Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

Working Paper nº 07/02

Cointegration in Fractional Systems with Unknown Integration Orders

Peter M. Robinson
Javier Hualde

Cointegration in Fractional Systems with Unknown Integration Orders
P.M. Robinson and J. Hualde
Working Paper No.07/02
November 2002
JEL Codes: C22

ABSTRACT

Cointegration of nonstationary time series is considered in a fractional context. Both the observable series and the cointegrating error can be fractional processes. The familiar situation in which the respective integration orders are 1 and 0 is nested, but these values have typically been assumed known. We allow one or more of them to be unknown real values, in which case Robinson and Marinucci (1997,2001) have justified least squares estimates of the cointegrating vector, as well as narrow-band frequency-domain estimates, which may be less biased. While consistent, these estimates do not always have optimal convergence rates, and they have non-standard limit distributional behaviour. We consider estimates formulated in the frequency domain, that consequently allow for a wide variety of (parametric) autocorrelation in the short memory input series, as well as time-domain estimates based on autoregressive transformation. Both can be interpreted as approximating generalized least squares and Gaussian maximum likelihood estimates. The estimates share the same limiting distribution, having mixed normal asymptotics (yielding Wald test statistics with χ^2 null limit distributions), irrespective of whether the integration orders are known or unknown, subject in the latter case to their estimation with adequate rates of convergence. The parameters describing the short memory stationary input series are $\sqrt{n}$-consistently estimable, but the assumptions imposed on these series are much more general than ones of autoregressive moving average type. A Monte Carlo study of finite-sample performance and an empirical application to testing the PPP hypothesis are included.

Keywords: Fractional cointegration; Unknown integration orders; System estimates; Mixed normal asymptotics.

Peter M. Robinson
London School of Economics
Department of Economics
Houghton Street
London WC2A 2AE, UK
p.m.robinson@lse.ac.uk

Javier Hualde
Universidad de Navarra
Facultad de Económicas
Campus Universitario
31080 Pamplona, Spain
jhualde@unav.es

ACKNOWLEDGMENTS

1. INTRODUCTION

Cointegration analysis has developed almost exclusively in the context of processes with non-fractional integration orders. Most popularly, observed series are assumed to have a single unit root, such that first differencing produces a weakly dependent, invertible stationary process, while cointegrating errors also satisfy the latter description. This basic setting has been greatly extended, to observed series in which twice differencing is required to produce stationary weak dependence, and to polynomial cointegration; polynomial time trends have also been introduced, and cointegration with respect to cyclic and seasonal frequencies has been examined. Methods of estimating cointegrating vectors have been developed which have optimal asymptotic properties, with a limiting mixed normal distribution, thereby generating Wald test statistics with a standard, χ^2, null limit distribution (see e.g. Phillips and Hansen, 1990, Phillips, 1991a,b, Johansen, 1991). The latter methods have been justified under the assumption that integration orders of observed series and cointegrating errors are correctly specified integers, though it is standard practice to test these integration orders, particularly by unit root tests against stationary autoregressive (AR) alternatives.

Cointegration can exist between much more general nonstationary (and indeed stationary) observations, with stationary or nonstationary cointegrating errors. The "optimal" methods referred to above lose their most desirable properties (such as the χ^2 hypothesis tests, for example) when integration orders on which they are based are misspecified, while methodology developed by Engle and Granger (1987) and subsequent authors is not designed to detect such cointegrating relationships. It is thus desirable to develop the topic in a broader context, nesting integer-order cases in a more general class and allowing integration orders to be unknown, and real-valued.

Recently, considerable interest in fractional processes has developed, and knowledge of their properties and statistical analysis has advanced to the extent that their role in cointegration analysis can be explored. We consider the following model for the bivariate observed series (y_t, x_t):

$$
\begin{aligned}
y_t &= \nu x_t + \Delta^{\beta-\delta} u_{1t}^{\#}, &\qquad (1.1)\\
x_t &= \Delta^{-\delta} u_{2t}^{\#}, &\qquad (1.2)
\end{aligned}
$$

for $t = 0, \pm1, \dots$. Throughout, the $\#$ superscript attached to a scalar or vector sequence v_t has the meaning

$$
v_t^{\#} = v_t 1(t > 0), \qquad (1.3)
$$

where $1(\cdot)$ is the indicator function. In (1.1), (1.2) we employ the difference operator $\Delta = 1 - L$, where L is the lag operator, and formally, for any real α, $\alpha \neq -1, -2, ..$,

$$
(1-z)^{-\alpha} = \sum_{j=0}^{\infty} a_j(\alpha) z^j, \quad a_j(\alpha) = \frac{\Gamma(j+\alpha)}{\Gamma(\alpha)\Gamma(j+1)}, \qquad (1.4)
$$

2

with Γ denoting the gamma function; with the prime denoting transposition, $u_t = (u_{1t}, u_{2t})'$ is a bivariate covariance stationary unobservable process with zero mean and spectral density matrix, $f(\lambda)$, satisfying

$$E(u_0 u_j') = \int_{-\pi}^{\pi} e^{ij\lambda} f(\lambda) d\lambda, \qquad (1.5)$$

that is at least nonsingular and continuous at all frequencies; and finally

$$\nu \quad \neq \quad 0, \qquad (1.6)$$

$$\delta \quad \geq \quad \beta > \frac{1}{2}, \qquad (1.7)$$

The truncation in (1.2) ensures that x_t has finite variance (albeit evolving at rate $t^{2\delta-1}$), and implies that $x_t = 0$, $t \leq 0$. The truncation in (1.1) is unnecessary if $\delta - \beta < 1/2$ ($y_t - \nu x_t$ is covariance stationary without it and "asymptotically covariance stationary" with it) but is imposed there also for the sake of a uniform treatment, implying that $y_t = 0$, $t \leq 0$. In common parlance, u_t is an $I(0)$ vector process, x_t is an $I(\delta)$ process, as is (due to (1.1), (1.2), (1.6), (1.7)) y_t, while the cointegrating error $y_t - \nu x_t$ is an $I(\gamma)$ process, where

$$\gamma = \delta - \beta, \qquad (1.8)$$

and we say that (x_t, y_t) is cointegrated of order (δ, β) $(CI(\delta, \beta))$. If $\beta = 0$, there is no cointegration and ν is not identified.

In (1.1), (1.2) the possibility that δ and/or β are known, but not necessarily integers, does not lack interest (in particular when $\delta = 1$ is fixed after pre-testing) but allowing both β and δ to be unknown, thereby avoiding complications and ambiguities due to pre-testing, may be attractive. Fractional values may be difficult to interpret economically, though aggregation explanations have been developed, mean-reversion is nicely described, in the present paper's context β and δ are just nuisance parameters, while fractional, like non-fractional, cointegration is a kind of dimensionality-reducing structure.

Simple estimates of ν not requiring knowledge of δ and/or β are readily available. For example ordinary least squares (OLS), with or without intercept, is $n^{\min(2\delta-1,\beta)_-}$ consistent (except in the case where $\delta > \beta$ and $2\delta - \beta = 1$, in which case it is $(n^\beta / \log n)$−consistent), as shown under mild conditions by Robinson and Marinucci (2001). In case $2\delta - 1 < \beta$, the rate of convergence can be improved upon by using a version of OLS in the frequency domain that focuses on a slowly degenerating band of low frequencies and thereby reduces the bias that is due to contemporaneous correlation between u_{1t}, u_{2t} (Robinson and Marinucci, 1997); these estimates were applied empirically by Marinucci and Robinson (2001). Both least squares and its narrowband counterpart have nonstandard limit distributions, which are unsuitable for use in statistical inference, while their rate of convergence seems capable of still further improvement over some regions of (δ, β)-space. In the present paper we develop and justify estimates of ν which have analogously optimal properties, in the presence of possibly unknown δ, β, to those previously established by, for example, Phillips and

Hansen (1990), Phillips (1991a,b), Johansen (1991) in case $\delta = \beta = 1$ is known. The estimates of ν are of generalized least squares (GLS) type, based on a constrained transformed bivariate regression model derived from (1.1), (1.2) and having the property that regressors are orthogonal to disturbances. We allow for very general forms of parametric autocorrelation in u_t, in which circumstances a frequency-domain form of estimate of ν is convenient and flexible, though we also consider a time-domain form based on autoregressive (AR) transformation.

The model (1.1), (1.2) is perhaps the simplest interesting one possible. Extensions are described in Section 6, but our treatment of (1.1), (1.2), with parametric autocorrelation, itself requires lengthy proofs, whose ideas are relevant to more general models but best conveyed in a relatively simple setting. Our model presumes the existence of cointegration. The question of establishing such existence, or non-existence, is itself especially difficult in our fractional context, with unknown integration orders. Recently, Robinson and Yajima (2001) have developed methods for determining fractional cointegrating rank in a multivariate extension of (1.1), (1.2) based on sequential testing, principal components analysis, and a model choice procedure, while Marinucci and Robinson (2001) proposed and empirically applied a Hausman-type test for determining the existence of cointegration in (1.1), (1.2).

Aside from work already mentioned pertaining to (1,1), (1.2), Dolado and Marmol (1996) considered fractional extension of the fully-modified OLS (FM-OLS) estimate of ν proposed by Phillips and Hansen (1990) in the $CI(1,1)$ case, but with non-parametric autocorrelation in u_t, and assuming knowledge of γ and δ. Some other work on fractional cointegration has employed an alternative definition of fractional nonstationarity, replacing (1.1), (1.2) by

$$\tilde{y}_t = \nu \tilde{x}_t + v_{1t}^{(\gamma)}, \qquad t \geq 1 \tag{1.9}$$

$$\tilde{x}_t = v_{21}^{(\delta)} + ... + v_{2t}^{(\delta)}, \qquad t \geq 1, \tag{1.10}$$

where $v_{1t}^{(\gamma)}$ and $v_{2t}^{(\delta)}$ are jointly stationary $I(\gamma)$ and $I(\delta - 1)$ processes, respectively, with $|\gamma| < 1/2$, $1/2 < \delta < \frac{3}{2}$. When $\gamma = 0$, $\delta = 1$, $v_t(\gamma, \delta) = \left(v_{1t}^{(\gamma)}, v_{2t}^{(\delta)} \right)' \equiv (u_{1t}, u_{2t})'$ implies $(\tilde{x}_t, \tilde{y}_t) \equiv (x_t, y_t)$, but more generally, with $v_t(\gamma, \delta)$ having spectral density matrix $\Lambda(\lambda; \gamma, \delta) f(\lambda) \Lambda(-\lambda; \gamma, \delta)$, for $\Lambda(\lambda; \gamma, \delta) = diag \left\{ (1 - e^{i\lambda})^{-\gamma}, (1 - e^{i\lambda})^{1-\delta} \right\}$, this is not the case. In particular, (1.10) provides an alternative definition of non-stationary $I(\delta)$ processes to (1.2). Marinucci and Robinson (1999) termed $\tilde{x}_t$ and x_t respectively Type I and Type II $I(\delta)$ processes; they converge weakly as $t \to \infty$ to different forms of fractional Brownian motion. Model (1.9), (1.10) covers a different range of γ, δ values from (1.1), (1.2), but higher δ can be involved by extending (1.10) to include two or more unit roots, while $\gamma \in (-1/2, 0)$ could be allowed in (1.1). Chan and Terrin (1995) developed asymptotic theory for OLS estimates in a general AR process with fractional innovations, including (1.10). Jeganathan (1999, 2001) considered ML estimation in (1.9), (1.10), stressing pure fractional $v_t(\gamma, \delta)$ (corresponding to white noise u_t in (1.1), (1.2)), having innovations with completely known, but not necessarily Gaussian, distribution. He obtained mixed normal asymptotics for his estimate of ν, in case γ and δ are known, though including some discussion of their estimation. Again in relation to (1.9), (1.10), with ν a matrix and both equations

4

vectors but depending still on only two integration orders γ and δ, Kim and Phillips (2000) consider an alternative extension of FM-OLS to that of Dolado and Marmol (1996), and its relation to Gaussian maximum likelihood (ML) estimation. They assume parametric autocorrelation in $v_t(\gamma, \delta)$, obtaining limit distribution theory that differs from that of Jeganathan (1999, 2001), and from ours, even after replacing their version of fractional Brownian motion by ours. They also consider estimation of nuisance parameters, and nonstationary $v_{1t}(\gamma)$. In a multivariate semiparametric version of (1.9), (1.10), and allowing also for the possibility of nonstationary $v_{1t}(\gamma)$, Velasco (2000) considered a tapered version of local Whittle estimation of ν, γ and δ, more particularly taking one Newton step from preliminary estimates with suitable convergence rates. This produces an estimate of ν which does not have optimal convergence rate but, unlike ours and those in the other references, is asymptotically normal. In a similar setting, Hassler, Marmol and Velasco(2002) focus on log periodogram estimation of γ and δ given preliminary estimation of ν, developing rules of asymptotic inference.

Our estimates of ν are described in the following section. Section 3 presents regularity conditions and the main results, also introducing simpler estimates that are asymptotically competitive when $\beta > 1$. In Appendix A we outline the proofs, which rest heavily on a series of propositions which are proved in Appendix B. Appendices C and D collect respectively some results used in the proofs of several propositions, and technical lemmas pertaining to properties of the $a_j(\alpha)$. Section 4 consists of a Monte Carlo study of finite-sample behaviour, Section 5 reports an empirical investigation of the purchasing power parity (PPP) hypothesis, and Section 6 discusses related topics.

2. ESTIMATES OF COINTEGRATING PARAMETERS

For any sequence $\{w_t\}$, and any $c \geq 0$, introduce the notation

$$w_t(c) = \Delta^c w_t^{\#}, \tag{2.1}$$

noting (1.3), (1.4). Also define, for $c \geq 0$, $d \geq 0$,

$$z_t(c, d) = (y_t(c), x_t(d))'. \tag{2.2}$$

Thus (1.1), (1.2) can be written

$$z_t(\gamma, \delta) = \zeta x_t(\gamma)\nu + u_t^{\#}, \tag{2.3}$$

where

$$\zeta = (1, 0)'. \tag{2.4}$$

In case u_t is white noise, with known, nonsingular covariance matrix Ω, and γ and δ are also known, GLS based on (2.3) and observations (x_t, y_t), $t = 1, ..., n$, is motivated by the orthogonality property $E\left(u_t' \Omega^{-1} \zeta x_t(\gamma)\right) = E\left(u_{2t} u_t'\right) \Omega^{-1} \zeta = 0$. More generally, GLS estimates can also be constructed in the presence of serial correlation in u_t, given

5

known $2n \times 2n$ covariance matrix Σ of $u = (u_1', ..., u_n')'$. If Σ is a known function of an unknown finite-dimensional parameter vector θ, we might hope that insertion of sufficiently good estimates of γ, δ and θ, producing a feasible GLS estimate of ν, will not affect limiting distributional properties. However, Σ and its estimate can be difficult to handle, both numerically and theoretically, so more convenient alternatives to such GLS or feasible GLS might be considered.

One such is based on AR transformation. Suppose u_t has an AR representation

$$B(L)u_t = \varepsilon_t, \tag{2.5}$$

where ε_t is a bivariate sequence that is at least (see Section 3 below) uncorrelated across t with nonsingular covariance matrix Ω, and

$$B(s) = I_2 - \sum_{j=1}^{\infty} B_j s^j, \tag{2.6}$$

where I_2 is the 2×2 identity matrix and the B_j are 2×2 matrices satisfying conditions prescribed below. Suppose further that we know functions $\Omega(h)$, $B_j(h)$, where $h \in \mathsf{R}^p$, $p \geq 1$, such that for some $\theta \in \mathsf{R}^p$, we have $\Omega = \Omega(\theta)$, $B_j \equiv B_j(\theta)$. Define

$$B(s; h) = I_2 - \sum_{j=1}^{\infty} B_j(h)s^j, \tag{2.7}$$

and then

$$\widetilde{a}(c, d, h) = \sum_t \left\{ B(L; h)\zeta x_t(c) \right\}' \Omega(h)^{-1} \left\{ B(L; h)z_t(c, d) \right\}, \tag{2.8}$$

$$\widetilde{b}(c, h) = \sum_t \left\{ B(L; h)\zeta x_t(c) \right\}' \Omega(h)^{-1} \left\{ B(L; h)\zeta x_t(c) \right\}, \tag{2.9}$$

where, throughout the paper, $\sum_t$ denotes $\sum_{t=1}^n$. Note that each of the AR transformations automatically entails a truncation since $x_t(c) = 0$, $z_t(c, d) = 0$, $t \leq 0$. Now write

$$\widetilde{\nu}(c, d, h) = \frac{\widetilde{a}(c, d, h)}{\widetilde{b}(c, h)}, \tag{2.10}$$

and consider as estimates of ν

$$\widetilde{\nu}(\gamma, \delta, \theta), \ \widetilde{\nu}(\gamma, \delta, \widehat{\theta}), \ \widetilde{\nu}(\widehat{\gamma}, \delta, \widehat{\theta}), \ \widetilde{\nu}(\gamma, \widehat{\delta}, \widehat{\theta}), \ \widetilde{\nu}(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}), \tag{2.11}$$

given estimates $\widehat{\gamma}$, $\widehat{\delta}$, $\widehat{\theta}$. The estimates (2.11) respectively consider the cases in which γ, δ and θ are all known, the integration orders γ and δ are known but θ is not, followed by the cases in which one or other and then both of γ, δ are unknown and θ is also unknown: $\widetilde{\nu}(\gamma, \delta, \widehat{\theta})$ covers situations familiar from the integer integration order cointegration literature, where for example $\gamma = 0$, $\delta = 1$ is known; $\widetilde{\nu}(\widehat{\gamma}, \delta, \widehat{\theta})$ extends this by assuming knowledge of the integration order of the observable x_t (say $\delta = 1$), but the order of the cointegrating error is not known to be 0; $\widetilde{\nu}(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta})$ expresses the situation of least knowledge.

6

The estimates (2.11) are computationally convenient when u_t is a finite-degree AR process, but less so otherwise, for example when u_t is a finite-degree moving average (MA) or autoregressive moving average (ARMA) sequence, when the $B_j(h)$, though recursively calculable, do not have a very neat closed form. On the other hand, the spectral density matrix $f(\lambda)$, defined in (1.5) has a neat form in such cases, so a frequency-domain approach might be preferred, as was considered by Phillips (1991a) in the case $\gamma = 0$, $\delta = 1$ is known, and one can construct parametric models for which the gap between tractability of the spectral density on the one hand, and AR coefficients (or indeed autocovariances) on the other, is even greater (see e.g. Bloomfield, 1972, Robinson, 1978). A frequency-domain approach also has the advantage of approaching a well-established form of semiparametric estimate in which $f(\lambda)$ is a nonparametric function (see, e.g. Hannan, 1963, in case of regression models, and Phillips, 1991b, in case of $CI(1,1)$ cointegration).

To define the frequency-domain estimates, first introduce $f(\lambda; h)$, a known function of $\lambda \in (-\pi, \pi]$ and $h \in \mathsf{R}^p$, such that $f(\lambda; \theta) = f(\lambda)$, see (1.5). In terms of the AR representation (2.5), we have

$$f(\lambda; h) = (2\pi)^{-1} B(e^{i\lambda}; h)^{-1} \Omega(h) B(e^{-i\lambda}; h)^{-1'}. \tag{2.12}$$

so $f(\lambda; h)$ is of simple form in the finite ARMA models, replacing $B(e^{i\lambda}; h)^{-1}$ by $B(e^{i\lambda}; h)^{-1} A(e^{i\lambda}; h)$, A and this B both being finite-degree matrix polynomials. (Our assumptions below guarantee the existence where necessary of matrix inverses.) Define the discrete Fourier transforms

$$w_{x(c)}(\lambda) = \frac{1}{(2\pi n)^{\frac{1}{2}}} \sum_t x_t(c) e^{it\lambda}, \quad w_{z(c,d)}(\lambda) = \frac{1}{(2\pi n)^{\frac{1}{2}}} \sum_t z_t(c,d) e^{it\lambda}. \tag{2.13}$$

Denoting

$$p(\lambda; h) = \zeta' f(\lambda; h)^{-1}, \quad q(\lambda; h) = \zeta' f(\lambda; h)^{-1} \zeta, \tag{2.14}$$

put

$$a(c, d, h) = \sum_j p(\lambda_j; h) w_{x(c)}(-\lambda_j) w_{z(c,d)}(\lambda_j), \tag{2.15}$$

$$b(c, h) = \sum_j q(\lambda_j; h) \left| w_{x(c)}(\lambda_j) \right|^2. \tag{2.16}$$

Define

$$\widehat{\nu}(c, d, h) = \frac{a(c, d, h)}{b(c, h)}. \tag{2.17}$$

Corresponding to the five estimates (2.11) we may consider also

$$\widehat{\nu}(\gamma, \delta, \theta), \ \widehat{\nu}(\gamma, \delta, \widehat{\theta}), \ \widehat{\nu}(\widehat{\gamma}, \delta, \widehat{\theta}), \ \widehat{\nu}(\gamma, \widehat{\delta}, \widehat{\theta}), \ \widehat{\nu}(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}). \tag{2.18}$$

From the orthogonality properties of the complex exponential function (see (B.3) below), it readily follows that when u_t is *a priori* white noise, so that $B_j(h) \equiv 0$, $j \geq 1$,

7

$f(\lambda; h) = (2\pi)^{-1}\Omega(h)$, we have $\widetilde{\nu}(c, d, h) \equiv \widehat{\nu}(c, d, h)$, so corresponding members of (2.11) and (2.18) are identical. Otherwise, when u_t is believed to be autocorrelated, they differ, but under regularity conditions all members of (2.11) and (2.18) have the same first-order asymptotic properties, as shown in Theorem 1 of the following section..

The $CI(1, 1)$ literature has stressed error-correction model (ECM) formulations, on which parameter estimation can be based. We can rewrite (2.3) as

$$\Delta^\delta z_t = -\zeta(1 - \Delta^\beta)\left\{\Delta^{\delta-\beta}(1, -\nu)z_t\right\} + v_t^\#, \tag{2.19}$$

with $z_t = z_t(0,0) = (y_t, x_t)'$ and $v_t^\# = \left(u_{1t}^\# + \nu u_{2t}^\#, u_{2t}^\#\right)'$. When $\delta = \beta = 1$, (2.19) reduces to the triangular ECM representation of Phillips (1991a) for the $CI(1, 1)$ case, on which he based a frequency-domain approximate Gaussian pseudo-ML estimate of ν. It is readily shown that this is equivalent to a corresponding Gaussian pseudo-ML estimate based on (2.3). In case u_t is known to be white noise, this is equivalent to the OLS estimate of ν in the extended regression $y_t(\gamma) = \nu x_t(\gamma) + \tau x_t(\delta) + w_t^\#$, where $\tau = E(u_{1t}u_{2t})/E(u_{2t}^2)$ and $w_t^\# = u_{1t}^\# - \tau u_{2t}^\#$, namely $\overline{\nu}(\gamma, \delta)$, where

$$\overline{\nu}(c, d) = \frac{\sum_t x_t^2(d)\sum_t x_t(c)y_t(c) - \sum_t x_t(c)x_t(d)\sum_t x_t(d)y_t(c)}{\sum_t x_t^2(c)\sum_t x_t^2(d) - \left\{\sum_t x_t(c)x_t(d)\right\}^2}, \tag{2.20}$$

to extend Phillips' (1991a) observation in the $CI(1, 1)$ case (though he derived from his ECM representation the OLS estimate of ν in $y_t(\gamma) = \nu\left\{x_t(\gamma) - x_t(\delta)\right\} + \chi x_t(\delta) + w_t^\#$, with $\chi = \nu + \tau$, which is identical to $\overline{\nu}(\gamma, \delta)$). Further, $\overline{\nu}(\gamma, \delta)$ can be shown to be equivalent to the GLS estimate $\widetilde{\nu}(\gamma, \delta, \overline{\theta}_I) = \widehat{\nu}(\gamma, \delta, \overline{\theta}_I)$, with $\overline{\theta}_I$ consisting of the three distinct elements of $\overline{\Omega}(\gamma, \delta)$, where

$$\overline{\Omega}(c, d) = n^{-1}\sum_t \left[y_t(c) - \overline{\nu}(c, d)x_t(c), x_t(d)\right]'\left[y_t(c) - \overline{\nu}(c, d)x_t(c), x_t(d)\right]. \tag{2.21}$$

Thus, our GLS approach can be seen to include Gaussian pseudo-ML estimation as a special case, where particular estimates of Ω are used, this interpretation continuing to apply when autocorrelation in u_t is incorporated. (Based on (2.19) in the $CI(1, 1)$ case, Phillips, 1991b, employed a semiparametric version of GLS, involving smoothed nonparametric estimation of $f(\lambda)$ across a coarser grid than the Fourier frequencies, following Hannan, 1963.)

3. CONDITIONS AND MAIN RESULTS

We present first a series of regularity conditions.

Assumption 1. *The process u_t, $t = 0, \pm 1, ...$, has representation*

$$u_t = A(L)\varepsilon_t, \tag{3.1}$$

where

$$A(s) = I_2 + \sum_{j=1}^\infty A_j s^j, \tag{3.2}$$

8

and the A_j are 2×2 matrices such that :

(i)

$$\det \{A(s)\} \neq 0, \ |s| = 1; \qquad (3.3)$$

(ii) $A(e^{i\lambda})$ is differentiable in λ with derivative in $Lip(\eta), \ \eta > \frac{1}{2}$;

and in addition, with $\|\cdot\|$ denoting the Euclidean norm:

(iii) the ε_t are independent and identically distributed vectors with mean zero, positive definite covariance matrix Ω, and $E \|\varepsilon_t\|^q < \infty, q \geq 4, q > 2/(2\beta - 1)$.

Notice that (ii) implies $\sum_{j=1}^{\infty} j \|A_j\| < \infty$, because the derivative of $A(e^{i\lambda})$ has Fourier coefficients jA_j, whence Zygmund (1977, p.240) can be applied. Further, this also implies $\sum_{j=1}^{\infty} j \|A_j\|^2 < \infty$, which, along with the condition in (iii), enables us to apply the functional limit theorem of Marinucci and Robinson (2000) (developing earlier work of Akonom and Gourieroux, 1987, Silveira, 1991) to the nonstationary process $x_t(\gamma)$, as is required to characterize the limit distribution of our estimates of ν. Further, due to (i), $B(e^{i\lambda})$ (see (2.6)) satisfies the same smoothness condition as $A(e^{i\lambda})$ in (ii), and thus

$$\sum_{j=1}^{\infty} j \|B_j\| < \infty, \qquad (3.4)$$

which implies the required conditions on the B_j in our other proofs, in particular of Propositions 1 and 2. It is Proposition 1 that requires the strongest conditions on the B_j, and this is possible by a lengthier proof under the milder requirement that $A(e^{i\lambda})$, and thus $B(e^{i\lambda})$, is boundedly differentiable, which itself implies (see Zygmund, 1977, p.251) $\sum_{j=1}^{\infty} j^{\frac{1}{2}} \|A_j\| < \infty$, and, from (i), $\sum_{j=1}^{\infty} j^{\frac{1}{2}} \|B_j\| < \infty$. However our present conditions seem satisfactorily mild, easily covering stationary and invertible ARMA systems. The moment assumption on ε_t is satisfied, for any $\beta > 1/2$, by Gaussianity.

The above assumption, with (1.1), (1.2), (1.6), (1.7), suffices in order to establish Theorem 1 below for the infeasible estimates $\widetilde{\nu}(\gamma, \delta, \theta)$ and $\widehat{\nu}(\gamma, \delta, \theta)$, but in order to insert estimated parameters further conditions are required. It is convenient to denote by Θ the set of all admissible values of $\widehat{\theta}$; often we may take Θ to be a bounded set, in part to satisfy stationarity conditions, while compactness of Θ would help to ensure existence of $\widehat{\theta}$.

Assumption 2.

(i) $f(\lambda; \theta) = f(\lambda)$.
(ii) $f(\lambda; h)$ has determinant bounded away from zero on $([-\pi, \pi] \times \Theta)$.
(iii) $f(\lambda; h)$ is boundedly differentiable in h on $([-\pi, \pi] \times \Theta)$, with derivative that is continuous in h at $h = \theta$ for all λ.
(iv) $f(\lambda; \theta)$ is differentiable in λ, with derivative satisfying a Lipschitz condition of order greater than $1/2$ in λ.

(v) $(\partial/\partial h) f(\lambda; h)$ is differentiable in λ at $h = \theta$, with derivative satisfying a Lipschitz condition of order greater than $1/2$ in λ.

Given correct specification (i), these assumptions seem innocuous, again being easily satisfied by standard stationary and invertible ARMA parameterizations, for example, and could be slightly relaxed at cost of greater proof detail.

Assumption 3.

(i) There exists $K < \infty$ such that

$$|\widehat{\gamma}| + \left|\widehat{\delta}\right| \leq K, \tag{3.5}$$

and $\rho > \max(0, 1 - \beta)$ such that

$$\widehat{\gamma} = \gamma + O_p\left(n^{-\rho}\right), \ \widehat{\delta} = \delta + O_p\left(n^{-\rho}\right). \tag{3.6}$$

(ii)

$$\widehat{\theta} = \theta + O_p(n^{-\frac{1}{2}}), \ \text{where} \ \theta \in \Theta. \tag{3.7}$$

Condition (3.5) is innocuous if $\widehat{\gamma}$ and $\widehat{\delta}$ optimize over compact sets, as is standard for implicitly defined estimates. The convergence rates required in Assumption 3 are all less than those achieved of estimates (2.11) and (2.18) of ν in Theorem 1 below. In fact (ii) could be relaxed to the rate on $\widehat{\gamma}$ and $\widehat{\delta}$ of (i) if $f(\lambda; h)$ is smoother in h than required in Assumption 2, in particular if it is analytic in h (as in the ARMA case). We prefer our milder Assumption 2, and the relatively brief proof that (ii) affords, because $n^{1/2}-$consistency of parameter estimates in short memory time series models is familiar, for example in case of Whittle estimates, see eg. Hannan (1973). On the other hand, we might be content to assume $\rho = 1/2$ in (3.6). The $n^{1/2}-$consistency and asymptotic normality of estimates of nonstationary integration orders (and indeed of parameters corresponding to θ in nonstationary fractional models), based on scalar series was established by Velasco and Robinson (2000), for Type I processes (see (1.10)). By bounding a measure of distance between Type I and Type II processes, Robinson (2002) showed that the same results hold for Type II processes, thereby checking (3.6) and (3.7) for estimates of δ and elements of θ identified by the u_{2t} process. Robinson (2002) likewise checked (3.6) and (3.7) for estimates (computed from residuals) of γ and elements of θ identified by $\{u_{1t}\}$, employing a preliminary estimate of ν, which satisfies a rate of convergence condition. This is satisfied by OLS when $\gamma + \delta \geq 1$, but not when $\gamma + \delta < 1$, where it is, however, satisfied by the narrow-band OLS estimate of Robinson and Marinucci (1997, 2001), using a bandwidth that increases sufficiently slowly; the strength of this rate condition is due in part to allowing the compact set of admissible values of γ to be arbitrarily large - if this is suitably reduced the condition can be relaxed so as to be satisfied by OLS even when $\gamma + \delta < 1$, so long as $\delta < 1/2$. The only gap left in demonstrating that Assumption 3 can be fully checked is that in general methods based on the bivariate

series z_t are appropriate in order to estimate part of θ. However the extension of Velasco and Robinson's (2000) theory to cover bivariate series, and the subsequent adaptation to our setting, seems straightforward, while if $B(s;\theta)$ is *a priori* diagonal the only parameter not estimated by two univariate procedures is the off-diagonal element of Ω, which is estimated by an obvious side calculation, to satisfy (ii). Unless β is close to 1/2, (3.6) is capable of being satisfied also by 'semiparametric' estimates of γ and δ, which might in any case be employed at an initial stage in determining the parametric model for f. On the other hand, from the viewpoint of a full cointegration analysis, efficient estimates of γ, δ and θ are desirable, suggesting construction of a Gaussian pseudo-ML approach, estimating all parameters jointly, which is computationally more onerous than the kind of step-by-step approach we have envisaged, but undoubtedly possible; asymptotic properties have yet to be explicitly derived, but the problem of differing convergence rates encountered by Saikkonen (1995) in a different setting can be avoided by concentrating out ν first.

We introduce notation to describe the limit distribution of our estimates. Denote by $W(r)$ the 2×1 vector Brownian motion with covariance matrix Ω, and define (Type II-see Marinucci and Robinson, 1999) fractional Brownian motion

$$W(r;\beta) = \int_0^r \frac{(r-s)^{\beta-1}}{\Gamma(\beta)} dW(s), \qquad (3.8)$$

and then define

$$\widetilde{W}(r;\beta) = \xi' B(1)^{-1} W(r;\beta), \qquad (3.9)$$

where

$$\xi = (0,1)'. \qquad (3.10)$$

By "$\Rightarrow$" we will mean convergence in the Skorohod J_1 topology of $D[0,1]$.

Theorem 1. *Let (1.1), (1.2), (1.6), (1.7) and Assumptions 1-3 hold. Then, denoting by ν^* any of the estimates in (2.11) or (2.18), we have as $n \to \infty$,*

$$n^\beta(\nu^* - \nu) \Rightarrow \left\{ q(0) \int_0^1 \widetilde{W}(r;\beta)^2 \, dr \right\}^{-1} 2\pi \zeta' B(1)' \Omega^{-1} \int_0^1 \widetilde{W}(r;\beta) \, dW(r). \quad (3.11)$$

The proof is outlined in Appendix A, by a series of propositions whose proofs appear in Appendix B. The rate of convergence in (3.11) is optimal for any regular parametric estimate in this model. Theorem 1 desirably implies that we can estimate ν as well, asymptotically, not knowing γ and/or δ and/or θ as knowing them, subject to the rate conditions of Assumption 3, with the implication that efficiency of estimation of γ, δ and θ does not matter.

The variates $\zeta' B(1)' \Omega^{-1} W(r)$ and $\widetilde{W}(r;\beta)$ are uncorrelated and thus, by Gaussianity, independent, so (3.11) indicates mixed normal asymptotics. As a consequence of this, and of the Propositions in Appendix A, we have

11

Corollary 1. *Denoting by b^* any of the quantities $\widetilde{b}(\gamma, \theta)$, $\widetilde{b}(\widehat{\gamma}, \theta)$, $\widetilde{b}(\gamma, \widehat{\theta})$, $\widetilde{b}(\widehat{\gamma}, \widehat{\theta})$, $b(\gamma, \theta)$, $b(\widehat{\gamma}, \theta)$, $b(\gamma, \widehat{\theta})$, $b(\widehat{\gamma}, \widehat{\theta})$, as $n \to \infty$, the Wald statistics*

$$b^* \left(\nu^* - \nu \right)^2 \to_d \chi_1^2. \tag{3.12}$$

The form of the limit distribution in (3.11), where spectral properties of u_t at only zero frequency are involved, and the nonstationarity of $x_t(\gamma)$, suggests simpler forms of estimate than (2.11), (2.18). We replace $p(\lambda_j; h)$, $q(\lambda_j; h)$ by $p(0; h)$, $q(0; h)$, and thence consider

$$\overline{\nu}(\gamma, \delta, \theta), \ \overline{\nu}(\gamma, \delta, \widehat{\theta}), \ \overline{\nu}(\widehat{\gamma}, \delta, \widehat{\theta}), \ \overline{\nu}(\gamma, \widehat{\delta}, \widehat{\theta}), \ \overline{\nu}(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}), \tag{3.13}$$

where

$$\overline{\nu}(c, d, h) = \frac{\overline{a}(c, d, h)}{\overline{b}(c, h)}, \tag{3.14}$$

in which

$$\overline{a}(c, d, h) = p(0; h) \sum_t z_t(c, d) x_t(c), \ \ \overline{b}(c, h) = q(0; h) \sum_t x_t^2(c), \tag{3.15}$$

after applying (B.3) below. If we act on the belief that u_t is white noise, (3.13) is identical to (2.11), (2.18), but to cover other circumstances we have:

Theorem 2. *Let (1.1), (1.2), (1.6), (1.7) and Assumptions 1-3 hold. Then, denoting by ν° any of the estimates in (3.13), we have as $n \to \infty$:*
(i) for $1/2 < \beta < 1$,

$$n^{2\beta-1} \left(\nu^\circ - \nu \right) \Rightarrow \left\{ q(0) \int_0^1 \widetilde{W}(r; \beta)^2 \, dr \right\}^{-1} p(0) \int_{-\pi}^{\pi} f(\lambda) \xi (1 - e^{-i\lambda})^{-\beta} d\lambda; \tag{3.16}$$

(ii) for $\beta = 1$,

$$n \left(\nu^\circ - \nu \right) \Rightarrow \left\{ q(0) \int_0^1 \widetilde{W}(r; \beta)^2 \, dr \right\}^{-1}$$

$$\times \left\{ p(0) \sum_{s=0}^{\infty} \psi_{-s} + 2\pi \zeta' B(1)' \Omega^{-1} \int_0^1 \widetilde{W}(r; 1) \, dW(r) \right\}, \tag{3.17}$$

12

where

$$\psi_s = E(u_0 u_s')\xi; \tag{3.18}$$

(iii) for $\beta > 1$,

$$n^\beta (\nu^\circ - \nu) \Rightarrow \left\{ q(0) \int_0^1 \widetilde{W}(r;\beta)^2 \, dr \right\}^{-1} 2\pi \zeta' B(1)' \Omega^{-1} \int_0^1 \widetilde{W}(r;\beta) \, dW(r). \tag{3.19}$$

If u_t is white noise, so $f(\lambda) \equiv f(0)$, we have $p(0)f(\lambda)\xi \equiv 0$ and (3.16) becomes $\nu^\circ = \nu + o_p(n^{1-2\beta})$, but Theorem 1 applies here, with the sharp result (3.11); also, $p(0)\sum_{s=0}^\infty \psi_{-s} = p(0)\psi_0 = 2\pi p(0)f(0)\xi = 0$, so (3.17) reduces to (3.11). For auto-correlated u_t, when $\beta > 1$, (3.19) indicates that (3.13) still does as well as (2.11), (2.18), but when $\beta = 1$ the convergence rate in (3.17) is as good but the desirable mixed-normal asymptotics are lacking, due to "second-order bias" (cf Phillips, 1991a,b) appearing as the first term in the second factor on the right of (3.17), and when $\beta < 1$, in (3.16), not only are mixed-normal asymptotics lacking but convergence is slower. Indeed, for $1/2 < \beta < 1$ (3.13) never converges faster, and nearly always converges slower, than OLS of y_t on x_t. From Propositions 6.1, 6.2 and 6.5 of Robinson and Marinucci (2001), OLS is $n^{2\delta-1}$-consistent when $\gamma + \delta = 2\delta - \beta < 1$, $n^{2\delta-1}/\log n$-consistent when $\gamma + \delta = 2\delta - \beta = 1$ and $\gamma > 0$, n-consistent when $\delta = 1$, $\gamma = 0$, and n^β-consistent when $\gamma + \delta = 2\delta - \beta > 1$, so over the intersection of these regions with $1/2 < \beta = \delta - \gamma < 1$ the rate in (3.13) is equalled when $\gamma = 0$ and exceeded when $\gamma > 0$, indicating that proper fractional differencing without proper accounting for $I(0)$ autocorrelation can do worse than simple methods based on unfiltered data.

Focusing more closely on $\gamma = 0$, where the central case (ii) is that of $I(1)$ x_t, while the widespread evidence of unit root behaviour based on tests against AR alternatives cannot be taken very seriously from a fractional viewpoint (see Diebold and Rudebusch (1991), Robinson (1994)), it might be reasonable to interpret this as suggesting that integration orders may often be close to 1, but either greater or less than 1, when the discontinuity in Theorem 2 at $\beta = 1$ makes use of (3.13) questionable. Even when $\beta > 1$, the detailed corrections for autocorrelation in (2.11) and (2.18) might be expected to produce better finite-sample properties than (3.13), which is based on an appeal to asymptotic theory due to a high degree of nonstationarity in $x_t(\gamma)$, while the extra computational burden of (2.11) and (2.18) does not seem prohibitive. Because this discussion indicates that it is less important than Theorem 1, and because its proof is in part embodied in that of Theorem 1 and in part straightforwardly uses Theorems 4.1, 4.3 and 4.4 of Robinson and Marinucci (2001), we have omitted the proof of Theorem 2. Theorem 4.3 of Robinson and Marinucci (2001) can also be applied to justify narrow-band frequency-domain versions of (3.13) which, at cost of introducing a user-chosen bandwidth, eliminate the second-order bias term in (3.17) and thereby achieve the asymptotics in (3.11), corresponding to an idea due to Phillips (1991b) in a semiparametric setting for the $CI(1,1)$ case $\beta = \delta = 1$.

4. MONTE CARLO EVIDENCE

With the main aim of studying the effect of estimating integration orders γ, δ on our estimates of ν and their distributional properties, a small Monte Carlo study was carried out in the simplest case where in (1.1), (1.2) we know that the u_t are serially uncorrelated, so $A_j \equiv 0$ *a priori* in (3.2). The treatment of autocorrelation in u_t looms large in the rest of the paper, and corresponding Monte Carlo analysis is warranted, but a reasonably representative analysis, perhaps looking at more than one time series model, with varying degrees of autocorrelation and comparing the performance of (2.11), (2.18) and (3.13), as well as varying scale and contemporaneous correlation parameters, would add considerable space to this already lengthy paper. There are two parts to our Monte Carlo investigation, the first comparing performance in fractional circumstances of estimates assuming both γ and δ are known with ones where both are estimated, and the second focusing on the standard case $(\gamma, \delta) = (0, 1)$, and considering also estimates in which one of γ or δ is estimated. We generated Gaussian u_t with covariance matrix Ω having ijth element ω_{ij}, varying the correlation $\rho = \omega_{12}/(\omega_{11}\,\omega_{22})^{1/2}$ (taking values 0, 0.5, -0.5, 0.75) and variance ratio $\tau = \omega_{22}/\omega_{11}$ (taking values 0.5, 1, 2). The parameter ρ heavily influences the "simultaneous equation bias" in (1.1), regressors and disturbances being orthogonal only when $\rho = 0$, while τ affects the signal-to-noise ratio in (1.1), with increase in τ generally being associated with an increase in precision in estimation of ν. Our estimates are invariant to $\nu \neq 0$ and also to a scale factor of Ω, and so we fixed $\nu = \omega_{11} = 1$ with no loss of generality.

In the first part of the study we employed all six (γ, δ) combinations of $\gamma = 0, 0.4$ with $\delta = 0.6, 1.2, 2$:

$$(\gamma, \delta) = (0, 0.6), (0, 1.2), (0, 2), (0.4, 0.6), (0.4, 1.2), (0.4, 2). \tag{4.1}$$

The fourth case, $(0.4, 0.6)$, does not satisfy (1.7), but is included to illustrate the case $\beta \leq 1/2$ discussed briefly in point 1 of Section 6 below. In the first case, $(0, 0.6)$, the bias of OLS is so strong as to determine the rate of convergence when $\rho \neq 0$ (see Robinson and Marinucci, 1997), while in the remaining four cases OLS achieves the optimal rate. Table I records the convergence rates of OLS when $\rho \neq 0$, OLS when $\rho = 0$, and the optimal rates (achieved in Theorem 1).

TABLE I
CONVERGENCE RATES:
OLS WITH $\rho \neq 0$, $\rho = 0$ AND OPTIMAL RATES

(γ, δ)	$(0, 0.6)$	$(0, 1.2)$	$(0, 2)$	$(0.4, 0.6)$	$(0.4, 1.2)$	$(0.4, 2)$
OLS, $\rho \neq 0$	$n^{.2}$	$n^{1.2}$	n^2	$n^{.2}/\log n$	$n^{.8}$	$n^{1.6}$
OLS, $\rho = 0$	$n^{.6}$	$n^{1.2}$	n^2	$n^{.2}$	$n^{.8}$	$n^{1.6}$
Optimal	$n^{.6}$	$n^{1.2}$	n^2	$n^{.5}$	$n^{.8}$	$n^{1.6}$

We generated 1000 series of lengths $n = 64, 128, 256$, and computed the Infeasible

estimate $\overline{\nu}_I$ and $\underline{\text{F}}$easible estimate $\overline{\nu}_F$, given by

$$\overline{\nu}_I = \widehat{\nu}(\gamma, \delta, \overline{\theta}_I) = \widetilde{\nu}(\gamma, \delta, \overline{\theta}_I), \qquad (4.2)$$

$$\overline{\nu}_F = \widehat{\nu}(\widehat{\gamma}, \widehat{\delta}, \overline{\theta}_F) = \widetilde{\nu}(\widehat{\gamma}, \widehat{\delta}, \overline{\theta}_F), \qquad (4.3)$$

for $\widehat{\gamma}, \widehat{\delta}$ to be described subsequently, and $\overline{\theta}_I, \overline{\theta}_F$ representing 3×1 vectors of estimates of $\theta = (\omega_{11}, \omega_{12}, \omega_{22})'$ given by $\overline{\Omega}_I = \overline{\Omega}(\gamma, \delta)$, $\overline{\Omega}_F = \overline{\Omega}(\widehat{\gamma}, \widehat{\delta})$, with the definition (2.21). Then we have $\overline{\nu}_I = \overline{\nu}(\gamma, \delta), \overline{\nu}_F = \overline{\nu}(\widehat{\gamma}, \widehat{\delta})$ (see (2.20)). Thus, we compare an optimal estimate $(\overline{\nu}_I)$ in case γ, δ are known (one that is familiar from the unit root cointegration literature in case $(\gamma, \delta) = (0, 1)$) with one $(\overline{\nu}_F)$ where γ, δ are unknown, and replaced by estimates.

We computed $\widehat{\delta}$ by variants of the univariate Whittle procedure of Velasco and Robinson (2000), using untapered x_t for $\delta < 1$, and for $\delta \geq 1$ using untapered Δx_t and adding back 1. The estimation of memory parameters of nonstationary series by means of integer-differenced stationary and invertible observations incurs no loss of efficiency (cf Robinson, 1994), but our use of the actual δ may favour $\overline{\nu}_F$. On the other hand, Velasco and Robinson's (2000) estimates based on untapered data are proved to be $n^{1/2}$-consistent only when the memory parameter is less than $3/4$, so our application of their procedure to first-differenced untapered data when $\delta = 2$ is not supported by their results, and may lead to inferior $\overline{\nu}_F$ compared to ones using memory parameter estimates which incorporate suitable tapering. We computed $\widehat{\gamma}$ from the same type of procedure based on the $y_t - \overline{\nu}_O x_t$, where $\overline{\nu}_O$ is the $\underline{\text{O}}$rdinary least squares estimate

$$\overline{\nu}_O = \frac{\Sigma_t x_t y_t}{\Sigma_t x_t^2}. \qquad (4.4)$$

Robinson and Marinucci (1997) have demonstrated how $\overline{\nu}_O$ can be improved upon by a narrow-band frequency domain OLS procedure. This would presumably lead to an improvement in $\widehat{\gamma}$, and thence in $\overline{\nu}_F$, but it involves choice of a bandwidth number, and in the purely "parametric" context of the current paper we prefer the more familiar and simpler $\overline{\nu}_O$, whose performance as an estimate of ν we also compare with $\overline{\nu}_I$ and $\overline{\nu}_F$.

Tables II and III respectively show the Monte Carlo bias (defined as the estimate minus ν averaged across replications) and standard deviation (SD) of $\overline{\nu}_I$, $\overline{\nu}_F$ and $\overline{\nu}_O$, across all cases (4.1), and for all three values of τ, but for $\rho = 0.5$ only. The tables for $\rho = 0, -0.5, 0.75$ have been omitted to conserve space, but our discussion reflects them equally. For all ρ, τ, bias tends to decrease in absolute value as β increases, as rates of convergence predict. Bias tends to vary inversely with τ, but this is very noticeable only in the cases $(0, 0.6)$, $(0.4, 0.6)$. Of the three estimates, in general $\overline{\nu}_I$ not surprisingly performs best, followed by $\overline{\nu}_F$ and then $\overline{\nu}_O$: it is reassuring that while $\overline{\nu}_F$ is damaged by nuisance parameter estimation, it nevertheless emerges as worthwhile relative to OLS, whose bias is unacceptably large in the cases $(0, 0.6)$, $(0.4, 0.6)$, even for $n = 256$, except, of course, when $\rho = 0$, when it tends to do better than $\overline{\nu}_F$ for $(0.4, 0.6)$. While the bias of $\overline{\nu}_I$ is virtually unaffected by varying ρ, there is evidence that the bias of $\overline{\nu}_F$ somewhat increases in absolute value with $|\rho|$, with sign opposite

to that of ρ. Biases tend to decrease with n, though some noticeable increases are observed for $\overline{\nu}_F$ when $(\gamma, \delta) = (0.4, 0.6)$. As anticipated, SD tends to decrease as τ and n increase. The SD of both $\overline{\nu}_I$ and $\overline{\nu}_F$ shows some tendency to decrease as $|\rho|$ increases, though it frequently increases in case of $\overline{\nu}_F$ when $(\gamma, \delta) = (0, 0.6)$ or $(0.4, 0.6)$, its much greater imprecision relative to $\overline{\nu}_I$ supporting in the latter case a conjecture offered in point 1 of Section 6 below. Otherwise, the close similarity in variability of $\overline{\nu}_I$ and $\overline{\nu}_F$ for $n = 256$ is encouraging. For $n = 64$, the change in sign of ρ is associated with some small improvement. Often $\overline{\nu}_O$ is more precise than $\overline{\nu}_F$, and even $\overline{\nu}_I$, when either n is small or $(\gamma, \delta) = (0, 0.6)$ or $(0.4, 0.6)$.

We next examine the accuracy of the large sample χ^2 approximation of Corollary 1, looking at the size of Wald tests. Define the Wald statistics $W_I = \overline{b}_I \left(\overline{\nu}_I - 1\right)^2$ and $W_F = \overline{b}_F \left(\overline{\nu}_F - 1\right)^2$, where

$$\overline{b}_I = b\left(\gamma, \overline{\theta}_I\right) = \widetilde{b}\left(\gamma, \overline{\theta}_I\right) = \overline{v}\left(\gamma, \delta\right), \tag{4.5}$$

$$\overline{b}_F = b\left(\widehat{\gamma}, \overline{\theta}_F\right) = \widetilde{b}\left(\widehat{\gamma}, \overline{\theta}_F\right) = \overline{v}\left(\widehat{\gamma}, \widehat{\delta}\right), \tag{4.6}$$

with

$$\overline{v}\left(c, d\right) = \frac{n\left\{\Sigma_t x_t^2(c)\Sigma_t x_t^2(d) - \left\{\Sigma_t x_t(c)x_t(d)\right\}^2\right\}}{\Sigma_t x_t^2(d)\Sigma_t \widehat{\varepsilon}_t^2\left(c, d\right)}, \tag{4.7}$$

where $\widehat{\varepsilon}_t\left(c, d\right)$ are residuals from the OLS regression of $y_t(c)$ on $x_t(c)$ and $x_t(d)$; $\overline{v}\left(c, d\right)$ is the usual OLS estimate of variance of the estimated coefficient of $x_t(c)$ in the OLS regression of $y_t(c)$ on $x_t(c)$ and $x_t(d)$. Tables IV-VII contain empirical sizes corresponding to nominal sizes $\alpha = 0.05, 0.10$, for the four values of ρ but for $\tau = 1$ only, the results for $\tau = 0.5$ and 2 being very similar. The results for W_I are on average too large, but only slightly, and performance here seems very satisfactory over all (γ, δ) and ρ. The empirical sizes of W_F do decrease significantly in n but are overall too large, worst when $(\gamma, \delta) = (0, 0.6)$, and tend to decrease in β for $|\rho| \geq 0.5$. Some results, for $\gamma = 0$, $\delta = 0.6, 1.2$ only, were also obtained for larger n; in particular when $n = 1024$ decreases of about 30% were typically achieved.

For the second part of the Monte Carlo study, we focus on the familiar case $(\gamma, \delta) = (0, 1)$, but include now also the "intermediate" estimates discussed in Section 2, employing prior knowledge of either γ or δ,

$$\overline{\nu}_\gamma = \overline{\nu}(0, \widehat{\delta}) = \widehat{\nu}(0, \widehat{\delta}, \overline{\theta}_\gamma) = \widetilde{\nu}(0, \widehat{\delta}, \overline{\theta}_\gamma), \tag{4.8}$$

$$\overline{\nu}_\delta = \overline{\nu}(\widehat{\gamma}, 1) = \widehat{\nu}(\widehat{\gamma}, 1, \overline{\theta}_\delta) = \widetilde{\nu}(\widehat{\gamma}, 1, \overline{\theta}_\delta), \tag{4.9}$$

where $\overline{\theta}_\gamma$, $\overline{\theta}_\delta$ consist of the appropriate elements of $\overline{\Omega}(0, \widehat{\delta})$, $\overline{\Omega}(\widehat{\gamma}, 1)$, respectively. In this case $\overline{\nu}_O$ has the same rate of convergence as $\overline{\nu}_I$, $\overline{\nu}_F$, $\overline{\nu}_\gamma$, $\overline{\nu}_\delta$, being n-consistent, but lacks the mixed normal asymptotics. We employed the same values of ρ and τ as before. Table VIII reports Monte Carlo bias and SD. The best and worst estimates, when $\rho \neq 0$, are again $\overline{\nu}_I$ and $\overline{\nu}_O$ respectively, but though $\overline{\nu}_\delta$ (which correctly assumes $\delta = 1$) is second-best, $\overline{\nu}_\gamma$ (which correctly assumes $\gamma = 0$) is inferior to $\overline{\nu}_F$; this is all the more surprising because γ is more problematic to estimate than δ as it uses

residuals. In $\overline{\nu}_F$ the contributions to bias from estimation of γ and δ may partly cancel, while even when $n = 64$ the bias of $\overline{\nu}_\gamma$ is never so large as to cause serious concern. As before, SD is much less variable. For $|\rho| \geq 0.5$, $\overline{\nu}_O$ clearly performs worst, but there is little difference between the optimal estimates, though for small n, $\overline{\nu}_I$ seems best, followed closely by $\overline{\nu}_\delta$, with almost identical values for $\overline{\nu}_\gamma$ and $\overline{\nu}_F$.

Table IX reports empirical sizes, including now results for $W_\gamma = \overline{b}_\gamma \left(\overline{\nu}_\gamma - 1 \right)^2$, $W_\delta = \overline{b}_\delta \left(\overline{\nu}_\delta - 1 \right)^2$, where

$$
\begin{aligned}
\overline{b}_\gamma &= b\left(0, \overline{\theta}_\gamma\right) = \widetilde{b}\left(0, \overline{\theta}_\gamma\right) = \overline{v}\left(0, \widehat{\delta}\right), &\qquad (4.10)\\
\overline{b}_\delta &= b\left(\widehat{\gamma}, \overline{\theta}_\delta\right) = \widetilde{b}\left(\widehat{\gamma}, \overline{\theta}_\delta\right) = \overline{v}\left(\widehat{\gamma}, 1\right), &\qquad (4.11)
\end{aligned}
$$

and

$$
W_O = \frac{n\left(\overline{\nu}_O - 1\right)^2 \Sigma_t x_t^2}{\Sigma_t \left(y_t - \overline{\nu}_O x_t\right)^2}, \qquad (4.12)
$$

though W_O does not have a limiting null χ_1^2 distribution. The empirical size of W_I is the most accurate, followed by W_γ, the discrepancy increasing with $|\rho|$. Even for $\rho \neq 0$, W_O often does better than W_δ and W_F, which perform quite similarly; the effect of estimating γ is dominant , and use of an improved preliminary estimate of ν, such as that proposed by Robinson and Marinucci (1997, 2001), or iteration, may be warranted.

5. EMPIRICAL INVESTIGATION OF THE PURCHASING POWER PARITY HYPOTHESES

Numerous empirical studies have cast significant doubt on the purchasing power parity (PPP) hypothesis with respect to the short run, but have yielded mixed evidence with respect to the long run (see e.g. Corbae and Ouliaris, 1988, Enders, 1988, Kim, 1990, Taylor, 1988). Cheung and Lai (1993) proposed a fractional version of the PPP specification, essentially (1.1), (1.2) with x_t representing the domestic price index and y_t the foreign price index, converted to domestic currency units. The coefficient ν in (1.1) is unity according to the absolute or homogeneous version of PPP, so this is testable by our Wald statistic of Corollary 1. Using unit root tests, Cheung and Lai (1993) failed to reject the hypothesis $\delta = 1$ and then, using differenced OLS residuals, they computed semiparametric log periodogram estimates of $\delta - \beta - 1$ and then tested the non-cointegration null hypothesis of $\beta = 0$ against the alternative $\beta > 0$, using critical values computed by simulation in view of the inapplicability of standard asymptotic theory in this case. They found evidence of cointegration in a number of bivariate series, but did not test $\nu = 1$. We employ a step-by-step approach, first testing whether the integration orders δ_x and δ_y of x_t and y_t are the same, then for the presence of cointegration, then for $\beta > 1/2$ and finally, given all these hurdles have been crossed, $\nu = 1$. In the first three steps we used semiparametric procedures (as did Cheung and Lai, 1993, Marinucci and Robinson, 2001), while in the final step,

which is most relevant to the material of the current paper, we identified parametric models for the autocorrelation in u_t and thence computed estimates of ν and Wald statistics.

The semiparametric estimates of integration orders were all Robinson's (1995) versions of log periodogram estimates, but without trimming, using first differences and then adding back 1. We estimated δ_x and δ_y separately, and then tested $\delta_x = \delta_y (= \delta)$ by an adaptation of Robinson and Yajima's (2001) statistic $\widehat{T}_{ab}$ to log periodogram estimation, with their trimming sequence $h(n)$ chosen as m^{-5-2i} for $i = 1, ..., 4$, with m the bandwidth used in the estimation. Given $\delta_x = \delta_y$ is not rejected, we performed the Hausman test for no-cointegration of Marinucci and Robinson (2001), comparing the estimate $\widetilde{\delta}_x$ of δ_x with the more efficient bivariate one of Robinson (1995), that uses the information $\delta_x = \delta_y$. Given cointegration is not rejected, the null $\beta = 1/2$ was rejected in favour of $\beta > 1/2$ if and only if a studentized $\widetilde{\delta}_x - \widetilde{\gamma} - 1/2$, was significantly large relative to the standard normal distribution, where $\widetilde{\gamma}$ is the estimate of γ using OLS residuals.

Using annual data of Obstfeld and Taylor (2002) (with $n = 123$), as is relevant to the long-run version of PPP, we applied the above methodology to four bivariate series, the US ('domestic') versus the 'foreign' countries Australia, Canada, Italy, UK. Strong evidence against equality of integration orders was found in case of Australia and Italy, and against cointegration in case of Canada. However, the UK 'passed' all three initial tests. Across the range $m = 10, ..., 29$, $(\widetilde{\delta}_x, \widetilde{\delta}_y)$ varied between the extremes $(1.341, 1.095)$ and $(1.572, 1.376)$, and across $m = 16, ..., 25$ and the four $h(n)$ choices, $\delta_x = \delta_y$ was rejected in only 9 out of 40 cases, and these all at the 10% level. For the same m, no-cointegration was rejected at 10% in all cases, at 5% in 4 cases, and at 1% in 3 cases, while $\beta = 1/2$ was rejected against $\beta > 1/2$ at the 1% level in all cases.

For the US-UK data, we identified parametric models for $f(\lambda)$ as follows. Throughout, $A(L)$ in (3.1) was diagonal, and u_{1t}, u_{2t} treated separately. They were proxied by $\Delta^{\widetilde{\gamma}}(y_t - \bar{\nu}_O x_t)$, $\Delta^{\widetilde{\delta}_x} x_t$, for each of the extreme $\widetilde{\gamma}, \widetilde{\delta}_x$, namely $\widetilde{\gamma} = .374, .698$ and $\widetilde{\delta}_x = 1.572, 1.341$, and then Box-Jenkins-type procedures identified models within the ARMA class. This resulted in $AR(1)$ and $ARMA(1,1)$ u_{1t} and white noise and $ARMA(1,1)$ u_{2t}, and we fitted all four combinations. We also fitted bivariate versions of Bloomfield's (1972) model, where $A(z) = diag\left\{\exp\left(\sum_{j=1}^{p} \theta_{1j} z^j\right), \exp\left(\sum_{j=1}^{p} \theta_{2j} z^j\right)\right\}$, for $p = 1, 2, 3$. For each model we applied the univariate Whittle procedure in Velasco and Robinson (2000), using untapered, differenced data and adding back 1. We summarize the seven models and the resulting $(\widehat{\delta}, \widehat{\gamma})$ as follows:

Model 1: u_{1t} is $AR(1)$ and u_{2t} is white noise. $(\widehat{\delta}, \widehat{\gamma}) = (1.612, .669)$.
Model 2: u_{1t} is $AR(1)$ and u_{2t} is $ARMA(1,1)$. $(\widehat{\delta}, \widehat{\gamma}) = (1.408, .669)$.
Model 3: u_{1t} is $ARMA(1,1)$ and u_{2t} is white noise. $(\widehat{\delta}, \widehat{\gamma}) = (1.612, .660)$.
Model 4: u_{1t} is $ARMA(1,1)$ and u_{2t} is $ARMA(1,1)$. $(\widehat{\delta}, \widehat{\gamma}) = (1.408, .660)$.
Model 5: u_t is bivariate Bloomfield with $p = 1$. $(\widehat{\delta}, \widehat{\gamma}) = (1.214, .710)$.
Model 6: u_t is bivariate Bloomfield with $p = 2$. $(\widehat{\delta}, \widehat{\gamma}) = (1.434, .701)$

Model 7: u_t is bivariate Bloomfield with $p = 3$. $\left(\widehat{\delta}, \widehat{\gamma}\right) = (1.323, .547)$.

The $\widehat{\gamma}$ seem very robust to the short memory specification, the $\widehat{\delta}$ rather less so.

Table X contains our estimates $\widehat{\nu}\left(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}\right) = \bar{\nu}_i$ and Wald statistics

$b(\widehat{\gamma}, \hat{\theta}) \left\{ \widehat{\nu}\left(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}\right) - 1 \right\}^2 = W_i$ for models $i = 1, ..., 7$, but also for ten different subsets of the observations, namely $t = j, ..., n$ for $j = 1, ..., 10$ (so that the numbers of observations were $n' = n - j$, $j = 1, ..., 10$) in order to explore sensitivity to starting value: if we drop the first observation the degree of filtering applied to all subsequent observations changes. Substantial variation is evident across the larger n', with all $\bar{\nu}_i$ exceeding 1 and the homogeneity hypothesis being strongly rejected when $n' = 123$, across all seven models, but as n' decreases, things stabilize. For $n' \leq 119$ some sensitivity to the u_{2t} specification was found, the white noise cases (Models 1 and 3) providing estimates of ν less than .9, whereas for the other models they all exceed .9, with the largest values for Model 7. For $n' \leq 122$ the homogeneity hypothesis $\nu = 1$ is never rejected even at the 10% level.

6. FINAL COMMENTS

Our treatment of a bivariate system in a parametric setting is quite general, in that a very wide range of models for the $I(0)$ input series u_t is covered, while our regularity conditions seem to afford little scope for relaxation. Nevertheless, there are significant aspects not explored in the paper.

1. Our case $\beta > 1/2$ includes the familiar $CI(1,1)$ setting, but $0 < \beta < 1/2$ is also of interest. As discussed by Hualde and Robinson (2001), $x_t(\gamma)$ is then "asymptotically stationary" and our estimates are $n^{1/2}$-consistent and asymptotically normal, with limiting variance that is affected by the estimation (and the efficiency of estimation) of one or more of γ, δ and θ, because the requirement $\rho > 1 - \beta$ on ρ in (3.6) still appears to be relevant when $\beta < 1/2$, but (3.6) is unachieveable then because $\widehat{\gamma}$, $\widehat{\delta}$ are at most $n^{1/2}$-consistent, no matter the values of γ and δ, see eg. Velasco and Robinson (2000).

2. In view of the literature on non-fractional cointegration, there would be empirical interest in incorporating also in (1.1) and/or (1.2) deterministic components. Modification of the theory to cover polynomial time trends seems relatively straightforward, though our fractional focus suggests allowing for possibly non-integral powers of t in studying the relative importance of stochastic and deterministic trends, as Robinson and Marinucci (2000) did in connection with OLS and its narrow-band modification, while if such powers are unknown the extension is decidedly non-trivial.

3. Extension of our methods and theory to vector y_t and x_t, and matrix ν, seems straightforward when there is no variation in integration orders across elements of x_t and $y_t - \nu x_t$. However, multivariate data invite consideration not only of multiple cointegrating relationships but also of observables and/or cointegrating

errors with differing integration orders, which would raise particular questions of identifiability and complicate estimation.

4. Our parametric treatment of autocorrelation in u_t follows a classical economic time series tradition and allows parsimony, but the unit root cointegration literature has stressed a nonparametric approach. Nonparametric estimation of $f(\lambda)$ should lead to the same outcomes as in Theorems 1 and 2, and corresponds in (2.11) to taking $B_j = 0$, $j > p$, but letting p go slowly to infinity in the asymptotic theory, while in (2.18) or (3.13) weighted autocovariance or periodogram estimation might be used. The forms (3.13) would be easiest to handle technically, while in (2.18), the variation in $f(\lambda_j)$ across the n Fourier frequencies might be dealt with by techniques like those used by Robinson (1991, pp.1354, 1355), or alternatively one can employ estimates which are constant over slowly degenerating bands, as proposed in Hannan (1963) and employed by Phillips (1991b) in the $CI(1,1)$ case. In any event, the slow convergence of nonparametric estimates of f is of concern because even the refinement of Assumption 3 (ii) mentioned in the discussion of that assumption requires a convergence rate that approaches arbitrarily close to $n^{-1/2}$ as $\beta \to 1/2$. In principle $n^{\kappa-1/2}-$consistent nonparametric spectral estimates can be found, for any $\kappa > 0$ (where, for example, κ depends on kernel order, see eg Cogburn and Davis, 1974), though, as β is unknown, one can never be sure that the κ achieved is sufficient.

APPENDIX A: OUTLINE OF PROOF OF THEOREM 1

Though the proof of (3.11) for the time-domain estimates (2.11) is not contained in that for the frequency-domain estimates (2.18), nevertheless the proof for the latter does involve approximation in the time domain so that many of the steps are similar. Thus, because it entails the greater technical challenge, computational elegance and generality, we give the proof only for (2.18).

Consider first the infeasible estimate $\widehat{\nu}(\gamma, \delta, \theta)$. We have

$$z_t(c, d) = \zeta x_t(c) \nu + v_t(c, d), \tag{A.1}$$

where

$$v_t(c, d) = (u_{1t}(c - \gamma), x_t(d))'. \tag{A.2}$$

Thus

$$\widehat{\nu}(c, d, h) - \nu = \frac{e(c, d, h)}{b(c, h)}, \tag{A.3}$$

where

$$e(c, d, h) = \sum_j p(\lambda_j; h) \, w_{x(c)}(-\lambda_j) \, w_{v(c,d)}(\lambda_j). \tag{A.4}$$

From (1.2), (A.2), $v_t(\gamma, \delta) = u_t^{\#}$, so that

$$\widehat{\nu}(\gamma, \delta, \theta) - \nu = \frac{e(\gamma)}{b(\gamma)}, \tag{A.5}$$

where

$$b(\gamma) = b(\gamma, \theta) = \sum_j q(\lambda_j) \left| w_{x(\gamma)}(\lambda_j) \right|^2, \tag{A.6}$$

$$e(\gamma) = e(\gamma, \delta, \theta) = \sum_j p(\lambda_j) w_{x(\gamma)}(-\lambda_j) w_u(\lambda_j), \tag{A.7}$$

with

$$w_u(\lambda) = \frac{1}{(2\pi n)^{\frac{1}{2}}} \sum_t u_t e^{it\lambda}, \tag{A.8}$$

$$p(\lambda) = p(\lambda; \theta), \quad q(\lambda) = q(\lambda; \theta). \tag{A.9}$$

Also define

$$e^*(\gamma) = \sum_m \left\{ \zeta x_m(\gamma) - \sum_{s=1}^{m-1} B_s \zeta x_{m-s}(\gamma) \right\}' \Omega^{-1} \varepsilon_m, \tag{A.10}$$

$$e^{**}(\gamma) = \zeta' B(1)' \Omega^{-1} \sum_m x_{m-1}(\gamma) \varepsilon_m, \tag{A.11}$$

$$b^*(\gamma) = \sum_m \left\{ \zeta x_m(\gamma) - \sum_{s=1}^{m-1} B_s \zeta x_{m-s}(\gamma) \right\}' \Omega^{-1} \left\{ \zeta x_m(\gamma) - \sum_{s=1}^{m-1} B_s \zeta x_{m-s}(\gamma) \right\}, \tag{A.12}$$

$$b^{**}(\gamma) = \frac{q(0)}{2\pi} \sum_m x_m^2(\gamma). \tag{A.13}$$

Now (3.11) for $\widehat{\nu}(\gamma, \delta, \theta)$ follows on establishing the following six propositions.

Proposition 1. *As* $n \to \infty$,

$$e(\gamma) - e^*(\gamma) = o_p(n^\beta). \tag{A.14}$$

Proposition 2. *As* $n \to \infty$,

$$e^*(\gamma) - e^{**}(\gamma) = o_p(n^\beta). \tag{A.15}$$

21

Proposition 3. *As* $n \to \infty$,

$$n^{-\beta} e^{**}(\gamma) \Rightarrow \zeta' B(1)' \Omega^{-1} \int_0^1 \widetilde{W}(r; \beta) \, dW(r). \qquad (A.16)$$

Proposition 4. *As* $n \to \infty$,

$$b(\gamma) - b^*(\gamma) = o_p\left(n^{2\beta}\right). \qquad (A.17)$$

Proposition 5. *As* $n \to \infty$,

$$b^*(\gamma) - b^{**}(\gamma) = o_p\left(n^{2\beta}\right). \qquad (A.18)$$

Proposition 6. *As* $n \to \infty$,

$$n^{-2\beta} b^{**}(\gamma) \Rightarrow \frac{q(0)}{2\pi} \int_0^1 \widetilde{W}(r; \beta)^2 \, dr, \qquad (A.19)$$

where the right side is almost surely positive.

To prove (3.11) for the remaining four estimates in (2.18), it suffices to consider only $\widehat{\nu}(\gamma, \delta, \widehat{\theta})$ and $\widehat{\nu}(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta})$ as the proof for the other, intermediate cases, will essentially be implied. It thus remains to show that

$$\widehat{\nu}(\gamma, \delta, \widehat{\theta}) - \widehat{\nu}(\gamma, \delta, \theta) = o_p\left(n^\beta\right), \qquad (A.20)$$

$$\widehat{\nu}(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}) - \widehat{\nu}(\gamma, \delta, \widehat{\theta}) = o_p\left(n^\beta\right). \qquad (A.21)$$

We have first

$$\widehat{\nu}(\gamma, \delta, \widehat{\theta}) - \nu = \frac{e(\gamma, \delta, \widehat{\theta})}{b(\gamma, \widehat{\theta})}, \qquad (A.22)$$

so that, from (A.3), the left side of (A.20) is

$$\frac{e(\gamma, \delta, \widehat{\theta}) - e(\gamma, \delta, \theta)}{b(\gamma, \widehat{\theta})} + e(\gamma, \delta, \theta) \left\{ \frac{1}{b(\gamma, \widehat{\theta})} - \frac{1}{b(\gamma, \theta)} \right\}. \qquad (A.23)$$

In view of Propositions 1-6, the proof of (A.20) follows on establishing the following two propositions.

Proposition 7. *As* $n \to \infty$,

$$e(\gamma, \delta, \widehat{\theta}) - e(\gamma, \delta, \theta) = o_p\left(n^\beta\right). \qquad (A.24)$$

Proposition 8. *As* $n \to \infty$,

$$b(\gamma, \widehat{\theta}) - b(\gamma, \theta) = o_p\left(n^{2\beta}\right). \qquad (A.25)$$

22

To prove (A.21), note that

$$\widehat{\nu}(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}) - \nu = \frac{e(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta})}{b(\widehat{\gamma}, \widehat{\theta})}, \tag{A.26}$$

so from (A.22) the left side of (A.21) is

$$\frac{e(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}) - e(\widehat{\gamma}, \widehat{\delta}, \theta) - e(\gamma, \delta, \widehat{\theta}) + e(\gamma, \delta, \theta)}{b(\widehat{\gamma}, \widehat{\theta})}$$
$$+ \frac{e(\widehat{\gamma}, \widehat{\delta}, \theta) - e(\gamma, \delta, \theta)}{b(\widehat{\gamma}, \widehat{\theta})} - \frac{e(\gamma, \delta, \widehat{\theta})}{b(\widehat{\gamma}, \widehat{\theta})b(\gamma, \widehat{\theta})} \{b(\widehat{\gamma}, \theta) - b(\gamma, \theta)\}$$
$$- \frac{e(\gamma, \delta, \widehat{\theta})}{b(\widehat{\gamma}, \widehat{\theta})b(\gamma, \widehat{\theta})} \left\{ b(\widehat{\gamma}, \widehat{\theta}) - b(\widehat{\gamma}, \theta) - b(\gamma, \widehat{\theta}) + b(\gamma, \theta) \right\}, \tag{A.27}$$

and (A.21) follows from Propositions 1-8 on establishing the following four propositions.

Proposition 9. *As* $n \to \infty$,

$$e(\widehat{\gamma}, \widehat{\delta}, \theta) - e(\gamma, \delta, \theta) = o_p\left(n^\beta\right). \tag{A.28}$$

Proposition 10. *As* $n \to \infty$,

$$e(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}) - e(\widehat{\gamma}, \widehat{\delta}, \theta) - e(\gamma, \delta, \widehat{\theta}) + e(\gamma, \delta, \theta) = o_p\left(n^\beta\right). \tag{A.29}$$

Proposition 11. *As* $n \to \infty$,

$$b(\widehat{\gamma}, \theta) - b(\gamma, \theta) = o_p\left(n^{2\beta}\right). \tag{A.30}$$

Proposition 12. *As* $n \to \infty$,

$$b(\widehat{\gamma}, \widehat{\theta}) - b(\widehat{\gamma}, \theta) - b(\gamma, \widehat{\theta}) + b(\gamma, \theta) = o_p\left(n^{2\beta}\right). \tag{A.31}$$

APPENDIX B: PROOFS OF PROPOSITIONS

Proof of Proposition 1

Write $e(\gamma)$ as

$$\frac{\zeta'}{n} \sum_j \sum_{l=-\infty}^{\infty} \overline{B}_l' e^{-il\lambda_{\mathrm{j}}} \Omega^{-1} \sum_{m=-\infty}^{\infty} \overline{B}_m e^{im\lambda_{\mathrm{j}}} \sum_s x_s\left(\gamma\right) e^{-is\lambda_{\mathrm{j}}} \sum_t u_t e^{it\lambda_{\mathrm{j}}}, \tag{B.1}$$

23

taking $\overline{B}_l = 0$, $l < 0$, $\overline{B}_0 = I_2$, $\overline{B}_l = -B_l$, $l > 0$. We can rewrite this as

$$\frac{\zeta'}{n} \sum_s \sum_t \sum_j \sum_{l=-\infty}^{\infty} \overline{B}'_{l-s} e^{-i(l-s)\lambda_j} \Omega^{-1} \sum_{m=-\infty}^{\infty} \overline{B}_{m-t} e^{i(m-t)\lambda_j} x_s(\gamma) e^{-is\lambda_j} u_t e^{it\lambda_j}$$

$$= \sum_{m=1}^{\infty} \sum_{r=-\infty}^{\infty} \left\{ \sum_s \overline{B}_{m-s+rn} \zeta x_s(\gamma) \right\}' \Omega^{-1} \sum_t \overline{B}_{m-t} u_t, \qquad (B.2)$$

because

$$\sum_j e^{it\lambda_j} = n, \ t = 0, \mathrm{mod}\,(n)\,; = 0, \ \text{otherwise}. \qquad (B.3)$$

The expectation of the absolute value of the difference between (B.2) and

$$\sum_{m=1}^{n} \sum_{r=-\infty}^{\infty} \left\{ \sum_s \overline{B}_{m-s+rn} \zeta x_s(\gamma) \right\}' \Omega^{-1} \sum_t \overline{B}_{m-t} u_t \qquad (B.4)$$

is bounded by

$$K \sum_{m=n+1}^{\infty} \left[E \left\| \sum_{r=-\infty}^{\infty} \sum_s \overline{B}_{m-s+rn} \sum_{v=1}^{s} a_{s-v} u_{2v} \right\|^2 E \left\| \sum_t \overline{B}_{m-t} u_t \right\|^2 \right]^{\frac{1}{2}}, \qquad (B.5)$$

with $a_t = a_t(\beta)$, where throughout K denotes a generic positive constant. The second expectation is bounded by

$$tr \left\{ \sum_t \sum_s \int_{-\pi}^{\pi} B_{m-t} f(\lambda) B'_{m-s} e^{i(s-t)\lambda} d\lambda \right\} \le K \int_{-\pi}^{\pi} \left\| \sum_t B_{m-t} e^{-it\lambda} \right\|^2 d\lambda$$

$$\le K \sum_t \|B_{m-t}\|^2 \le K \sum_{t=m-n}^{\infty} \|B_t\|^2 \qquad (B.6)$$

for $m > n$. The first expectation in (B.5) is bounded by

$$tr \left\{ \int_{-\pi}^{\pi} \sum_{r=-\infty}^{\infty} \sum_s \sum_{v=1}^{s} \overline{B}_{m-s+rn} a_{s-v} e^{-iv\lambda} f_{22}(\lambda) \sum_{q=-\infty}^{\infty} \sum_t \sum_{w=1}^{t} \overline{B}'_{m-t+qn} a_{t-w} e^{iw\lambda} d\lambda \right\}$$

24

$$\leq \quad K \int_{-\pi}^{\pi} \left\| \sum_{r=-\infty}^{\infty} \sum_{s} \sum_{v=1}^{s} \overline{B}_{m-s+rn} a_{s-v} e^{-iv\lambda} \right\|^2 d\lambda$$

$$\leq \quad K \sum_{r=-\infty}^{\infty} \sum_{s} \sum_{q=-\infty}^{\infty} \sum_{t} \left\| \overline{B}_{m-s+rn} \right\| \left\| \overline{B}_{m-t+qn} \right\| \sum_{v=1}^{\min(s,t)} a_{s-v} a_{t-v},$$

$$\text{(B.7)}$$

where $f_{ii}(\lambda)$ is the $(i, i)th$ element of $f(\lambda)$, and thus is bounded. From Lemma D.2, (B.7) is bounded by

$$Kn^{2\beta-1} \left(\sum_{l=0}^{\infty} \|B_l\| \right)^2 = O\left(n^{2\beta-1}\right), \tag{B.8}$$

using (3.4). It follows that (B.5) is bounded by

$$Kn^{\beta-\frac{1}{2}} \sum_{m=n+1}^{\infty} \left(\sum_{t=m-n}^{\infty} \|B_t\|^2 \right)^{\frac{1}{2}} \quad \leq \quad Kn^{\beta-\frac{1}{2}} \sum_{m=1}^{\infty} \left(\sum_{t=m}^{\infty} \|B_t\|^2 \right)^{\frac{1}{2}}$$

$$\leq \quad Kn^{\beta-1/2} \sum_{m=1}^{\infty} \sum_{t=m}^{\infty} \|B_t\|$$

$$\leq \quad Kn^{\beta-\frac{1}{2}} \sum_{j=1}^{\infty} j \|B_j\| = O(n^{\beta-\frac{1}{2}}), \quad \text{(B.9)}$$

again using (3.4).

Next, the expectation of the absolute value of the difference between (B.4) and

$$\sum_{m=1}^{n} \sum_{r=-\infty}^{\infty} \left\{ \sum_{s} \overline{B}_{m-s+rn} \zeta x_s(\gamma) \right\}' \Omega^{-1} \varepsilon_m \tag{B.10}$$

is bounded by

$$K \sum_{m=1}^{n} \left[E \left\| \sum_{r=-\infty}^{\infty} \sum_{s} \overline{B}_{m-s+rn} \sum_{v=1}^{s} a_{s-v} u_{2v} \right\|^2 E \left\| \sum_{t=-\infty}^{0} \overline{B}_{m-t} u_t \right\|^2 \right]^{\frac{1}{2}}. \tag{B.11}$$

Proceeding as in (B.6), the second expectation is bounded by $K \sum_{t=m}^{\infty} \|B_t\|^2$, so since the first expectation is bounded by (B.8), it follows that (B.11) is bounded by

$$Kn^{\beta-\frac{1}{2}} \sum_{m=1}^{\infty} \left(\sum_{t=m+1}^{\infty} \|B_t\|^2 \right)^{\frac{1}{2}} = O(n^{\beta-\frac{1}{2}}), \tag{B.12}$$

25

as in (B.9). The expectation of the absolute value of the difference between (B.10) and $e^*(\gamma)$ is bounded by

$$K \sum_{m=1}^{n} \left[E \left\| \sum_{r>0} \sum_s \overline{B}_{m-s+rn} \sum_{v=1}^{s} a_{s-v} u_{2v} \right\|^2 \right]^{\frac{1}{2}} \leq Kn^{\beta-\frac{1}{2}} \sum_{m=1}^{n} \left[\sum_{r>0} \sum_s \left\| \overline{B}_{m-s+rn} \right\|^2 \right]^{\frac{1}{2}}$$

$$\leq Kn^{\beta-\frac{1}{2}} \sum_{m=1}^{n} \left[\sum_{t=m}^{\infty} \|B_t\|^2 \right]^{\frac{1}{2}}, \quad \text{(B.13)}$$

which is $O(n^{\beta-\frac{1}{2}})$, to complete the proof.

Proof of Proposition 2

Consider first the difference

$$\zeta' \sum_m \sum_{s=1}^{m} \overline{B}'_{m-s} d_{m-1,s}(\gamma) \Omega^{-1} \varepsilon_m, \quad \text{(B.14)}$$

where $d_{m-1,s}(\gamma) = x_{m-1}(\gamma) - x_s(\gamma)$. Because there is a contribution to the mean only when $s = m$, (B.14) has expectation

$$-\sum_m \zeta' \Omega^{-1} E\left[\varepsilon_m \varepsilon'_m\right] \xi = -n\zeta'\xi = 0. \quad \text{(B.15)}$$

(B.14) has variance $c_1 + c_2 + c_3$, where

$$c_1 = \sum_m \sum_q \sum_{s=1}^{m} \sum_{t=1}^{q} \zeta' \overline{B}'_{m-s} \Omega^{-1} E\left[\varepsilon_m \varepsilon'_q\right] \Omega^{-1} \overline{B}_{q-t} \zeta E\left[d_{m-1,s}(\gamma) d_{q-1,t}(\gamma)\right], \quad \text{(B.16)}$$

$$c_2 = \sum_m \sum_q \sum_{s=1}^{m} \sum_{t=1}^{q} \zeta' \overline{B}'_{m-s} \Omega^{-1} E\left[\varepsilon_m d_{q-1,t}(\gamma)\right] E\left[\varepsilon'_q d_{m-1,s}(\gamma)\right] \Omega^{-1} \overline{B}_{q-t} \zeta, \quad \text{(B.17)}$$

and c_3 is a fourth cumulant term to be described subsequently. We have

$$d_{m-1,s}(\gamma) = u_{2,m-1}(-\beta) - u_{2,s}(-\beta)$$

$$= \sum_{v=1}^{s} (a_{m-1-v} - a_{s-v}) u_{2v} + \sum_{v=s+1}^{m-1} a_{m-1-v} u_{2v} 1 (s \leq m-2),$$

(B.18)

with $a_t = a_t(\beta)$, $a_{-1} = 0$.

Considering first c_1, there is a contribution only when $q = m$, and then $|E\left[d_{m-1,s}(\gamma) d_{q-1,t}(\gamma)\right]|$ is

$$\left| \int_{-\pi}^{\pi} f_{22}(\lambda) r_{sm}(-\lambda) r_{tm}(\lambda) d\lambda \right| \leq K \left\{ \int_{-\pi}^{\pi} f_{22}(\lambda) |r_{sm}(\lambda)|^2 d\lambda \int_{-\pi}^{\pi} f_{22}(\lambda) |r_{tm}(\lambda)|^2 d\lambda \right\}^{1/2}$$

$$\leq K(r_{sm} r_{tm})^{1/2}, \quad \text{(B.19)}$$

26

writing

$$r_{sm}(\lambda) \;=\; \sum_{v=1}^{s}(a_{m-1-v}-a_{s-v})e^{iv\lambda}+\sum_{v=s+1}^{m-1}a_{m-1-v}e^{iv\lambda}1\,(s\le m-2)\,,$$

(B.20)

$$r_{sm} \;=\; \sum_{v=1}^{s}(a_{m-1-v}-a_{s-v})^2+\sum_{v=s+1}^{m-1}a_{m-1-v}^2 1\,(s\le m-2)\,. \qquad \text{(B.21)}$$

Then (B.19) is bounded by $K\left\{|m-s-1|\,|m-t-1|\right\}^{\frac{1}{2}}m^{\max(0,2\beta-2)}$, on taking $t=m-2$ in Lemma D.3 for $s\le m-2$, then noting that $r_{m-1,m}=0$, and that $r_{mm}=\sum_{v=1}^{m-1}(a_{m-v}-a_{m-1-v})^2+1=O(m^{\max(0,2\beta-2)})$, on applying Lemma D.3 with $s=m-1,t=m$. It follows that

$$\begin{aligned}|c_1| &\le K\sum_m m^{\max(0,2\beta-2)}\left\{\sum_{j=0}^{m}j^{\frac{1}{2}}\,\|B_j\|\right\}^2\\ &= O\,(n)\,1(1/2<\beta\le 1)+O\left(n^{2\beta-1}\right)1(\beta>1)\,.\end{aligned}$$

(B.22)

Next, note that c_2 is zero unless $m=q=s=t$, so $c_2=O\,(n)=o(n^{2\beta})$. Finally, the fourth cumulant term, c_3, involves the fourth cumulant of $\varepsilon_m,\varepsilon_q,x_{m-1}(\gamma)-x_s(\gamma)$, $x_{q-1}(\gamma)-x_t(\gamma)$, which is easily seen to be zero unless $m=q=s=t$, so that $c_3=O\,(n)$ also.

It remains to show that

$$\zeta'\sum_m\left\{B\,(1)-\sum_{s=1}^{m}\overline{B}_{m-s}\right\}' x_{m-1}(\gamma)\,\Omega^{-1}\varepsilon_m = o_p\left(n^{\beta}\right)\,. \qquad \text{(B.23)}$$

Clearly the left side has mean zero. Its variance is, from arguments similar to those above, bounded by

$$K\sum_m\left\|B\,(1)-\sum_{s=0}^{m-1}\overline{B}_s\right\|^2 Ex_{m-1}^2(\gamma)\,E\,\|\varepsilon_m\|^2 \le K\sum_m\left(\sum_{s=m}^{\infty}\|B_s\|\right)^2 m^{2\beta-1}, \quad \text{(B.24)}$$

because $Ex_m^2(\gamma)=O(m^{2\beta-1})$ from Robinson and Marinucci (2001). Then, (B.24) is $o\left(n^{2\beta}\right)$ from the Toeplitz lemma, to complete the proof.

Proof of Proposition 3. Note that $\zeta'B\,(1)'\,\Omega^{-1}\varepsilon_m$ has mean zero and variance $q\,(0)/2\pi$; in view of Theorem 1 of Marinucci and Robinson (2000) and Assumption 1, the proof follows by Theorem 2.2 of Kurtz and Protter (1991).

Proof of Proposition 4. Omitted, as it is similar to the proof of Proposition 1 but significantly easier, especially in view of the norming $n^{-2\beta}$ rather than $n^{-\beta}$.

Proof of Proposition 5. This is likewise omitted due to its similarity to, and simplicity relative to, the proof of Proposition 2.

27

Proof of Proposition 6. Follows straightforwardly from Marinucci and Robinson (2000), the continuous mapping theorem and Assumption 1, and the fact that $\widetilde{W}(r;\beta)$ is almost surely nonzero, from (3.8), (3.9).

Proof of Proposition 7. By the mean-value theorem, $p(\lambda;\widehat{\theta}) - p(\lambda;\theta) = (\widehat{\theta} - \theta)'\overline{P}(\lambda)$, where $\overline{P}(\lambda)$ is the matrix $P(\lambda;h) = \partial p(\lambda;h)/\partial h$, with columns evaluated respectively at $\overline{\theta}^{(1)}$, $\overline{\theta}^{(2)}$, where $\left\|\overline{\theta}^{(i)} - \theta\right\| \leq \left\|\widehat{\theta} - \theta\right\|$, $i = 1, 2$. Writing $P(\lambda) = P(\lambda;\theta)$,

$$
\begin{aligned}
\sup_\lambda \left\|\overline{P}(\lambda) - P(\lambda)\right\| &\leq 2\sup_{h\in N_\epsilon} \sup_\lambda \|P(\lambda;h) - P(\lambda)\| \\
&\quad + 4\sup_{h\in\Theta} \sup_\lambda \|P(\lambda;h)\| \, 1\left(\left|\widehat{\theta} - \theta\right| \geq \epsilon\right),
\end{aligned}
\tag{B.25}
$$

where $\epsilon > 0$ and $N_\epsilon = \{h : \|h - \theta\| < \epsilon\}$. Noting Assumption 2 parts (ii) and (iii), since continuity in h for all λ implies uniform continuity on the compact set $[-\pi, \pi]$, the first term on the right of (B.25) tends to 0 as $\epsilon \to 0$. The second term is $o_p(1)$ as $n \to \infty$ for $\epsilon > 0$ from Assumption 2 (ii) and (iii) and Assumption 3 (ii). It follows that

$$
\begin{aligned}
\left\|\sum_j \{\overline{P}(\lambda_j) - P(\lambda_j)\} w_{x(\gamma)}(-\lambda_j)w_u(\lambda_j)\right\| &= o_p\left(\sum_j \|w_{x(\gamma)}(-\lambda_j)w_u(\lambda_j)\|\right) \\
&= o_p\left(\left\{\sum_t x_t^2(\gamma) \sum_t \|u_t\|^2\right\}^{\frac{1}{2}}\right),
\end{aligned}
\tag{B.26}
$$

which is $o_p\left(n^{\beta+\frac{1}{2}}\right)$, where we use the Cauchy inequality, (B.3), $\sum_t \|u_t\|^2 = O_p(n)$ and

$$
\sum_t x_t(\gamma)^2 = O_p\left(n^{2\beta}\right),
\tag{B.27}
$$

from Robinson and Marinucci (2001). Thus, noting Assumption 3 (ii), it remains to show that

$$
\sum_j P(\lambda_j)w_{x(\gamma)}(-\lambda_j)w_u(\lambda_j) = o_p\left(n^{\beta+\frac{1}{2}}\right).
\tag{B.28}
$$

Denote by $P_L(\lambda)$ the partial sum, to L terms, of the Fourier series of $P(\lambda)$, so

$$
P_L(\lambda) = \sum_{l=-L}^{L} P_l e^{-il\lambda}, \quad P_l = \frac{1}{2\pi}\int_{-\pi}^{\pi} P(\lambda)e^{il\lambda}d\lambda.
\tag{B.29}
$$

From Assumption 2 (ii) and (v), and Zygmund (1977, p.64),

$$
\sup_\lambda \|P(\lambda) - P_L(\lambda)\| = O\left(\frac{\log L}{L}\right),
\tag{B.30}
$$

28

as $L \to \infty$. Thus

$$\left\| \sum_j \{P(\lambda_j) - P_L(\lambda_j)\} w_{x(\gamma)}(-\lambda_j) w_u(\lambda_j) \right\| \le K \frac{\log L}{L} \left\{ \sum_t x_t(\gamma)^2 \sum_t \|u_t\|^2 \right\}^{\frac{1}{2}},$$
(B.31)

proceeding as in (B.26). With $L = [n^{\frac{1}{2}}]$, (B.31) is $O_p\left((\log n) n^\beta\right) = o_p\left(n^{\beta+\frac{1}{2}}\right).$

On the other hand, for $L < n$,

$$\sum_j P_L(\lambda_j) w_{x(\gamma)}(-\lambda_j) w_u(\lambda_j) = \frac{1}{2\pi} \sum_{l=-L}^{L} P_l \left\{ \sum_{t(l)}{}' x_t(\gamma) u_{t+l} \right.$$

$$\left. + \sum_{t(l)}{}'' x_t(\gamma) u_{t+l+n} + \sum_{t(l)}{}''' x_t(\gamma) u_{t+l-n} \right\},$$
(B.32)

where

$$\sum_{t(l)}{}' = \sum_{1 \le t, t+l \le n}, \quad \sum_{t(l)}{}'' = \sum_{1 \le t, t+l+n \le n}, \quad \sum_{t(l)}{}''' = \sum_{1 \le t, t+l-n \le n},$$
(B.33)

on applying (B.3). Looking first at the second and third terms in (B.32), we note that $1 \le t, t+l+n \le n$ and $1 \le t, t+l-n \le n$ are equivalent, respectively, to $1 \le t \le -l$, for $-L \le l \le -1$, and $1+n-l \le t \le n$, for $1 \le l \le L$. Then

$$E \left\| \sum_{t(l)}{}'' x_t(\gamma) u_{t+l+n} + \sum_{t(l)}{}''' x_t(\gamma) u_{t+l-n} \right\| \le K |l| \sum_{s=0}^{n-1} |a_s(\beta)| \le K |l| n^\beta,$$
(B.34)

from Lemma D.1. Thus, because Assumption 2 (ii) and (v) implies

$$\sum_{l=-\infty}^{\infty} |l| \|P_l\| < \infty,$$
(B.35)

(Zygmund, 1977, p.240), the contribution from the final two terms of (B.32) is $O_p\left(n^\beta\right).$ Finally

$$\sum_{t(l)}{}' x_t(\gamma) u_{t+l} = O_p\left(n^{\max(\beta,1)}\right),$$
(B.36)

uniformly in l, from Lemmas C.1 and C.2, which, with (B.35) and Assumption 3 (ii), completes the proof of (B.28).

Proof of Proposition 8. Follows similarly to, but more easily than, the proof of Proposition 7.

Proof of Proposition 9. The left side of (A.28) is

$$\sum_j p(\lambda_j) \left\{ w_{x(\widehat{\gamma})}(-\lambda_j) - w_{x(\gamma)}(-\lambda_j) \right\} \left\{ w_{v(\widehat{\gamma},\widehat{\delta})}(\lambda_j) - w_u(\lambda_j) \right\} \quad \text{(B.37)}$$

$$+ \sum_j p(\lambda_j) w_{x(\gamma)}(-\lambda_j) \left\{ w_{v(\widehat{\gamma},\widehat{\delta})}(\lambda_j) - w_u(\lambda_j) \right\} \quad \text{(B.38)}$$

$$+ \sum_j p(\lambda_j) \left\{ w_{x(\widehat{\gamma})}(-\lambda_j) - w_{x(\gamma)}(-\lambda_j) \right\} w_u(\lambda_j). \quad \text{(B.39)}$$

Consider first (B.39)). Noting Assumption 2 (ii) and (iv) and proceeding as in the proof of Proposition 7, define

$$p_L(\lambda) = \sum_{l=-L}^{L} p_l e^{-il\lambda}, \quad p_l = \frac{1}{2\pi} \int_{-\pi}^{\pi} p(\lambda) e^{il\lambda} d\lambda, \quad \text{(B.40)}$$

where

$$\sup_\lambda \|p(\lambda) - p_L(\lambda)\| = O\left(\frac{\log L}{L}\right), \quad \sum_{l=-\infty}^{\infty} |l| \|p_l\| < \infty. \quad \text{(B.41)}$$

Thus

$$\sum_j \left\{ p(\lambda_j) - p_L(\lambda_j) \right\} \left\{ w_{x(\widehat{\gamma})}(-\lambda_j) - w_{x(\gamma)}(-\lambda_j) \right\} w_u(\lambda_j) \quad \text{(B.42)}$$

is bounded in norm by

$$\frac{K \log L}{L} \left\{ \sum_t \{x_t(\widehat{\gamma}) - x_t(\gamma)\}^2 \sum_t \|u_t\|^2 \right\}^{\frac{1}{2}}, \quad \text{(B.43)}$$

using the Cauchy inequality and (B.3) again. Now choosing $L = [n^{\frac{1}{2}}]$ and taking $c = \delta - \gamma = \beta$, $\widehat{c} = \delta - \widehat{\gamma}$ in Lemma C.5, (B.43) is $O_p\left((\log n)^2 n^{\beta - \rho}\right) = o_p(n^\beta)$.

On the other hand, for $L < n$,

$$\sum_j p_L(\lambda_j) \left\{ w_{x(\widehat{\gamma})}(-\lambda_j) - w_{x(\gamma)}(-\lambda_j) \right\} w_u(\lambda_j) = \frac{1}{2\pi} \sum_{l=-L}^{L} p_l \left[\sum_{t(l)}' \{x_t(\widehat{\gamma}) - x_t(\gamma)\} u_{t+l} \right.$$

$$\left. + \sum_{t(l)}'' \{x_t(\widehat{\gamma}) - x_t(\gamma)\} u_{t+l+n} + \sum_{t(l)}''' \{x_t(\widehat{\gamma}) - x_t(\gamma)\} u_{t+l-n} \right]. \quad \text{(B.44)}$$

30

As in the proof of Lemma C.5, we can write, for any $R \geq 2$,

$$
\begin{aligned}
x_t(\widehat{\gamma}) - x_t(\gamma) &= u_{2t}(\widehat{\gamma} - \delta) - u_{2t}(-\beta) \\
&= \sum_{r=1}^{R-1} \frac{(\gamma - \widehat{\gamma})^r}{r!} g^{(r)}\left(u_{2t}; \beta\right) + \frac{(\gamma - \widehat{\gamma})^R}{R!} g^{(R)}\left(u_{2t}; \delta - \overline{\gamma}\right),
\end{aligned}
$$

(B.45)

where, for a vector or scalar sequence φ_t, and real $b \geq 0$,

$$
g^{(r)}(\varphi_t; b) = \sum_{s=1}^{t-1} a_s^{(r)}(b) \varphi_{t-s},
$$

(B.46)

with $a_s^{(r)}(b) = (d^r/db^r)a_s(b)$ and $|\overline{\gamma} - \gamma| \leq |\widehat{\gamma} - \gamma|$. Applying (C.14) of Lemma C.4 with $r = R$, $c = \beta$, $\widehat{c} = \delta - \overline{\gamma}$, and Assumption 3 (i), indicates that the final term in (B.45) is uniformly $O_p\left(n^{-R\rho} t^{\beta+\epsilon}\right)$, for any $\epsilon > 0$. Thus, the contribution of this term to (B.44) is, by the Cauchy inequality and (B.41), $O_p\left(n^{\beta+\epsilon+1-R\rho}\right)$, which is $o_p\left(n^{\beta}\right)$ on choosing R large enough.

Next, as in (B.34), we have

$$
E \left\| \sum_{t(l)}'' g^{(r)}\left(u_{2t}; \beta\right) u_{t+l+n} + \sum_{t(l)}''' g^{(r)}\left(u_{2t}; \beta\right) u_{t+l-n} \right\| \leq K |l| (\log n)^r n^{\beta}, \quad \text{(B.47)}
$$

applying again Lemma D.1, so on taking account of the $(\gamma - \widehat{\gamma})^r$ factors and invoking Assumption 3 (i) and (B.41), the contribution of the sums $\sum_{t(l)}''$ and $\sum_{t(l)}'''$ to (B.44) is $O_p\left((\log n)^r n^{\beta-\rho}\right) + o_p\left(n^{\beta}\right) = o_p\left(n^{\beta}\right)$. It remains to consider the quantities

$$
(\gamma - \widehat{\gamma})^r \sum_{l=-L}^{L} p_l \sum_{t(l)}' g^{(r)}\left(u_{2t}; \beta\right) u_{t+l}, \quad 1 \leq r \leq R-1.
$$

(B.48)

From (C.2) of Lemma C.1 and (C.8) of Lemma C.2 the sum over $\sum_{t(l)}'$ is $O_p\left((\log n)^r n^{\max(\beta,1)}\right)$, and thus, using (B.41) and Assumption 3 (i), (B.48) is $O_p\left(n^{\max(\beta,1)-\rho} \log n\right)$ for $\rho > \max(0, 1-\beta)$, that is, $o_p\left(n^{\beta}\right)$. This completes the proof that (B.39) is $o_p\left(n^{\beta}\right)$.

We next consider (B.38), and again wish to replace $p(\lambda)$ by $p_L(\lambda)$. First

$$
\left\| \sum_j \{p(\lambda_j) - p_L(\lambda_j)\} w_{x(\gamma)}(-\lambda_j) \left\{ w_{v(\widehat{\gamma}, \widehat{\delta})}(\lambda_j) - w_u(\lambda_j) \right\} \right\|
$$

(B.49)

is bounded by

$$
\frac{K \log L}{L} \left\{ \sum_t x_t(\gamma)^2 \sum_t \left\| v_t(\widehat{\gamma}, \widehat{\delta}) - u_t \right\|^2 \right\}^{\frac{1}{2}}.
$$

(B.50)

31

Noting that $v_t(\widehat{\gamma}, \widehat{\delta}) = \left(u_{1t}(\widehat{\gamma} - \gamma), u_{2t}(\widehat{\delta} - \delta) \right)'$ the second factor in braces is $\sum_t \left\| v_t(\widehat{\gamma}, \widehat{\delta}) - u_t \right\|^2 = O_p\left(n^{1-2\rho}\right)$ from Lemma C.5, so that, choosing $L = [n^{\frac{1}{2}}]$, and using (B.27), (B.49) is $O_p\left((\log n)\, n^{\beta - \rho}\right) = o_p\left(n^\beta\right)$.

Next, proceeding as above, for $R \geq 2$,

$$\sum_j p_L(\lambda_j)\, w_{x(\gamma)}(-\lambda_j) \left\{ w_{v(\widehat{\gamma}, \widehat{\delta})}(\lambda_j) - w_u(\lambda_j) \right\}$$

$$= \frac{1}{2\pi} \sum_{l=-L}^{L} p_l \sum_{r=1}^{R-1} \frac{1}{r!} \begin{pmatrix} (\widehat{\gamma} - \gamma)^r & 0 \\ 0 & (\widehat{\delta} - \delta)^r \end{pmatrix} \sideset{}{'}\sum_{t(l)} x_t(\gamma) g^{(r)}(u_{t+l}; 0) + o_p\left(n^\beta\right),$$

(B.51)

and the leading term is $o_p\left(n^\beta\right)$ from (C.3) of Lemma C.1 and (C.9) of Lemma C.2, (B.41) and Assumption 3 (i).

We are left with (B.37). It is clear from its structure, which involves both the differences appearing in (B.38) and (B.39), that application of similar arguments to those above will show it is $o_p\left(n^\beta\right)$, so we omit the details.

Proof of Proposition 10. The left side of (A.29) has norm bounded by

$$K \sup_\lambda \left\| f(\lambda; \widehat{\theta})^{-1} - f(\lambda; \theta)^{-1} \right\| \left[\left\{ \sum_j \left| w_{x(\widehat{\gamma})}(\lambda_j) \right|^2 \sum_j \left\| w_{v(\widehat{\gamma}, \widehat{\delta})}(\lambda_j) - w_u(\lambda_j) \right\|^2 \right\}^{\frac{1}{2}} \right.$$

$$\left. + \left\{ \sum_j \left| w_{x(\widehat{\gamma})}(\lambda_j) - w_{x(\gamma)}(\lambda_j) \right|^2 \sum_j \left\| w_u(\lambda_j) \right\|^2 \right\}^{\frac{1}{2}} \right],$$

(B.52)

and this is clearly $O_p\left(n^{\beta - \rho + \epsilon}\right)$ for any $\epsilon > 0$, from earlier arguments.

Proof of Proposition 11. Omitted, being similar to but easier than the proof of Proposition 9.

Proof of Proposition 12. Omitted, in view of the remarks about the proofs of Propositions 10 and 11.

APPENDIX C: TECHNICAL LEMMAS

Lemma C.1. *Uniformly in $l \in [-L, L]$, $L < n$,*

$$E \left\{ \sideset{}{'}\sum_{t(l)} x_t(\gamma) u_{t+l} \right\} = O\left(n^{\max(\beta, 1)}\right),$$

(C.1)

32

$$E\left\{\sideset{}{'}\sum_{t(l)} g^{(r)}(u_{2t}; \beta)u_{t+l}\right\} = O\left((\log n)^r n^{\max(\beta,1)}\right), \tag{C.2}$$

$$E\left\{\sideset{}{'}\sum_{t(l)} x_t(\gamma)g^{(r)}(u_{t+l}; 0)\right\} = O\left((\log n)^r n^{\max(\beta,1)}\right). \tag{C.3}$$

Proof. The proofs are very similar, and in fact are possible under milder conditions following techniques of Robinson and Marinucci (2001), and we just discuss the proof of (C.3), which is slightly the most complicated. Writing $\Gamma_s = E\left(u_{2t}u_{t+s}\right)$, the left side is

$$\sideset{}{'}\sum_{t(l)} \sum_{s=1}^{t-1} a_s(\beta) \sum_{q=r}^{t+l-1} a_q^{(r)}(0)\Gamma_{s+l-q}, \tag{C.4}$$

which has norm bounded by

$$\sum_t \sum_{q=r}^{n} \left|a_q^{(r)}(0)\right| \sum_s \|\Gamma_s\| = O\left((\log n)^r n\right) \tag{C.5}$$

for $\beta < 1$, uniformly in l, and by

$$n^{\beta-1}\sum_t \sum_{q=r}^{n} \left|a_q^{(r)}(0)\right| \sum_s \|\Gamma_s\| = O\left((\log n)^r n^\beta\right) \tag{C.6}$$

for $\beta \geq 1$, by Lemma D.4 and Assumption 2, to complete the proof.

Lemma C.2. *Uniformly in $l \in [-L, L]$, $L < n$,*

$$V\left\{\sideset{}{'}\sum_{t(l)} x_t(\gamma)u_{t+l}\right\} = O\left(n^{2\beta}\right), \tag{C.7}$$

$$V\left\{\sideset{}{'}\sum_{t(l)} g^{(r)}(u_{2t}; \beta)u_{t+l}\right\} = O\left((\log n)^{2r} n^{2\beta}\right), \tag{C.8}$$

$$V\left\{\sideset{}{'}\sum_{t(l)} x_t(\gamma)g^{(r)}(u_{t+l}; 0)\right\} = O\left(n^{2\beta+\eta}\right). \tag{C.9}$$

for any $\eta > 0$.

Proof. The results follow from minor modifications of the proof of Theorem 5.1 of Robinson and Marinucci (2001). There are only two differences. The first is that the sums in the latter reference are over $t \in [1, n]$, whereas the Lemma requires uniformity in l for sums over $t(l)$. But because the $t(l)$ are just a subset of $[1, n]$, this follows easily. The second difference is that in (C.8) and (C.9) (though not in (C.7)), the weights $a_s^{(r)}(\beta)$ and $a_s^{(r)}(0)$ that are involved are not covered by the weights of Robinson and Marinucci (2001), due to the presence of log factors. But allowance for such log factors is readily made, and they contribute the $(\log n)^{2r}$ and n^η factors in (C.8) and (C.9). We observe that the regularity conditions of Robinson and Marinucci (2001) are noticeably weaker than those on u_t in the present paper.

Lemma C.3. *For $i = 1, 2$, and uniformly in $r \geq 1$ and $t \geq 2$,*

$$E\left\{g^{(r)}(u_{it}; 0)^2\right\} = O(1), \tag{C.10}$$

and for $c > 1/2$

$$E\left\{g^{(r)}(u_{it}; c)^2\right\} = O((\log t)^{2r} t^{2c-1}). \tag{C.11}$$

Proof. For any $c \geq 0$,

$$
\begin{aligned}
E\left\{g^{(r)}(u_{it}; c)^2\right\} &= \sum_{s=1}^{t-1}\sum_{v=1}^{t-1} a_s^{(r)}(c) a_v^{(r)}(c) \int_{-\pi}^{\pi} f_{ii}(\lambda) e^{i(s-v)\lambda} d\lambda \\
&= \int_{-\pi}^{\pi} f_{ii}(\lambda) \left|\sum_{s=1}^{t-1} a_s^{(r)}(c) e^{is\lambda}\right|^2 d\lambda \leq K \int_{-\pi}^{\pi} \left|\sum_{s=1}^{t-1} a_s^{(r)}(c) e^{is\lambda}\right|^2 d\lambda \\
&\leq K \sum_{s=1}^{t-1} a_s^{(r)}(c)^2. \tag{C.12}
\end{aligned}
$$

From Lemmas D.1 and D.4, this is bounded by the right sides of (C.10) and (C.11), for $c = 0$ and $c > 1/2$ respectively.

Lemma C.4. *For $i = 1, 2$, $\rho > 0$, uniformly in $t \in [1, n]$, $r \geq 1$,*

$$g^{(r)}(u_{it}; \widehat{c}) = O_p(t^{\frac{1}{2}}) \tag{C.13}$$

if $\widehat{c} = O_p(n^{-\rho})$, and

$$g^{(r)}(u_{it}; \widehat{c}) = O_p(t^{c+\epsilon}) \tag{C.14}$$

for any $\epsilon > 0$, if $\widehat{c} = c + O_p(n^{-\rho})$, $c > 1/2$.

Proof. By the Cauchy inequality, for any $c \geq 0$,

$$\left|g^{(r)}(u_{it}; \widehat{c})\right| \leq \left\{\sum_{s=1}^{t-1} a_s^{(r)}(\widehat{c})^2 \sum_{s=1}^{t-1} u_{is}^2\right\}^{\frac{1}{2}}. \tag{C.15}$$

From Lemma D.5, for $\epsilon > 0$,

$$\sum_{s=0}^{t-1} a_s^{(r)}(\widehat{c})^2 = O_p\left(\sum_{s=0}^{t-1} \{\log(s+1)\}^{2r} (s+1)^{2(c+\epsilon-1)}\right), \tag{C.16}$$

where $c = 0$ or $c > 1/2$. Thus, with $\sum_{s=1}^{t-1} u_{is}^2 = O_p(t)$, the bounds (C.13) and (C.14) follow.

Lemma C.5. *For $i = 1, 2$, if $\widehat{c} = c + O_p(n^{-\rho})$, $\rho > 0$, uniformly in $t \in [1, n]$, as $n \to \infty$*

$$u_{it}(-\widehat{c}) - u_{it} = O_p(n^{-\rho}), \quad c = 0, \tag{C.17}$$

$$u_{it}(-\widehat{c}) - u_{it}(-c) = O_p\left(n^{-\rho}t^{c-\frac{1}{2}}\log t\right), \quad c > \frac{1}{2}. \tag{C.18}$$

Proof. We have, for $c \geq 0$,

$$u_{it}(-\widehat{c}) - u_{it}(-c) = \sum_{s=1}^{t-1} \{a_s(\widehat{c}) - a_s(c)\} u_{i,t-s}, \tag{C.19}$$

with $u_{it}(0) = u_{it}$. By Taylor's theorem, for any $R \geq 2$,

$$a_s(\widehat{c}) - a_s(c) = \sum_{r=1}^{R-1} a_s^{(r)}(c)\frac{(\widehat{c}-c)^r}{r!} + a_s^{(R)}(\overline{c})\frac{(\widehat{c}-c)^R}{R!}, \tag{C.20}$$

where $|\overline{c} - c| \leq |\widehat{c} - c|$, so we can write (C.19) as

$$\sum_{r=1}^{R-1} \frac{(\widehat{c}-c)^r}{r!} g^{(r)}(u_{it}; c) + \frac{(\widehat{c}-c)^R}{R!} g^{(R)}(u_{it}; \overline{c}). \tag{C.21}$$

Taking $c = 0$, (C.10) and (C.13) indicate that (C.21) is $O_p(n^{-\rho}) + O_p\left(n^{-R\rho}t^{\frac{1}{2}}\right)$, whence (C.17) is proved by choosing R large enough and observing that $t \leq n$. In the same way, (C.18) is proved because (C.21) is $O_p\left(n^{-\rho}t^{c-\frac{1}{2}}\log t\right) + O_p\left(n^{-R\rho}t^{c+\eta}\right)$ for $\eta > 0$, due to (C.11) and (C.14).

APPENDIX D: LEMMAS CONCERNING THE a_s WEIGHTS

Lemma D.1. *For $c \in [c_0, C_0]$, $c_0 > 0$, $C_0 < \infty$, $s \geq 0$,*

$$|a_s(c)| \leq K_0(1+s)^{c-1}, \tag{D.1}$$

$$|a_s(c) - a_{s+1}(c)| \leq K_0(1+s)^{c-2}, \tag{D.2}$$

$$\left|a_s^{(r)}(c)\right| \leq K_{0R}(\log(1+s))^r (1+s)^{c-1}, \ 1 \leq r \leq R, \tag{D.3}$$

where $K_0 < \infty$ depends only on c_0 and C_0 and $K_{0R} < \infty$ depends only on c_0, C_0 and R.

Proof. First, (D.1) is familiar from Stirling's approximation, or derivable by induction, while (D.2) follows easily from the identity $a_{s+1}(c) = \{(s+c)/(s+1)\} a_s(c)$. To prove (D.3), introduce the digamma function and its derivatives

$$\psi(x) = \frac{d}{dx} \log \Gamma(x), \quad \psi^{(r)}(x) = \frac{d^r \psi(x)}{dx^r}, \tag{D.4}$$

which exist for $r \geq 1$ and $x > 0$. We deduce from the chain rule that

$$a_s^{(r)}(c) = \sum_{i=0}^{r-1} \tau_i \left\{ \psi^{(i)}(s+c) - \psi^{(i)}(c) \right\} a_s^{(r-1-i)}(c), \tag{D.5}$$

with the convention that $\psi^{(0)}(\cdot) = \psi(\cdot)$, $a^{(0)}(\cdot) = a(\cdot)$, and for finite constants τ_i, $0 \leq i \leq r-1$. Now from Gradshteyn and Ryzhik (1994, p.95), for $x > 0$

$$\psi(x) = \sum_{i=0}^{\infty} \frac{x-1}{(i+1)(x+i)} - \eta, \tag{D.6}$$

where η is Euler's constant. Thus for $x > 0$

$$\begin{aligned} |\psi(x)| &\leq \sum_{i=0}^{[x]} (i+1)^{-1} + |x-1| \sum_{i=[x]+1}^{\infty} i^{-2} + \eta \\ &\leq \log(x+1) + 1 + \eta \leq K \log(x+1), \end{aligned} \tag{D.7}$$

where $[.]$ denotes integer part and K is independent of x. Also, for $l \geq 1$,

$$\psi^{(l)}(x) = (-1)^{l+1} l! \sum_{i=0}^{\infty} (x+i)^{-l-1}, \tag{D.8}$$

so that

$$\left| \psi^{(l)}(x) \right| \leq l!(x^{-l-1} + \frac{x^{-l}}{l}) \leq K_{0R}(1+x)^{-l}, \tag{D.9}$$

$1 \leq l \leq r \leq R$, for $x \geq c_0$. The proof is completed by applying (D.5) recursively, (D.9), and noting that $|\log(s+c+1)| \leq K_0 \log(s+1)$.

Lemma D.2. *Uniformly in $s,t \in [1,n]$, for $c > \frac{1}{2}$*

$$\sum_{v=1}^{\min(s,t)} a_{s-v}(c) a_{t-v}(c) = O\left(n^{2c-1}\right). \tag{D.10}$$

Proof. From (D.1), the left side of (D.10) is bounded in absolute value by $K \sum_{v=1}^{n} v^{c-1} (v + |s - t|)^{c-1}$. Since $(v + |s - t|)^{c-1} \leq v^{c-1}$ for $c \leq 1$ and $(v + |s - t|)^{c-1} \leq K n^{c-1}$ for $c > 1$, (D.10) readily follows.

Lemma D.3. *For $1 \leq s \leq t - 1$, $c > 1/2$,*

$$\sum_{v=1}^{s} \{a_{t-v}(c) - a_{s-v}(c)\}^2 + \sum_{v=s+1}^{t} a_{t-v}^2(c) \leq K(t-s)t^{\max(0, 2c-2)}. \qquad (D.11)$$

Proof. Writing $a_s = a_s(c)$, for $1 \leq v \leq s$, $a_{t-v} - a_{s-v} = 0$, $c = 1$, while for $c \neq 1$ we have from (D.2)

$$|a_{t-v} - a_{s-v}| \leq \sum_{r=s+1}^{t} |a_{r-v} - a_{r-1-v}| \leq K \sum_{r=s+1}^{t} (r - v)^{c-2}. \qquad (D.12)$$

Now (D.12) is bounded on the one hand by $K (s + 1 - v)^{c-1} 1(c < 1) + K t^{c-1} 1(c > 1)$, and on the other by $K (t - s) \left\{ (s + 1 - v)^{c-2} 1(c < 2) + t^{c-2} 1(c \geq 2) \right\}$. It follows that (D.12) is also bounded by

$$K (t - s)^{\frac{1}{2}} (s + 1 - v)^{c - \frac{3}{2}}, \quad \frac{1}{2} < c < 1, \qquad (D.13)$$

$$K (t - s)^{\frac{1}{2}} t^{\frac{c-1}{2}} (s + 1 - v)^{\frac{c}{2} - 1}, \quad 1 < c < 2, \qquad (D.14)$$

$$K (t - s)^{\frac{1}{2}} t^{c - \frac{3}{2}}, \quad c \geq 2. \qquad (D.15)$$

Thus $\sum_{v=1}^{s} \{a_{t-v}(c) - a_{s-v}(c)\}^2$ is bounded by

$$K (t - s) \sum_{v=1}^{s} (s + 1 - v)^{2c-3} \leq K (t - s), \quad \frac{1}{2} < c < 1, \qquad (D.16)$$

$$K (t - s) t^{c-1} \sum_{v=1}^{s} (s + 1 - v)^{c-2} \leq K (t - s) t^{2(c-1)}, \quad 1 < c < 2, \qquad (D.17)$$

$$K (t - s) t^{2c-3} s \leq K (t - s) t^{2(c-1)}, \quad c \geq 2, \qquad (D.18)$$

that is by $K (t - s) t^{\max(0, 2c-2)}$, $c > \frac{1}{2}$. On the other hand, for all $c > 1/2$

$$\sum_{v=s+1}^{t} a_{t-v}^2 \leq K (t - s)^{2c-1}, \qquad (D.19)$$

whence the result immediately follows.

Lemma D.4. *For $r \geq 1$*

$$a_s^{(r)}(0) = 0, \quad s < r \qquad (D.20)$$

37

and

$$\left| a_s^{(r)}(0) \right| \leq \frac{K_r \left(\log\left(s+1 \right) \right)^{r-1}}{\left(s-r+1 \right)}, \quad s \geq r, \tag{D.21}$$

where $K_r < \infty$ depends only on r.

Proof. On taking logs in (1.4) and differentiating with respect to α we have

$$-\log\left(1-z \right) \left(1-z \right)^{-\alpha} = \sum_{s=0}^{\infty} a_s^{(1)}(\alpha) z^s. \tag{D.22}$$

Evaluating this expression at $\alpha = 0$ gives $a_0^{(1)}(0) = 0$ and $a_s^{(1)}(0) = s^{-1}$, $s \geq 1$. This proves the lemma for $r = 1$. For $r > 1$ we differentiate (D.22) $r-1$ times and evaluate at $\alpha = 0$ to get

$$\left\{ -\log\left(1-z \right) \right\}^r = \sum_{s=0}^{\infty} a_s^{(r)}(0) z^s. \tag{D.23}$$

Clearly $a_s^{(r)}(0) = 0$, $s < r$. Also, we have the recursion

$$\sum_{s=0}^{\infty} a_s^{(r)}(0) z^s = -\log\left(1-z \right) \sum_{s=0}^{\infty} a_s^{(r-1)}(0) z^s, \quad r \geq 2. \tag{D.24}$$

It follows that

$$a_s^{(r)}(0) = \frac{a_{r-1}^{(r-1)}(0)}{s-r+1} + \frac{a_r^{(r-1)}(0)}{s-r} + \ldots + a_{s-1}^{(r-1)}(0), \quad s \geq r > 1. \tag{D.25}$$

If (D.21) is true with r replaced by $r-1$ we have

$$\begin{aligned}
\left| a_s^{(r)}(0) \right| &\leq K_{r-1} \left(\log\left(s+1 \right) \right)^{r-2} \left\{ \frac{1}{1.\left(s-r+1 \right)} + \frac{1}{2\left(s-r \right)} + \ldots + \frac{1}{1.\left(s-r+1 \right)} \right\} \\
&\leq 2K_{r-1} \left\{ \log\left(s+1 \right) \right\}^{r-2} \frac{\log\left(s+1 \right)}{s-r+1} \leq K_r \frac{\left(\log\left(s+1 \right) \right)^{r-1}}{s-r+1} \tag{D.26}
\end{aligned}$$

for $K_r \geq 2K_{r-1}$. The proof thus follows by induction.

Lemma D.5. *Let $\widehat{c} = c + O_p\left(n^{-\rho} \right)$, $\rho > 0$ such that $0 \leq c < K$ and $\left| \widehat{c} \right| \leq K$ for some $K < \infty$, and suppose $\overline{c}$ satisfies $\left| \overline{c} - c \right| \leq \left| \widehat{c} - c \right|$. Then uniformly in $s \in [0, n)$ as $n \to \infty$, and for any $\epsilon > 0$,*

$$a_s^{(r)}(\overline{c}) = O_p\left(\left(\log\left(s+1 \right) \right)^r \left(s+1 \right)^{c+\epsilon-1} \right) \tag{D.27}$$

as $n \to \infty$.

Proof. From Lemma D.1 and Lemma D.4 we have, for any $\epsilon > 0$

$$
\begin{aligned}
\left| a_s^{(r)}\left(\overline{c}\right) \right| &\leq \left| a_s^{(r)}\left(\overline{c}\right) \right| 1\left(|\widehat{c} - c| \leq \epsilon\right) + \left| a_s^{(r)}\left(\overline{c}\right) \right| 1\left(|\widehat{c} - c| > \epsilon\right) \\
&\leq K\left(\log\left(s+1\right)\right)^r \left(\left(s+1\right)^{c+\epsilon-1} + \left(s+1\right)^{K-1} \frac{|\widehat{c} - c|^M}{\epsilon^M} \right) \\
&\leq K\left(\log\left(s+1\right)\right)^r \left(\left(s+1\right)^{c+\epsilon-1} + \left(s+1\right)^{K-1} n^{-M\rho} \right) \qquad \text{(D.28)}
\end{aligned}
$$

for any $M \geq 1$. We may choose $M \geq (K - c - \epsilon)/\rho$ which, with $s \leq n$, completes the proof.

REFERENCES

AKONOM, J., AND C. GOURIEROUX (1987): "A Functional Central Limit Theorem for Fractional Processes," preprint, CEPREMAP, Paris.

BLOOMFIELD, P. (1972): "An Exponential Model for the Spectrum of a Scalar Time Series," *Biometrika,* 60, 217-226.

CHAN, N.H., AND N. TERRIN (1995): "Inference for Unstable Long-Memory Processes With Applications to Fractional Unit Root Autoregressions," *Annals of Statistics,* 23, 1662-1683.

CHEUNG, Y. W., AND K. S. LAI (1993): "A Fractional Cointegration Analysis of Purchasing Power Parity," *Journal of Business and Economic Statistics,* 11, 103-112.

COGBURN, I., AND H. T. DAVIS (1974): "Periodic Splines and Spectral Estimation," *Annals of Statistics,* 2, 1108-1126.

CORBAE, D., AND S. OULIARIS (1988): "Cointegration and Tests of Purchasing Power Parity," *Review of Economics and Statistics*, 70, 508-511.

DIEBOLD, F., AND C. RUDEBUSCH (1991): "On the Power of Dickey-Fuller Tests Against Fractional Alternatives," *Economics Letters*, 35, 155-160.

DOLADO, J., AND F. MARMOL (1996): "Efficient Estimation of Cointegrating Relationships Among Higher Order and Fractionally Integrated Processes," preprint, Banco de Espana-Servicio de Estudios, Madrid.

ENDERS, W. (1988): "Arima and Cointegration Tests of PPP under Fixed and Flexible Exchange Rate Regimes," *The Review of Economics and Statistics*, 70, 504-508.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Cointegration and Error Correction: Representation, Estimation, and Testing," *Econometrica,* 55, 251-276.

GRADSHTEYN, I. S., AND I.M. RYZHIK (1994): *Table of Integrals, Series, and Products.* Boston: Academic Press.

HANNAN, E. J. (1963): "Regression for Time Series," in *Time Series Analysis,* (M. Rosenblatt, Ed). New York: Wiley, pp.17-32.

HANNAN, E. J. (1973): "The Asymptotic Theory of Linear Time Series Models," *Journal of Applied Probability,* 10, 130-145.

HASSLER, U., F. MARMOL, AND C. VELASCO (2002): "Residual Log-Periodogram Inference for Long-Run Relationships," preprint, Universidad Carlos III, Madrid.

HUALDE, J., AND P.M. ROBINSON (2001): "Root-n-Consistent Estimation of Weak Fractional Cointegration," preprint, London School of Economics, London.

JEGANATHAN, P. (1999): "On Asymptotic Inference in Cointegrated Time Series With Fractionally Integrated Errors," *Econometric Theory,* 15, 583-621.

JEGANATHAN, P. (2001): "Correction to 'On Asymptotic Inference in Cointegrated Time Series With Fractionally Integrated Errors'," preprint, University of Michigan, Ann Arbor.

JOHANSEN, S. (1991): "Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models," *Econometrica,* 59, 1551-1580.

KIM, C. S., AND P. C. B. PHILLIPS (2000): "Fully Modified Estimation of Fractional Cointegration Models," preprint, Yale University, New Haven.

KIM, Y. (1990): "Purchasing Power Parity: Another Look at the Long-Run Data," *Economic Letters*, 32, 334-339.

KURTZ, T. G., AND P. PROTTER (1991): "Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations," *Annals of Probability,* 19, 1035-1070.

MARINUCCI, D., AND P. M. ROBINSON (1999): "Alternative Forms of Fractional Brownian Motion," *Journal of Statistical Planning and Inference,* 80, 111-122.

——— (2000): "Weak Convergence of Multivariate Fractional Processes," *Stochastic Processes and their Applications,* 86, 103-120.

——— (2001): "Semiparametric Fractional Cointegration Analysis," *Journal of Econometrics,* 105, 225-247.

OBSTFELD, M., AND A.M. TAYLOR (2002): "Global Capital Markets: Integration, Crisis and Growth," forthcoming, *Japan-U.S. Center Sanwa Monographs on International Financial Markets*, Cambridge University Press, Cambridge.

PHILLIPS, P. C. B. (1991a): "Optimal Inference in Cointegrated Systems," *Econometrica,* 59, 283-306.

———— (1991b): "Spectral Regression for Cointegrated Time Series," , in W.A Barnett, J. Powell and G. Tauchen (eds.) *Nonparametric and Semiparametric Methods in Econometrics and Statistics,* Cambridge: Cambridge University Press.

PHILLIPS, P. C. B., AND B.E. HANSEN (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies,* 57, 99-125.

ROBINSON, P. M. (1978): "Alternative Models for Stationary Stochastic Processes," *Stochastic Processes and their Applications,* 8, 141-152.

———— (1991): "Automatic Frequency-Domain Inference on Semiparametric and Nonparametric Models," *Econometrica,* 59, 1329-1363.

———— (1994): "Efficient Tests of Nonstationary Hypotheses", *Journal of the American Statistical Association,* 89, 1420-1437.

———— (1995): "Log-periodogram Regression of Time Series with Long Range Dependence," *Annals of Statistics,* 23, 1048-1072.

———— (2002): "The Distance Between Rival Nonstationary Fractional Processes," preprint, London School of Economics, London.

ROBINSON, P. M., AND D. MARINUCCI (1997): "Semiparametric Frequency-Domain Analysis of Fractional Cointegration," preprint, London School of Economics, London.

———— (2000): "The Averaged Periodogram for Nonstationary Vector Time Series," *Statistical Inference for Stochastic Processes,* 3, 149-160.

———— (2001): "Narrow-Band Analysis of Nonstationary Processes," *Annals of Statistics,* 29, 947-986.

ROBINSON, P. M., AND Y. YAJIMA (2001): "Determination of Cointegrating Rank in Fractional Systems," *Journal of Econometrics,* 106, 217-241.

SAIKKONEN, P. (1995): "Problems With the Asymptotic Theory of Maximum Likelihood in Integrated and Cointegrated Systems," *Econometric Theory,* 11, 888-911.

SILVEIRA, G. (1991): *Contributions to Strong Approximations in Time Series With Applications in Nonparametric Statistics and Functional Central Limit Theorems,* Ph.D Thesis, University of London.

TAYLOR, M.P. (1988): "An Empirical Examination of Long Run Purchasing Power Parity Using Cointegration Techniques," *Applied Economics,* 20, 1369-1381.

VELASCO, C. (2000): "Gaussian Semiparametric Estimation of Fractional Cointegration," preprint, Universidad Carlos III, Madrid.

VELASCO, C., AND P. M. ROBINSON (2000): "Whittle Pseudo-Maximum Likelihood Estimation for Nonstationary Time Series," *Journal of the American Statistical Association,* 95, 1229-1243.

ZYGMUND, A. (1977): *Trigonometric Series.* Cambridge: Cambridge University Press.

TABLE II
MONTE CARLO BIAS OF $\overline{\nu}_I, \overline{\nu}_F, \overline{\nu}_O$ FOR $\rho = 0.5$, 1000 REPLICATIONS

τ	γ	δ	$n = 64$			$n = 128$			$n = 256$		
			$\overline{\nu}_I$	$\overline{\nu}_F$	$\overline{\nu}_O$	$\overline{\nu}_I$	$\overline{\nu}_F$	$\overline{\nu}_O$	$\overline{\nu}_I$	$\overline{\nu}_F$	$\overline{\nu}_O$
.5	0	.6	.002	-.056	.269	.004	-.019	.223	.002	-.013	.185
	0	1.2	.001	-.002	.010	.000	-.001	.003	.000	.000	.001
	0	2	.000	.000	-.001	.000	.000	.000	.000	.000	.000
	.4	.6	-.002	-.119	.472	.013	-.136	.441	.008	-.119	.404
	.4	1.2	.002	-.013	.052	.002	-.003	.030	.001	-.002	.016
	.4	2	.000	-.001	.000	.000	.000	.000	.000	.000	.000
1	0	.6	.001	-.040	.194	.003	-.014	.160	.001	-.010	.133
	0	1.2	.001	-.002	.007	.000	-.001	.002	.000	.000	.001
	0	2	.000	.000	-.001	.000	.000	.000	.000	.000	.000
	.4	.6	-.001	-.040	.341	.009	-.073	.318	.006	-.086	.291
	.4	1.2	.001	-.009	.038	.001	-.002	.022	.001	-.001	.012
	.4	2	.000	.000	.000	.000	.000	.000	.000	.000	.000
2	0	.6	.001	-.029	.137	.002	-.009	.113	.001	-.007	.094
	0	1.2	.000	-.001	.005	.000	.000	.002	.000	.000	.000
	0	2	.000	.000	.000	.000	.000	.000	.000	.000	.000
	.4	.6	-.001	-.060	.241	.007	-.070	.225	.004	-.061	.206
	.4	1.2	.001	-.007	.027	.001	-.002	.015	.000	-.001	.008
	.4	2	.000	.000	.000	.000	.000	.000	.000	.000	.000

TABLE III
MONTE CARLO SD OF $\overline{\nu}_I, \overline{\nu}_F, \overline{\nu}_O$ FOR $\rho = 0.5$, 1000 REPLICATIONS

τ	γ	δ	$n = 64$			$n = 128$			$n = 256$		
			$\overline{\nu}_I$	$\overline{\nu}_F$	$\overline{\nu}_O$	$\overline{\nu}_I$	$\overline{\nu}_F$	$\overline{\nu}_O$	$\overline{\nu}_I$	$\overline{\nu}_F$	$\overline{\nu}_O$
.5	0	.6	.142	.345	.140	.081	.119	.105	.048	.064	.079
	0	1.2	.030	.033	.031	.012	.013	.013	.005	.005	.005
	0	2	.004	.004	.004	.001	.001	.001	.000	.000	.000
	.4	.6	.599	3.27	.284	.387	2.48	.244	.262	.968	.196
	.4	1.2	.083	.110	.094	.043	.050	.056	.023	.026	.031
	.4	2	.011	.011	.013	.003	.003	.004	.001	.001	.001
1	0	.6	.101	.247	.100	.058	.085	.075	.034	.046	.057
	0	1.2	.021	.024	.022	.009	.009	.009	.004	.004	.004
	0	2	.003	.003	.003	.001	.001	.001	.000	.000	.000
	.4	.6	.426	2.39	.203	.276	1.80	.174	.187	.741	.140
	.4	1.2	.059	.079	.067	.031	.036	.040	.016	.018	.022
	.4	2	.008	.008	.009	.002	.002	.003	.001	.001	.001
2	0	.6	.072	.176	.071	.041	.061	.053	.025	.033	.040
	0	1.2	.015	.017	.016	.006	.007	.006	.003	.003	.003
	0	2	.002	.002	.002	.000	.000	.001	.000	.000	.000
	.4	.6	.305	1.67	.145	.197	1.26	.124	.134	.493	.100
	.4	1.2	.042	.056	.048	.022	.026	.029	.012	.013	.016
	.4	2	.005	.006	.007	.002	.002	.002	.001	.001	.001

43

TABLE IV
EMPIRICAL SIZES OF W_I AND W_F FOR $\rho = 0$, 1000 REPLICATIONS

		$\alpha = .05$						$\alpha = .10$					
	n	64	64	128	128	256	256	64	64	128	128	256	256
γ	δ	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F
0	.6	.072	.194	.049	.125	.052	.090	.131	.261	.099	.166	.116	.137
0	1.2	.059	.198	.062	.136	.048	.097	.113	.260	.110	.208	.122	.161
0	2	.054	.184	.057	.122	.058	.102	.109	.255	.108	.199	.120	.167
.4	.6	.077	.154	.055	.097	.062	.076	.126	.234	.095	.159	.110	.132
.4	1.2	.060	.193	.050	.115	.051	.076	.125	.254	.109	.176	.099	.131
.4	2	.051	.177	.071	.133	.059	.104	.108	.238	.123	.201	.121	.157

TABLE V
EMPIRICAL SIZES OF W_I AND W_F FOR $\rho = 0.5$, 1000 REPLICATIONS

		$\alpha = .05$						$\alpha = .10$					
	n	64	64	128	128	256	256	64	64	128	128	256	256
γ	δ	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F
0	.6	.064	.238	.054	.152	.052	.116	.128	.322	.113	.224	.105	.178
0	1.2	.067	.203	.057	.132	.053	.097	.122	.289	.108	.202	.104	.157
0	2	.065	.201	.055	.133	.059	.108	.116	.272	.112	.193	.111	.160
.4	.6	.055	.255	.057	.180	.051	.153	.137	.338	.115	.274	.107	.231
.4	1.2	.067	.231	.051	.153	.049	.110	.127	.312	.102	.207	.092	.168
.4	2	.065	.184	.055	.114	.058	.095	.122	.254	.114	.187	.111	.149

TABLE VI
EMPIRICAL SIZES OF W_I AND W_F FOR $\rho = -0.5$, 1000 REPLICATIONS

		$\alpha = .05$						$\alpha = .10$					
	n	64	64	128	128	256	256	64	64	128	128	256	256
γ	δ	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F
0	.6	.062	.227	.059	.166	.059	.129	.128	.311	.120	.231	.109	.203
0	1.2	.047	.209	.074	.161	.052	.095	.105	.292	.129	.225	.100	.149
0	2	.049	.199	.073	.163	.063	.112	.110	.264	.129	.222	.109	.157
.4	.6	.070	.263	.057	.190	.062	.163	.108	.332	.117	.268	.122	.239
.4	1.2	.056	.238	.061	.167	.050	.109	.120	.318	.117	.222	.103	.174
.4	2	.049	.186	.074	.146	.066	.094	.097	.248	.134	.214	.105	.152

TABLE VII
EMPIRICAL SIZES OF W_I AND W_F FOR $\rho = 0.75$, 1000 REPLICATIONS

		$\alpha = .05$						$\alpha = .10$					
	n	64	64	128	128	256	256	64	64	128	128	256	256
γ	δ	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F
0	.6	.069	.332	.050	.259	.052	.247	.120	.416	.107	.337	.104	.327
0	1.2	.066	.231	.054	.144	.053	.100	.127	.311	.099	.217	.112	.158
0	2	.054	.221	.042	.144	.064	.104	.122	.293	.104	.208	.112	.150
.4	.6	.065	.430	.058	.372	.060	.339	.108	.502	.115	.451	.108	.428
.4	1.2	.065	.292	.048	.195	.057	.141	.130	.383	.110	.278	.111	.199
.4	2	.064	.210	.054	.130	.060	.097	.123	.267	.110	.193	.112	.148

TABLE VIII
MONTE CARLO BIAS(SD) OF $\overline{\nu}_I, \overline{\nu}_\gamma, \overline{\nu}_\delta, \overline{\nu}_F, \overline{\nu}_O$, FOR $\delta = 1$, $\gamma = 0$, 1000 REPLICATIONS

ρ	τ n	1 64	1 256	2 64	2 256	.5 64	.5 256
0	$\overline{\nu}_I$	-.002(.041)	.000(.009)	-.001(.029)	.000(.007)	-.003(.058)	.000(.013)
	$\overline{\nu}_\gamma$	-.002.(043)	.000(.009)	-.001(.031)	.000(.007)	-.002(.060)	-.001(.013)
	$\overline{\nu}_\delta$	-.002(.043)	.000(.009)	-.001(.030)	.000(.007)	-.002(.060)	.000(.013)
	$\overline{\nu}_F$	-.001(.044)	.000(.009)	-.001(.031)	.000(.007)	-.002(.061)	.000(.013)
	$\overline{\nu}_O$	-.002(.040)	.000(.009)	-.002(.029)	.000(.007)	-.003(.056)	.000(.013)
.25	$\overline{\nu}_I$	.000(.004)	.000(.009)	.000(.029)	.000(.006)	.000(.057)	.000(.012)
	$\overline{\nu}_\gamma$	-.005(.044)	.000(.009)	-.004(.031)	.000(.006)	-.007(.062)	-.001(.012)
	$\overline{\nu}_\delta$	.002(.042)	.000(.009)	.001(.030)	.000(.006)	.002(.058)	.000(.013)
	$\overline{\nu}_F$	-.002(.044)	.000(.009)	-.002(.031)	.000(.006)	-.003(.062)	-.001(.013)
	$\overline{\nu}_O$	.014(.040)	.004(.009)	.010(.029)	.003(.006)	.020(.056)	.005(.013)
.5	$\overline{\nu}_I$	.001(.035)	.000(.008)	.001(.025)	.000(.006)	.001(.049)	.000(.011)
	$\overline{\nu}_\gamma$	-.010(.043)	-.001(.008)	-.007(.030)	.000(.006)	-.014(.060)	-.001(.011)
	$\overline{\nu}_\delta$	.004(.037)	.000(.008)	.003(.026)	.000(.006)	.005(.052)	.000(.012)
	$\overline{\nu}_F$	-.005(.043)	.000(.008)	-.003(.031)	.000(.006)	-.006.(060)	-.001(.012)
	$\overline{\nu}_O$	.030(.040)	.007(.010)	.021(.028)	.005(.007)	.041(.056)	.010(.014)
-.5	$\overline{\nu}_I$	.000(.033)	.000(.008)	.000(.024)	.000(.006)	.000(.046)	.000(.011)
	$\overline{\nu}_\gamma$	.009(.040)	.001(.009)	.007(.029)	.001(.006)	.013(.056)	.001(.012)
	$\overline{\nu}_\delta$	-.003(.035)	.000(.008)	-.002(.025)	.000(.006)	-.004(.049)	.000(.012)
	$\overline{\nu}_F$	.004(.039)	.001(.009)	.003(.028)	.000(.006)	.005(.055)	.001(.012)
	$\overline{\nu}_O$	-.028(.039)	-.007(.010)	-.020(.028)	-.005(.007)	-.039(.054)	-.010(.014)
.75	$\overline{\nu}_I$	.001(.026)	.000(.006)	.000(.019)	.000(.004)	.001(.037)	.000(.009)
	$\overline{\nu}_\gamma$	-.016(.042)	-.001(.007)	-.012(.030)	-.001(.005)	-.023(.059)	-.001(.010)
	$\overline{\nu}_\delta$	.004(.031)	.000(.007)	.003(.022)	.000(.005)	.005(.043)	.000(.010)
	$\overline{\nu}_F$	-.008(.042)	-.001(.007)	-.005(.030)	-.001(.005)	-.010(.059)	-.001(.010)
	$\overline{\nu}_O$	.044(.043)	.011(.011)	.031(.030)	.008(.007)	.061(.060)	.015(.015)

45

TABLE IX

EMPIRICAL SIZES OF $W_I, W_\gamma, W_\delta, W_F, W_O$ FOR $\delta = 1$, $\gamma = 0$, 1000 REPLICATIONS

ρ	α / n	.05					.10				
		W_I	W_γ	W_δ	W_F	W_O	W_I	W_γ	W_δ	W_F	W_O
	64	.061	.055	.199	.200	.058	.122	.125	.267	.264	.122
0	128	.053	.053	.126	.126	.052	.107	.107	.191	.191	.113
	256	.048	.048	.090	.090	.046	.118	.115	.154	.153	.109
	64	.069	.077	.192	.204	.075	.124	.130	.270	.275	.137
.25	128	.051	.056	.130	.130	.069	.116	.117	.203	.197	.114
	256	.051	.051	.092	.089	.064	.102	.105	.143	.149	.119
	64	.066	.106	.199	.218	.126	.127	.175	.266	.297	.213
.5	128	.056	.067	.137	.140	.126	.113	.129	.209	.202	.201
	256	.053	.064	.085	.095	.107	.095	.117	.146	.159	.180
	64	.047	.104	.196	.223	.131	.110	.174	.274	.309	.210
-.5	128	.068	.086	.145	.159	.121	.114	.148	.221	.218	.205
	256	.045	.061	.093	.100	.119	.101	.123	.148	.156	.199
	64	.066	.185	.211	.254	.212	.122	.262	.280	.333	.331
.75	128	.052	.116	.153	.156	.204	.099	.190	.217	.224	.330
	256	.056	.094	.102	.115	.197	.109	.159	.170	.170	.306

TABLE X

PPP EMPIRICAL EXAMPLE: ESTIMATES OF ν AND WALD TESTS OF $\nu = 1$ FOR
MODELS 1-7 COMPUTED FROM THE LAST $n' = 113, ..., 123$ OBSERVATIONS OF
US/UK DATA

n'	123	122	121	120	119	118	117	116	115	114	113
$\overline{\nu}_1$	1.139	1.050	1.014	.952	.889	.875	.871	.867	.864	.875	.875
W_1	26.23	.352	.017	.163	.759	.940	.986	1.035	1.082	.903	.890
$\overline{\nu}_2$	1.294	.959	1.030	.995	.949	.941	.941	.938	.936	.944	.943
W_2	117.3	.231	.078	.002	.159	.208	.206	.226	.243	.181	.182
$\overline{\nu}_3$	1.113	1.084	1.017	.955	.889	.871	.866	.863	.859	.871	.868
W_3	18.64	1.070	.027	.161	.823	1.079	1.138	1.196	1.251	1.051	1.059
$\overline{\nu}_4$	1.290	.966	1.028	.997	.950	.939	.939	.936	.934	.942	.939
W_4	122.6	.178	.078	.001	.170	.241	.240	.263	.281	.212	.227
$\overline{\nu}_5$	1.274	1.042	1.025	.986	.940	.933	.932	.931	.929	.939	.936
W_5	112.2	.225	.055	.014	.230	.283	.283	.296	.306	.223	.239
$\overline{\nu}_6$	1.278	.960	1.015	.983	.939	.932	.931	.930	.927	.937	.935
W_6	114.9	.211	.019	.020	.241	.292	.292	.306	.325	.246	.255
$\overline{\nu}_7$	1.298	.999	1.048	1.024	.975	.961	.962	.956	.956	.963	.958
W_7	116.9	.000	.279	.052	.047	.109	.105	.138	.136	.096	.122