ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Inference on Nonstationary Time Series with Moving Mean

Jiti Gao and Peter M. Robinson

July 2013

Working Paper 15/13

Inference on Nonstationary Time Series with Moving Mean

Jiti Gao and Peter M Robinson*

Monash University and London School of Economics

July 23, 2013

Abstract

A semiparametric model is proposed in which a parametric filtering of a nonstationary time series, incorporating fractionally differencing with short memory correction, removes correlation but leaves a nonparametric deterministic trend. Estimates of the memory parameter and other dependence parameters are proposed, and shown to be consistent and asymptotically normally distributed with parametric rate. Unit root tests with standard asymptotics are thereby justified. Estimation of the trend function is also considered. We include a Monte Carlo study of finite-sample performance.

Keywords: fractional time series; fixed design nonparametric regression; nonstationary time series; unit root tests.

JEL Classifications: C14, C22.

Proposed running head: Nonstationary Time Series

....

*Corresponding author. E-mail: p.m.robinson@lse.ac.uk

1 INTRODUCTION

A long-established vehicle for smoothing a deterministically-trending time series y_t, $t = 1, ..., T$, is the fixed-design nonparametric regression model given by

$$y_t = g\left(\frac{t}{T}\right) + u_t,\ t = 1, ..., T, \tag{1}$$

where $g(x)$, $x \in [0, 1]$, is an unknown, smooth, nonparametric function, and u_t is an unobservable sequence of random variables with zero mean. The dependence on sample size T of $g(t/T)$ in (1) is to ensure sufficient accumulation of information to enable consistent estimation of $g(\tau)$ at any $\tau \in (0, 1)$. A more basic trend function is a polynomial in t of given degree, as still frequently employed in various econometric models. A more general class of models than polynomials (and having analogy with the fractional stochastic trends we will employ in the current paper) involves fractional powers, i.e.

$$y_t = \beta_0 + \beta_1 t^{\gamma_1} + ... + \beta_p t^{\gamma_p} + u_t,\quad t = 1, ..., T, \tag{2}$$

where all the β_i and γ_i are unknown and real-valued. Subject to identifiability and other restrictions, these parameters can be estimated consistently and asymptotically normally, e.g. by nonlinear least squares (Robinson (2012a)). Models such as (2) can be especially useful in modest sample sizes. However, and as with any parametric function of t, misspecification leads to inconsistent estimation, and a nonparametric treatment affords greater flexibility when T is large (recognizing that nonparametric estimates converge more slowly than parametric ones).

With independent and identically distributed (iid) u_t, with finite variance, various kernel-type estimates of g in (1) were developed by Gasser and Mueller (1979), Priestley and Chao (1972), with statistical properties established; in particular, under regularity conditions kernel estimates of $g(\tau)$ are consistent and asymptotically normally distributed as $T \to \infty$ (see e.g. Benedetti (1977)). A suitable choice of kernel (and bandwidth) is an important ingredient in this theory, although kernel estimates are essentially an elaboration on simple moving window averages, which have a much longer history in empirical work. More recent empirical uses of (1) include Starica and Granger (2005) in modelling stock price series.

The iid assumption on u_t is very restrictive, but similar asymptotic properties result when u_t has weak dependence, for example is a covariance stationary process, generated by a linear process or satisfying suitable mixing conditions, and having finite and positive spectral density at degree zero (see e.g. Roussas, Tran and Ioannides (1992), Tran, Roussas, Yakowitz and Truong Van (1996)). The rate of convergence of kernel estimates is unaffected by this level of serial correlation, though the asymptotic variance differs from that in the iid case (unlike in the stochastic-design model in which the argument of g in (1) is instead a weakly dependent stationary stochastic sequence).

Long-range dependence in u_t has a greater impact on large-sample inference. If u_t is a stationary and invertible fractional process, for example

$$(1-L)^{\delta_0} u_t = \varepsilon_t, \quad |\delta_0| < 1/2, \tag{3}$$

L being the lag operator and the ε_t forming an iid sequence, or if u_t has a "semiparametric" specification with spectral density $f(\lambda)$ having rate $\lambda^{-2\delta_0}$ as frequency λ approaches zero from above, then the convergence rate of kernel estimates of $g(\tau)$ is slower when $\delta_0 > 0$ and faster when $\delta_0 < 0$. References dealing with (1) for such u_t include Beran and Feng (2002), Csorgo and Mielniczuk (1995), Deo (1997), Guo and Koul (2007), Robinson (1997), Zhao, Zhang and Li (2013). The asymptotic variance of the kernel estimates depends on δ_0 and any other time series parameters; for the "semiparametric" specification Robinson (1997) justified studentization using local Whittle estimates of δ_0.

The restriction $\delta_0 < 1/2$ implies stationarity of u_t, so that y_t given by (1) is nonstationary only in the mean. Stochastic trends are also an important source of nonstationarity in many empirical time series. However, a nonstationary stochastic trend in y_t generated by a nonstationary u_t, for example one having a unit root, would render $g\left(t/T\right)$ undetectable. An alternative, semiparametric, model which both incorporates a possibly nonstationary stochastic trend and enables estimation of a nonparametric deterministic trend is

$$\Delta_t^{\delta_0} y_t = g\left(\frac{t}{T}\right) + u_t, \quad t = 1, ..., T, \tag{4}$$

where u_t is a sequence of uncorrelated, homoscedastic random variables and, for any real δ, Δ_t^δ is the truncated fractional differencing operator,

$$\Delta_t^\delta = \sum_{j=0}^{t-1} \alpha_j(\delta) L^j, \ t \geq 1, \tag{5}$$

the $\alpha_j(\delta)$ being coefficients in the (possibly formal) expansion

$$(1-z)^\delta = \sum_{j=0}^{\infty} \alpha_j(\delta) z^j,$$

namely

$$\alpha_j(\delta) = \frac{\Gamma(j-\delta)}{\Gamma(-\delta)\Gamma(j+1)}. \tag{6}$$

The truncation in (5) reflects non-observability of y_t when $t \leq 0$, and avoids explosion of the moving average representation of (4) when $\delta_0 \geq 1/2$, the nonstationary region for δ_0; it is this region with which we will be concerned.

One such δ_0 has assumed wide empirical importance in connection with a variety of econometric models, the unit root case $\delta_0 = 1$, when (4) becomes

$$(1-L)y_t = g\left(\frac{t}{T}\right) + u_t, \ t = 1, ..., T. \tag{7}$$

The bulk of the econometric literature nests the unit root in autoregressive structures, which suggests treating (7) as a special case of

$$(1-\rho L)y_t = g\left(\frac{t}{T}\right) + u_t, \quad t = 1, ..., T, \tag{8}$$

rather than (4). The autoregressive unit root literature suggests that estimates of ρ in (8) will have a nonstandard limit distribution under (7), but a normal one in the "stationary" region $|\rho| < 1$. By contrast we can anticipate, for example from literature concerning (4) with $g(x)$ *a priori* constant, that estimates of δ_0 such as ones optimizing an approximate pseudo-Gaussian likelihood, and Wald and other test statistics, will enjoy standard asymptotics, with the usual parametric convergence rate, $\sqrt{T}$, whatever the value of δ_0, due essentially to smoothness properties of the fractional operator; tests are also expected to have the classical local efficiency properties. While (4) cannot, unlike (8), describe "explosive" behaviour (occurring when $|\rho| > 1$), it describes a continuum of stochastic trends indexed by δ_0. A consequence of the $T-$dependence in $g(t/T)$ is that the left side of (4) is also T-dependent, so the $y_t = y_{tT}$ in fact form a triangular array, but in common with the bulk of literature concerning versions of (1) we suppress reference to T. The model (4) (which nests (1) with iid u_t on taking $\delta_0 = 0$) supposes that the fractional filtering of y_t successfully eliminates correlation, but possibly leaves a trend which we are not prepared to parameterize.

To provide greater generality than (4), the paper in fact considers the extended model

$$\xi_t\left(L;\delta_0,\theta_0\right)y_t = g\left(\frac{t}{T}\right) + u_t, \quad t = 1,...,T, \tag{9}$$

where θ_0 is an unknown $p-$dimensional column vector and

$$\xi_t\left(z;\delta,\theta\right) = \sum_{j=0}^{t-1}\xi_j(\delta,\theta)z^j, \ t \geq 1,$$

where the $\xi_j(\delta,\theta)$ are coefficients in the possibly formal expansion

$$\xi\left(z;\delta,\theta\right) = \sum_{j=0}^{\infty}\xi_j(\delta,\theta)z^j,$$

such that

$$\xi\left(z;\delta,\theta\right) = (1-z)^{\delta}\zeta\left(z;\theta\right),$$

where

$$\zeta\left(z;\theta\right)=\sum_{j=0}^{\infty}\zeta_{j}(\theta)z^{j}$$

is a known function of z and θ that is at least continuous, and nonzero for z on or inside the unit circle in the complex plane. When $\zeta\left(z;\theta\right)\equiv 1$, we have $\xi\left(z;\delta,\theta\right)=(1-z)^{\delta}$. Leading examples of $\zeta\left(z;\theta\right)$ are stationary and invertible autoregressive moving average operators of known degree, for example the first order autoregressive operator $\zeta\left(z;\theta\right)=1-\theta z$, with θ here a scalar such that $|\theta|<1$. In general ζ leaves the essential memory or degree of nonstationarity δ_0 unchanged but allows otherwise richer dependence structure.

It would be possible to consider in effect a nonparametric ζ, $\zeta\left(z\right)$, satisfying smoothness assumptions only near $z=1$, and hence a "semiparametric" operator on y_t. This would lead to an estimate of δ_0 with only a nonparametric convergence rate. However, establishing the parametric, $\sqrt{T}$, rate for estimating δ_0 and θ_0 seems actually more challenging and delicate, because of the presence of the nonparametric $g\left(t/T\right)$ in (9), estimates of which converge more slowly than $\sqrt{T}$. In particular, proving consistency requires establishing that certain (stochastic and deterministic) contributions to residuals, whose squares make up the objective function minimized by the parameter estimates, are negligible uniformly over the parameter space; these contributions are of larger order than would be the case with a parametric trend (and this fact also explains why we find ourselves unable to choose the parameter space for δ_0 as large as is possible with a parametric trend). Then, corresponding contributions to scores evaluated at δ_0, θ_0 are also of larger order than in the parametric trend case and have to be shown to be negligible after being normalized by $\sqrt{T}$, rather than by a slower, nonparametric, rate, in order to prove asymptotic normality of the parameter estimates with $\sqrt{T}$ rate. Of course, the strong dependence in y_t also impacts on the conditions, due to non-summability of certain weight sequences.

The following section proposes estimates of δ_0 and θ_0, and establishes their consistency and asymptotic normality, the proofs appearing in Appendices A and B. Section 3 develops unit root tests based on Wald, pseudo log likelihood ratio and Lagrange multiplier principles. Section 4 proposes estimates of $g(x)$ and establishes their asymptotic properties. A small Monte Carlo study of finite-sample performance is contained in Section 5. Section 6 concludes by describing further issues that might be considered.

2. ESTIMATION OF DEPENDENCE PARAMETERS

Were $g(x) \equiv 0$ in (9) *a priori*, a natural method of estimating δ_0 and θ_0 would be conditional-sum-of-squares, which approximates Gaussian pseudo-maximum-likelihood estimation. We modify this method by employing residuals, which requires preliminary estimation of $g(x)$. Note that under the conditions imposed below, $g(t/T) - g((t-1)/T) = O(T^{-1}),\ 2 \leq t \leq T$, so we could instead consider proceeding by first differencing in (9) as this would effectively eliminate the deterministic trend; however this also induces a moving average unit root on the right hand side.

Let $k(x), x \in \mathcal{R}$, be a user-chosen kernel function and h a user-chosen positive bandwidth number. For any δ, θ write $\xi_{t\delta\theta}(z) = \xi_t(z; \delta, \theta)$ and introduce

$$\hat{g}_{\delta\theta}(x) = \sum_{s=1}^{T} \xi_{s\delta\theta}(L)\, y_s k\left(\frac{x - s/T}{h}\right) / \sum_{s=1}^{T} k\left(\frac{x - s/T}{h}\right), \tag{10}$$

for any $x \in [0,1]$. The corresponding estimate of Priestley and Chao (1972) type replaces the denominator by Th, but we prefer to use weights (of the $\xi_{s\delta\theta}(L)\, y_s$) that exactly sum to 1 for all x. Define residuals

$$\hat{u}_t(\delta,\theta) = \xi_{t\delta\theta}(L)\, y_t - \hat{g}_{\delta\theta}(t/T) = \frac{\sum_{s=1}^{T} \left(\xi_{t\delta\theta}(L)\, y_t - \xi_{s\delta\theta}(L)\, y_s\right) k_{ts}}{\sum_{s=1}^{T} k_{ts}}, \tag{11}$$

where $k_{ts} = k\left((t-s)/Th\right)$. Denote by ∇_1, ∇_2 chosen real numbers such that $\nabla_1 < \nabla_2$, write $\nabla = [\nabla_1, \nabla_2]$, and let Θ be a suitably chosen compact subset of $\mathcal{R}^p$. We estimate δ_0 and θ_0 by

$$\left(\widehat{\delta}, \widehat{\theta}\right) = \arg \min_{\delta \in \nabla; \theta \in \Theta} Q(\delta, \theta), \tag{12}$$

where

$$Q(\delta, \theta) = \sum_{t=1}^{T} \hat{u}_t^2(\delta, \theta). \tag{13}$$

We first establish consistency of $\widehat{\delta}$, $\widehat{\theta}$, under the following regularity conditons.

Assumption 1

The u_t are stationary and ergodic with finite fourth moment, and

$$E\left(u_t | \mathcal{F}_{t-1}\right) = 0, \quad E\left(u_t^2 \middle| \mathcal{F}_{t-1}\right) = \sigma^2$$

almost surely, where $\mathcal{F}_t$ is the σ-field of events generated by ε_s, $s \leq t$, and conditional on $\mathcal{F}_{t-1}$ 3rd and 4th moments of u_t equal corresponding unconditional moments.

Assumption 2

(i) $\theta_0 \in \Theta$;

(ii) $|\zeta(z;\theta)| \neq |\zeta(z;\theta_0)|$ *,for all $\theta \neq \theta_0$, $\theta \in \Theta$, on a $z-$set of positive Lebesgue measure;*

(iii) *for all $\theta \in \Theta$ and real λ, $\zeta\left(e^{i\lambda};\theta\right)$ is differentiable in λ with derivative in $Lip(\varsigma)$, $\varsigma > 1/2$;*

(iv) *for all λ, $\zeta\left(e^{i\lambda};\theta\right)$ is continuous in θ;*

(v) *for all $\theta \in \Theta$, $|\zeta(z;\theta)| \neq 0$, $|z| \leq 1$.*

Assumption 1 is weaker than imposing independence and identity of distribution of u_t, and Assumption 2 is standard from the literature on parametric short memory models since Hannan (1973), ensuring identifiability of θ_0 and easily covering stationary and invertible moving averages. These assumptions correspond to ones of Hualde and Robinson (2011), who established consistency of the same kind of estimates when

$g(x) \equiv 0$ in (9) *a priori.* In that setting they were able to choose the set of admissible memory parameters (our $[\nabla_1, \nabla_2]$) arbitrarily large, to simultaneously cover stationary, nonstationary, invertible and non-invertible values. This is more difficult, and perhaps imposssible, to achieve in the presence of the unknown, nonparametric g in (9), which can only be estimated with a slow rate of convergence, and we impose:

Assumption 3

$$\nabla_1 > 3/4, \nabla_2 < 5/4, \tag{14}$$

$$\delta_0 \in \nabla. \tag{15}$$

As can be inferred from the proof of Theorem 1, strictly what is required instead of (14) is the weaker condition $\nabla_2 - \delta_0 < 1/2$, but since δ_0 is known from (15) only to be no less than ∇_1 the restriction $\nabla_2 - \nabla_1 < 1/2$ implied by (14) is appropriate. Inspection of our proofs indicates that they go through with ∇ in Assumption 3 replaced by $[\varkappa, \varkappa + \omega]$, for any real $\varkappa$ and for $\omega \in (0, 1/2)$ (for example a subset of the stationary and invertible region $(-1/2, 1/2)$), but for the sake of clarity we fix on (14), which seems among the more empirically realistic possibilities, and covers the unit root case $\delta_0 = 1$.

We also need conditions on g, k and h.

Assumption 4

The function $g(x)$ is twice boundedly differentiable on $[0, 1]$ and $g(0) = 0$.

Assumption 5

The function $k(x)$ is even, differentiable at all but possibly finitely many x, with derivative $k'(x)$, and

$$\int_{\mathcal{R}} k(x)dx = 1,$$

$$k(x) = O((1 + x^{2+\eta})^{-1}),\ k'(x) = O((1 + |x|^{1+\eta})^{-1}),\ \textit{some}\ \eta > 0.$$

Assumption 6

As $T \to \infty$, the positive-valued sequence $h = h_T$ satisfies:

$$(Th)^{-1} + T^{2(\nabla_2 - \nabla_1)} h^3 \to 0. \tag{16}$$

Assumption 5 is virtually costless, covering many of the usual kernel choices. Assumption 6, however, represents a trade-off with Assumption 3: in the latter, $\nabla_2 - \nabla_1$ is desirably as close to $1/2$ as possible, but as it approaches $1/2$ from below the range of h satisfying Assumption 6 reduces to $(Th)^{-1} + Th^3 \to 0$.

Theorem 1

Let (9) and Assumptions 1-6 hold. Then as $T \to \infty$,

$$\widehat{\delta} \to_p \delta_0,\ \widehat{\theta} \to_p \theta_0.$$

The proof is in Appendix A. Asymptotic normality requires two further assumptions.

Assumption 7

(i) $\delta_0 \in (\nabla_1, \nabla_2)$; θ_0 *is an interior point of* Θ.

(ii) *for all real λ, $\zeta\left(e^{i\lambda}; \theta\right)$ is twice continuously differentiable in θ on a closed neighbourhood of radius $< 1/2$ about θ_0;*

(iii) *the matrix*

$$\Omega = \begin{pmatrix} \pi^2/6 & -\sum_{j=1}^{\infty} \psi_j' (\theta_0)/j \\ -\sum_{j=1}^{\infty} \psi_j (\theta_0)/j & \sum_{j=1}^{\infty} \psi_j (\theta_0) \psi_j' (\theta_0) \end{pmatrix}$$

is non-singular, where

$$\psi_j(\theta) = \sum_{k=0}^{j-1} \phi_k(\theta) \frac{\partial \zeta_{j-k}(\theta)}{\partial \theta},$$

the $\phi_j(\theta)$ being coefficients in the expansion

$$\phi_j(z;\theta) = \zeta(z;\theta)^{-1} = \sum\nolimits_{j=0}^{\infty} \phi_j(\theta) z^j.$$

This condition again is based on one of Hualde and Robinson (2011), but is similar to others in the literature, and practically unrestrictive. However we have to strengthen the first component of Assumption 6 on h.

Assumption 8

As $T \to \infty$, $Th^2/(\log T)^2 \to \infty$.

Theorem 2

Let (9) and Assumptions 1-8 hold. Then as $T \to \infty$

$$T^{1/2}\begin{pmatrix} \widehat{\delta} - \delta_0 \\ \widehat{\theta} - \theta_0 \end{pmatrix} \to_d \mathcal{N}(0, \Omega^{-1}).$$

The proof is in Appendix B. Note that the same limit distributions results when g is known or replaced by a parametric function. In the special case (4) of (9), we deduce that as $T \to \infty$

$$T^{1/2}\left(\widehat{\delta} - \delta_0\right) \to_d \mathcal{N}(0, 6/\pi^2).$$

3. UNIT ROOT TESTING

We first establish Wald tests for $\delta_0 = 1$ in (9), based on Theorem 2. Define

$$\Omega(\theta) = \begin{pmatrix} \pi^2/6 & -\sum_{j=1}^{\infty}\psi_j'(\theta)/j \\ -\sum_{j=1}^{\infty}\psi_j(\theta)/j & \sum_{j=1}^{\infty}\psi_j(\theta)\psi_j'(\theta) \end{pmatrix}$$

and denote by $\widehat{\Omega}^{(1,1)}$ the element in the top left hand corner of $\Omega\left(\widehat{\theta}\right)^{-1}$. Put

$$W = T^{1/2}\left(\widehat{\delta} - 1\right)/\widehat{\Omega}^{(1,1)1/2}.$$

Theorem 3

Let $\delta_0 = 1$ *in (9) and let Assumptions 1-8 hold. Then as* $T \to \infty$,

$$W \to_d \mathcal{N}(0,1). \tag{17}$$

The theorem follows from Theorem 2 and $\Omega\left(\widehat{\theta}\right) \to_p \Omega$, where the latter is implied by the proof of Theorem 2. We can reject the unit root null against more non-stationary alternatives when W falls in the appropriate upper tail of the standard normal density, and reject against less nonstationary alternatives when it falls in the appropriate lower tail.

Pseudo-log likelihood ratio tests can also be constructed. Define

$$\widetilde{\theta} = \arg\min_{\theta\in\Theta} Q_T(1,\theta), \tag{18}$$

and

$$LR = \log \frac{Q_T(1,\widetilde{\theta})}{Q_T(\widehat{\delta},\widehat{\theta})}.$$

Theorem 4

Let $\delta_0 = 1$ in (9) and let Assumptions 1-8 hold. Then as $T \to \infty$,

$$LR \to_d \chi_1^2.$$

The proof is standard, given Theorem 2 and a central limit theorem for $\widetilde{\theta}$ (see e.g. Hannan (1973), or implied by Hualde and Robinson(2011)).

Though it of course does not use $\widehat{\delta}$, $\widehat{\theta}$, for completeness we also present a Lagrange multiplier-type test, as it and the Wald and pseudo-log likelihood tests are expected to have equal local power. Robinson (1994) developed Lagrange multiplier tests for unit root and other hypotheses against fractional alternatives for the disturbances in multiple linear regression models. The stress there was on frequency-domain tests, but starting from an initial time-domain statistic, and to avoid introducing considerable additional notation we stay in the time domain here.

Writing $\partial = \partial/\partial(\delta,\theta')'$, from (13)

$$\partial Q(\delta,\theta) = \frac{2}{T}\sum_{t=1}^{T} \hat{u}_t(\delta,\theta)\partial\hat{u}_t(\delta,\theta), \tag{19}$$

where

$$\partial\hat{u}_t(\delta,\theta) = \xi_{t\delta\theta}(L)\, y_t - \partial\hat{g}_{\delta\theta}(t/T) = \frac{\sum_{s=1}^{T}\left(\partial\xi_{t\delta\theta}(L)\, y_t - \partial\xi_{s\delta\theta}(L)\, y_s\right) k_{ts}}{\sum_{s=1}^{T} k_{ts}}, \tag{20}$$

in which

$$\partial\xi_{t\delta\theta}(z) = \sum_{j=0}^{t-1} \partial\xi_j(\delta,\theta)z^j, \ \partial\xi_j(\delta,\theta) = \sum_{l=0}^{j} \begin{pmatrix} \left(\partial\alpha_l(\delta)/\partial\delta\right) \zeta_{j-l}(\theta)) \\ \alpha_l(\delta)\partial\zeta_{j-l}(\theta)/\partial\theta \end{pmatrix}. \tag{21}$$

In fact

$$\partial\xi_{t\delta\theta}(1,\theta) = \begin{pmatrix} \sum_{j=0}^{t-1} \left(\sum_{l=1}^{j} \frac{\zeta_{j-l}(\theta)}{l} \right) (y_{t-j} - y_{t-j-1}) \\ \sum_{j=0}^{t-1} \left(\partial\zeta_j(\theta)/\partial\theta\right) (y_{t-j} - y_{t-j-1}) \end{pmatrix}.$$

Define

$$LM = \frac{T}{4}\partial Q(1,\widetilde{\theta})'\Omega\left(\widetilde{\theta}\right)^{-1}\partial Q(1,\widetilde{\theta}),$$

with $\widetilde{\theta}$ given by (18). The proof of the following theorem is straightforward given the sentence after Theorem 3.

Theorem 5

Let $\delta_0 = 1$ in (9) and let Assumptions 1-8 hold. Then as $T \to \infty$,

$$LM \to_d \chi_1^2.$$

4. NONPARAMETRIC REGRESSION ESTIMATION

We can base estimation of $g(x)$ on our estimates of $\widehat{\delta}$, $\widehat{\theta}$ and (10), but in view of the stringent conditions on the bandwidth h in Theorems 1 and 2 we allow use of a possibly different bandwidth, b, in

$$\widetilde{g}_{\delta\theta}(x) = \sum_{s=1}^{T} \xi_{s\delta\theta}(L)\, y_s k\left(\frac{x - s/T}{b}\right) / \sum_{s=1}^{T} k\left(\frac{x - s/T}{b}\right), \tag{22}$$

We provide a multivariate central limit theorem for $\widetilde{g}_{\widehat{\delta}\widehat{\theta}}(\tau_i)$, $i = 1, 2, .., q$, where τ_i, $i = 1, 2, .., q$, are distinct fixed points, imposing also:

Assumption 9

As $T \to \infty$, $(bT)^{-1} + b^5 T \to 0$.

The proof of the following theorem is omitted as univariate and multivariate central limit theorems for the $\widetilde{g}_{\delta_0\theta_0}(x_i)$ are already in the literature (see e.g. Benedetti (1977), Robinson (1997)) and from Theorem 2 it is readily shown that $\widetilde{g}_{\widehat{\delta}\widehat{\theta}}(x) - \widetilde{g}_{\delta_0\theta_0}(x) = O_p(T^{-1/2})$ for all x.

Theorem 6

Let (9) and Assumptions 1-9 hold. Then as $T \to \infty$, *the* $(bT)^{1/2}\left(\widetilde{g}_{\widehat{\delta}\widehat{\theta}}(\tau_i) - g(\tau_i)\right)$, $i = 1, 2, ..., q$, *converge in distribution to independent* $N\left(0, \sigma^2 \int\limits_{\mathcal{R}} k(x)^2 dx\right)$ *random variables, where* σ^2 *is consistently estimated by*

$$\widehat{\sigma}^2 = Q\left(\widehat{\delta}, \widehat{\theta}\right).$$

This is the same limit distribution as results if δ_0 and θ_0 are known, i.e. the same as in the model *(1)* with iid u_t.

5. FINITE-SAMPLE PERFORMANCE

A small Monte Carlo study was carried out to investigate the finite-sample behaviour of our parameter estimates, and of one of our unit root tests. To generate data, in (9) we took $g(x) = \sin(2\pi x)$, $p = 1$, $\zeta(z; \theta) = 1 - \theta z$ (so y_t was a FARIMA$(1, \delta_0, 0)$), for various values of δ_0 and θ_0, and ε_t standard normally distributed. Throughout, parameter estimates were obtained taking k to be the standard normal kernel. All results are based on 1000 replications.

Tables 1-3 contain Monte Carlo biases b_δ and b_θ of $\widehat{\delta}$ and $\widehat{\theta}$ and coresponding Monte Carlo standard deviations s_δ and s_θ , for two (δ_0, θ_0) choices each, the sets ∇ and Θ varying with these. In Table 1 $(\delta_0, \theta_0) = \left(\frac{5}{8}, \frac{1}{2}\right)$ $(\nabla = [.4, 1.2]\,, \Theta = [.1, .9])$ and $(\delta_0, \theta_0) = \left(\frac{7}{8}, 0\right)$ $(\nabla = [.4, 1.1]\,, \Theta = [-.5, .5])$; in Table 2 $(\delta_0, \theta_0) = \left(1, \frac{1}{2}\right)$ $(\nabla = [.5, 1.5]\,, \Theta = [.1, .9])$ and $(\delta_0, \theta_0) = (1, 0)$ $(\nabla = [.5, 1.5]\,, \Theta = [-.5, .5])$; in Table

3 $(\delta_0, \theta_0) = \left(\frac{13}{8}, \frac{1}{2}\right)$ ($\nabla = [.9, 2.1]$, $\Theta = [.1, .9]$) and $(\delta_0, \theta_0) = \left(\frac{9}{8}, 0\right)$ ($\nabla = [.9, 1.5]$, $\Theta = [-.5, .5]$). All but one of these δ_0, and none of these ∇, satisfy Assumption 3. Note that in the cases $\theta_0 = 0$, one of which is included in each table, y_t reduces to a FARIMA$(0, \delta_0, 0)$, but we suppose that the practitioner does not know this. Three different (T, Th) combinations were employed: $(250, 20)$, $(600, 50)$ and $(1000, 80)$; these choices only partly represent an h sequence obeying our assumptions, because though Th increases with T, h takes successive values 12.5, 12, 12.5. In the tables, the excessive biases for the smaller sample sizes whenever $\theta_0 \neq 0$ may partly be due to the overly large ∇, as well as the deleterious effect of the nonparametric estimation, and the estimation procedure having some difficulty in distinguishing the long and short memory effects. The biases when $\theta_0 = 0$ are much smaller, and generally both biases and standard deviations diminish with increasing T, while there is a high stability across corresponding elements of the tables, especially Tables 1 and 2.

Table 1: Bias and standard deviation of $\left(\widehat{\delta}, \widehat{\theta}\right)$, $(\delta_0, \theta_0) = \left(\frac{5}{8}, \frac{1}{2}\right)$, $\left(\frac{7}{8}, 0\right)$.

(δ_0, θ_0)	$\left(\frac{5}{8}, \frac{1}{2}\right)$				$\left(\frac{7}{8}, 0\right)$			
T	b_δ	s_δ	b_θ	s_θ	b_δ	s_δ	b_θ	s_θ
250	0.3770	0.0537	-0.3081	0.0828	0.0422	0.0655	-0.0264	0.0952
600	0.1905	0.0436	-0.1719	0.0672	0.0272	0.0396	-0.0184	0.0579
1000	0.0669	0.0424	-0.0624	0.0581	0.0175	0.0327	-0.0131	0.0466

Table 2: Bias and standard deviation of $\left(\widehat{\delta}, \widehat{\theta}\right)$, $(\delta_0, \theta_0) = \left(1, \frac{1}{2}\right)$, $(1, 0)$.

(δ_0, θ_0)	$\left(1, \frac{1}{2}\right)$				$(1, 0)$			
T	b_δ	s_δ	b_θ	s_θ	b_δ	s_δ	b_θ	s_θ
250	0.3713	0.0555	-0.3000	0.0841	0.0380	0.0609	-0.0226	0.0911
600	0.1925	0.0437	-0.1733	0.0668	0.0252	0.0396	-0.0187	0.0603
1000	0.0674	0.0443	-0.0632	0.0594	0.0159	0.0317	-0.0136	0.0460

Table 3: Bias and standard deviation of $\left(\widehat{\delta},\widehat{\theta}\right)$, $(\delta_0,\theta_0)=\left(\frac{13}{8},\frac{1}{2}\right)$, $\left(\frac{9}{8},0\right)$.

(δ_0,θ_0)	$\left(\frac{13}{8},\frac{1}{2}\right)$				$\left(\frac{9}{8},0\right)$			
T	b_δ	s_δ	b_θ	s_θ	b_δ	s_δ	b_θ	s_θ
250	0.2666	0.0488	-0.1790	0.0801	0.0376	0.0636	-0.0178	0.0946
600	0.1853	0.0485	-0.1667	0.0735	0.0251	0.0393	-0.0171	0.0597
1000	0.0612	0.0457	-0.0568	0.0592	0.0170	0.0315	-0.0126	0.0461

Table 4 contains Monte Carlo sizes and powers for the LR unit root test described in Section 3, based on nominal 1% and 5% levels. Sizes were obtained using $(\delta_0,\theta_0)=\left(1,\frac{1}{2}\right)$ and $(T,Th)=(250,12)$, $(600,25)$, $(1000,52)$, powers using $(\delta_0,\theta_0)=\left(\frac{15}{16},\frac{1}{2}\right)$ and $\left(\frac{17}{16},\frac{1}{2}\right)$, with $(T,Th)=(250,20)$, $(600,50)$, $(1000,80)$. Considering the serious biases found in Tables 1-3 when $\theta_0=\frac{1}{2}$, the sizes in Table 4 do not seem bad, and they do improve with increasing T. Only one alternative in either direction from the unit root null is considered, but given that these are both close to 1 the differences between powers and corresponding sizes seem quite satisfactory, with slightly the greater sensitivity when $\delta_0=\frac{15}{16}$, and again there is improvement as T increases.

Table 4: Sizes and powers at nominal 1% and 5% levels

H_0	Size, $(\delta_0,\theta_0)=\left(1,\frac{1}{2}\right)$			
T	$\alpha=1\%$		$\alpha=5\%$	
250	0.019		0.068	
600	0.014		0.061	
1000	0.012		0.045	
H_1	Power, $(\delta_0,\theta_0)=\left(\frac{15}{16},\frac{1}{2}\right)$		Power, $(\delta_0,\theta_0)=\left(\frac{17}{16},\frac{1}{2}\right)$	
T	$\alpha=1\%$	$\alpha=5\%$	$\alpha=1\%$	$\alpha=5\%$
250	0.133	0.284	0.102	0.215
600	0.303	0.562	0.211	0.439
1000	0.504	0.725	0.479	0.708

6. FINAL REMARKS

The paper has justified large sample inference on the fractional and short memory parameters and nonparametric regression function in a semiparametric model incorporating nonstationary stochastic and deterministic trends. For parametric inference, the restrictions on the admissible memory parameter interval and the range of bandwidths are relatively strong, due to the presence of the nonparametric function and the extent of the time series dependence; possibly one or both these restrictions could be relaxed by means of higher-order kernels, as have been used elsewhere in the semiparametric econometric literature. A variety of further avenues might be explored. As always when nonparametric estimation is involved, bandwidth choice is a practical issue, though possibly a less acute one than in the stochastic design setting in which the density of explanatory variables varies over their support. In our Monte Carlo study only one value of h was used for each T, but sensitivity of estimates and tests to h, and b, can be gauged by carrying out the computations over a range of choices. With respect to automatic rules, in the model (1) a cross–validation choice of b is known to minimize mean integrated squared error (MISE), and we can extend this property to our setting, using $\widehat{\delta}$, $\widehat{\theta}$, though as usual the minimum-MSE rate does not quite satisfy conditions (our Assumption 9) for asymptotic normality about g; for h, as is familiar in the semiparametric literature the minimum-MISE rate is clearly excluded by the conditions (our Assumption 8) for asymptotic normality of parameter estimates, and a more appropriate goal may be to make a selection that matches the orders of the next two terms after the normal distribution function in an Edgeworth expansion for distribution function of $\widehat{\delta}$, $\widehat{\theta}$, and thereby minimizes the departure from the normal limit and leads to better-sized tests and more accurate interval estimates; in some settings this problem has a neat solution, but we do not know whether this is the case in ours. Bootstrapping is also likely to improve finite-sample properties.

Inference issues that might be investigated include testing constancy or other parametric restrictions on $g(x)$. Possible model extensions that require non-trivial further work include adding a nonparametric function of explanatory variables to $g(t/T)$ in (9) , and allowing for unconditional or conditional heteroscedasticity in u_t. Our work might also be extended to a panel setting, including individual effects and possible cross–sectional dependence.

ACKNOWLEDGMENTS

The first author acknowledges the Australian Research Council Discovery Grants Program for its support under Grant number: DP1096374. The second author's research was supported by ESRC Grant ES/J007242/1, and he thanks the Department of Econometrics and Business Statistics at Monash University for hosting and supporting his visits in March 2011 and September 2012. Thanks also go to Dr Jiying Yin for computing assistance.

APPENDIX A

Proof of Theorem 1

The model (9) refers to y_t u_t and $g(t/T)$ only for $t \geq 1$ so we can set $y_t = u_t = 0$, $t \leq 0$ and $g(x) = 0$, $x < 0$. Then for $t = 1, ..., T$

$$\xi_t \left(L; \delta, \theta\right) y_t = \xi \left(L; \delta, \theta\right) y_t$$

and

$$y_t = \xi \left(L; \delta_0, \theta_0\right)^{-1} \left(g \left(\frac{t}{T} \right) + u_t \right),$$

so

$$\begin{aligned} \xi_t \left(L; \delta, \theta\right) y_t &= \xi \left(L; \delta, \theta\right) \xi \left(L; \delta_0, \theta_0\right)^{-1} y_t \\ &= (1 - L)^{\delta - \delta_0} \tau \left(z; \theta\right) \left(g \left(\frac{t}{T} \right) + u_t \right), \end{aligned}$$

where

$$\tau(z;\theta) = \zeta(z;\theta)\,\zeta(z;\theta_0)^{-1} = \sum_{j=0}^{\infty} \tau_j(\theta) z^j.$$

From Zygmund (1977, p. 46), Assumption 2 implies that the Fourier coefficients $\tau_j(\theta)$ satisfy

$$\sup_{\Theta} |\tau_j(\theta)| = O(j^{-1-\varsigma}). \tag{A.1}$$

The Fourier coeficients $\chi_j(\delta,\theta)$ of

$$(1-z)^{\delta-\delta_0}\tau(z;\theta) = \sum_{j=0}^{\infty} \chi_j(\delta,\theta) z^j$$

are given by

$$\chi_j(\delta,\theta) = \sum_{l=0}^{j} \tau_l(\theta)\alpha_{j-l}(\delta-\delta_0).$$

(Note that $\alpha_j(0) = \tau_j(\theta_0) = \chi_j(\delta_0,\theta_0) \equiv 0,\ j \geq 1$). From (6), uniformly in $\delta \in [\nabla_1,\nabla_2] \setminus \{\delta_0\}$

$$\alpha_j(\delta-\delta_0) = O(j^{\delta_0-\delta-1}),\ \text{as } j \to \infty, \tag{A.2}$$

and so, using also (A.1)

$$\begin{aligned}
\left|\chi_j(\delta,\theta)\right| &\leq \left|\sum_{l=0}^{[j/2]} \tau_l(\theta)\alpha_{j-l}(\delta-\delta_0)\right| + \left|\sum_{l=[j/2]+1}^{j} \tau_l(\theta)\alpha_{j-l}(\delta-\delta_0)\right| \\
&\leq Kj^{\delta_0-\delta-1}\sum_{l=0}^{\infty}|\tau_l(\theta)| + Kj^{-1-\varsigma}\left|\sum_{l=0}^{j}\alpha_l(\delta-\delta_0)\right| \\
&\leq Kj^{\delta_0-\delta-1} + Kj^{\delta_0-\delta-1-\varsigma} \leq Kj^{\delta_0-\delta-1}
\end{aligned} \tag{A.3}$$

uniformly in δ,θ, where K throughout denotes a generic finite, positive constant. Also for future use note that from (6), uniformly in $\delta \in [\nabla_1,\nabla_2] \setminus \{\delta_0\}$, $\theta \in \Theta$,

$$|\alpha_j(\delta-\delta_0) - \alpha_{j+1}(\delta-\delta_0)| = O(j^{\delta_0-\delta-2}),\ \text{as } j \to \infty, \tag{A.4}$$

$$\begin{aligned}\left|\chi_j(\delta,\theta)-\chi_{j+1}(\delta,\theta)\right| &\leq \left|\sum_{l=0}^{j}\tau_l(\theta)\left(\alpha_{j-l}(\delta-\delta_0)-\alpha_{j+1-l}(\delta-\delta_0)\right)\right|+|\tau_{j+1}(\theta)| \\ &\leq Kj^{\delta_0-\delta-2}\sum_{l=0}^{\infty}|\tau_l(\theta)|+Kj^{-1-\varsigma}\sum_{l=1}^{\infty}l^{\delta_0-\delta-2}+Kj^{-1-\varsigma} \\ &\leq Kj^{\max(\delta_0-\delta,1-\varsigma)-2}. \qquad (A.5)\end{aligned}$$

With the abbreviations

$$\chi_{t\delta\theta}=\sum_{j=0}^{t-1}\chi_j(\delta,\theta)L^j,\ \ g_t=g\left(\frac{t}{T}\right),\ \ k_t=\frac{1}{Th}\sum_{s=1}^{T}k_{ts}$$

we have from (11)

$$\begin{aligned}\hat{u}_t\left(\delta,\theta\right) &= \frac{1}{Th}\sum_{s=1}^{T}\left(\chi_{t\delta\theta}g_t-\chi_{s\delta\theta}g_s\right)k_{ts}/k_t+\frac{1}{Th}\sum_{s=1}^{T}\left(\chi_{t\delta\theta}u_t-\chi_{s\delta\theta}u_s\right)k_{ts}/k_t \\ &= \chi_{t\delta\theta}u_t+D_{t\delta\theta}-S_{t\delta\theta},\end{aligned}$$

where

$$D_{t\delta\theta}=\frac{1}{Th}\sum_{s=1}^{T}\left(\chi_{t\delta\theta}g_t-\chi_{s\delta\theta}g_s\right)k_{ts}/k_t$$

and

$$S_{t\delta\theta}=\frac{1}{Th}\sum_{s=1}^{T}\chi_{s\delta\theta}u_sk_{ts}/k_t$$

are respectively the deterministic and stochastic errors contributing to the residual, that are absent when $g(t/T)\equiv 0$ in (9). Thus

$$\begin{aligned}Q(\delta,\theta) &= \frac{1}{T}\sum_{t=1}^{T}\left(\chi_{t\delta\theta}u_t+D_{t\delta\theta}-S_{t\delta\theta}\right)^2 \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\chi_{t\delta\theta}u_t\right)^2+\frac{1}{T}\sum_{t=1}^{T}D_{t\delta\theta}^2+\frac{1}{T}\sum_{t=1}^{T}S_{t\delta\theta}^2 \\ &\quad +\frac{2}{T}\sum_{t=1}^{T}\left(\chi_{t\delta\theta}u_t\right)D_{t\delta\theta}-\frac{2}{T}\sum_{t=1}^{T}\left(\chi_{t\delta\theta}u_t\right)S_{t\delta\theta} \\ &\quad -\frac{2}{T}\sum_{t=1}^{T}D_{t\delta\theta}S_{t\delta\theta}. \qquad (A.6)\end{aligned}$$

Hualde and Robinson (2011) show that the estimates minimizing

$$\frac{1}{T}\sum_{t=1}^{T}\left(\chi_{t\delta\theta}u_t\right)^2 \tag{A.7}$$

are consistent for δ_0, θ_0. From their proof it suffices to show that as $T \to \infty$,

$$\begin{aligned}
\sup \frac{1}{T}\sum_{t=1}^{T} D_{t\delta\theta}^2 \quad &\to \quad 0, && \text{(A.8)}\\
\sup \frac{1}{T}\sum_{t=1}^{T} S_{t\delta\theta}^2 \quad &\to_p \quad 0, && \text{(A.9)}\\
\sup \left|\frac{1}{T}\sum_{t=1}^{T}\left(\chi_{t\delta\theta}u_t\right) D_{t\delta\theta}\right| \quad &\to_p \quad 0, && \text{(A.10)}\\
\sup \left|\frac{1}{T}\sum_{t=1}^{T}\left(\chi_{t\delta\theta}u_t\right) S_{t\delta\theta}\right| \quad &\to_p \quad 0, && \text{(A.11)}\\
\sup \left|\frac{1}{T}\sum_{t=1}^{T} D_{t\delta\theta}S_{t\delta\theta}\right| \quad &\to_p \quad 0, && \text{(A.12)}
\end{aligned}$$

where the suprema here and subsequently are over $\delta \in [\nabla_1, \nabla_2]$, $\theta \in \Theta$. Given (A.8) and (A.9), and using the Cauchy inequality, (A.10)-(A.12) follow from the fact, implied by the proof of Hualde and Robinson (2011), that (A.7) is uniformly $O_p(1)$.

To prove (A.8) note first that Lemma 3 of Robinson (2012) gives, for all sufficiently large T,

$$\inf_t |k_t| \geq \frac{1}{8}. \tag{A.13}$$

Suppressing reference to δ, θ in $\chi_j = \chi_j(\delta, \theta)$,

$$\begin{aligned}
\sum_{s=1}^{T}\left(\chi_{t\delta\theta}g_t - \chi_{s\delta\theta}g_s\right) k_{ts} \quad &= \quad \sum_{s=1}^{T}\left(\sum_{j=0}^{t-1}\chi_j g_{t-j} - \sum_{j=0}^{s-1}\chi_j g_{s-j}\right) k_{ts}\\
&= \quad \sum_{j=0}^{T-1}\chi_j \sum_{s=1}^{T}\left(g_{t-j} - g_{s-j}\right) k_{ts}, && \text{(A.14)}
\end{aligned}$$

with the convention already adopted that $g(x) = 0$, $x < 0$, and $g(0) = 0$ from

Assumption 4. Then

$$\sup\left|\sum_{s=1}^{T}\left(\chi_{t\delta\theta}g_t-\chi_{s\delta\theta}g_s\right)k_{ts}\right|\leq\sum_{j=0}^{T-1}\sup\left|\chi_j\right|\left|\sum_{s=1}^{T}\left(g_{t-j}-g_{s-j}\right)k_{ts}\right|. \tag{A.15}$$

From (A.3) and Assumption 3

$$\sup\left|\chi_j\right|\leq Kj^{\nabla_2-\nabla_1-1}. \tag{A.16}$$

Applying Assumption 4 and with $\dot{g}_t$ denoting the derivative of $g(x)$ at $x=t/T$,

$$\sum_{s=1}^{T}\left(g_{t-j}-g_{s-j}\right)k_{ts}=\dot{g}_{t-j}\sum_{s=1}^{T}(\frac{t-s}{T})k_{ts}+O\left(\sum_{s=1}^{T}(\frac{t-s}{T})^2\left|k_{ts}\right|\right), \tag{A.17}$$

where $\dot{g}_t=0$, $t\leq 0$. By Lemma 2 of Robinson (2012)

$$\sum_{s=1}^{T}(\frac{t-s}{T})k_{ts}=Th^2\left(\frac{1}{Th}\sum_{s=1}^{T}(\frac{t-s}{Th})k_{ts}-\int uk(u)du\right)=O\left(h\right) \tag{A.18}$$

uniformly in $t\in(Th,T-Th)$. Uniformly in $t\leq Th$, $t\geq T-Th$,

$$\sum_{s=1}^{T}(\frac{t-s}{T})k_{ts}=O(Th^2) \tag{A.19}$$

from Lemma 1 of Robinson (2012). By the same lemma,

$$\sum_{s=1}^{T}(\frac{t-s}{T})^2\left|k_{ts}\right|=O(Th^3) \tag{A.20}$$

uniformly in t. Thus from (A.17)-(A.20),

$$\begin{aligned}\max_j\left|\sum_{s=1}^{T}\left(g_{t-j}-g_{s-j}\right)k_{ts}\right| &= O\left(h+Th^3\right),t\in(Th,T-Th)\\ &= O\left(Th^2\right),t\in(Th,T-Th),\end{aligned} \tag{A.21}$$

uniformly. Using also (A.13), (A.15) and (A.16),

$$\begin{aligned}\sup_{\delta,\theta}\left|D_{t\delta\theta}\right| &\leq K(Th)^{-1}\left(h+Th^3\right)\sum_{j=0}^{T-1}(1+j)^{\delta_0-\nabla_1-1}\\ &\leq K(T^{\nabla_2-\nabla_1-1}+T^{\nabla_2-\nabla_1}h^2),\ t\in(Th,T-Th),\end{aligned}$$

and

$$\begin{aligned}\sup_{\delta,\theta}|D_{t\delta\theta}| &\leq K(Th)^{-1}Th^2\sum_{j=0}^{T-1}(1+j)^{\nabla_2-\nabla_1-1} \\ &\leq KT^{\nabla_2-\nabla_1}h,\ t\leq Th, t\geq T-Th,\end{aligned}$$

uniformly over the stated ranges of t. Thus

$$\sup\frac{1}{T}\sum_{t=1}^{T}D_{t\delta\theta}^2\leq K(T^{2(\nabla_2-\nabla_1-1)}+T^{2(\nabla_2-\nabla_1)}h^4+T^{2(\nabla_2-\nabla_1)}h^3)\to 0$$

by Assumption 6, verifying (A.8).

To prove (A.9), we have

$$\sum_{s=1}^{T}\chi_{s\delta\theta}u_s k_{ts}=\sum_{j=0}^{T-1}\chi_j c_{tj}=\sum_{j=0}^{[Th]}\chi_j c_{tj}+\sum_{j=[Th]+1}^{T-1}\chi_j c_{tj}$$

where

$$c_{tj}=\sum_{r=1}^{T-j}u_r k_{t,r+j},$$

so, using (A.3),

$$\sup\left|\sum_{j=0}^{[Th]}\chi_j c_{tj}\right|\leq\sum_{j=0}^{[Th]}\left(\sup\left|\chi_j\right|\right)|c_{tj}|\leq K\sum_{j=1}^{[Th]}j^{\nabla_2-\nabla_1-1}|c_{tj}|$$

and thus

$$E\left(\sup\left|\sum_{j=0}^{[Th]}\chi_j c_{tj}\right|\right)^2\leq K\sum_{j=1}^{[Th]}\sum_{l=1}^{[Th]}j^{\nabla_2-\nabla_1-1}l^{\nabla_2-\nabla_1-1}\left(Ec_{tj}^2Ec_{tl}^2\right)^{1/2}. \tag{A.22}$$

Now

$$Ec_{tj}^2=\sigma^2\sum_{r=1}^{T-j}k_{t,r+j}^2=O(Th)$$

by Assumption 6, so (A.22) is $O((Th)^{2(\nabla_2-\nabla_1)+1})=o((Th)^2)$ uniformly in t, by Assumption 3.

By summation-by-parts

$$\sum_{j=[Th]+1}^{T-1} \chi_j c_{tj} = \sum_{j=[Th]+1}^{T-2} \left(\chi_j - \chi_{j+1}\right) d_{tj} + \chi_{T-1} d_{t,T-1}. \tag{A.23}$$

where

$$d_{tj} = \sum_{l=0}^{j} c_{tl}.$$

Now (A.23) is bounded uniformly by

$$\begin{aligned} & \sum_{j=[Th]+1}^{T-2} \left(\sup\left|\chi_j - \chi_{j+1}\right|\right) \left|d_{tj}\right| + \left(\sup\left|\chi_{T-1}\right|\right) \left|d_{t,T-1}\right| \\ \leq \; & K \sum_{j=[Th]+1}^{T-2} j^{\gamma-1} \left|d_{tj}\right| + KT^{\gamma} \left|d_{t,T-1}\right| \end{aligned} \tag{A.24}$$

using (A.3) and (A.5) and writing $\gamma = \max\left(\nabla_2 - \nabla_1, 1 - \varsigma\right) - 1$. Rearranging,

$$d_{tj} = \sum_{r=1}^{T} u_r \left(\sum_{s=r}^{\min(r+j,T)} k_{ts} \right),$$

so

$$Ed_{tj}^2 = \sigma^2 \sum_{r=1}^{T} \left(\sum_{s=r}^{\min(r+j,T)} k_{ts} \right)^2 \leq Kj \left(\sum_{s=1}^{T} \left|k_{ts}\right| \right)^2 \leq Kj \left(Th\right)^2$$

and (A.24) has second moment bounded by

$$\begin{aligned} & K\left(Th\right)^2 \sum_{j=[Th]+1}^{T-2} \sum_{l=[Th]+1}^{T-2} j^{\gamma-1/2} l^{\gamma-1/2} + K\left(Th\right)^2 T^{2\gamma+1} \\ \leq \; & K\left(Th\right)^{2\gamma+3} = o(\left(Th\right)^2) \end{aligned}$$

uniformly in t, since $\gamma < -1/2$. We have established that $E\sup S_{t\delta\theta}^2 = o(1)$ uniformly in t, whence follows (A.9), to complete the proof of the theorem.

APPENDIX B

Proof of Theorem 2

Writing $\partial = \partial/\partial(\delta, \theta')$, by the mean value theorem

$$0 = \partial Q(\widehat{\delta}, \widehat{\theta})/2 = \partial Q(\delta_0, \theta_0)/2 + \overline{\Omega} \begin{pmatrix} \widehat{\delta} - \delta_0 \\ \widehat{\theta} - \theta_0 \end{pmatrix},$$

where $\partial Q(\delta, \theta)$ is given by (19)-(21) and $\overline{\Omega}$ is obtained from the matrix $\partial^2 Q(\delta, \theta)/2 = \partial\partial' Q(\delta, \theta)/2$ by evaluating each row at a generally different $\widetilde{\delta}, \widetilde{\theta}$ such that $\left\|\widetilde{\delta} - \delta_0, \widetilde{\theta}' - \theta_0'\right\| \leq \left\|\widehat{\delta} - \delta_0, \widehat{\theta}' - \theta_0'\right\|$. The theorem follows if

$$\frac{T^{1/2}}{2} \partial Q(\delta_0, \theta_0) \rightarrow_d \mathcal{N}(0, \Omega), \tag{B.1}$$

$$\overline{\Omega} \rightarrow_p \Omega. \tag{B.2}$$

From (11)

$$\begin{aligned} \hat{u}_t(\delta_0, \theta_0) &= u_t + D_t - S_t \\ &= u_t + \sum_{s=1}^{T} (g_t - g_s) k_{ts} - \sum_{s=1}^{T} u_s k_{ts} \end{aligned}$$

where

$$D_t = \sum_{s=1}^{T} (g_t - g_s) k_{ts}/k_t, \quad S_t = \sum_{s=1}^{T} u_s k_{ts}/k_t.$$

From (20), (21)

$$\partial \hat{u}_t(\delta, \theta) = \partial \chi_{t\delta\theta} u_t + \partial D_{t\delta\theta} - \partial S_{t\delta\theta}.$$

Write

$$\pi_j = \partial \chi_j \left|_{\delta_0, \theta_0}\right. = \sum_{l=0}^{j} \begin{pmatrix} \tau_l(\theta_0) \partial \alpha_{j-l}(\delta_0)/\partial \delta \\ \alpha_{j-l}(\delta_0) \partial \tau_l(\theta_0)/\partial \theta \end{pmatrix} = \begin{pmatrix} j^{-1} \\ \psi_j(\theta_0) \end{pmatrix}, \ j \geq 1$$

Then

$$\partial \hat{u}_t(\delta_0, \theta_0) = v_t + \partial D_t - \partial S_t,$$

where

$$\begin{aligned}
v_t &= \partial\chi_{t\delta\theta}u_t\,|_{\delta_0,\theta_0} = \sum_{j=1}^{t-1}\pi_j u_{t-j}, \\
\partial D_t &= \partial D_{t\delta\theta}\,|_{\delta_0,\theta_0} = \sum_{s=1}^{T}\left(\sum_{j=1}^{t-1} g_{t-j}\pi_j - \sum_{j=1}^{s-1} g_{s-j}\pi_j\right) k_{ts}/k_t, \\
\partial S_t &= \partial S_{t\delta\theta}\,|_{\delta_0,\theta_0} = \sum_{s=1}^{T}\sum_{j=1}^{s-1} u_{s-j}\pi_j k_{ts}/k_t.
\end{aligned}$$

Thus from (19)

$$\partial Q(\delta_0,\theta_0) = \frac{2}{T}\sum_{t=1}^{T}\left(u_t + D_t - S_t\right)\left(v_t + \partial D_t - \partial S_t\right).$$

Hualde and Robinson (2011) show that

$$T^{-1/2}\sum_{t=1}^{T} u_t v_t \;\to_d\; \mathcal{N}(0,\Omega).$$

Thus (B.1) holds if all the following are $o_p(1)$:

$$\begin{aligned}
&T^{-1/2}\sum_{t=1}^{T} u_t\partial D_t,\; T^{-1/2}\sum_{t=1}^{T} u_t\partial S_t,\; T^{-1/2}\sum_{t=1}^{T} D_t v_t,\; T^{-1/2}\sum_{t=1}^{T} S_t v_t, \\
&T^{-1/2}\sum_{t=1}^{T} D_t\partial D_t,\; T^{-1/2}\sum_{t=1}^{T} D_t\partial S_t,\; T^{-1/2}\sum_{t=1}^{T} S_t\partial D_t,\; T^{-1/2}\sum_{t=1}^{T} S_t\partial S_t. \quad \text{(B.3)}
\end{aligned}$$

Note first that, as in (A.14), and using (A.13), (A.21) and $\|\pi_j\| = O\left(j^{-1}\right)$,

$$\begin{aligned}
\|\partial D_t\| &\le \frac{K}{Th}\left\|\sum_{j=1}^{T}\pi_j \sum_{s=1}^{T}\left(g_{t-j} - g_{s-j}\right)k_{ts}\right\| \\
&\le \frac{K}{Th}\max_{t,j}\left|\sum_{s=1}^{T}\left(g_{t-j} - g_{s-j}\right)k_{ts}\right|\sum_{j=1}^{T} j^{-1} \\
&= O\left(\left(T^{-1} + h^2\right)\log T\right), \quad t \in (Th, T-Th) \\
&= O\left(h\log T\right),\; t \le Th, t \ge T - Th, \qquad \text{(B.4)}
\end{aligned}$$

uniformly over the stated ranges of t. Similarly but more easily, we derive, uniformly,

$$\begin{aligned} D_t &= O\left(\left(T^{-1}+h^2\right)\right), \quad t \in (Th, T-Th) \\ &= O\left(h\right), \; t \leq Th, t \geq T-Th. \end{aligned} \tag{B.5}$$

We check each claim of (B.3) in turn; for notational convenience, when $j \leq 0$ we take $\pi_j = 0$ and interpret $1/j$ to be 0.

First, using (B.4) and Assumption 6,

$$\begin{aligned} E\left\|T^{-1/2}\sum_{t=1}^{T} u_t \partial D_t\right\|^2 &\leq \frac{K}{T}\sum_{t=1}^{T}\left\|\partial D_t\right\|^2 \\ &\leq \frac{K}{T}\left(T\left(T^{-1}+h^2\right)^2\log^2 T + Th^3\log^2 T\right) \\ &= O\left(T^{-2}\log^2 T + h^4\log^2 T + h^3\log^2 T\right) = o(1). \end{aligned}$$

Next,

$$\begin{aligned} \left\|T^{-1/2}\sum_{t=1}^{T} u_t \partial S_t\right\|^2 &= \frac{1}{T}E\left\{\sum_{t=1}^{T} u_t \sum_{r=1}^{T} u_r \sum_{s=1}^{T}\sum_{j=1}^{s-1} u_{s-j}\pi_j' \frac{k_{ts}}{k_t}\sum_{q=1}^{T}\sum_{l=1}^{q-1} u_{q-l}\pi_l \frac{k_{rq}}{k_r}\right\} \\ &= \frac{\sigma^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{s-1}\sum_{q=1}^{T}\pi_j'\pi_{q-s+j}\frac{k_{ts}k_{tq}}{k_t k_r} \\ &\quad + \frac{\sigma^4}{T}\sum_{t=1}^{T}\sum_{r=1}^{T}\sum_{s=1}^{T}\sum_{q=1}^{T}\pi_{s-t}'\pi_{q-r}\frac{k_{ts}}{k_t}\frac{k_{rq}}{k_r} \\ &\quad + \frac{\sigma^4}{T}\sum_{t=1}^{T}\sum_{r=1}^{T}\sum_{s=1}^{T}\sum_{q=1}^{T}\pi_{s-r}'\pi_{q-t}\frac{k_{ts}k_{rq}}{k_t k_r} \\ &\quad + \frac{Eu_t^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{q=1}^{T}\pi_{s-t}'\pi_{q-t}\frac{k_{ts}k_{tq}}{k_t k_r}. \end{aligned}$$

By boundedness of k, the final term in the last displayed expression is bounded by

$$\frac{K}{T^3h^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{q=1}^{T}\frac{1}{s-t}\frac{1}{q-t} = O\left(\frac{\log^2 T}{T^2h^2}\right),$$

while the other terms are bounded by

$$
\begin{aligned}
& \frac{K}{T^3h^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{s-1}\sum_{q=1}^{T}\frac{|k_{ts}|}{j}\frac{|k_{tq}|}{q-s+j}+\frac{K}{T^3h^2}\sum_{t=1}^{T}\sum_{r=1}^{T}\sum_{s=1}^{T}\sum_{q=1}^{T}\frac{|k_{ts}|}{s-r}\frac{|k_{rq}|}{q-t} \\
\leq\ & \frac{K}{T^3h^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{s-1}\sum_{q=1}^{T}\frac{|k_{ts}|}{j}\frac{1}{q-s+j}+\frac{K}{T^3h^2}\sum_{t=1}^{T}\sum_{r=1}^{T}\sum_{s=1}^{T}\sum_{q=1}^{T}\frac{1}{s-r}\frac{|k_{rq}|}{q-t} \\
\leq\ & \frac{KT^2h\log^2 T}{T^3h^2}=O\left(\frac{\log^2 T}{Th}\right)=o(1),
\end{aligned}
$$

by Assumption 8.

Next, noting that

$$
\sum_{t=1}^{T}D_t v_t=\sum_{t=1}^{T}\ u_t\sum_{s=t+1}^{T}D_s\pi_{s-t},
$$

from (B.5)

$$
\begin{aligned}
E\left\|T^{-1/2}\sum_{t=1}^{T}D_t v_t\right\|^2 &= \sigma^2 T^{-1}E\sum_{t=1}^{T}\left\|\sum_{s=t+1}^{T}D_s\pi_{s-t}\right\|^2 \\
&\leq KT^{-1}h^2\sum_{t=1}^{T}\left(\sum_{s=t+1}^{T}(s-t)^{-1}\right)^2 \\
&\leq Kh^2\log^2 T=o\left(1\right),
\end{aligned}
$$

by Assumption 6.

Next, using (B.4) and (B.5),

$$
\begin{aligned}
T^{-1/2}\sum_{t=1}^{T}D_t\partial D_t &= O\left(T^{-1/2}T\left(T^{-1}+h^2\right)^2\log T+T^{-1/2}Th^3\log T\right) \\
&= O\left(T^{-3/2}\log T+T^{1/2}h^4\log T+T^{1/2}h^3\log T\right)=o\left(1\right),
\end{aligned}
$$

by Assumption 6.

Next,

$$
E\left\|T^{-1/2}\sum_{t=1}^{T}S_t v_t\right\|^2=T^{-1}E\left\{\sum_{t=1}^{T}\sum_{s=1}^{T}u_s\frac{k_{ts}}{k_t}\sum_{j=1}^{t-1}\pi_j' u_{t-j}\sum_{r=1}^{T}\sum_{q=1}^{T}u_q\frac{k_{rq}}{k_r}\sum_{l=1}^{r-1}\pi_l u_{r-l}\right\}
$$

which equals

$$\begin{aligned}
&\sigma^4 T^{-1}\left(\sum_{t=1}^{T}\sum_{s=1}^{t-1}\frac{k_{ts}}{k_t}\pi'_{t-s}\right)^2\\
&+\sigma^4 T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\frac{k_{ts}}{k_t}\frac{k_{rs}}{k_r}\sum_{j=1}^{t-1}\pi'_j\pi_{r-t+j}\\
&+\sigma^4 T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}\frac{k_{ts}}{k_t}\frac{k_{rq}}{k_r}\pi'_{t-q}\pi_{r-s}\\
&+Eu_t^4\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\frac{k_{ts}}{k_t}\frac{k_{rs}}{k_r}\pi'_{t-s}\pi_{r-s},
\end{aligned}$$

and using boundedness of k this is bounded by $KT^{-1}(Th)^{-2}T^2\log^2 T + KT^{-1}(Th)^{-2}T\log^2 T \leq K\log^2 T/(Th^2) = o(1)$, by Assumption 8.

Next,

$$\begin{aligned}
E\left\|T^{-1/2}\sum_{t=1}^{T}D_t\partial S_t\right\|^2 &= \sigma^2 T^{-1}\sum_{t=1}^{T}\sum_{r=1}^{T}D_tD_r\sum_{s=1}^{T}\sum_{q=1}^{T}\sum_{j=1}^{s-1}\pi'_j\pi_{q-s+j}\frac{k_{ts}}{k_t}\frac{k_{rq}}{k_r}\\
&\leq KT^{-1}(Th)^{-2}h^2\sum_{t=1}^{T}\sum_{r=1}^{T}\sum_{s=1}^{T}|k_{ts}|\sum_{q=1}^{T}\|\pi_{q-s+j}\|\sum_{j=1}^{s-1}\|\pi_j\|\\
&\leq KT^{-1}(Th)^{-2}h^2T^2(Th)\log^2 T\\
&\leq Kh\log^2 T = o(1),
\end{aligned}$$

by (B.5) and Assumption 6.

Next,

$$\begin{aligned}
E\left\|T^{-1/2}\sum_{t=1}^{T}S_t\partial D_t\right\|^2 &= E\left\|T^{-1/2}\sum_{t=1}^{T}\partial D_t\sum_{s=1}^{T}u_s\frac{k_{ts}}{k_t}\right\|^2\\
&= \sigma^2 T^{-1}\sum_{t=1}^{T}\sum_{r=1}^{T}\partial D'_t\partial D_r\sum_{s=1}^{T}\frac{k_{ts}k_{rs}}{k_tk_r}\\
&\leq KT^{-1}(Th)^{-2}\ h^2\log^2 T\sum_{t=1}^{T}\sum_{r=1}^{T}\sum_{s=1}^{T}|k_{ts}k_{rs}|\\
&\leq KT^{-1}(Th)^{-2}h^2\left(\log^2 T\right)T(Th)^2 \leq Kh^2\log^2 T = o(1),
\end{aligned}$$

by Assumption 6.

Finally,

$$
\begin{aligned}
& E\left\|T^{-1/2}\sum_{t=1}^{T}S_t\partial S_t\right\|^2 \\
= \;& T^{-1}E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}u_s\frac{k_{ts}}{k_t}\sum_{r=1}^{T}\sum_{j=1}^{r-1}u_{r-j}\pi_j'\frac{k_{tr}}{k_t}\sum_{q=1}^{T}\sum_{p=1}^{T}u_p\frac{k_{qp}}{k_q}\sum_{n=1}^{T}\sum_{l=1}^{n-1}u_{n-l}\pi_l\frac{k_{qn}}{k_q}\right),
\end{aligned}
$$

which equals

$$
\begin{aligned}
& \sigma^4T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{k_{ts}}{k_t}\sum_{r=1}^{T}\pi_{r-s}'\frac{k_{tr}}{k_t}\sum_{q=1}^{T}\sum_{p=1}^{T}\frac{k_{qp}}{k_q}\sum_{n=1}^{T}\pi_{n-p}\frac{k_{qn}}{k_q} \\
& +\sigma^4T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{k_{ts}}{k_t}\sum_{r=1}^{T}\sum_{j=1}^{r-1}\pi_j'\frac{k_{tr}}{k_t}\sum_{q=1}^{T}\frac{k_{qs}}{k_q}\sum_{n=1}^{T}\pi_{n-r+j}\frac{k_{qn}}{k_q} \\
& +\sigma^4T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{k_{ts}}{k_t}\sum_{r=1}^{T}\sum_{p=1}^{T}\pi_{p-r}'\frac{k_{tr}}{k_t}\sum_{q=1}^{T}\frac{k_{qp}}{k_q}\sum_{n=1}^{T}\pi_{n-s}\frac{k_{qn}}{k_q} \\
& +Eu_t^4T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{k_{ts}}{k_t}\sum_{r=1}^{T}\pi_{r-s}'\frac{k_{tr}}{k_t}\sum_{q=1}^{T}\frac{k_{qs}}{k_q}\sum_{n=1}^{T}\pi_{n-s}\frac{k_{qn}}{k_q},
\end{aligned}
$$

which is bounded by

$$
\begin{aligned}
& KT^{-1}(Th)^{-4}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}k_{ts}\sum_{r=1}^{T}\|\pi_{r-s}\|\right)^2 \\
& +KT^{-1}(Th)^{-4}\sum_{t=1}^{T}\sum_{s=1}^{T}|k_{ts}|\sum_{r=1}^{T}\sum_{j=1}^{r-1}\|\pi_j\|\sum_{q=1}^{T}|k_{qs}|\sum_{n=1}^{T}\|\pi_{n-r+j}\| \\
& +KT^{-1}\;(Th)^{-4}\sum_{t=1}^{T}\sum_{s=1}^{T}|k_{ts}|\sum_{r=1}^{T}\sum_{p=1}^{T}\|\pi_{p-r}\|\sum_{q=1}^{T}|k_{qp}|\sum_{n=1}^{T}\|\pi_{n-s}\| \\
& +KT^{-1}(Th)^{-4}\sum_{t=1}^{T}\sum_{s=1}^{T}|k_{ts}|\sum_{r=1}^{T}\|\pi_{r-s}\|\sum_{q=1}^{T}|k_{qs}|\sum_{n=1}^{T}\|\pi_{n-s}\| \\
\leq\;& KT^{-1}(Th)^{-4}\left(T^2h\log T\right)^2+KT^{-1}(Th)^{-4}T\,(Th\log T)^2 \\
\leq\;& K\left(Th^2\right)^{-1}\log^2 T=o(1),
\end{aligned}
$$

by Assumption 8.

This completes the proof of (B.3), and thus of (B.1).

Finally (B.2) follows if

$$\frac{1}{2}\partial^2 Q(\delta_0,\theta_0) \to_p \Omega \tag{B.6}$$

and, given Theorem 1, for a neighbourhood $\mathcal{N}$ of δ_0, θ_0,

$$\sup_{\mathcal{N}} \left\|\partial^2 Q(\delta,\theta) - \partial^2 Q(\delta_0,\theta_0)\right\| \to_p 0. \tag{B.7}$$

The proof of (B.6) partly employs Theorem 2.2 of Hualde and Robinson (2011) and partly methods used above to deal with contributions from deterministic and stochastic errors, where these are less delicate than in Theorem 1's proof because $\nabla_2 - \nabla_1$ is replaced by an arbitrarily small positive number, and less delicate than the proof of Theorem 1 because $T^{-1/2}$ norming is replaced by T^{-1} norming. The proof of (B.7) uses that of Theorem 2.2 of Hualde and Robinson (2011) and standard techniques. The full details of the proofs of (B.6) and (B.7) are very lengthy but straightforward relative to what has gone before and are thus omitted.

REFERENCES

Benedetti, J.K., 1977. On the nonparametric estimation of regression functions. Journal of the Royal Statistical Society, Series B 39, 248-253.

Beran, J., Feng, Y., 2002. Local polynomial fitting with long memory, short memory and antipersistent errors. Annals of the Institute of Statisitical Mathematics 54, 291-311.

Csorgo, S., Mielniczuk, J., 1995. Nonparametric regression under long-range dependent normal errors. Annals of Statistics 23, 1000-1014.

Deo, R., 1997.Nonparametric regression with long-memory errors. Statistics and Probability Letters 33, 89-94.

Gasser, Th., Mueller, H.G., 1979. Kernel estimation of regression functions. In

Smoothing Techniques for Curve Estimation (eds. Th. Gasser and M. Rosenblatt) pp. 23-68. Springer, Heidelberg.

Guo, H., Koul, H. L., 2007. Nonparametric regression with heteroscedastic long memory errors. Journal of Statistical Planning and Inference 137, 379 – 404.

Hannan, E.J., 1973. The asymptotic theory of linear time-series models. Journal of Applied Probability 10, 130-145.

Hualde, J., Robinson, P.M., 2011. Gaussian pseudo-maximum likelihood estimation of fractional time series models, Annals of Statistics 39, 3152-3181.

Priestley, M.B., Chao, M.T., 1972. Non-parametric function fitting. Journal of the Royal Statistical Society, Series B 34, 385-392.

Robinson, P.M.,1994. Efficient tests of nonstationary hypotheses. Journal of the American Statistical Association 89, 1420-1437.

Robinson, P.M.,1997. Large-sample inference for nonparametric regression with dependent errors. Annals of Statistics 25, 2054-2083.

Robinson, P.M., 2012a. Inference on power law spatial trends", Bernoulli 18, 644-677.

Robinson, P.M., 2012b. Nonparametric trending regression with cross-sectional dependence. Journal of Econometrics 169, 4-14.

Roussas, G., Tran, L.T., Ioannides, O.A., 1992. Fixed design regression for time series: asymptotic normality. Journal of Multivariate Analysis 40, 262-291.

Starica, C., Granger, C., 2005. Nonstationarities in stock returns. Review of Economics and Statistics 87, 503-522.

Tran, L.T., Roussas, G., Yakowitz, S., Truong Van, B., 1996. Fixed-design regression for linear time series. Annals of Statistics 24, 975-991.

Zhao, Z., Zhang, Y., Li, R., 2013. Nonparametric estimation under strong dependence. Preprint.

Zygmund, A., 1979. Trigonometric Series, Cambridge University Press, Cambridge.