The Distance Between Rival Nonstationary
Fractional Processes

by

Peter M Robinson
London School of Economics and Political Science

Contents:

Abstract

Asymptotic inference on nonstationary fractional time series models, including cointegrated ones, is proceeding along two routes, determined by alternative definitions of nonstationary processes. We derive bounds for the mean squared error of the difference between (possibly tapered) discrete Fourier transforms under two regimes. We apply the results to deduce limit theory for estimates of memory parameters, including ones for cointegrated errors, with mention also of implications for estimates of cointegrating coefficients.

Contact address: Professor P M Robinson, Department of Economics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. Email: p.m.robinson@lse.ac.uk

1. INTRODUCTION

The treatment of pre-sample observations has long been an issue in time series analysis. In stationary autoregressive (AR) models, different initial value conventions lead to parameter estimates which typically share the same first-order asymptotic properties but have different finite sample ones. In computer generation of such series the exact autocovariance structure is achievable, but often zeros or the sample mean initiate the series, with early observations then thrown away. Different conventions have also been followed in nonstationary series with a unit AR root. In all these situations the treatment of only a few pre-sample observations (as many as the AR order) is in question. In stationary fractional processes, whose AR and moving average representations have infinite degree, infinitely many pre-sample values have to be chosen, so the potential divergence between rival methods of computer generation, and between rival parameter estimates, is greater, even though first-order asymptotic properties are again robust.

In the above examples there is often consensus about the underlying process, and the differences in pre-sample treatment appear only in the working model employed in obtaining parameter estimates for a finite stretch of data, for example manifesting some form of truncation (as in many "time domain" estimates) or approximating by a circulant (as in "frequency domain" ones). In nonstationary fractional series, however, at least two notably different definitions have been employed. To describe these, define

$$\Delta^{-a} = \sum_{j=0}^{\infty} \phi_j(a) L^j, \qquad \phi_j(a) = \frac{\Gamma(j+a)}{\Gamma(a)\Gamma(j+1)}, \qquad (1.1)$$

where L is the lag operator, $\Delta = 1 - L$ is the difference operator and Γ is the Gamma function with $\Gamma(a) = \infty$ for $a = 0, -1, ...,$ and $\Gamma(0)/\Gamma(0) = 1$. Let $\{\eta_t, t = 0, \pm 1, ...\}$ be a zero-mean covariance stationary process, with spectral density $f(\lambda)$ that is

bounded and bounded away from zero; we call η_t a short-memory process. For $d \in [-bd, bd)$

$$\zeta_t = \Delta^{-d} \eta_t, \qquad t = 0, \pm 1, ..., \tag{1.2}$$

is thus covariance stationary and, for $d > -bd$, invertible, having long memory when $d > 0$, negative memory when $d < 0$, and short memory when $d = 0$. Define the truncated ζ_t and η_t processes

$$\zeta_t^{\#} = \zeta_t 1(t \geq 1), \qquad \eta_t^{\#} = \eta_t 1(t \geq 1), \qquad t = 0, \pm 1, ..., \tag{1.3}$$

where 1 is the indicator function. For integer $q \geq 0$

$$X_t = \Delta^{-q} \zeta_t^{\#}, \qquad t = 0, \pm 1, ..., \tag{1.4}$$

is called a *Type I $I(q + d)$ process* and

$$Y_t = \Delta^{-q-d} \eta_t^{\#}, \qquad t = 0, \pm 1, ..., \tag{1.5}$$

is called a *Type II $I(q + d)$ process*.

When $d = 0$, (1.2) and (1.3) indicate that $\zeta_t^{\#} \equiv \eta_t^{\#}$, and thence from (1.4) and (1.5), $X_t \equiv Y_t$, so the two definitions are equivalent in non-fractional series, that have one or more unit roots ($q \geq 1$) or are stationary ($q = 0$). The definitions differ when $d \neq 0$. With $q \geq 1$, the Type I definition has been used by Sowell (1990), Hurvich and Ray (1995), Chan and Terrin (1995), Jeganathan (1999), Velasco (1999a,b), Marinucci (2000), Velasco and Robinson (2000) and others, whilst the Type II definition has been used by Robinson and Marinucci (2001), Kim and Phillips (2000) and others. (Robinson and Marinucci's (2001) $I(q + d)$ definition involves weights that generalize (1.1).) Type I and Type II processes were compared by Marinucci and Robinson (1999), who observed that when $q = 0$

$$E(X_t - Y_t)^2 = O(t^{2d-1}) \to 0 \qquad \text{as } t \to \infty, \tag{1.6}$$

2

and noted, from work of Mandelbrot and Van Ness (1968), Davydov (1970), Gorodet-skii (1977) (in connection with X_t), and of Akonom and Gourieroux (1987), Silveira (1991) and Marinucci and Robinson (2000) (in connection with Y_t) that empirical processes formed from X_t and Y_t converge to different forms of fractional Brownian motion; these latter were respectively termed "Type I" and "Type II" fractional Brownian motion by Marinucci and Robinson (1999), and our designation of X_t and Y_t corresponds.

The (albeit slow) convergence in (1.6) suggests that the choice between X_t and Y_t (generated from the same η_t sequence) may not always impact on asymptotic properties. This would be a desirable outcome as there seems no way of reliably determining whether an observed time series is Type I or Type II. Information on the "distance" between X_t and Y_t is useful in bounding the difference between limiting behaviour of statistics under the two regimes (1.4) and (1.5). If the limit distribution under one regime is already known, and this difference is of suitably small order, we can conclude the limit distribution under the other regime, without having to derive it independently. It seems more useful to consider the difference between discrete Fourier transforms (DFTs) of X_t and Y_t, than $X_t - Y_t$ itself. Frequency-domain methods have loomed large in memory parameter estimation. For semiparametric models (where, in the stationary case, the spectrum of a long memory process is parameterized only near zero frequency, or the autocovariance function only at long lags) log periodogram and local Whittle estimates have come to the fore, due to their intuitive appeal, computational simplicity and desirable asymptotic properties, being asymptotically normal for all d with simple, d-free, asymptotic variance. For parametric models, a form of Whittle estimation has computational advantages due to the fast Fourier transform and the convenient explicit form of the spectrum for many processes, can deal readily with mean-correction, and is $\sqrt{n}$-consistent and asymptotically normal, with limit variance matrix largely uninfluenced by the parent

innovation distribution, and asymptotically efficient in the stationary Gaussian case. Tapering of the data (see e.g. Hurvich and Ray, 1995) has helped in extending these methods to cover nonstationary series. DFTs of the data, or tapered versions, are the basic building blocks of frequency-domain methods.

The following section bounds the mean squared error of the difference between the DFTs of X_t and Y_t, and of tapered versions. Section 3 applies our results to show that tapered Whittle estimates for parametric nonstationary Type I models investigated by Velasco and Robinson (2000) can have the same limit distribution in the Type II case. This helps to verify unprimitive conditions on rates of convergence of memory parameter estimates that were employed by Robinson and Hualde (2003) in deriving the limit behaviour of cointegration estimates in the presence of unknown memory. These conditions include ones on the memory of (possible asymptotically stationary) cointegrating errors, which can only be estimated from residuals, and Section 4 derives the limiting distribution of such memory estimates, for both Type I and Type II processes. The concluding Section 5 briefly discusses some other applications of our results. Proofs appear in four appendices.

2. THE DISTANCE BETWEEN DISCRETE FOURIER TRANSFORMS

With C throughout denoting a generic positive constant, we call $h_t = h((t-0.5)/n)$,

$t = 1, ..., n$, a taper of (integer) order $p \geq 0$ if

$$h(1 - t) = h(t), \qquad 0 \leq t \leq 0.5,$$

$$|h(t)| \leq Ct^p, \qquad 0 \leq t \leq 0.5, \tag{2.1}$$

$$\left|\Delta^k h_t\right| \leq C\frac{t^{p-k}}{n^p}, \qquad 1 \leq k \leq p, \qquad k+1 \leq t \leq [(n+1)/2], \tag{2.2}$$

$$\sum_{t=p+1}^{n} \left|\Delta^{p+1} h_t\right| \leq Cn^{-p}, \tag{2.3}$$

$$\sum_{t=1}^{n} h_t^2 \geq n/C.$$

The simplest example of a taper of order $p = 0$ is

$$h_t \equiv 1, \tag{2.4}$$

corresponding to no tapering at all; for simplicity we take $p = 0$ to imply (2.4). Tapers of any chosen order $p > 0$ can be generated by the proposal of Kolmogorov (see Zhurbenko (1986)): when n is of form $n = (p+1)(r-1) + 1$ we choose h_t proportional to the coefficient of z^{t-1} in the expansion of $\{(1 - z^r)/(1 - z)\}^{p+1}$. Then (2.4) occurs when $p = 0$ (so $r = n$), but for $p = 1$ we have weights $h_t \propto t$, and for general p, $h_t \propto \Gamma(p+t)/\{\Gamma(t)\Gamma(p+1)\}$, $1 \leq t \leq [(n+1)/2]$. As p increases, h_t decays more and more smoothly towards zero as t approaches 1 and n.

Define the tapered DFTs

$$w_X(\lambda) = \left(2\pi \sum_{t=1}^{n} h_t^2\right)^{-bd} \sum_{t=1}^{n} h_t X_t e^{it\lambda}, \; w_Y(\lambda) = \left(2\pi \sum_{t=1}^{n} h_t^2\right)^{-bd} \sum_{t=1}^{n} h_t Y_t e^{it\lambda}. \tag{2.5}$$

When $q = 0$, untruncated, stationary, versions of X_t, Y_t have spectral densities proportional to λ^{-2d} as $\lambda \to 0+$, whilst when $q > 0$ we have a "pseudo spectrum" proportional to λ^{-2d-2q} as $\lambda \to 0+$. Thus, we expect appropriately normalized versions of (2.5) to be $\lambda^{d+q} w_X(\lambda)$, $\lambda^{d+q} w_Y(\lambda)$.

Theorem *When $p = q = 0$, $d \in (-bd, bd]$,*

$$E\left\{\lambda^{2d} |w_Y(\lambda) - w_X(\lambda)|^2\right\} \leq C |\log \lambda|^{1(d=-bd)} (n\lambda)^{-1}, \qquad 0 < \lambda \leq \pi. \tag{2.6}$$

5

When $p \geq \max(q, 1)$, $q \geq 0$, $d \in (-bd, bd]$,

$$E\left\{\lambda^{2d+2q} |w_Y(\lambda) - w_X(\lambda)|^2\right\} \leq C(\log n)^{1(d=-bd)}(n\lambda)^{2d-2}, \qquad 0 < \lambda \leq \pi. \quad (2.7)$$

The proof is in Appendix B, employing a sequence of lemmas established in Appendix A.

Remark 1 In connection with the restriction on λ, note that $|w_Y(\lambda) - w_X(\lambda)|^2$ is even and periodic of period 2π.

Remark 2 The Theorem holds under mild conditions on η_t (indeed it does not use the property that $f(\lambda)$ is everywhere positive, so holds also if X_t, Y_t are $I(c)$, $c < q+d$) but much stronger conditions are needed in the applications that follow.

Remark 3 The difference between DFTs seems intuitively a more meaningful measure of distance than the raw difference $X_t - Y_t$ (cf (1.6)) because DFTs can be approximately uncorrelated across frequencies that can be suitably separated.

Remark 4 We can discuss (2.6) in relation to approximations to the variance of the normalized DFT (e.g. Robinson, 1995a) for Type I stationary processes, in particular for $|d| < bd$, $f(\lambda)^{-1}\lambda^{2d}E |w_X(\lambda)|^2 = 1 + O\left((n\lambda)^{-1}\log(n\lambda + 1)\right)$ so (2.6) suggests that $w_Y(\lambda)$ approximates the leading term of $w_X(\lambda)$ if $n\lambda/\log n \to \infty$ as $n \to \infty..$

Remark 5 For $q > 0$ the above Type I variance approximation worsens for the untapered DFT (see e.g. Velasco and Robinson, 2000, Theorem A.1) but an improvement is achieved by tapering (see e.g. Velasco and Robinson, 2000, Theorem A.3). Their error bound can be dominated by a local smoothness property of f (unlike ours for the difference, so this effect apparently cancels). Ignoring this, their approximation error for the variance of the normalized tapered DFT is $O\left((n\lambda)^{2d-2}\log(n\lambda + 1)\right)$ when $p = q$, $|d| < bd$, so tapering helps here to an extent similar to that in (2.7). Both results improve when $d \downarrow -bd$ and worsen when $d \uparrow bd$.

Remark 6 These variance approximations were actually derived only for Fourier frequencies $\lambda = \lambda_j = 2\pi j/n$, $j = 1, 2, ..., [n/2]$. For such λ the bounds in (2.6) and (2.7) become $O(j^{-1})$ and $O(j^{2d-2})$ respectively for $|d| < bd$. It is dealing with λ close to zero (i.e. $j = o(n)$) that poses a distinctive difficulty in asymptotics for fractional series, but with λ bounded away from zero the bounds are $O(n^{-1})$ and $O(n^{2d-2})$.

Remark 7 It is evident how multivariate extensions of X_t and Y_t can be formed, perhaps with memory parameters varying across elements, but our Theorem applies, in element-by-element fashion, to multivariate series, so it could be used in multivariate extensions of Propositions 1-4 of Sections 3 and 4.

Remark 8 The Theorem continues to hold if both X_t and Y_t are corrupted by the same trend.

Remark 9 Our Type I and Type II definitions are only two of several possibilities, stressed here for simplicity of exposition, for their relatively marked discrepancy, and for their popularity. It is easily shown that the Theorem holds if Y_t is replaced by $\Delta^{-q-d}\{\eta_t 1(t \geq -\tau)\}$ for $\tau > 0$, indeed if τ increases fast enough relative to n the results can be improved. In empirical work, Hualde and Robinson (2001) found evidence of sensitivity to choice of τ.

3. ESTIMATION OF MEMORY PARAMETERS FROM RAW DATA

Parametric models are important in describing time series of moderate length, and in prediction. Suppose we know a function $\ell(\lambda; \theta^{(-)})$ of λ and the $(a-1) \times 1$ vector $\theta^{(-)}$, such that $\int_{-\pi}^{\pi} \log \ell(\lambda; \theta^{(-)}) d\lambda = 0$ for all $\theta^{(-)}$ and, for $f(\lambda; \theta^{(-)}, \sigma^2) = (\sigma^2/2\pi) \ell\left(\lambda; \theta^{(-)}\right)$ and some unknown $\sigma_0^2 > 0$, $\theta_0^{(-)}$, we have $f(\lambda) = f\left(\lambda; \theta_0^{(-)}, \sigma_0^2\right)$.

For example, $f\left(\lambda;\theta^{(-)},\sigma^2\right)$ can be a "standard parameterization" of an ARMA spectrum, where, with $\delta_0 = q + d$, $\left|1 - e^{i\lambda}\right|^{-2\delta_0} f\left(\lambda;\theta_0^{(-)},\sigma_0^2\right)$ is the spectrum of a fractional ARIMA process when $q = 0$ and the "pseudo-spectrum" when $q > 0$. The processes in question are Type I, but corresponding Type II processes are defined by employing the same $f(\lambda)$ and δ_0 with respect to (1.5). Some attempts at asymptotic theory for Whittle-type estimates of $\theta_0 = \left(\delta_0, \theta_0^{(-)'}\right)'$ (for Type II processes) when q is unknown are incomplete because they did not demonstrate uniform convergence of the objective function over a suitably broad interval of δ-values. Velasco and Robinson (2000) (hereafter VR) achieved this in case of Type I processes using tapering, establishing $\sqrt{n}$-consistency and asymptotic normality (and thus standard asymptotics, as foreshadowed in Robinson's (1994) treatment of score testing in non-stationary fractional models based on time- and frequency-domain approximations to a Gaussian likelihood).

We apply the Theorem to show that the same results hold in case of a corresponding Type II process. Define $\theta = \left(\delta, \theta^{(-)'}\right)'$, $k(\lambda;\theta) = \left|1 - e^{i\lambda}\right|^{-2\delta} \ell(\lambda;\theta^{(-)})$. We estimate θ_0 by $\tilde{\theta} = \arg\min_\Theta \tilde{Q}_n(\theta)$, where $\Theta = [\nabla_1, \nabla_2] \times \Theta^{(-)}$ for $\nabla_1 > -bd$, $\nabla_2 < \infty$ and $\Theta^{(-)}$ a compact subset of $\mathbb{R}^{a-1}$, and

$$\tilde{Q}_n(\theta) = \frac{2\pi(p+1)}{n} \sum_{j(p)} \frac{\tilde{I}(\lambda_j)}{k(\lambda_j;\theta)}, \quad \tilde{I}(\lambda) = |w_Y(\lambda)|^2,$$

where $\sum_{j(p)}$ denotes a sum over $j = p+1, 2(p+1), ..., n-p+1$, p being the order of the taper in w_Y. The estimate studied by VR, based on X_t, is $\hat{\theta} = \arg\min_\Theta Q_n(\theta)$, where

$$Q_n(\theta) = \frac{2\pi(p+1)}{n} \sum_{j(p)} \frac{I(\lambda_j)}{k(\lambda_j;\theta)}, \qquad I(\lambda) = |w_X(\lambda)|^2.$$

Both Q_n and $\tilde{Q}_n$ result from concentrating out σ^2 from tapered frequency-domain approximations to a Gaussian likelihood. As in VR, the skipping of frequencies when $p > 0$ avoids correlation across neighbouring Fourier frequencies induced by tapering,

8

and it corrects for corruption of X_t or Y_t by a polynomial trend of degree no greater than p.

The asymptotic theory of VR requires regularity conditions which are relatively standard and are not repeated here - they permit Y_t to be a Gaussian fractional ARIMA for example. Our conditions on h imply VR's, noting that their p is our $p+1$.

Proposition 1 *If h is of order $p \geq q$ then under Assumptions A.1-A.4 of VR, $\tilde{\theta} \rightarrow_p \theta_0$ as $n \rightarrow \infty$.*

Appendix C contains the proof of this and of the next proposition, which indicates that $\tilde{\theta}$ has the same limit distribution as VR's $\hat{\theta}$.

Define

$$
\begin{aligned}
\Sigma &= \int_{-\pi}^{\pi} \left\{ \frac{\partial}{\partial \theta} \log k(\lambda; \theta_0) \right\} \left\{ \frac{\partial}{\partial \theta'} \log k(\lambda; \theta_0) \right\} d\lambda, \\
\Phi_p &= \lim_{n \to \infty} \left\{ \sum_{t=1}^{n} h_t^2 \right\}^{-2} \sum_{j(p)} \left\{ \sum_{t=1}^{n} h_t^2 \cos t\lambda_j \right\}^2 .
\end{aligned}
$$

Proposition 2 *If h is of order $p \geq 1$ for $q = 0$ or of order $p \geq q$ for $q > 0$, then under Assumptions A.1-A.4 and A.6-A.9 of VR, as $n \rightarrow \infty$ $n^{bd}(\tilde{\theta} - \theta_0) \rightarrow_d N\left(0, 4\pi(p+1)\Phi_p\Sigma^{-1}\right).$*

4. ESTIMATION OF MEMORY PARAMETERS FROM RESIDUALS

An important application of parametric memory estimation arises in fractional cointegration analysis, where the memory of observables and/or cointegrating errors is unknown and possibly fractional, and limitations on the length of macroeconomic

series support parametric modelling of dynamics. Consider the "Type II" cointegrated system

$$y_t = \nu x_t + \Delta^{-\delta_0}\eta_{1t}^{\#}, \qquad x_t = \Delta^{-\gamma_0}\eta_{2t}^{\#}, \tag{4.1}$$

where $\eta_{it}^{\#} = \eta_{it}1(t \geq 1)$, $i = 1,2$, and (η_{1t}, η_{2t}) is a bivariate covariance stationary process with spectral density matrix $F(\lambda)$ that is at least bounded and positive definite for all λ, whilst

$$\delta_0 \geq 0, \qquad \gamma_0 - \delta_0 > bd; \tag{4.2}$$

ν is the unknown cointegrating coefficient. Robinson and Hualde (2003) established desirable asymptotic properties for estimates of ν that have the apparently optimal convergence rate $n^{\gamma_0 - \delta_0}$, in the presence of unknown γ_0, δ_0 (whose meaning they reverse). If the second condition in (4.2) is reversed the optimal rate appears to be n^{bd} and asymptotic inference is substantially affected. Robinson and Hualde (2003) assumed $F(\lambda)$ is parametric, depending on an unknown vector ϕ_0, e.g. (η_{1t}, η_{2t}) is a bivariate ARMA. The asymptotic properties of their estimates of ν require certain rates of convergence of $\gamma_0, \delta_0, \phi_0$; $\sqrt{n}$-consistency suffices, so Proposition 2 implies the desired $\sqrt{n}$-consistency of the estimates of γ_0 and the parameters explaining autocorrelation in η_{2t}. However, Proposition 2 does not apply to estimates of δ_0 or parameters explaining autocorrelation in η_{1t} because ν is unknown, so $Y_t = y_t - \nu x_t$ is unobservable. Given a preliminary estimate $\tilde{\nu}$, estimates can be based on the $\hat{Y}_t = y_t - \tilde{\nu}x_t$. Assuming the spectral density of η_{1t} can be parameterized as $\sigma^2 \ell(\lambda; \theta^{(-)})/2\pi$, then with $k(\lambda; \theta)$ as in Section 3 we consider

$$\bar{\theta} = \arg\min_{\Theta}\bar{Q}_n(\theta) \qquad \bar{Q}_n(\theta) = \frac{2\pi(p+1)}{n}\sum_{j(p)}\frac{\hat{I}(\lambda_j)}{k(\lambda_j; \theta)},$$

$$\hat{I}(\lambda) = |w_{\hat{Y}}(\lambda)|^2, \qquad w_{\hat{Y}}(\lambda) = (2\pi n)^{-bd}\sum_{t=1}^{n}h_t\hat{Y}_te^{it\lambda}.$$

10

Denote by r the positive integer such that $-bd \leq \gamma_0 - r < bd$. In employing assumptions of VR in the following propositions we mean that their X_t denotes the corresponding Type I process to the Type II process $Y_t = y_t - \nu x_t$.

Proposition 3 *If h is of order $p \geq r$, and*

$$\hat{\nu} - \nu = O_p(n^{\delta_0 - \gamma_0 + \rho}), \qquad \text{any } \rho > 0, \tag{4.3}$$

then under (4.1) and (4.2) and Assumptions A.1-A.4 of VR, $\bar{\theta} \to_p \theta_0$ as $n \to \infty$.

Proposition 4 *If h is of order $p \geq r$ and (4.3) holds, then under (4.1) and (4.2) and Assumptions A.1-A.4 and A.6-A.9 of VR, $n^{bd}(\bar{\theta} - \theta_0) \to_d N\left(0, 4\pi(p+1)\Phi_p \Sigma^{-1}\right)$ as $n \to \infty$.*

The proofs are in Appendix D. Because $r > \gamma_0 - bd > \delta_0$ implies $r \geq q$, the degree of tapering required may be greater than in Propositions 1 and 2, but versions of Propositions 1 and 2 applying to estimates of γ_0, using x_t, would entail $p \geq r$, suggesting that the same taper be used in both equations of (4.1). Robinson and Marinucci (2001, Propositions 6.1 and 6.2) showed that if $\gamma_0 + \delta_0 \geq 1$ (4.3) holds when $\tilde{\nu}$ is the least squares estimate $\tilde{\nu}_{LS}$, and that if $\gamma_0 + \delta_0 < 1$ it can hold when $\tilde{\nu}$ is the narrow-band least squares estimate $\tilde{\nu}_{NBLS}$ if the number of Fourier frequencies used increases slowly enough, say like $\log n$. The strength of (4.3) in part reflects the need to deal with a desirably large admissible parameter set in the consistency proof. We may infer from Propositions 1-4 that Propositions 3 and 4 hold for a Type I version of the cointegration model (4.1), applying the Theorem in the opposite direction.

5. FINAL COMMENTS

The estimation of ν in (4.1) is one area in which Type I and II processes may lead to different limit distributions. For untapered estimates in the Type I case

with $0 \leq \delta_0 < bd$, $1 \leq \gamma_0 < 3/2$, Marinucci (2000) obtained the same rate of convergence as Robinson and Marinucci (2001) did under Type II, but a different limit distribution (due to the different forms of fractional Brownian motion). Over the wider region $\gamma_0 + \delta_0 > 1$, limit distributions of tapered $\tilde{\nu}_{LS}$ and $\tilde{\nu}_{NBLS}$ differ between Type I and Type II processes (while converging at the same rate, that for untapered estimates), but (using the Theorem) are identical when $\gamma_0 + \delta_0 < 1$. The Theorem should also be useful in other time series regression settings. Returning to memory parameter estimation, the results of Robinson (1995a,b) and Velasco (1999a,b) for semiparametric estimates under Type I should apply, via the Theorem, to Type II also.

APPENDIX A: TECHNICAL LEMMAS

The proof of the Theorem is facilitated by the following lemmas.

Lemma 1 *Defining, for $t > 1$,*

$$b_{t\ell}(r, d) = \sum_{k=0}^{t-1} \phi_k(r)\phi_{t+\ell-k}(d),$$

we have the representations

$$X_t = \sum_{\ell=0}^{t-1} \phi_\ell(q + d)\eta_{t-\ell} + \sum_{\ell=0}^{\infty} b_{t\ell}(q, d)\eta_{-\ell}, \qquad Y_t = \sum_{\ell=0}^{t-1} \phi_\ell(q + d)\eta_{t-\ell},$$

so that

$$X_t - Y_t = \sum_{\ell=0}^{\infty} b_{t\ell}(q, d)\eta_{-\ell}. \tag{A.1}$$

Proof: The representation of Y_t is immediate given (1.1), (1.3) and (1.5) and the identity $\Delta^a \Delta^b = \Delta^{a+b}$. From (1.2) $\zeta_t = \sum_{j=0}^{\infty} \phi_j(d)\eta_{t-j}$, so that

$$\begin{aligned}
X_t &= \sum_{k=0}^{t-1} \phi_k(q) \sum_{j=0}^{\infty} \phi_j(d)\eta_{t-j-k} \\
&= \sum_{\ell=0}^{t-1} \sum_{j=0}^{\ell} \phi_j(q)\phi_{\ell-j}(d)\eta_{t-\ell} + \sum_{\ell=t}^{\infty} \sum_{j=0}^{t-1} \phi_j(q)\phi_{\ell-j}(d)\eta_{t-\ell},
\end{aligned}$$

which gives the desired result, on noting that $\phi_j(a+b) = \sum_{k=0}^{j} \phi_k(a)\phi_{j-k}(b)$. $\qquad\square$

Lemma 2

$$b_{t\ell}(r - 1, d) = b_{t\ell}(r, d) - b_{t-1,\ell}(r, d), \qquad t > 1. \tag{A.2}$$

Proof: The right hand side of (A.2) is

$$\phi_{t+\ell}(d) + \sum_{k=1}^{t-1} \left\{ \phi_k(r) - \phi_{k-1}(r) \right\} \phi_{t+\ell-k}(d),$$

13

which, due to (1.1), equals the left side in view of the identity

$$\phi_k(r-1) = \phi_k(r) - \phi_{k-1}(r), \qquad k \geq 1, \qquad \phi_0(r) \equiv 1. \qquad \square$$

Lemma 3 *For $t \geq 1, \ell \geq 0$*

$$
\begin{align}
b_{t\ell}(r,0) &= 0, & r \geq -1, & \qquad\qquad\text{(A.3)}\\
b_{t\ell}(-1,d) &\leq C(t+\ell)^{d-2}, & d \neq 0, & \qquad\qquad\text{(A.4)}\\
b_{t\ell}(0,d) &\leq C(t+\ell)^{d-1}, & d \neq 0, & \qquad\qquad\text{(A.5)}
\end{align}
$$

and for $k \geq 1$,

$$
\begin{align}
b_{t\ell}(k,d) &\leq Ct^{k-1+d}, & \ell \leq t, & \quad d \geq 0, & \text{(A.6)}\\
&\leq Ct^{k-1}(\ell+1)^d, & \ell \leq t, & \quad d < 0, & \text{(A.7)}\\
&\leq Ct^k \ell^{d-1}, & \ell > t, & \quad d < 1. & \text{(A.8)}
\end{align}
$$

Proof: The relation (A.3) is immediate, whilst (A.4) and (A.5) follow from Stirling's formula. From Lemma 2 and (A.5) we deduce (A.6)-(A.8) when $k = 1$, and thence, recursively, when $k > 1$. $\qquad\qquad\square$

Lemma 4 *For a sequence $\Delta_t = \Delta_t^{(0)}$, $t = 1, 2, ...$, define, for integer $r \geq 1$,*

$$\Delta_t^{(r)} = \Delta_t^{(r-1)} - \Delta_{t-1}^{(r-1)}, \qquad t > r. \qquad\qquad\text{(A.9)}$$

Then for any integer $q \geq 1$

$$\sum_{t=1}^{n} \Delta_t e^{it\lambda} = \sum_{r=1}^{q} g(\lambda)^r \left\{ \Delta_r^{(r-1)} - \Delta_n^{(r-1)} e^{i(n-r+1)\lambda} \right\} \qquad\text{(A.10)}$$

$$+ g(\lambda)^q \sum_{t=1}^{n-q} \Delta_{t+q}^{(q)} e^{it\lambda}, \qquad\qquad\text{(A.11)}$$

14

where $g(\lambda) = e^{i\lambda}(1 - e^{i\lambda})^{-1}$.

Proof: Define $v_r(\lambda) = \sum_{t=1}^{n-r} \Delta_{t+r}^{(r)} e^{it\lambda}$. By summation-by-parts, for $r \geq 1$

$$v_{r-1}(\lambda) = \sum_{t=1}^{n-r} \left\{ \Delta_{t+r-1}^{(r-1)} - \Delta_{t+r}^{(r-1)} \right\} D_{1t}(\lambda) + \Delta_n^{(r-1)} D_{1,n-r+1}(\lambda),$$

where $D_{st}(\lambda) = \sum_{r=s}^{t} e^{ir\lambda}$. Noting (A.9), and $D_{1t}(\lambda) = g(\lambda)(1 - e^{it\lambda})$, we have

$$
\begin{aligned}
v_{r-1}(\lambda) &= g(\lambda) \left\{ v_r(\lambda) - \sum_{t=1}^{n-r} \Delta_{t+r}^{(r)} + \Delta_n^{(r-1)} \left(1 - e^{i(n-r+1)\lambda} \right) \right\} \\
&= g(\lambda) \left\{ v_r(\lambda) + \Delta_r^{(r-1)} - \Delta_n^{(r-1)} e^{i(n-r+1)\lambda} \right\}.
\end{aligned}
$$

Applying this recursion successively for $r = 1, ..., q$ completes the proof. $\qquad \square$

Lemma 5 *For $1 \leq s \leq t \leq n$*

$$\left| \sum_{r=s}^{t} \Delta^k h_r e^{ir\lambda} \right| \leq \frac{C}{n^k |\lambda|}, \qquad 0 < |\lambda| \leq \pi, \qquad k = 0, ..., p. \qquad (A.12)$$

Proof: By summation by parts,

$$\sum_{r=s}^{t} \Delta^k h_r e^{ir\lambda} = - \sum_{r=s}^{t-1} \Delta^{k+1} h_r D_{sr}(\lambda) + \Delta^k h_t D_{st}(\lambda).$$

From Zygmund (1977, p.51)

$$|D_{st}(\lambda)| \leq C |\lambda|^{-1}, \qquad O < |\lambda| \leq \pi, \qquad (A.13)$$

so from (2.2)

$$\left| \sum_{r=s}^{t} \Delta^k h_r e^{ir\lambda} \right| \leq \frac{C}{|\lambda|} \sum_{r=s}^{t-1} \frac{r^{p-k-1}}{n^p} + \frac{C}{n^k |\lambda|} \leq \frac{C}{n^k |\lambda|}$$

for $k = 0, ..., p - 1$, whilst for $k = p$ (A.12) follows using (2.3). $\qquad \square$

APPENDIX B: PROOF OF THEOREM

The proof for $d = 0$ is trivial, indeed the left sides of (2.6) and (2.7) are actually zero. Consider $|d| < bd$. From (A.1)

$$\sum_{t=1}^{n} h_t \left(X_t - Y_t\right) e^{it\lambda} = \sum_{\ell=0}^{\infty} \eta_{-\ell} w_\ell(q, d), \tag{B.1}$$

where $w_\ell(q, d) = \sum_{t=1}^{n} h_t b_{t\ell}(q, d) e^{it\lambda}$. Now (B.1) has mean zero and variance

$$\sum_{\ell=0}^{\infty} \sum_{m=0}^{\infty} E(\eta_{-\ell} \eta_{-m}) w_\ell(q, d) w_m(q, d) = \int_{-\pi}^{\pi} f(\omega) \left| \sum_{\ell=0}^{\infty} e^{-i\ell\omega} w_\ell(q, d) \right|^2 d\omega$$

$$\leq C \int_{-\pi}^{\pi} \left| \sum_{\ell=0}^{\infty} e^{-i\ell\omega} w_\ell(q, d) \right|^2 d\omega \leq C \sum_{\ell=0}^{\infty} |w_\ell(q, d)|^2. \tag{B.2}$$

It suffices to show that

$$\sum_{\ell=0}^{\infty} |w_\ell(q, d)|^2 \leq C\lambda^{-2d-1}, \qquad p = q = 0, \tag{B.3}$$

$$\leq Cn^{2d-1}\lambda^{-2q-2}, \qquad q \geq 0, \qquad p \geq \max(q, 1). \tag{B.4}$$

To estimate $w_\ell(q, d)$ when $q > 0$ we apply Lemma 4, taking $\Delta_t = \Delta_{t\ell} = h_t b_{t\ell}(q, d)$. We consider first the terms on the right side of (A.10). For sequences c_t, d_t we may write $\Delta^{r-1}(c_s d_t)$, $1 \leq r \leq q$, as a linear combination of terms $\Delta^k c_{s-i} \Delta^{r-1-k} d_{t-j}$, for $k = 0, ..., r - 1$, and finitely many integers i, j satisfying $0 \leq i, j \leq r - 1$. Thus

$$\Delta_t^{(r-1)} = \Delta_{t\ell}^{(r-1)} = O\left(\sum_{k=0}^{r-1} |\Delta^k h_t| |\Delta^{r-1-k} b_{t\ell}(q, d)|\right)$$

$$= O\left(\sum_{k=0}^{r-1} |\Delta^k h_t| b_{t\ell}(q - r + 1 + k, d)\right)$$

from Lemma 2. From (2.1), (2.2) and Lemma 3,

$$\left|\Delta_{n\ell}^{(r-1)}\right| \leq C\frac{(\ell+1)^d}{n^p} \sum_{k=0}^{r-1} n^{q-r+k} \leq C\frac{(\ell+1)^d}{n^{p-q+1}}, \qquad 0 \leq \ell \leq n, \qquad d \leq 0;$$

$$\leq \frac{C}{n^p} \sum_{k=0}^{r-1} n^{q-r+k+d} \leq Cn^{q-p+d-1}, \qquad \ell > n, \qquad d \geq 0;$$

$$\leq C\frac{\ell^{d-1}}{n^p} \sum_{k=0}^{r-1} n^{q-r+k+1} \leq C\frac{\ell^{d-1}}{n^{p-q}}, \qquad \ell > n.$$

16

Thus for $d \leq 0$

$$\sum_{\ell=0}^{\infty} \Delta_{n\ell}^{(r)2} \leq Cn^{2(q-p)} \left\{ \frac{1}{n^2} \sum_{\ell=0}^{n} (\ell+1)^{2d} + \sum_{\ell=n}^{\infty} \ell^{2d-2} \right\} \leq Cn^{2(q-p+d)-1},$$

with the same result for $d > 0$. Thus from (A.4) and $|g(\lambda)| \leq C |\lambda|^{-1}$, the contribution of this term to (B.2) is bounded by $C\lambda^{-2q} n^{2d-1}$, which is bounded by the right sides of both (B.3) and (B.4). Clearly the contributions of the $\Delta_{r\ell}^{(r-1)}$ are of no higher order.

To consider (B.4) for $q > 0$, and correspondingly for $q = 0$ with $\Delta_t^{(0)} = \Delta_t$, note that $\Delta_t^{(q)}$ is a linear combination of terms $\Delta^k h_{t-i} \Delta^{q-k} b_{t-j,\ell}(q,d) = \Delta^k h_{t-i} b_{t-j,\ell}(k,d)$, for $k = 0, ..., q$. For fixed positive, negative or zero i, and n_1, n_2 satisfying $C \geq n_1 \geq \max(1, i+1)$, $n - C \leq n_2 \leq \min(n, n+i)$ consider, for $0 \leq k \leq q$, $r_{k\ell} = \sum_{t=n_1}^{n_2} b_{t\ell}(k,d) \Delta^k h_{t-i} e^{it\lambda}$. We estimate $r_{k\ell}$ by applying, as appropriate, (2.1), (2.2), Lemmas 2, 3 and 5, and summation-by-parts. We find that $|r_{0\ell}|$ is bounded by:

$$\frac{C}{\lambda} \sum_{t=n_1}^{n_2-1} (t+\ell)^{d-2} + \frac{C(n_2+\ell)^{d-1}}{\lambda} \quad \leq \quad \frac{C\ell^{d-1}}{\lambda}, \qquad p \geq 0;$$

$$C \sum_{t=1}^{n} (t+\ell)^{d-1} \quad \leq \quad C(\ell+1)^d, \qquad d < 0, \ \ p = 0;$$

$$C \sum_{t=n_1}^{m-1} (t+\ell)^{d-1} + \frac{C}{\lambda} \sum_{t=m}^{n_2-1} (t+\ell)^{d-2} + \frac{C(n_2+\ell)^{d-1}}{\lambda}$$

$$\leq C(m+\ell)^d + \frac{C(m+\ell)^{d-1}}{\lambda} \leq C\lambda^{-d}, \qquad \ell \leq \lambda^{-1}, \ \ d > 0, \ \ p = 0,$$

(on taking $m \sim 1/\lambda$);

$$\frac{C}{\lambda} \sum_{t=n_1}^{n_2-1} \left\{ |b_{t+1,\ell}(-1,d) h_{t-j} + b_{t+1,\ell}(0,d) |\Delta h_{t+1-j}|| \right\} + \frac{C}{\lambda} b_{n_2,\ell}(0,d) h_{n_2-j}$$

$$\leq \quad \frac{C}{n^p \lambda} \sum_{t=n_1}^{n_2-1} \left\{ (t+\ell)^{d-2} t^p + (t+\ell)^{d-1} t^{p-1} \right\} + \frac{C(n_2+\ell)^{d-1}}{\lambda n^p}$$

$$\leq \quad \frac{C(\ell+1)^d}{n\lambda}, \qquad \ell \leq n, \ \ d \leq 0, \ \ p \geq 1;$$

$$\leq \quad \frac{Cn^{d-1}}{\lambda}, \qquad \ell \leq n, \ \ d > 0, \ \ p \geq 1;$$

17

We find that $|r_{1\ell}|$ is bounded by

$$\frac{C}{n\lambda} \sum_{t=n_1}^{n_2-1} (t+\ell)^{d-1} + \frac{C(\ell+1)^d}{n\lambda} \leq \frac{C(\ell+1)^d}{n\lambda}, \qquad \ell \leq n,\ d < 0,\ p \geq 1.$$

We find that $|r_{k\ell}|$ is bounded by

$$\frac{C\ell^{d-1}}{n^k\lambda} \sum_{t=n_1}^{n_2-1} t^{k-1} + \frac{C\ell^{d-1}}{\lambda} \leq \frac{C\ell^{d-1}}{\lambda}, \qquad \ell > n,\ 1 \leq k \leq p;$$

$$\frac{C\ell^{d-1}}{n^k\lambda} \sum_{t=1}^{\ell} t^{k-1} + \frac{C(\ell+1)^d}{n^k\lambda} \sum_{t=\ell+1}^{n-1} t^{k-2} + \frac{(n+\ell)^d}{n\lambda} \leq \frac{C(\ell+1)^d}{n\lambda}, \ell \leq n,\ d < 0,\ 2 \leq k \leq p;$$

$$\frac{C\ell^{d-1}}{n^k\lambda} \sum_{t=1}^{\ell} t^{k-1} + \frac{C}{n^k\lambda} \sum_{t=\ell+1}^{n-1} t^{k-2+d} + \frac{Cn^{d-1}}{\lambda} \leq \frac{Cn^d}{\lambda}, \qquad \ell \leq n,\ d > 0,\ 1 \leq k \leq p.$$

Thus $\Sigma_{\ell=0}^{\infty} r_{0\ell}^2$ is bounded by

$$C \sum_{\ell \leq 1/\lambda} (\ell+1)^{2d} + \frac{C}{\lambda^2} \sum_{\ell > 1/\lambda} \ell^{2d-2} \leq C\lambda^{-2d-1}, \qquad d < 0,\ p = 0, \qquad \text{(B.5)}$$

$$C \sum_{\ell \leq 1/\lambda} \lambda^{-2d} + \frac{C}{\lambda^2} \sum_{\ell > 1/\lambda} \ell^{2d+1} \leq C\lambda^{-2d-1}, \qquad d > 0,\ p = 0,$$

whilst $\Sigma_{\ell=0}^{\infty} r_{k\ell}^2$ is bounded by

$$\frac{C}{n^2\lambda^2} \sum_{\ell \leq n} (\ell+1)^{2d} + \frac{C}{\lambda^2} \sum_{\ell > n} \ell^{2d-2} \leq Cn^{2d-1}\lambda^{-2}, \quad d < 0,\ 0 \leq k \leq p,\ p \geq 1;\text{(B.6)}$$

$$\frac{C}{n^2\lambda^2} \sum_{\ell \leq n} n^{2d-2} + \frac{C}{\lambda^2} \sum_{\ell > n} \ell^{2d-2} \leq Cn^{2d-1}\lambda^{-2}, \quad d > 0,\ 0 \leq k \leq p,\ p \geq 1.$$

The bounds (B.3) and (B.4) immediately follow. The proof for $d = -bd$ follows much as above, the bounds for the $r_{k\ell}$ still holding and (B.5) and (B.6) becoming $C|\log\lambda|$ and $C(\log n)(n\lambda)^{-2}$, respectively. $\qquad\square$

APPENDIX C: PROOFS OF PROPOSITIONS 1 AND 2

Proof of Proposition 1: By a standard argument for consistency of implicitly-defined extremum estimates, for $\eta > 0$

$$P\left(\left\|\tilde{\theta} - \theta_0\right\| > \eta\right) \leq P\left(\inf_{\bar{\mathcal{N}}} \tilde{R}_n(\theta) \leq 0\right), \qquad \text{(C.1)}$$

where $\bar{\mathcal{N}}$ is the complement in Θ of an η-neighbourhood of θ_0 and $\tilde{R}_n(\theta) = \tilde{Q}_n(\theta) - \tilde{Q}_n(\theta_0)$. As in Robinson (1995b), when $\nabla_1 > \delta_0 - bd$ write $\Theta_1^{(1)} = \{\delta : \nabla_1 \leq \delta \leq \nabla_2\}$ and take $\Theta_2^{(1)}$ to be empty, and otherwise take $\Theta_1^{(1)} = \{\delta : \delta_0 - bd + \varepsilon \leq \delta < \nabla_2\}$, $\Theta_2^{(1)} = \{\delta : \nabla_1 \leq \delta < \delta_0 - bd + \varepsilon\}$, for $\varepsilon \in (0, bd)$, so that in either case $\Theta^{(1)} = \{\delta : \nabla_1 \leq \delta \leq \nabla_2\}$. Writing $\Theta_1 = \Theta_1^{(1)} \times \Theta^{(-)}$, $\Theta_2 = \Theta_2^{(1)} \times \Theta^{(-)}$ we thus bound (C.1) by

$$P\left(\inf_{\bar{\mathcal{N}} \cap \Theta_1} \tilde{R}_n(\theta) \leq 0\right) + P\left(\inf_{\Theta_2} \tilde{R}_n(\theta) \leq 0\right), \tag{C.2}$$

noting that $\theta_0 \in \Theta_1$. Writing $R_n(\theta) = Q_n(\theta) - Q_n(\theta_0)$ as $S(\theta) - U(\theta)$, with the definitions in the proof of Theorem 1 of VR, and defining also

$$a_n = \inf_{\bar{\mathcal{N}} \cap \Theta_1} R_n(\theta) - \inf_{\bar{\mathcal{N}} \cap \Theta_1} \tilde{R}_n(\theta), \qquad b_n = \sup_{\Theta_1} \left|Q_n(\theta) - \tilde{Q}_n(\theta)\right|,$$

we bound the first term of (C.2) by

$$P\left(\sup_{\Theta_1} |U(\theta)| + a_n \geq \inf_{\bar{\mathcal{N}} \cap \Theta_1} S(\theta)\right). \tag{C.3}$$

We have $a_n \leq 2b_n$, on taking infima over $\bar{\mathcal{N}} \cap \Theta_1$ on both sides of the inequality $Q_n(\theta) \leq \tilde{Q}_n(\theta) + b_n$, which holds for all $\theta \in \Theta_1$. On the other hand, VR show in the "First Step" of the proof of their Theorem 1 that $\sup_{\Theta_1} |U(\theta)| \to_p 0$ and $\inf_{\bar{\mathcal{N}} \cap \Theta_1} S(\theta) > 0$. It follows that (C.3) is $o(1)$ if b_n is $o_p(1)$. From the triangle inequality,

$$|I_Y(\lambda) - I_X(\lambda)| \leq |u(\lambda)|^2 + 2|u(\lambda)||w_X(\lambda)|, \tag{C.4}$$

where $u(\lambda) = w_Y(\lambda) - w_X(\lambda)$. Applying also the Cauchy inequality,

$$b_n \leq \sup_{\Theta} P_n(\theta) + 2\left\{\sup_{\Theta} P_n(\theta) \sup_{\Theta} Q_n(\theta)\right\}^{bd},$$

where

$$P_n(\theta) = \frac{2\pi(p+1)}{n} \sum_{j(p)} \frac{|u(\lambda_j)|^2}{k(\lambda_j; \theta)}.$$

Now

$$\sup_{\Theta} P_n(\theta) \leq \frac{C}{n} \sum_{j(p)} \left| \lambda_j^{\delta_0} u(\lambda_j) \right|^2 \lambda_j^{2\varepsilon - 1}.$$

From the Theorem, the right hand side has expectation bounded by

$$\frac{C}{n} \sum_{j=1}^{n} (\log j) j^{-1} \left(\frac{j}{n} \right)^{2\varepsilon - 1} \leq C n^{-2\varepsilon} \sum_{j=1}^{\infty} (\log j) j^{2\varepsilon - 2} \to 0,$$

as $n \to \infty$. The "First Step" of the proof of Theorem 1 of VR shows that $Q_n(\theta)$ converges uniformly on Θ_1 to a bounded limit. Thus $b_n = o_p(1)$. To deal with the second term in (C.2), the triangle inequality gives

$$
\begin{aligned}
\inf_{\Theta_2} \tilde{Q}_n(\theta) \;\; &\geq \;\; \frac{1}{n} \sum_{j(p)} \tilde{I}(\lambda_j) \lambda_j^{2(\delta_0 + \varepsilon - bd)} \\
&\geq \;\; \frac{1}{n} \sum_{j(p)} I(\lambda_j) \lambda_j^{2(\delta_0 + \varepsilon - bd)} \quad\quad\quad\quad\quad\quad\text{(C.5)} \\
&\quad - \left| \frac{1}{n} \sum_{j(p)} \left\{ \tilde{I}(\lambda_j) - I(\lambda_j) \right\} \lambda_j^{2(\delta_0 + \varepsilon - bd)} \right|. \quad\quad\text{(C.6)}
\end{aligned}
$$

The "Second Step" of the proof of Theorem 1 of VR shows that (C.5) tends in probability to a limit that can be made arbitrarily large on letting $\varepsilon \to 0$, whereas the above proof has shown that (C.6) is $o_p(1)$. Since $\theta_0 \in \Theta_1$ it follows from the above proof and the "First Step" of the proof of Theorem 1 of VR that $\tilde{Q}_n(\theta_0)$ has a finite probability limit. Thus the second probability in (C.2) is $o(1)$ as $n \to \infty$. $\square$

Proof of Proposition 2: By the usual mean value theorem argument, $n^{bd}(\tilde{\theta} - \theta_0) = \tilde{\Xi}_n^{-1} \tilde{\xi}_n$, where

$$\tilde{\xi}_n = n^{-bd} \sum_{j(p)} \frac{\partial}{\partial \theta} k(\lambda_j; \theta_0)^{-1} \tilde{I}(\lambda_j),$$

and $\tilde{\Xi}_n$ is the second derivative matrix of $\tilde{Q}_n(\theta)$ with each row evaluated at some point on the line segment joining θ_0 and $\tilde{\theta}$. Denote by Ξ_n the matrix obtained by replacing

the $I_Y(\lambda_j)$ by the $I_X(\lambda_j)$ in $\tilde{\tilde{\Xi}}_n$. In view of Proposition 1 and its proof, $\tilde{\tilde{\Xi}}_n - \Xi_n = o_p(1)$, where, by Proposition 1 and Lemma A.7 of VR, Ξ_n converges in probability to a positive definite matrix. The result follows from Theorem 2 of VR if $\tilde{\xi}_n - \xi_n \to_p 0$, where ξ_n is defined like $\tilde{\xi}_n$ with I replacing $\tilde{I}$. Since $(\partial/\partial\theta)k(\lambda;\theta_0)^{-1} = O(\lambda^{2\delta_0-\eta})$ for arbitrary $\eta \in (0, bd)$, applying (C.4) again,

$$E\left\|\tilde{\xi}_n - \xi_n\right\| \le Cn^{-bd}\sum_{j(p)}\lambda_j^{2\delta_0-\eta}\left[E\left|u(\lambda_j)\right|^2 + \left\{E\left|u(\lambda_j)\right|^2 EI(\lambda_j)\right\}^{bd}\right]. \tag{C.7}$$

From Theorem A.3 of VR, $\lambda_j^{2\delta_0}EI(\lambda_j) \le C$. Thus, applying the Theorem, (C.7) is bounded by

$$Cn^{\eta-bd}\left\{\log n\sum_{j(p)}j^{2d-2-\eta} + (\log n)^{bd}\sum_{j(p)}j^{d-1-\eta}\right\}.$$

Choosing $\eta \in (\max(0,d), bd)$, this is $O((\log n)^{bd}n^{\eta-bd}) \to 0$ as $n \to \infty$, to complete the proof. $\square$

APPENDIX D: PROOFS OF PROPOSITIONS 3 AND 4

Proof of Proposition 3: An analogous argument to that in the proof of Proposition 1 indicates that it suffices to show that

$$\bar{Q}_n(\theta) - \tilde{Q}_n(\theta) \to_p 0 \tag{D.1}$$

uniformly in Θ_1, as we can then infer from that proof that $\inf_{\Theta_2} \bar{Q}(\theta)$, with probability approaching 1, exceeds an arbitrarily large quantity. The left side of (D.1) is bounded in absolute value by

$$\frac{C\left|\tilde{\nu} - \nu\right|}{n}\left\{\left|\sum_{j(p)}\frac{w_x(\lambda_j)w_Y(-\lambda_j)}{k(\lambda_j;\theta)}\right|\right\} + \frac{C(\tilde{\nu}-\nu)^2}{n}\left\{\sum_{j(p)}\frac{\left|w_x(\lambda_j)\right|^2}{k(\lambda_j;\theta)}\right\}, \tag{D.2}$$

where $w_x(\lambda) = (\sum_{t=1}^n h_t^2)^{-bd}\sum_{t=1}^n h_t x_t e^{it\lambda}$. Now if x_t and y_t were Type I processes we would deduce $E\left|w_x(\lambda_j)\right|^2 \le C\lambda_j^{-2\gamma_0}$, $E\left|w_Y(\lambda_j)\right|^2 \le C\lambda_j^{-2\delta_0}$. In view of (C.4) and the

Theorem it is readily seen that these bounds apply to our Type II processes. Thus the suprema over Θ_1 of the expressions in braces in (D.2) have, for small enough ε, expectations bounded by

$$C \sum_{j(p)} \lambda_j^{2\delta - \delta_0 - \gamma_0} \;\leq\; C \sum_{j=1}^{n} \lambda_j^{\delta_0 - \gamma_0 - 1 + 2\varepsilon} \leq C n^{\gamma_0 - \delta_0 + 1 - 2\varepsilon},$$

$$C \sum_{j(p)} \lambda_j^{2(\delta - \gamma_0)} \;\leq\; C \sum_{j=1}^{n} \lambda_j^{2(\delta_0 - \gamma_0) - 1 + 2\varepsilon} \leq C n^{2(\gamma_0 - \delta_0) + 1 - 2\varepsilon},$$

respectively. The proof is completed by applying (4.3). $\qquad\square$

Proof of Proposition 4: For reasons given in the proof of Proposition 2, we need discuss only the proof of $\tilde{\xi}_n - \hat{\xi}_n \to_p 0$, with $\hat{\xi}_n$ defined like $\tilde{\xi}_n$ with $\tilde{I}$ replaced by $\hat{I}$. In view of the proofs of Propositions 2 and 3,

$$\left\| \tilde{\xi}_n - \hat{\xi}_n \right\| \;\leq\; \frac{C\,|\tilde{\nu} - \nu|}{n^{bd}} \left\{ \sum_{j(p)} \lambda_j^{2\delta_0 - \eta} \left| w_x(\lambda_j) w_Y(-\lambda_j) \right| \right\} \tag{D.3}$$

$$+ \frac{C(\tilde{\nu} - \nu)^2}{n^{bd}} \left\{ \sum_{j(p)} \lambda_j^{2\delta_0 - \eta} \left| w_x(\lambda_j) \right|^2 \right\}, \tag{D.4}$$

for any $\eta \in (0, bd)$. The expressions in braces have expectations bounded by

$$C \sum_{j(p)} \lambda_j^{\delta_0 - \gamma_0 - \eta} \leq C n^{\gamma_0 - \delta_0 + \eta} \sum_{j=1}^{n} j^{\delta_0 - \gamma_0 - \eta}, \tag{D.5}$$

$$C \sum_{j(p)} \lambda_j^{2(\delta_0 - \gamma_0) - \eta} \leq C n^{2(\gamma_0 - \delta_0) + \eta} \sum_{j=1}^{n} j^{2(\delta_0 - \gamma_0) - \eta}, \tag{D.6}$$

respectively. In view of (4.2), (D.6) is $O\left(n^{2(\gamma_0 - \delta_0) + \eta}\right)$ so that (D.4) is $O_p\left(n^{\eta + 2\rho - bd}\right) = o_p(1)$. Because (D.5) is $O(n)$ for $\delta_0 - \gamma_0 - \eta > -1$, $O(n \log n)$ for $\delta_0 - \gamma_0 - \eta = -1$ and $O\left(n^{\gamma_0 - \delta_0 + \eta}\right)$ for $\delta_0 - \gamma_0 - \eta < -1$, it follows that in these three cases the left side of (D.3) is respectively $O_p\left(n^{bd + \delta_0 - \gamma_0 + \rho}\right)$, $O_p(n^{\eta + \rho - bd} \log n)$ and $O_p\left(n^{\eta + \rho - bd}\right)$, all of which are $o_p(1)$. $\qquad\square$

REFERENCES

[1] Akonom, J., Gourieroux, C., 1987. A functional limit theorem for fractional processes, Preprint.

[2] Chan, N.H., Terrin, N., 1995. Inference for unstable long-memory processes with applications to fractional unit root autoregression, Annals of Statistics 23, 1662-1683.

[3] Davydov, Y.A., 1970. The invariance principle for stationary processes, Theory of Probability and its Applications 15, 487-498.

[4] Gorodetskii, V.V., 1977. On convergence to semi-stable Gaussian processes, Theory of Probability and its Applications 22, 498-508.

[5] Hurvich, C.M., Ray, B.K., 1995. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes, Journal of Time Series Analysis 16, 17-42.

[6] Hualde, J., Robinson, P.M., 2001. Root-n-consistent estimation of weak fractional cointegration, Preprint.

[7] Jeganathan, P., 1999. On asymptotic inference in cointegrated time series with fractionally integrated errors, Econometric Theory 15, 583-621.

[8] Kim, C.S., Phillips, P.C.B., 2000. Modified log periodogram regression, Preprint.

[9] Mandelbrot, B.B., Van Ness, J.W., 1968. Fractional Brownian motion, fractional noises and applications, SIAM Review 10, 422-437.

[10] Marinucci, D., 2000. Spectral regression for cointegrated time series with long-memory innovations. Journal of Time Series Analysis 21, 685-705.

[11] Marinucci, D., Robinson, P.M., 1999. Alternative forms of fractional Brownian motion, Journal of Statistical Planning and Inference 80, 111-122.

[12] Marinucci, D., Robinson, P.M., 2000. Weak convergence of multivariate fractional processes, Stochastic Processes and their Applications 86, 103-120.

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

[14] Robinson, P.M., 1995a. Log-periodogram regression of time series with long range dependence, Annals of Statistics 23, 1048-1072.

[15] Robinson, P.M., 1995b. Gaussian semiparametric estimation of long-range dependence, Annals of Statistics 23, 1630-1661.

[16] Robinson, P.M., Hualde, J., 2000. Cointegration in fractional systems with unknown integration orders, Econometrica 71, 1727-1766.

[17] Robinson, P.M., Marinucci, D., 2001. Narrow-band analysis of nonstationary processes, Annals of Statistics 29, 947-986.

[18] Silveira, G., 1991. Contributions to Strong Approximations in Time Series with Applications in Nonparametric Statistics and Functional Limit Theorems, Ph.D. thesis, University of London.

[19] Sowell, F.B., 1990. The fractional unit root distribution, Econometrica 58, 495-404.

[20] Velasco, C., 1999a. Non-stationary log-periodogram regression, Journal of Econometrics 91, 325-371.

[21] Velasco, C., 1999b. Gaussian semiparametric estimation for non-stationary time series. Journal of Time Series Analysis 20, 87-127.

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

[23] Zhurbenko, I.G., 1986. The Spectral Analysis of Time Series (North-Holland, Amsterdam).

[24] Zygmund, A., 1977. Trigonometric Series (Cambridge University Press, Cambridge).