	0.25	.84	.76	.65	.74	.77	.67	.60	.54
	0.75	.85	.79	.70	.79	.80	.78	.69	.63
	1.25	1.01	.96	.81	.82	.91	.83	.74	.68

Table 4: $\varepsilon_t \sim$ (scaled) Laplace

Monte Carlo MSE($\hat{\xi}$)/MSE($\tilde{\xi}$) with $n = 64$ and 1000 replications

		$\phi(s) = s$				$\phi(s) = s(1+s^2)^{-\frac{1}{2}}$			
	L	1	2	3	4	1	2	3	4
ξ_0	-0.25	1.07	.85	.92	.96	1.04	.90	.60	.61
	0.25	.89	.60	.58	.87	.78	.62	.65	.67
	0.75	.56	.52	.55	.81	.51	.53	.53	.54
	1.25	.28	.23	.23	.86	.32	.26	.28	.38

Table 5: $\varepsilon_t \sim$ (scaled) t_5

Monte Carlo MSE($\hat{\xi}$)/MSE($\tilde{\xi}$) with $n = 64$ and 1000 replications

		$\phi(s) = s$				$\phi(s) = s(1+s^2)^{-\frac{1}{2}}$			
	L	1	2	3	4	1	2	3	4
ξ_0	-0.25	.58	.54	.53	.65	.55	.53	.55	.60
	0.25	.56	.56	.57	.74	.51	.54	.55	.58
	0.75	.58	.58	.62	.75	.51	.56	.57	.61
	1.25	.63	.61	.60	.69	.54	.55	.52	.53

REFERENCES

[1] BERAN, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *R. Statist. Soc. Ser. B* **57** 659-672.

[2] BERAN, R. (1976). Adaptive estimates for autoregressive processes. *Ann. Inst. Statist. Math.* **26** 77-89.

[3] BICKEL, P. (1982). On adaptive estimation. *Ann. Statist.* **10** 647-671.

[4] BLOOMFIELD, P. (1972). An exponential model for the spectrum of a scalar time series. *Biometrika* **60** 217-226.

[5] BOX, G.E.P. AND JENKINS, G.M. (1971). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

[6] DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** 1749-1766.

[7] DAHLHAUS, R. (1995). Efficient location and regression estimation for long range dependent regression models. *Ann. Statist.* **23** 1029-1048.

[8] DROST, F.L., KLASSEN, C.A.J. AND WERKER, B.J.M. (1997). Adaptive estimation in time series models. *Ann. Statist.* **25** 786-818.

[9] FOX, R. AND TAQQU, M.S. (1986). Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517-532.

[10] FREUD, G. (1971). *Orthogonal Polynomials.* Pergamon Press, Oxford.

[11] GIRAITIS, L. AND SURGAILIS, D. (1990). A central limit theorem for quadratic forms in strongly dependent random variables and its application to asymptotic normality of Whittle's estimate. *Probab. Theory Related Fields* **86** 87-104.

[12] HAJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 128-194. University of California Press, Berkeley.

[13] HALL, P. AND HEYDE, C.C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

[14] HALLIN, M. AND SERROUKH, A. (1999). Adaptive estimation of the lag of a long-memory process. *Statist. Inference Stochastic Process.* **1** 1-19.

[15] HALLIN, M., TANIGUCHI, M., SERROUKH, A. AND CHOY, K. (1999). Local asymptotic normality for regression models with long-memory disturbance. *Ann. Statist.* **27** 2054-2080.

[16] HANNAN, E.J. (1973). The asymptotic theory of linear time series models. *J. Appl. Probab.* **10** 130-145.

[17] KNUTH, D.E. (1968). *The Art of Computer Programming.* Vol.1 *Fundamental Algorithms*, Addison-Wesley, Boston, MA.

[18] KOUL, H.L. AND SCHICK, A. (1997). Efficient estimation in nonlinear autoregressive models. *Bernoulli* **3** 247-277.

[19] KREISS, J.-P. (1987). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112-133.

[20] LE CAM, L. (1960). Local asymptotically normal families of distributions. *Univ. California Publ. Statist.* **3** 27-98.

[21] LI, W.K. AND McLEOD, A.I. (1986). Fractional time series modelling. *Biometrika.* **73** 217-221.

[22] LING, S. (2003). Adaptive estimators and tests of stationary and nonstationary short and long memory ARFIMA-GARCH models. *J. Amer. Statist. Assoc.* **97** 955-967.

[23] LOEVE, M. (1977). *Probability Theory 1.* Springer-Verlag, New York.

[24] MARTIN, R.D. (1982). The Cramer-Rao bound and robust M-estimates for autoregressions. *Biometrika* **69** 437-442.

[25] NEWEY, W.K. (1988). Adaptive estimation of regression models via moment restrictions. *J. Econometrics* **38** 301-339.

[26] ROBINSON, P.M. (1994). Efficient tests of nonstationary hypotheses. *J. Amer. Statist. Assoc.* **89** 1420-1437.

[27] SCOTT, D.J. (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Adv. Appl. Probab.* **5** 119-137.

[28] STONE, C.J. (1975). Adaptive maximum-likelihood estimation of a location parameter. *Ann. Statist.* **3** 267-284.

[29] VELASCO, C. AND ROBINSON, P.M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *J. Amer. Statist. Assoc.* **95** 1229-1243.

[30] WHITTAKER, E.T. AND WATSON, G.N. (1940). *A Course of Modern Analysis.* Cambridge University Press, Cambridge.

[31] YAJIMA, Y. (1988). On estimation of a regression model with long-memory stationary errors. *Ann. Statist.* **16** 791-807.

[32] YAJIMA, Y. (1991). Asymptotic properties of the LSE in a regression model with long memory stationary errors. *Ann. Statist.* **19** 158-177.