SEMIPARAMETRIC FREQUENCY DOMAIN ANALYSIS OF FRACTIONAL COINTEGRATION*

P M Robinson and D Marinucci
London School of Economics and Political Science

Contents:

Abstract
1. Introduction
2. Frequency Domain Least Squares
3. Stationary Cointegration
4. The Averaged Periodogram in Nonstationary Environments
5. Nonstationary Fractional Cointegration
6. Monte Carlo Evidence
7. Empirical Examples
8. Final Comments
Appendix
References
Tables 1 – 9

Discussion Paper
No. EM/98/348
March 1998

The Suntory Centre
Suntory and Toyota International Centres for Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.:020 –8955 6698

* Research supported by ESRC grant no.R000235892.

ABSTRACT

The concept of cointegration has principally been developed under the assumption that the raw data vector z_t is $I(1)$ and the cointegrating residual e_t is $I(0)$, but is also of interest in more general, including fractional, circumstances, where z_t is stationary with long memory and e_t is stationary with less memory, or where z_t is nonstationary while e_t is either less nonstationary or stationary, possibly with long memory. Inference rules based on estimates of the cointegrating vector that have been developed in the $I(1)/I(0)$ case appear to lose validity in the above circumstances, while the estimates themselves, including ordinary least squares, are typically inconsistent when z_t is stationary. Partitioning z_t into a scalar y_t and vector x_t, we consider a narrow-band frequency domain least squares estimate of y_t on x_t. This estimate is consistent under stationarity of z_t, whereas least squares is inconsistent due to correlation between x_t and e_t. This correlation does not prevent consistency of least squares when z_t is nonstationary, but it produces a larger second order bias relative to the frequency domain estimate in the $I(1)/I(0)$ case, and a slower rate of convergence in many circumstances in which z_t exhibits less-than-$I(1)$ nonstationarity. When e_t is itself nonstationary, the two estimates have a common limit distribution. Our conclusions in the $I(1)/I(0)$ case are supported by Monte Carlo simulations. A semiparametric methodology for fractional cointegration analysis is applied to data analyzed by Engle and Granger (1987) and Campbell and Shiller (1987).

Keywords: Fractional cointegration; narrow-band frequency analysis.
JEL No.: C22

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

1. INTRODUCTION

Cointegration analysis has developed as a major theme of time series econometrics since the article of Engle and Granger (1987), much applied interest prompting considerable methodological and theoretical development during the past decade. Numerous empirical studies have investigated the possibility of cointegration in areas of economics in which a long-run relationship can be conjectured between nonstationary variables, including stock prices and dividends, consumption and income, wages and prices, short- and long-run interest rates, monetary aggregates and nominal GNP, exchange rates and prices, GNP and public debt, and many others.

The bulk of theoretical and applied work has focused on what might be called the "$I(1)/I(0)$" paradigm. We say a scalar process u_t, $t = 0, 1, ...,$ is $I(0)$ ($u_t \equiv I(0)$) if it is covariance stationary and has spectral density that is finite and positive at zero frequency, and, for a scalar process v_t, that $v_t \equiv I(d)$, $d > 0$, if

$$(1-L)^d v_t = u_t 1(t > 0), \; t > 0, \tag{1.1}$$

where $1(.)$ is the indicator function and formally

$$(1-L)^d = \sum_{j=1}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} L^j, \; \Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx.$$

For $d > 0$, v_t generated by (1.1) is said to have long memory, d measuring the extent of the "memory". The initial condition that the innovation in (1.1) is zero for $t \leq 0$ implies that v_t is nonstationary for all $d > 0$. However, for $0 < d < \frac{1}{2}$ v_t is asymptotically stationary, while for $d \geq \frac{1}{2}$ it is not. In the former case we might replace the right hand side of (1.1) by u_t, to make v_t covariance stationary (and for convenience we do so in Section 3). In much of the existing literature the definition of $I(0)$ effectively employed relaxes covariance stationarity to a form of asymptotic stationarity or stable heterogeneity because limit theorems for relevant statistics are available in such settings (though the mixing conditions employed are typically in other respects stronger than ours, implying that the limiting analogue of the spectral density is bounded at all frequencies). The $I(1)/I(0)$ paradigm envisages a vector of economic variables z_t which are all $I(1)$ and are cointegrated if there exists a linear combination $e_t = \alpha' z_t$ which is $I(0)$, the prime denoting transposition. Some approaches employ parametric time series models, such as autoregression for z_t and white noise for e_t (e.g. Johansen (1988, 1991)), while others adopt a nonparametric characterization of $I(0)$ to cover a wide range of stationary behaviour (e.g. Phillips (1991a)). In both approaches, it is standard to test such assumptions on z_t and e_t. At the same time, the $I(1)$ and $I(0)$ classes are clearly highly specialized forms of, respectively, nonstationary and stationary processes, for example when nested in the $I(d)$ processes, for real-valued d.

To define fractional cointegration for a $p \times 1$ vector z_t whose i-th element $z_{it} \equiv I(d_i)$, $d_i > 0$, $i = 1, ..., p$, we say $z_t \equiv FCI(d_1, ..., d_p; d_e)$ if there exists a $p \times 1$ vector

$\alpha \neq 0$ such that $e_t = \alpha' z_t \equiv I(d_e)$ where $0 \leq d_e < \min_{1 \leq i \leq p} d_i$. This property is possible and meaningful if and only if $d_i = d_j$, some $i \neq j$; moreover, a necessary condition for α to be a cointegrating vector is that its i-th component be equal to zero if $d_i > d_j$ for all $j \neq i$.

In case $d_1 = ... = d_p = d$ it is usual to write $z_t \equiv CI(d, b)$, where $b = d - d_e$ measures the strength of the cointegrating relationship. Here the possibility of fractional (i.e. non-integer) d or d_e was mentioned in the original paper of Engle and Granger (1987). That paper focused, however, on the $FCI(1, ..., 1; 0)$ or $CI(1, 1)$ case, which we hereafter abbreviate to $CI(1)$, and several explanations can be suggested for the overwhelming interest of the subsequent literature in this case. First, unit roots can sometimes be viewed as a consequence of economic theory, for example the efficient markets hypothesis and the random walk hypothesis for consumption. Second, standard tests have failed to reject the unit root hypothesis in very many time series. Third, the computational implications of the unit root hypothesis, that simple differencing removes nonstationarity, are attractive. Fourth, asymptotic theory for statistics based on $I(0)$ sequences was much better developed than that for stationary $I(d)$ series with $d \neq 0$. Finally, rules of inference relating to stationary and nonstationary fractional $I(d)$ processes were not well developed.

On the other hand, as argued by Sims (1988), unit root theory in economics typically rests on strong assumptions, or provides only very approximate justification. The power of unit root tests has often been criticised, and the bulk of these have been directed against $I(0)$ alternatives (for example, stationary autoregressive ones), and data that seem consistent with the $I(1)$ hypothesis might well also be consistent with $I(d)$ behaviour on some interval of d values. Fractional differencing is not a prohibitive computational drawback nowadays, and in recent years progress has been made on rigorously justifying large sample inference for stationary $I(d)$ processes, in both parametric and nonparametric directions. Indeed the relative "smoothness" of the $I(d)$ class for real d can lead to standard limit distribution theory and optimality theory in situations where the $I(1)/I(0)$ approach entails nonstandard and discontinuous asymptotics; for example, tests of a unit root null against fractional alternatives were shown by Robinson (1994b) to have a standard null local distribution and optimal asymptotic local power (and to extend immediately to test the $I(d)$ hypothesis for any other value of d) whereas statistics for testing the value of an autoregressive parameter have nonstandard asymptotics at the unit root, as evaluated by Dickey and Fuller (1979) and others, but standard asymptotics at stationary points arbitrarily close to the unit circle.

It is important to recognize that the $CI(1)$ setting has entailed challenges which have been surmounted with ingenuity and difficulty. It is also important to appreciate that $CI(1)$-based inference rules are largely invalidated when in fact $(d_1, ..., d_p; d_e) \neq (1, ..., 1; 0)$. It is possible to imagine how aspects of the $CI(1)$ methodology and theory can be extended to more general $FCI\,(d_1, ..., d_p; d_e)$ situations where the d_i and d_e are, while not necessarily one and zero, known values, especially in view of limit theory for nonstationary fractional processes of Akonom and Gourieroux (1987), Silveira

(1991) and Chan and Terrin (1995). Indeed Dolado and Marmol (1996) have recently pursued this line of study. However, when non-integral d_i and d_e are envisaged, assuming their values seems somehow more arbitrary than stressing the $CI(1)$ case in an autoregressive setting. Moreover for $\frac{1}{2} < d_i < \frac{3}{2}$ Dolado and Marmol's (1996) definition of a nonstationary fractionally integrated process differs from ours, being the partial sum of stationary, long memory innovations, and leading to "cointegration with fractionally integrated errors" (as in Jeganathan (1996)). The parametrization we adopt for the nonstationary case (which corresponds to theirs for $d_i > \frac{3}{2}$) relies more directly on the linear expansion of the fractional differencing operator; see Section 4. With an empirical focus, the issue of fractional cointegration is also dealt with by Cheung and Lai (1993) and others. It seems of greatest interest to study the problem in the context of unknown orders of integration d_i and d_e in the observed and cointegrated processes, possibly less or greater than unity. For example, in circumstances where $CI(1)$ cointegration has been rejected it may be possible to find evidence of $FCI\,(d_1, ..., d_p; d_e)$ cointegration for some $(d_1, ..., d_p; d_e) \neq (1, ..., 1; 0)$. Our allowance for some "memory" remaining in the cointegrating residual e_t (i.e., $d_e > 0$), is appealing, especially recalling how d_e can be linked to the speed of convergence to long run equilibrium (compare for instance Diebold and Rudebusch (1989)).

Cointegration is commonly thought of as a stationary relation between nonstationary variables (so that $d_i \geq \frac{1}{2}$, for all i, $d_e < \frac{1}{2}$). Other circumstances covered by our definition of cointegration are also worth entertaining. One case is $d_e \geq \frac{1}{2}$, when both z_t and e_t are nonstationary. Another is $0 \leq d_i < \frac{1}{2}$, for all i, when both z_t and e_t are stationary.

The latter situation was considered by Robinson (1994a), as an application of limit theory for averages of periodogram ordinates on a degenerating frequency band in stationary long memory series. Ordinary least squares estimates (OLS) (and other "full-band" estimates such as generalized least squares) are inconsistent due to the usual simultaneous equation bias. Robinson (1994a) showed, in case of bivariate z_t, that a narrow-band frequency domain least squares (FDLS) estimate of (a normalized) α can be consistent. It is possible that some macroeconomic time series that have been modelled as nonstationary with a unit root could arise from stationary $I(d)$ processes with d near $\frac{1}{2}$, say, and interest in the phenomenon of cointegration of stationary variates has recently emerged in a finance context. Moreover, it is likely to be extremely difficult in practice to distinguish a z_t with unit root from one, say, composed additively of a stationary autoregression with a root near the unit circle, and a stationary long memory process.

FDLS is defined in the following section, after which in Section 3 we extend Robinson's (1994a) results to a more general stationary vector setting, with rates of convergence. We go on in Section 4 to establish results of some independent interest on the approximation of sample moments of nonstationary sequences by narrow-band periodogram averages. These results, which constitute the main technical innovation of the paper, are exploited in Section 5 to demonstrate the usefulness of FDLS for nonstationary z_t: correlation between z_t and e_t does not prevent consistency of OLS, but it produces a larger second order bias relative to FDLS in the $CI(1)$ case, and a slower rate of convergence in many circumstances in which z_t exhibits less-than-$I(1)$ nonstationarity. When e_t is itself nonstationary, the two estimates share a common limit distribution. Our theoretical result for the $CI(1)$ case is supported in finite samples by Monte Carlo simulations in Section 6. Section 7 describes a semiparametric methodology for investigating the question of cointegration in possibly fractional conditions, and applies it to series that were studied in the early papers of Engle and Granger (1987) and Campbell and Shiller (1987). Section 8 mentions possibilities for further work. Proofs are collected in an Appendix.

OLS by no means represents the state of the art in $CI(1)$ analysis. A number of more elaborate estimates have been proposed and shown to have advantages over OLS, such as Engle and Granger's (1987) two step regression; Johansen's (1988, 1991) maximum likelihood estimate for the error-correction mechanism (ECM); Phillips and Hansen's (1990) fully modified least squares; Phillips' (1991b) spectral regression for the ECM; Stock's (1987) nonlinear least squares; and Bossaerts' (1988) canonical correlation approach. Many of these methods make use of OLS at an early stage, so one implication of our results for the $CI(1)$ case is that FDLS be substituted here. These methods are all specifically designed for the $CI(1)$ case, and in more general settings the validity and optimality of the associated inference procedures will be lost, and they may have no obvious advantage over OLS. Moreover, like OLS, they will not even be consistent in the stationary case. For the computationally simple FDLS procedure, our paper demonstrates a consistency-robustness not achieved by OLS and other procedures, a matching of limit distributional properties in some cases, and superiority in others, including the standard $CI(1)$ case.

2. FREQUENCY DOMAIN LEAST SQUARES

For a sequence of column vectors a_t, $t = 1, ..., n$, define the discrete Fourier transform

$$w_a(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_t a_t e^{it\lambda},$$

where the sample mean of the a_t is given by $\bar{a} = \sqrt{(2\pi/n)}w_a(0)$ and $\sum_t$ will throughout denote $\sum_{t=1}^n$. If b_t, $t = 1, ..., n$, is also a sequence of column vectors, define the (cross-) periodogram matrix

$$I_{ab}(\lambda) = w_a(\lambda) w_b^*(\lambda),$$

where * indicates transposition combined with complex conjugation. Further, for $\lambda_j = 2\pi j/n$, define the (real part of the) averaged periodogram

$$\hat{F}_{ab}(k, \ell) = \frac{2\pi}{n} \sum_{j=k}^{\ell} \mathrm{Re}\left\{I_{ab}(\lambda_j)\right\}, \ 1 \leq k \leq \ell \leq n-1.$$

In case $\widehat{F}_{ab}(k,\ell)$ is a vector we shall denote its i-th element $\widehat{F}_{ab}^{(i)}(k,\ell)$, and in case it is a matrix we shall denote its (i,j)-th element $\widehat{F}_{ab}^{(i,j)}(k,\ell)$; analogously we will use $I_{ab}^{(i)}(\lambda)$ and $I_{ab}^{(i,j)}(\lambda)$ to denote respectively the i-th element of the vector $I_{ab}(\lambda)$ and the (i,j)-th element of the matrix $I_{ab}(\lambda)$.

Now suppose we observe vectors $z_t = (x_t', y_t)'$, $t = 1, ..., n$, where y_t is real-valued and x_t is a $(p-1)\times 1$ vector with real-valued elements. Consider, for various m, the statistic

$$\widehat{\beta}_m = \widehat{F}_{xx}(1,m)^{-1}\widehat{F}_{xy}(1,m), \tag{2.1}$$

assuming the inverse exists. We can interpret $\widehat{\beta}_m$ as estimating the unknown β in the "regression model"

$$y_t = \beta' x_t + e_t,\ t = 1, 2, ... \ . \tag{2.2}$$

Notice that

$$\widehat{F}_{xx}(1,n-1) = \frac{1}{n}\sum_t (x_t - \overline{x})(x_t - \overline{x})',\ \widehat{F}_{xy}(1,n-1) = \frac{1}{n}\sum_t (x_t - \overline{x})(y_t - \overline{y}). \tag{2.3}$$

Thus $\widehat{\beta}_{n-1}$ is the OLS estimate of β with allowance for a non-zero mean in the unobservable e_t. Our main interest is in cases $1 < m < n-1$, where, because $w_a(\lambda)$ has complex conjugate $\overline{w}_a(2\pi - \lambda)$, we restrict further to $1 < m < n/2$. Then we call $\widehat{\beta}_m$ an FDLS estimate. Properties of e_t will be discussed subsequently, but these permit it to be correlated with x_t as well as y_t, $\widehat{\beta}_m$ being consistent for β due to $\widehat{F}_{ee}(1,m)$ being dominated by $\widehat{F}_{xx}(1,m)$ in a sense to be indicated. This can happen when z_t is stationary with long memory and e_t is stationary with less memory, if

$$\frac{1}{m} + \frac{m}{n} \to 0, \text{ as } n \to \infty, \tag{2.4}$$

which rules out OLS. Under (2.4) $\widehat{\beta}_m$ can be termed a "narrow-band" FDLS estimator. It can also happen when x_t is nonstationary while e_t is stationary or nonstationary with less memory, if only

$$m < n, \qquad m \to \infty, \text{ as } n \to \infty, \tag{2.5}$$

which includes OLS. In both situations $z_t \equiv FCI(d_1, ..., d_p; d_e)$ and the focus on low frequencies is thus natural. Notice that when $\lim(m/n) = \theta \in \left(0, \frac{1}{2}\right)$ (so that $\widehat{\beta}_m$ is not narrow-band), $\widehat{\beta}_m$ is a special case of the estimate introduced by Hannan (1963) and developed by Engle (1974) and others. However, while such m satisfy (2.5), our primary interest is in the narrow-band case (2.4) where $\widehat{\beta}_m$ is based on a degenerating band of frequencies and its superiority over OLS can be established under wider circumstances. It is the stationary case which we first discuss.

3. STATIONARY COINTEGRATION

The covariance stationary processes with which we shall be concerned will always be assumed to have absolutely continuous spectral distribution function. Thus, for jointly covariance stationary column vector processes a_t, b_t, $t = 1, 2, ...$, define the (cross) spectral density matrix $f_{ab}(\lambda)$ to satisfy

$$E\left(a_0 - E\left(a_0\right)\right)\left(b_j - E\left(b_0\right)\right)' = \int_{-\pi}^{\pi} f_{ab}(\lambda) e^{ij\lambda} d\lambda,\ j = 0, \pm 1, ...$$

We impose the following condition on z_t introduced earlier. For two matrices A and B, of equal dimension and possibly complex-valued elements, we say that $A \sim B$ if, for each (i,j), the ratio of the (i,j)-th elements of A and B tends to unity.

Assumption A The vector process z_t is covariance stationary with

$$f_{zz}(\lambda) \sim \Lambda G \Lambda, \text{ as } \lambda \to 0^+,$$

where G is a real matrix whose leading $(p-1) \times (p-1)$ submatrix has full rank and

$$\Lambda = diag\left\{\lambda^{-d_1}, ..., \lambda^{-d_p}\right\},$$

for $0 < d_i < \frac{1}{2}$, $1 \le i \le p$, and there exists a $p \times 1$ vector $\alpha \neq 0$, and a $c \in (0, \infty)$ and $d_e \in [0, \widetilde{d})$, such that

$$\alpha' f_{zz}(\lambda)\alpha \sim c\lambda^{-2d_e}, \text{ as } \lambda \to 0^+.$$

Assumption A is similar to that introduced by Robinson (1995a), where it is shown to hold for vector stationary and invertible fractional ARIMA processes (we could allow here, as there, for negative orders of integration greater than $-\frac{1}{2}$). However, there G was positive definite, whereas if Assumption A is imposed it has reduced rank, because otherwise

$$\alpha' f_{zz}(\lambda)\alpha \sim \left(\alpha'\Lambda\right) G \left(\Lambda\alpha\right) \ge \widetilde{c}\lambda^{-2\widetilde{d}} \text{ as } \lambda \to 0^+,$$

for $0 < \widetilde{c} < \infty$. Nevertheless G must be non-negative definite because $f_{zz}(\lambda)$ is, for all λ. The rank condition on G is a type of no-multicollinearity one on x_t. Notice that $z_{it} \equiv I(d_i)$, $i = 1, ..., p$ and $e_t \equiv I(d_e)$ if we adopt the stationary definition (replacing $u_t 1(t > 0)$ by u_t in (1.1)) of an $I(d)$ process. Notice then that Assumption A follows if $z_t \equiv FCI\left(d_1, ..., d_p; d_e\right)$ with $d_e < \widetilde{d}$. We adopt the normalization given by $\alpha = (-\beta', 1)'$, so the cointegrating relation is given by (2.2) if $e_t \equiv I(d_e)$. We stress that Assumption A does not restrict the spectrum of e_t away from frequency zero, because it is only local properties that matter since here we consider $\widehat{\beta}_m$ under (2.4). Asymptotic properties of $\widehat{\beta}_m$ require an additional regularity condition, such as

Assumption B

$$z_t = \mu_z + \sum_{j=0}^{\infty} A_j \varepsilon_{t-j},\ \sum_{j=0}^{\infty} \|A_j\|^2 < \infty, \tag{3.1}$$

where $\mu_z = E(z_0)$, and the $p \times 1$ vectors ε_t satisfy

$$E\left(\varepsilon_t \mid \Im_{t-1}\right) = 0,\ E\left(\varepsilon_t \varepsilon_t' \mid \Im_{t-1}\right) = \sum, \text{ a.s.},$$

for a constant, full rank matrix $\sum$, $\Im_t$ being the σ-field of events generated by ε_s, $s \leq t$, and $\|.\|$ denoting Euclidean norm, and the $\varepsilon_t \varepsilon_t'$ are uniformly integrable.

This assumption is a generalization of that of Robinson (1994a), the square summability of the A_j only confirming, in view of the other assumptions, the finite variance of z_t implied by Assumption A. We could replace the martingale difference assumption on ε_t and $\varepsilon_t \varepsilon_t' - \sum$ by fourth moment conditions, as in Robinson (1994c). Notice that it would be equivalent to replace z_t by $(x_t', e_t)'$ with e_t given by (2.2). When z_t satisfies both Assumptions A and B, the A_j are restricted by the requirement that $\| \alpha' A(e^{i\lambda}) \| \sim c^* \lambda^{-d_e}$ as $\lambda \to 0^+$, for $0 < c^* < \infty$. A very simple model covered by Assumptions A and B is (2.2) and $x_t = \varphi x_{t-1} + u_t$, with $p = 2$, $0 < \varphi < 1$, $u_t \equiv I(d_1)$ (implying $x_t \equiv I(d_1)$), and $e_t \equiv I(d_e)$, $0 \leq d_e < d_1 < \frac{1}{2}$. When u_t and e_t are not orthogonal OLS is of course inconsistent for β, as indeed is any other standard cointegration estimator, notwithstanding the fact that for φ close enough to unity x_t is indistinguishable for any practical purpose from a unit root process.

Theorem 3.1 Under Assumption A with $\alpha = (-\beta', 1)'$, Assumption B and (2.4), as $n \to \infty$

$$\widehat{\beta}_{im} - \beta_i = O_p\left(\left(\frac{n}{m}\right)^{d_e - d_i}\right),\ i = 1, ..., p-1,$$

where $\widehat{\beta}_{im}$ and β_i are the i-th elements of, respectively, $\widehat{\beta}_m$ and β.

It follows that if there is cointegration, so $\widetilde{d} > d_e$, $\widehat{\beta}_m$ is consistent for β. In case the d_i are identical there is a common stochastic order $O_p((n/m)^{-b})$, varying inversely with the strength $b = d_i - d_e$ of the cointegrating relation. We conjecture that Theorem 3.1 is sharp and that under suitable additional conditions the $(n/m)^{d_i - d_e}\left(\widehat{\beta}_{im} - \beta_i\right)$ will jointly converge in probability to a non-null constant vector. We conjecture also that after bias-correction and with a different normalization the limit distribution will be normal in some proper subset of stationary $(d_1, ..., d_{p-1}, d_e)$-space, and non-normal elsewhere (cf the derivation of Lobato and Robinson (1996) of the limit distribution of the scalar averaged periodogram). A proper study of this issue would take up considerable space, however, whereas our principle purpose here is to establish consistency, with rates, as an introduction to a study of $\widehat{\beta}_m$ in nonstationary environments.

4. THE AVERAGED PERIODOGRAM IN NONSTATIONARY ENVIRONMENTS

In order to analyze FDLS in case x_t, and possibly e_t, is nonstationary, we provide some basic properties of the averaged (cross-) periodogram that are of more general interest. For these, it suffices to consider a bivariate sequence (a_{1t}, a_{2t}), $t = 1, 2, ...$, given by

$$a_{it} = \sum_{j=1}^{t} \varphi_{i,t-j}\eta_{ij},\ i = 1, 2, \tag{4.1}$$

where we impose the following assumptions.

Assumption C (η_{1t}, η_{2t}), $t = 0, \pm 1, ...$, is a jointly covariance stationary process with zero mean and bounded spectral density matrix.

Assumption D For $i = 1, 2$, $0 \leq \gamma_2 \leq \gamma_1$, $\gamma_1 > \frac{1}{2}$, $\gamma_2 \neq \frac{1}{2}$, the sequences φ_{it} satisfy $\varphi_{it} = \varphi_t(\gamma_i)$, where for $t \geq 0$,

$$\begin{aligned} \varphi_t(\gamma) &= 1(t=0),\ \gamma = 0, \\ &= O\left((1+t)^{\gamma-1}\right),\ \gamma > 0, \\ &= 1,\ \gamma = 1, \\ \left|\varphi_t(\gamma) - \varphi_{t+1}(\gamma)\right| &= O\left(\frac{|\varphi_t(\gamma)|}{t}\right),\ \gamma > 0. \end{aligned}$$

As we shall see in the next section, Assumption D covers cases where a_{1t} is nonstationary whereas a_{2t} is either $I(0)$ $(\gamma_2 = 0)$, has asymptotically stationary long memory $(0 < \gamma_2 < \frac{1}{2})$, or is nonstationary $(\gamma_2 > \frac{1}{2})$.

We consider here not only the statistic $\widehat{F}_{aa}^{(1,2)}(1, m)$, but also $\widehat{F}_{aa}^{(1,2)}(m+1, M)$, where $0 < m < M \leq n/2$. The latter arises as follows. We have

$$\begin{aligned} \widehat{F}_{aa}^{(1,2)}(1,m) &= \frac{\pi}{n}\sum_{j=1}^{m}\left\{I_{aa}^{(1,2)}(\lambda_j) + I_{aa}^{(1,2)}(\lambda_{n-j})\right\} \\ &= \frac{1}{2}\left\{\widehat{F}_{aa}^{(1,2)}(1,m) + \widehat{F}_{aa}^{(1,2)}(n-m, n-1)\right\} \\ &= \frac{1}{2}\widehat{F}_{aa}^{(1,2)}(1, n-1) - \frac{1}{2}\widehat{F}_{aa}^{(1,2)}(m+1, n-m-1). \end{aligned} \tag{4.2}$$

For n odd (4.2) is

$$\frac{1}{2}\widehat{F}_{aa}^{(1,2)}(1, n-1) - \widehat{F}_{aa}^{(1,2)}(m+1, (n-1)/2),$$

and for n even it is

$$\frac{1}{2}\widehat{F}_{aa}^{(1,2)}(1, n-1) - \frac{1}{2}\widehat{F}_{aa}^{(1,2)}(m+1, n/2) - \frac{1}{2}\widehat{F}_{aa}^{(1,2)}(m+1, n/2-1),$$

where

$$\widehat{F}_{aa}^{(1,2)}(1,n-1) = \frac{1}{n}\sum_t (a_{1t}-\overline{a}_1)(a_{2t}-\overline{a}_2).$$

This development follows Robinson (1994c).

When $\gamma_1+\gamma_2>1$ we will deduce that

$$\widehat{F}_{aa}^{(1,2)}(m+1,M) = o_p\left(n^{\gamma_1+\gamma_2-1}\right)$$

by showing that both the mean and the standard deviation of the left side are $o(n^{\gamma_1+\gamma_2-1})$. Thus if $\frac{1}{2}n^{1-\gamma_1-\gamma_2}\widehat{F}_{aa}^{(1,2)}(1,n-1)$ has a nondegenerate limit distribution, $n^{1-\gamma_1-\gamma_2}\widehat{F}_{aa}^{(1,2)}(1,m)$ shares it. For $\gamma_1+\gamma_2\le 1$, only the standard deviation of $\widehat{F}_{aa}^{(1,2)}(m+1,M)$ is $o(n^{\gamma_1+\gamma_2-1})$ and it is necessary to estimate $E\left(\widehat{F}_{aa}^{(1,2)}(1,m)\right)$, which differs non-negligibly from $E\left(\widehat{F}_{aa}^{(1,2)}(1,[(n-1)/2])\right)$. We first consider the means.

Proposition 4.1 Under (4.1), Assumptions C and D and (2.5), for $0< m< M\le n/2$

$$E\left(\widehat{F}_{aa}^{(1,2)}(m+1,M)\right) = o\left(n^{\gamma_1+\gamma_2-1}\right),\ \gamma_1+\gamma_2>1, \tag{4.3}$$

and

$$E\left(\widehat{F}_{aa}^{(1,2)}(1,m)\right) = O\left(n^{\gamma_1+\gamma_2-1}\right),\ \gamma_1+\gamma_2>1, \tag{4.4}$$

$$= O\left(\left(\frac{n}{m}\right)^{\gamma_1+\gamma_2-1}\right),\ \gamma_1+\gamma_2<1. \tag{4.5}$$

The proof of this, and of Proposition 4.2 below, is contained in the Appendix. To consider the variances we impose the additional:

Assumption E (η_{1t},η_{2t}) is fourth order stationary with bounded fourth-order cross-cumulant spectrum $f(\mu_1,\mu_2,\mu_3)$ satisfying

$$k(j,k,l) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(\mu_1,\mu_2,\mu_3)\exp(ij\mu_1+ik\mu_2+il\mu_3)\prod_{k=1}^{3}d\mu_k,$$

where $k(j,k,l)$ is the fourth order cumulant of $\eta_{10},\eta_{2j},\eta_{1,j+k},\eta_{2,j+k+l}$ for $j,k,l=0,\pm1,\ldots$.

Proposition 4.2 Under (4.1), Assumptions C, D and E and (2.5), for $0<m<M\le n/2$, as $n\to\infty$

$$Var\left(\widehat{F}_{aa}^{(1,2)}(m+1,M)\right) = o\left(n^{2(\gamma_1+\gamma_2-1)}\right) \tag{4.6}$$

and

$$Var\left(\widehat{F}_{aa}^{(1,2)}(1,M)\right) = O\left(n^{2(\gamma_1+\gamma_2-1)}\right). \tag{4.7}$$

In view of (2.3), $I_{aa}^{(1,2)}(\lambda_j)$ distributes the sample covariance $\widehat{F}_{aa}^{(1,2)}(1,n-1)$ across the Fourier frequencies λ_j, $j=1,...,n-1$. Propositions 4.1 and 4.2 suggest that $\widehat{F}_{aa}^{(1,2)}(1,n-1)$ is dominated by the contributions from a possibly degenerating frequency band $(0,\lambda_m)$ when the collective memory in a_{1t}, a_{2t} is sufficiently strong $(\gamma_1+\gamma_2>1)$ while otherwise $\widehat{F}_{aa}^{(1,2)}(1,m)-\frac{1}{2}\widehat{F}_{aa}^{(1,2)}(1,n-1)$ is estimated by its mean, in view of (4.5). These results are crucial to the derivation of the asymptotic behaviour of $\widehat{F}_{xx}(1,m)$ and $\widehat{\beta}_m$ introduced in Section 2.

5. NONSTATIONARY FRACTIONAL COINTEGRATION

For z_t nonstationary $\left(d_i>\frac{1}{2},\, i=1,2,...,p\right)$, we find it convenient to stress a linear representation for $w_t=(x_t',e_t)'$ in place of that for $z_t=(x_t',y_t)'$ in Assumption B. Introduce the diagonal fractional operator

$$D(L)=diag\left\{(1-L)^{-d_1},...,(1-L)^{-d_{p-1}},(1-L)^{-d_e}\right\}$$

and

Assumption F The vector sequence w_t is given by

$$D^{-1}(L)(w_t-\mu_w)=u_t1(t>0) \tag{5.1}$$

for a fixed vector μ_w with p-th element zero, where

$$d_i>\frac{1}{2},\, i=1,...,p-1,\, d_e\geq 0,$$

$$u_t=C(L)\varepsilon_t,\quad C(L)=\sum_{j=0}^{\infty}C_jL^j, \tag{5.2}$$

$$\det\{C(1)\}\neq 0, \tag{5.3}$$

$$\sum_{j=0}^{\infty}\left(\sum_{k=j}^{\infty}\|C_k\|^2\right)^{\frac{1}{2}}<\infty, \tag{5.4}$$

and the ε_t are independent and identically distributed $p\times 1$ vectors such that

$$E(\varepsilon_t)=0,\, E(\varepsilon_t\varepsilon_t')=\sum,\, rank(\sum)=p, \tag{5.5}$$

$$E\|\varepsilon_t\|^{\theta}<\infty,\, \theta>\max\left(4,\frac{2}{2d_{\min}-1}\right). \tag{5.6}$$

Assumption F strengthens the requirements on ε_t of Assumption B. Under (5.4)

$$\sum_{j=0}^{\infty}\|C_j\|\leq\sum_{j=0}^{\infty}\left(\sum_{k=j}^{\infty}\|C_k\|^2\right)^{\frac{1}{2}}<\infty \tag{5.7}$$

so (5.3) and (5.4) imply that all elements of u_t are in $I(0)$, whereas with reference to (1.1), for $i = 1, ..., p-1$ the i-th element of w_t (and thus of x_t) is in $I(d_i)$, while its p-th element, e_t, is in $I(d_e)$, so in particular w_t could be a vector fractional ARIMA process. We have allowed for an unknown intercept, μ_w, in w_t.

From (5.1) and (5.2), we can write

$$w_t = \mu_w + \sum_{j=0}^{\infty} B_{jt}\varepsilon_{t-j}, \tag{5.8}$$

where

$$B_{jt} = \sum_{i=0}^{\min(j,t-1)} D_i C_{j-i} \tag{5.9}$$

with D_j given by the formal (binomial) expansion

$$D(L) = \sum_{j=0}^{\infty} D_j L^j.$$

Defining the nonsingular matrix

$$P = \begin{bmatrix} I_{p-1} & O_{p-1} \\ -\beta' & 1 \end{bmatrix}$$

where I_j and O_j are respectively the j-rowed identity matrix and the $j \times 1$ vector of zeroes, we find that (5.1) is equivalent to

$$z_t = \mu_z + \sum_{j=0}^{\infty} A_{jt}\varepsilon_{t-j}, \tag{5.10}$$

where $\mu_z = P^{-1}\mu_w$, $A_{jt} = P^{-1}B_{jt}$. The representation (5.10) can be compared with the time-invariant one (3.1) for the stationary case (in which $A_j = P^{-1}B_{j\infty}$).

Define $d = (d_1, ..., d_{p-1})'$ and

$$\Delta(d) = diag\left\{n^{\frac{1}{2}-d_1}, ..., n^{\frac{1}{2}-d_{p-1}}\right\}, \; G(r,d) = diag\left\{r^{d_1-1}, ..., r^{d_{p-1}-1}\right\}.$$

Let Ω be the leading $(p-1) \times (p-1)$ submatrix of $C(1)\sum C(1)'$, and $B(r,\Omega)$ be $(p-1)$-dimensional Brownian motion with covariance matrix Ω (which has full rank under Assumption F). Let

$$W(r;d,\Omega) = \int_0^r G(r-s;d)dB(s;\Omega), \; W(d,\Omega) = \int_0^1 W(r;d,\Omega)dr,$$

$$V(d,\Omega) = \int_0^1 \{W(r;d,\Omega)W'(r;d,\Omega) - W(d,\Omega)W(d,\Omega)'\}\,dr.$$

We call $W(r;d,\Omega)$ a multivariate fractional Brownian motion, following Akonom and Gourieroux (1987) and Gourieroux, Maurel and Monfort (1989). A somewhat different definition of (scalar) fractional Brownian motion was proposed by Mandelbrot and Van Ness (1968), and has since prevailed in the probability literature (though the latter authors also mentioned $W(r;d,\Omega)$). Note that the components of $W(r;d,\Omega)$ are continuous Gaussian processes with zero means and variances that grow like r^{2d_i-1}, $i=1,...,p-1$. Let $\Rightarrow$ denote weak convergence and $\overline{x}=n^{-\frac{1}{2}}\sum_t x_t$.

Theorem 5.1 Under Assumption F, as $n\to\infty$

$$\Delta(d)x_{[nr]}\Rightarrow W(r;d,\Omega),\ 0<r\leq 1, \tag{5.11}$$

$$\Delta(d)\overline{x}\Rightarrow W(d,\Omega), \tag{5.12}$$

$$\Delta(d)\widehat{F}_{xx}(1,n-1)\Delta(d)\Rightarrow V(d,\Omega). \tag{5.13}$$

The proof of (5.11), which relates to results of Akonom and Gourieroux (1987), Gourieroux, Maurel and Monfort (1989) and Silveira (1991), is given by Marinucci and Robinson (1997), whence (5.12) and (5.13) follow from the continuous mapping theorem. For $d_1=...=d_{p-1}=1$, fractional Brownian motion reduces to classical Brownian motion and so (5.11) includes a multivariate invariance principle for $I(1)$ processes, as can be found for instance in Phillips and Durlauf (1986). (5.13) provides an invariance principle for the sample covariance matrix of x_t (see (2.3)), and due to the following lemma Propositions 4.1 and 4.2 can be applied to deduce one for $\widehat{F}_{xx}(1,m)$.

Lemma 5.1 Let Assumption F hold. Then with the choices

$$(a_{1t},a_{2t})=(x_{it},e_t),\ \gamma_1=d_i,\ \gamma_2=d_e,\ i=1,...,p-1 \tag{5.14}$$

or

$$(a_{1t},a_{2t})=(x_{it},x_{jt}),\ \gamma_1=d_i,\ \gamma_2=d_j,\ i,j=1,...,p-1, \tag{5.15}$$

it follows that Assumptions C, D and E are satisfied.

Lemma 5.2 Under Assumption F and (2.5), as $n\to\infty$

$$\Delta(d)\widehat{F}_{xx}(1,m)\Delta(d)\Rightarrow V(d,\Omega).$$

We can now proceed to investigate asymptotic behaviour of OLS and FDLS in various of the cases that arise when x_t is nonstationary, and e_t has short memory or stationary or nonstationary long memory.

Case I: $d_i+d_e<1,\ i=1,...,p-1.$

Here not only does x_t possess less-than-unit-root nonstationarity, but the collective memory in x_t and e_t is more limited than in the $CI(1)$ case. It corresponds to $\gamma_1+\gamma_2<1$ of Section 4, and we require first a more precise result than (4.5) in case $m=n-1$. Let b'_{ij} be the i-th row of $B_j=B_{j\infty}=\sum_{i=0}^{j}D_iC_{j-i}$ given by (5.9). Define

$$\xi_i=\sum_{j=0}^{\infty}b'_{ij}\sum b_{pj},\ i=1,...,p-1.$$

Lemma 5.3 Under Assumption F with $d_i+d_e<1$, $d_i>\frac{1}{2}$, $i=1,...,p-1$,

$$\lim_{n\to\infty}E\left\{\widehat{F}_{xx}^{(i)}(1,n-1)\right\}=\xi_i,\ i=1,...,p-1,$$

where the right hand side is finite.

Lemma 5.3 (and (4.5) of Proposition 4.1) are of some independent interest in that they indicate how sample covariances between a nonstationary and a stationary sequence can be stochastically bounded and have the same structure as when both sequences are stationary, so long as the memory parameters sum to less than 1, as automatically applies in the fully stationary case.

Define ξ to be the $(p-1)\times 1$ vector with i-th element ξ_i if $d_i=d_{\min}$ and zero if $d_i>d_{\min}$, where $d_{\min}=\min_{1\le i\le p-1}d_i$.

Theorem 5.2 Let Assumption F hold with $d_i+d_e<1$, $d_e<d_i$, $i=1,...,p-1$, and

$$rank\{V(d,\Omega)\}=p-1,\ \text{a.s.} \tag{5.16}$$

Then as $n\to\infty$

$$n^{d_{\min}-\frac{1}{2}}\Delta(d)^{-1}\left(\widehat{\beta}_{n-1}-\beta\right)\Rightarrow V(d,\Omega)^{-1}\xi, \tag{5.17}$$

and under (2.4)

$$n^{d_{\min}-\frac{1}{2}}\Delta(d)^{-1}\left(\widehat{\beta}_m-\beta\right)=O_p\left(\left(\frac{n}{m}\right)^{d_{\min}+d_e-1}\right)=o_p(1). \tag{5.18}$$

Theorem 5.2 indicates that so long as ξ is non-null the $n^{d_{\min}+d_i-1}\left(\widehat{\beta}_{i,n-1}-\beta_i\right)$ have a nondegenerate limit distribution, whereas when the interval $(0,\lambda_m)$ degenerates $\widehat{\beta}_{im}-\beta_i=o_p\left(n^{1-d_{\min}-d_i}\right)$, $i=1,...,p-1$, so that FDLS converges faster than OLS. In view of the "global" nature of $\widehat{\beta}_{m-1}$ and the "local" nature of $\widehat{\beta}_m$ this outcome is at first sight surprising, but it is due to the bias of $\widehat{\beta}_m$ becoming negligible relative to that of $\widehat{\beta}_{n-1}$. Notice that the rate of convergence of $\widehat{\beta}_{n-1}$ is independent of d_e.

Case II: The $CI(1)$ case ($d_i=1$, $i=1,...,p-1$, $d_e=0$).

Now we consider the case considered in the bulk of the cointegration literature, where z_t has a unit root and the cointegrating error is $I(0)$. Write ι for the $(p-1)\times 1$ vector of units, so that in the present case $d = \iota$ and $W(r;d,\Omega) = B(r;\Omega)$. Let $B_*(r;\Omega_*)$ be p-dimensional Brownian motion with covariance matrix $\Omega_* = C(1)\sum C(1)'$, and thus write

$$B_*(r;\Omega_*) = \begin{pmatrix} B(r;\Omega) \\ b(r;\sigma^2) \end{pmatrix},$$

where σ^2 is the (p,p)-th element of Ω_*, and in general $B(r;\Omega)$ and $b(r;\sigma^2)$ are correlated and thus in effect depend not only on Ω and σ^2 but on the other elements of Ω_* also. Write

$$U(\Omega_*) = \int_0^1 B(r;\Omega)db(r;\sigma^2).$$

Denote by γ_j the $(p-1)\times 1$ vector with i-th element $E\left(u_{it}e_{t+j}\right)$, recalling that $d_e = 0$ implies $e_t = u_{pt}$. Now define

$$\Gamma_j = \sum_{\ell=|j|}^{\infty} \gamma_{\ell sign(\ell)},\ j = 0, \pm 1, ..., \tag{5.19}$$

so that $\Gamma_j = \sum_{\ell=j}^{\infty}\gamma_\ell$ for $j \geq 0$ and $\Gamma_j = \sum_{\ell=-\infty}^{j}\gamma_\ell$ for $j < 0$, and the sum (5.19) converges absolutely for all j under Assumption F. Let $h(\lambda)$ be the vector function with Fourier coefficients given by

$$\Gamma_{|j|} - \Gamma_{-|j|-1} = \int_{-\pi}^{\pi} h(\lambda)e^{ij\lambda}d\lambda,\ j = 0, \pm 1, ...$$

Assumption G $h(\lambda)$ is continuous at $\lambda = 0$, and integrable.

Assumption G is implied by $\sum_{j=0}^{\infty} \| \Gamma_{|j|} - \Gamma_{-|j|-1} \| < \infty$, which is in turn implied by

$$\sum_{j=0}^{\infty}(j+1) \| \gamma_j - \gamma_{-j-1} \| < \infty, \tag{5.20}$$

in which case we may write

$$h(0) = \frac{1}{2\pi}\sum_{j=0}^{\infty}(2j+1)(\gamma_j - \gamma_{-j-1}).$$

Of course (5.20) is itself true if $\sum_{j=-\infty}^{\infty}(|j|+1)\left\|\gamma_j\right\| < \infty$ for which a sufficient condition in terms of (5.2) is

$$\sum_{j=0}^{\infty} (j+1) \left\|C_j\right\| < \infty, \tag{5.21}$$

which is stronger than (5.4), while holding when u_t is a stationary ARMA process.

Lemma 5.4 Under Assumption G and (2.4)

$$\lim_{n\to\infty} E\left(\frac{n}{m}\widehat{F}_{xe}(1,m)\right) = \frac{1}{2}h(0).$$

Theorem 5.3 Let Assumption F hold with $d = \iota$, $d_e = 0$. Then as $n \to \infty$

$$n\left(\widehat{\beta}_{n-1} - \beta\right) \Rightarrow V(\iota,\Omega)^{-1}\left\{U(\Omega) + \Gamma_0\right\} \tag{5.22}$$

and if also Assumption G and (2.4) hold

$$n\left(\widehat{\beta}_m - \beta\right) \Rightarrow V(\iota,\Omega)^{-1}U(\Omega). \tag{5.23}$$

Thus in the $CI(1)$ case $\widehat{\beta}_{n-1}$ and $\widehat{\beta}_m$ have the same rate of convergence but under (2.4) $\widehat{\beta}_m$ does not suffer from the "second-order bias" term Γ_0 incurred by $\widehat{\beta}_{n-1}$. More precisely, as the proof of Theorem 5.3 indicates, there is a second-order bias of order $O(m/n^2)$ in $\widehat{\beta}_m$ which is thus too small to contribute to (5.23), by comparison with the $O(n^{-1})$ second-order bias in (5.22). Phillips (1991b) considered a form of narrow-band spectral regression in the $CI(1)$ case, albeit stressing a system type of estimate which has superior limit distributional properties to $\widehat{\beta}_m$, assuming the $CI(1)$ hypothesis is correct. However his proof is based on weighted autocovariance spectrum estimates, rather than our averaged periodogram ones. As is well known, in many stationary environments these two types of estimate are very close asymptotically, but in the $CI(1)$ case the weighted autocovariance version of $\widehat{\beta}_m$ turns out to exhibit second-order bias due to correlation between u_t and e_t (specifically, to their cross-spectrum at zero frequency).

Case III: The Case $d_i + d_e > 1$, $i = 1, ..., p-1$, $d_e < \frac{1}{2}$.

We now look at the case where the collective memory in each (x_{it}, u_t) combination exceeds that of the previous two cases, yet e_t is still stationary. Thus x_t could have less than unit root stationarity but in that case the memory in e_t must compensate suitably. On the other hand x_t could exhibit nonstationarity of arbitrarily high degree.

Theorem 5.4 Let Assumption F hold with $d_i + d_e > 1$, $0 \le d_e < \frac{1}{2} < d_i$, $i = 1, ..., p-1$, and let (5.16) hold. Then for $i = 1, ..., p-1$, as $n \to \infty$

$$\widehat{\beta}_{i,n-1} - \beta_i = O_p\left(n^{d_e - d_i}\right), \tag{5.24}$$

and if also (2.5) holds

$$\widehat{\beta}_{im} - \widehat{\beta}_{i,n-1} = o_p\left(n^{d_e - d_i}\right), \tag{5.25}$$

$$\widehat{\beta}_{im} - \beta_i = O_p\left(n^{d_e - d_i}\right). \tag{5.26}$$

The results (5.24) and (5.26) only bound the rates of convergence of OLS and FDLS, and we have been unable to characterize even the exact rate of convergence of OLS in the present case, due to the fact that on the one hand e_t is stationary so that the continuous mapping theorem does not suffice, whereas on the other hand e_t cannot be approximated by a semi-martingale, unlike in the short-memory case $d_e = 0$ (where in fact an exact rate and limit distribution can be derived, as it was in the $CI(1)$ case). We conjecture, however, that at least under some additional conditions the rate in (5.24) is exact, whereupon (5.25) implies immediately that $\widehat{\beta}_m$ shares the same rate and limit distribution as $\widehat{\beta}_{n-1}$.

Case IV: The case $d_e > \frac{1}{2}$.

Now we suppose that cointegration does not account for all the nonstationarity in z_t, so that e_t is nonstationary, as is motivated by some of the empirical experience to be described in Section 7. Write $d_* = (d', d_e)'$ and

$$\begin{aligned}
\Delta_*(d_*) &= diag\left\{n^{\frac{1}{2}-d_1}, ..., n^{\frac{1}{2}-d_{p-1}}, n^{\frac{1}{2}-d_e}\right\},\ G_*(r; d_*) = diag\left\{r^{d_1-1}, ..., r^{d_{p-1}-1}, r^{d_e-1}\right\}, \\
W_*(r; d_*, \Omega_*) &= \int_0^r G_*(r-s; d_*)dB_*(s; \Omega_*) = \begin{bmatrix} W(r; d, \Omega) \\ w(r; d_e, \sigma^2) \end{bmatrix}, \\
W_*(d_*, \Omega_*) &= \int_0^1 W_*(r; d_*, \Omega_*)dr, \\
U(d_*, \Omega_*) &= \int_0^1 \{W(r; d, \Omega) - W(d, \Omega)\}\, w(r; d_e, \sigma^2)dr.
\end{aligned}$$

Let $\overline{w} = n^{-1}\sum_t w_t$. The following theorem is analogous to Theorem 5.1 and needs no additional explanation.

Theorem 5.5 Under Assumption F and $\frac{1}{2} < d_e < d_i$, $i = 1, ..., p-1$, as $n \to \infty$

$$\begin{aligned}
\Delta_*(d_*)w_{[nr]} &\Rightarrow W_*(r; d_*, \Omega_*), \\
\Delta_*(d_*)\overline{w} &\Rightarrow W(d_*, \Omega_*), \\
n^{\frac{1}{2}-d_e}\Delta(d)\widehat{F}_{xe}(1, n-1) &\Rightarrow U(d_*, \Omega_*).
\end{aligned} \tag{5.27}$$

Theorem 5.6 Under Assumption F, $\frac{1}{2} < d_e < d_i$, $i = 1, ..., p-1$, and (5.16), as $n \to \infty$

$$n^{\frac{1}{2}-d_e}\Delta(d)^{-1}\left(\widehat{\beta}_{n-1} - \beta\right) \Rightarrow V(d, \Omega)^{-1}U(d_*, \Omega_*), \tag{5.28}$$

and if also (2.5) holds

$$n^{\frac{1}{2}-d_e}\Delta(d)^{-1}\left(\widehat{\beta}_m - \beta\right) \Rightarrow V(d, \Omega)^{-1}U(d_*, \Omega_*). \tag{5.29}$$

Now so long as m is regarded as increasing with n, the limit distribution is unaffected however many frequencies we omit from $\widehat{\beta}_m$. Notice that in case the d_i are all equal the rate of convergence reflects the cointegrating strength b defined in Section 1, such that $\widehat{\beta}_m$ is n^b-consistent.

6. MONTE CARLO EVIDENCE

Because OLS is often used as a preliminary step in $CI(1)$ analysis, the previous section suggests that even if fractional possibilities are to be ignored, FDLS might be substituted at this stage. To compare the performance of FDLS with OLS in moderate sample sizes a small Monte Carlo study in the $CI(1)$ case was conducted. The models we employed are as follows. For $i = 1, 2$, let u_{it} be a sequence of $N(0, i)$ random variables, independent across t.

Model A: AR(1) cointegrating error, $p = 2$, in (2.2) with

$$\begin{aligned}(1-L)x_t &= u_{1t},\\ (1-\varphi L)e_t &= u_{2t},\\ E\left(u_{1t}u_{2t}\right) &= 1,\ \varphi = 0.8, 0.6, 0.4, 0.2.\end{aligned}$$

Model B: AR(2) cointegrating error, $p = 2$, in (2.2) with

$$\begin{aligned}(1-L)x_t &= u_{1t},\\ (1-\varphi_1 L-\varphi_2 L^2)e_t &= u_{2t},\\ E\left(u_{1t}u_{2t}\right) &= 1,\ \varphi_2 = -0.9; \varphi_1 = .947, .34, -.34, -.947.\end{aligned}$$

We fix $\varphi_2 = -0.9$ in Model B to obtain a spectral peak for e_t in the interior of $(0, \pi)$, in particular at $\lambda^* = \arccos\left(-\varphi_1(1+\varphi_2)/4\varphi_2\right)$, that is at $\lambda^* = \pi/3, 4\pi/9, 5\pi/9$ and $2\pi/3$, respectively, for the four φ_1. On the other hand in Model A e_t always has a spectral peak at zero frequency.

Series of lengths $n = 64$, 128 and 256 were generated. $\widehat{\beta}_{n-1}$ and $\widehat{\beta}_m$, for $m = 3, 4, 5$ were computed, as were an estimate superior to OLS in the $CI(1)$ case, the fully-modified least squares estimate (FM-OLS, denoted $\widetilde{\beta}_{FM}$) of Phillips and Hansen (1990) which uses OLS residuals at a first step, and also a modified version of this (denoted $\widetilde{\beta}^*_{FM}$), using FDLS residuals. Bartlett nonparametric spectral estimation was used in $\widetilde{\beta}_{FM}$ and $\widetilde{\beta}^*_{FM}$, with lag numbers $\widetilde{m} = 4, 6, 8$, for $n = 64, 128, 256$ respectively.

Monte Carlo bias and mean squared error (MSE), based on 5000 replications, are reported in Tables 1 and 2 for Models A and B respectively. For each m, FDLS is superior to OLS in terms of both bias and MSE in every single case, often significantly. In fact $\widehat{\beta}_m$ is best for the smallest $m, 3$. Our modified version $\widetilde{\beta}^*_{FM}$ of FM-OLS improves on the standard one $\widetilde{\beta}_{FM}$ in 23 out of 24 cases in terms of bias, and 16 out of 24 in terms of MSE, with 8 ties. The intuition underlying FDLS is that on the smallest frequencies cointegration implies a high signal-to-noise ratio, so it is not surprising that FDLS performs better for AR(2) e_t than AR(1) e_t, especially as λ^* increases. It is possible to devise an e_t with such power around $\lambda = 0$ that, in finite samples, FDLS performs worse than OLS, for example when $\varphi_1 \simeq 2$, $\varphi_2 \simeq -1$, $\varphi_2 + \varphi_1 < 1$, so e_t is "near-$I(2)$" and in small samples e_t dominates x_t. However, given the intuition underlying the concept of cointegration, we believe this could be described as a "pathological" case.

7. EMPIRICAL EXAMPLES

Our empirical work employs the data of Engle and Granger (1987) and Campbell and Shiller (1987). We consider seven bivariate series, denoting by y the variable chosen to be "dependent" and by x the "independent" one in (2.2), and by d_y, d_x integration orders. We describe the methodology used in three steps.

1) Memory of raw data

A necessary condition for cointegration is $d_x = d_y$, which can be tested using estimates of d_x and d_y. Three types of estimate were computed, and one test statistic. The estimates are all "semiparametric", based only on a degenerating band of frequencies around zero frequency and assuming only a local-to-zero model for the spectral density (cf Assumption A) rather than a parametric model over all frequencies. The semiparametric estimates are inefficient when the parametric model is correct, but are consistent more generally and seem natural in the context of the present paper. Their asymptotic properties were established by Robinson (1995a,b) under the assumption of stationarity and invertibility (having integration order between -$\frac{1}{2}$ and $\frac{1}{2}$) and so because our raw series seem likely to be nonstationary, and quite possibly with integration orders between $\frac{1}{2}$ and $\frac{3}{2}$, we first-differenced them prior to d estimation, and then added unity. The stationarity assumption is natural in view of the motivation of these estimates, but recently Hurvich and Ray (1995), Velasco (1997a,b) have shown that they can still be consistent and have the same limit distribution as in Robinson (1995a,b) under nonstationarity (although with a different definition of $I(d)$ nonstationarity from ours), at least if a data taper is used. We thus estimated d_x and d_y directly from the raw data also, but as the results were similar they are not reported.

Denote by Δz_t either Δx_t or Δy_t where Δ is the difference operator. We describe the estimation and testing procedures as follows

(i) Log-periodogram regression. For $z = x, y$ we report in the tables $\widehat{d}_z = 1 + \widehat{\delta}_z$, where $\widehat{\delta}_z$ is the slope estimate obtained by regressing $\log(I_{\Delta z \Delta z}(\lambda_j))$ on $-2\log(\lambda_j)$ and an intercept, for $j = 1, ..., \ell$, where ℓ is a bandwidth number, tending to infinity slower than n. This is the version proposed by Robinson (1995a) rather than the original one of Geweke and Porter-Hudak (1983), except that we do not trim out any frequencies; recent evidence of Hurvich, Deo and Brodsky (1998) suggests that this is not necessary for nice asymptotic properties.

(ii) Test of $d_x = d_y$. We report the Wald statistic, denoted W in the tables, of Robinson (1995a,b), based on the difference $\widehat{d}_x - \widehat{d}_y = \widehat{\delta}_z - \widehat{\delta}_y$. The significance of W is judged by comparison with the upper tail of the χ_1^2 distribution, the 5% and 1% points being respectively 3.78 and 5.5

(iii) GLS log-periodogram regression. Given that $d_x = d_y$ we estimate the common value by $\widehat{d}_{x=y} = 1 + \widehat{\delta}_{x=y}$ where $\widehat{\delta}_{x=y}$ is the generalized least squares (GLS) log-periodogram etimate of Robinson (1995a) based on the bivariate series $(\Delta x_t, \Delta y_t)$, using residuals from the regression in (i), $\widehat{\delta}_{x=y}$ is asymptotically more efficient than $\widehat{\delta}_x$ and $\widehat{\delta}_y$ when $d_x = d_y$.

(iv) Gaussian estimation. For $z = x, y$ we report $\widetilde{d}_z = 1 + \widetilde{\delta}_z$ where $\widetilde{\delta}_z$ minimizes

$$\log\left(\sum_{j=1}^{\ell} \lambda_j^{2d} I_{\Delta z \Delta z}(\lambda_j)\right) - \frac{2d}{\ell}\sum_{j=1}^{\ell}\log(\lambda_j),$$

which is a concentrated narrow-band Gaussian pseudo-likelihood, see Künsch (1987), Robinson (1995b). As shown by Robinson (1995b), $\widetilde{\delta}_z$ is asymptotically more efficient than $\widehat{\delta}_z$.

For the estimates in (iii) and (iv) we report also approximate 95% confidence intervals (denoted CI in the tables) based on the (normal) asymptotic distribution theory developed by Robinson (1995a,b). Robinson (1995a) assumed Gaussianity in establishing consistency and asymptotic normality of the estimates in (i) and (iii), but recent work of Velasco (1997b) suggests that this can be relaxed. For the estimates in (iv), Robinson (1995b) assumed a linear filter of martingale differences satisfying mild moment conditions. Although progress is currently being made on the choice of bandwidth ℓ in log-periodogram and Gaussian estimation, we have chosen a grid of three arbitrary values for each data set analyzed in order to judge sensitivity to ℓ. Note that the estimates are $\ell^{\frac{1}{2}}$-consistent.

2) Cointegration analysis

We report $\widehat{\beta}_m$ and also a "high-frequency" estimate

$$\widehat{\beta}_{-m} = \frac{\widehat{F}_{xy}(m+1, [(n-1)/2])}{\widehat{F}_{xx}(m+1, [(n-1)/2])}$$

based on the remaining frequencies, substantial deviations between $\widehat{\beta}_m$ and $\widehat{\beta}_{-m}$ suggesting that a full-band estimate such as OLS could be distorted by misspecification at high frequencies which is irrelevant to the essentially low-frequency concept of cointegration.

The tables include results for three values of m for each data set. These are much smaller than the bandwidths ℓ used in inference on d_x and d_y due to the anticipation of nonstationarity in the raw data; for stationary x_t, y_t optimal rules of bandwidth choice would lead to m that are more comparable with the ℓ we have used. After computing residuals $\widehat{e}_t = y_t - \widehat{\beta}_m x_t$, we obtained the low- and high-frequency R^2 quantities

$$R_m^2 = 1 - \frac{\widehat{F}_{\widehat{e}\widehat{e}}(1,m)}{\widehat{F}_{yy}(1,m)},\ R_{-m}^2 = 1 - \frac{\widehat{F}_{\widehat{e}\widehat{e}}(m+1,[(n-1)/2])}{\widehat{F}_{yy}(m+1,[(n-1)/2])}.$$

We can judge the fit of a narrow-band regression by R_m^2 and by comparing this with R_{-m}^2 see to what extent this semiparametric fit compares with a parametric one.

For each m we report also the fractions

$$r_{xx,m} = \frac{\widehat{F}_{xx}(1,m)}{\widehat{F}_{xx}(1,[(n-2)/2])},\ r_{xy,m} = \frac{\widehat{F}_{xy}(1,m)}{\widehat{F}_{xy}(1,[(n-2)/2])},$$

their closeness to unity indicating directly the empirical, finite sample relevance of Propositions 4.1 and 4.2 (though note that $r_{xy,m}$ need not lie in $[0,1]$.)

3) Memory of cointegrating error

We estimated d_e first by $\widehat{d}_e$ and $\widetilde{d}_e$, which are respectively the log-periodogram and Gaussian estimates of (i) and (iv) above, based on first differences of the $\widehat{e}_t$ and then adding unity. We also report $\widehat{d}_e^*$ and $\widetilde{d}_e^*$ which use the raw $\widehat{e}_t$ and do not add unity, because in general we have little prior reason for believing e_t is either stationary or nonstationary. In addition we report 95% confidence intervals based on the asymptotic theory of Robinson (1995a,b), though strictly this has not been justified in case of the residuals $\widehat{e}_t$.

Tables 3-9 report empirical results based on several data sets.

a) Consumption (y) and income (x) (quarterly data), 1947Q1-1981Q2

Engle and Granger (1987) found evidence of $CI(1)$ cointegration in these data. Table 3 tends to suggest an integration order very close to one for both variables, the estimates ranging from .89 to 1.08 for income and from 1.04 to 1.13 for consumption. The Wald statistic is at most 1.06, so we can safely not reject the $d_x = d_y$ null. Exploiting this information, one obtains GLS estimates ranging from .953 to 1.02; but with confidence intervals all so narrow as to exclude unity. The $\widehat{\beta}_m$ are about .232, which is close to OLS (.229), but the high frequency estimates $\widehat{\beta}_{-m}$ are closer to .20. The unexplained variability is four times smaller around frequency zero $(1 - R_m^2)$ than at short run frequencies $(1 - R_{-m}^2)$. Variability concentrates rapidly around frequency zero, 85.1% of the variance of income being accounted for by the three smallest periodogram ordinates, less than 5% of the total. This proportion rises to 92.6% for 6 frequencies, and is even greater for the cross-periodogram, confirming the high coherency of the two series at low frequencies. The residual diagnostics are less clear-cut, but in only one case out of 12 does the confidence interval for d_e include zero, providing strong evidence against weak dependence. The estimates of d_e vary quite noticeably with ℓ and the procedure adopted, ranging from .2 to .87.

b) Stock prices (y) and dividends (x) (annual data), 1871-1986.

The idea that these might be cointegrated follows mainly from a present value model, which asserts that an asset price is linear in the present discounted value of future dividends, $y_t = \theta(1-\delta)\sum_{i=0}^{\infty}\delta^i E_t(x_{t+i}) + c$, where δ is the discount factor; see Campbell and Shiller (1987). In Table 4, the estimates of d_x, d_y appear close to unity, although now the hypothesis that dividends are mean-reverting ($d_x < 1$) appears to be supported. The Wald statistics for testing $d_x = d_y$ are always manifestly insignificant. A marked difference between $\widehat{\beta}_m$ and $\widehat{\beta}_{-m}$ is found, the former oscillating around 33 and the latter below 24. The spectral R^2 still indicate a much better fit at low frequencies, but empirical evidence of cointegration is extremely weak. Notice in particular that if y and x are not cointegrated, $d_e = \max(d_x, d_y)$, as is amply confirmed by the Gaussian estimates, where one gets identical estimates of d_x and $d_e = (1.04, .91, .90)$ for $\ell = 22, 30, 40$. The results of Campbell and Shiller on this data set were, in their own words, inconclusive; our findings confirm those of Phillips and Ouliaris (1988), who were unable to reject the null of no cointegration at the 10% level.

c) Log prices (y) and wages (x) (monthly data), 1960M1-1979M12

The results in Table 5 tend to develop those of Engle and Granger (1987) by supporting an absence of a cointegrating relationship of any order. Where our conclusions differ is in the integration orders of x and y, in particular of log prices, which appear not to be unity, ranging from 1.54 to 1.60, while confidence intervals never include unity. This is not very surprising in that the inflation rate might plausibly be characterized as a stationary long memory process. W is always above 5.8, so we reject also at the 1% level the hypothesis that $d_x = d_y$, and so because this necessary condition for cointegration is not satisfied the analysis is taken no further.

d) Quantity theory of money (quarterly data): log M1, M2, M3 or L (y) and log GNP (x), where L denotes total liquid assets, 1959Q1 - 1981Q2.

Engle and Granger (1987) found the classical equation $MV = PY$ of the quantity theory of money to hold for $M = M2$, but not $M1, M3, L$. This is somewhat unsatisfactory since the latter monetary aggregates are linked with M2 in the long run, so that there might exist cointegration (albeit of different orders) between more than one of these aggregates and GNP. For $\log L$ in Table 6 we reject at the 1% level the hypothesis that GNP shares the same integration order. For M2, in Table 7, the necessary condition for cointegration is met, GLS confidence intervals tending to suggest integration orders around 1.3, which seems unsurprising since both aggregates are nominal. The $\widehat{\beta}_m$ are not noticeably influenced by m and are indeed the same as OLS (.99). Estimates of the d_e are strongly inconsistent with stationarity, ranging from 1.02 to 1.23, the confidence intervals excluding values below .88. Overall, it seems very difficult to draw reliable conclusions about the existence of fractional cointegration between these variables given such a small sample. The relationship of nominal GNP with M1 and M3, in Tables 8 and 9, appears much closer to that with M2 than Engle and Granger concluded, exploiting the greater flexibility of our framework. In particular, the common integration order for the bivariate raw data is again estimated via GLS to be 1.31, 1.39, 1.29, (for m = 16, 22, 30) for nominal GNP and M1 and 1.33, 1.44, 1.42 for nominal GNP and M3; estimates of d_e range from .76 to 1.20 for the former case and from .88 to 1.08 for the latter.

8. FINAL COMMENTS

The paper demonstrates that OLS estimates of a cointegrating vector are asymptotically matched or bettered in a variety of stationary and nonstationary cases by a narrow-band frequency domain estimate, FDLS. The overall superiority of FDLS relies on correlation between the cointegrating errors and regressors; in the absence of such correlation FDLS is inferior to OLS for stationary data, and comparable for nonstationary data. The finite-sample advantages of FDLS in correlated situations are observed in a small Monte Carlo study. FDLS is incorporated in a semiparametric methodology for investigating the possibility of fractional cointegration, which is applied to bivariate macroeconomic series.

The paper leaves open numerous avenues for further research. It is possible that the whole of x_t does not satisfy the conditions of either Section 3 or one of the cases I, II or III/IV of Section 5, but rather that subsets of x_t are classified differently. It is straightforward to extend our results to cover such situations, and we have not done so for the sake of simplicity, and because the case $p = 2$ is itself of practical importance. A more challenging development would cover such omitted cases as when x_t has integration order $\frac{1}{2}$, on the boundary between stationarity and nonstationarity, though this can be thought of as occupying a measure-zero subset of the parameter space. From a practical viewpoint a significant deficiency of our treatment of nonstationarity is the lack of allowance for deterministic trends, such as (possibly nonintegral) powers of t, but if these are suitably dominated by the stochastic trends it appears that the results of Sections 4 and 5 continue to hold.

A more challenging area for study is the extent to which we can improve on FDLS in our semiparametric context, with unknown integration orders, to mirror the improvement of OLS by various estimates in the $CI(1)$ case. There is also a need, as in the $CI(1)$ case, to allow for the possibility of more than one cointegrating relation, where we might wish to permit these to have different integration orders. Certainly it seems clear that results such as Propositions 4.1 and 4.2 and Lemma 5.1 can be established for more general quadratic forms, and so the extensive asymptotic theory for quadratic forms of stationary long memory series can be significantly extended in a nonstationary direction. The choice of bandwidth m in $\widehat{\beta}_m$ seems less crucial under nonstationarity than under stationarity, but nevertheless some criterion must be given to practitioners. For the stationary case, which seems of interest in financial applications, bandwidth theory of Robinson (1994c) can be developed, but there is a need also to develop asymptotic distribution theory for FDLS, useful application of which is likely to require bias-correction due to correlation between x and e. For the relatively short macroeconomic series analyzed in the present paper the semiparametric approach employed, while based on very mild assumptions, will not produce as reliable estimates of integration orders as correctly specified parametric time series models, and it is possible to analyze narrow-band β estimates in such a parametric framework also. In this connection a recent treatment of parametric inference in multivariate stationary long memory is given by Hosoya (1997).

APPENDIX

Proof of Theorem 3.1 From (2.1), (2.2) we have

$$\widehat{\beta}_m - \beta = \widehat{F}_{xx}(1,m)^{-1}\widehat{F}_{xe}(1,m). \tag{A.1}$$

By the Cauchy inequality, as in Robinson (1994a)

$$|\, \widehat{F}_{xe}^{(i)}(1,m) \,| \leq \left\{\widehat{F}_{xx}^{(i,i)}(1,m)\widehat{F}_{ee}(1,m)\right\}^{\frac{1}{2}}. \tag{A.2}$$

For any non-null $p \times 1$ vectors γ and δ, by Assumption A

$$\gamma' \left\{\widehat{F}_{zz}(1,m) - F_{zz}(\lambda_m)\right\}\delta = o_p\left(\{\gamma' F_{zz}(\lambda_m)\gamma\delta' F_{zz}(\lambda_m)\delta\}^{\frac{1}{2}}\right), \text{ as } n \to \infty \tag{A.3}$$

by a straightforward multivariate extension of Theorem 1 of Robinson (1994a) (see also Lobato, 1997), with

$$F_{zz}(\lambda) = \int_0^\lambda \mathrm{Re}\left\{f_{zz}(\mu)\right\} d\mu \sim G(\lambda), \text{ as } \lambda \to 0^+, \tag{A.4}$$

where $G(\lambda)$ has (i,j)-th element

$$G_{ij}(\lambda) = \frac{G_{ij}\lambda^{1-d_i-d_j}}{1-d_i-d_j},$$

G_{ij} being the (i,j)-th element of G. Applying (A.3), (A.4) and (2.4),

$$\Lambda_m^{-1}\left\{\widehat{F}_{zz}(1,m) - F_{zz}(\lambda_m)\right\}\Lambda_m^{-1} = o_p(\lambda_m),$$

where $\Lambda_m = diag\{\lambda_m^{-d_1}, ..., \lambda_m^{-d_p}\}$, so

$$\widehat{F}_{xx}^{(i,j)}(1,m) = G_{ij}(\lambda_m) + o_p\left(\lambda_m^{1-d_i-d_j}\right),$$

$$\begin{aligned}\widehat{F}_{ee}(1,m) &= \alpha'\widehat{F}_{zz}(1,m)\alpha = \alpha' F_{zz}(\lambda_m)\alpha + o_p(\alpha' F_{zz}(\lambda_m)\alpha) \\ &= O_p(\lambda_m^{1-2d_e}),\end{aligned}$$

because

$$\alpha' F_{zz}(\lambda)\alpha = \int_0^\lambda \alpha' f_{zz}(\mu)\alpha d\mu \sim c\int_0^\lambda \mu^{-2d_e}d\mu = c\frac{\lambda^{1-2d_e}}{1-2d_e}, \text{ as } \lambda \to 0^+.$$

Denote by $\widetilde{\Lambda}_m$, $\widetilde{G}(\lambda)$ the leading $(p-1)\times(p-1)$ submatrices of Λ_m, $G(\lambda)$. For any $(p-1)\times 1$ non-null vector $\nu=(\nu_1,...,\nu_{p-1})'$

$$\begin{aligned}\nu'\widetilde{\Lambda}_m^{-1}F_{xx}(\lambda_m)\widetilde{\Lambda}_m^{-1}\nu &\sim \nu'\widetilde{\Lambda}_m^{-1}\widetilde{G}(\lambda_m)\widetilde{\Lambda}_m^{-1}\nu \geq \omega\int_0^{\lambda_m}\sum_{j=1}^{p-1}\left\{(\lambda_m/\lambda)^{d_j}\nu_j\right\}^2 d\lambda \\ &= \omega\lambda_m\sum_{j=1}^{p-1}\frac{\nu_j^2}{1-2d_j},\end{aligned}$$

where ω is the smallest eigenvalue of the leading $(p-1)\times(p-1)$ submatrix of G, which is positive definite by Assumption A. It follows that $\widetilde{\Lambda}_m\widehat{F}_{xx}^{-1}(\lambda_m)\widetilde{\Lambda}_m = O_p(n/m)$, whence the proof is completed by elementary manipulation. □

To assist proof of Propositions 4.1 and 4.2, we introduce the following Lemma. In the sequel, C denotes a generic positive constant.

Lemma A.1 Under Assumption D,

$$S_{uv}(\lambda,\gamma)\overset{def}{=}\sum_{t=u}^{v}e^{it\lambda}\varphi_t(\gamma)$$

satisfies, for $0\leq u<v$, $0\leq|\lambda|\leq\pi$,

$$\begin{aligned}S_{uv}(\lambda,0) &= 1,\ u=0,\\ &= 0,\ u>0,\end{aligned}$$

and for $\gamma>0$

$$\begin{aligned}|S_{uv}(\lambda,\gamma)| &\leq C\min\left(v^{\gamma},\ \frac{(u+1)^{\gamma-1}}{|\lambda|},\ \frac{1}{|\lambda|^{\gamma}}\right),\ 0<\gamma\leq 1,\\ |S_{uv}(\lambda,\gamma)| &\leq C\min\left(v^{\gamma},\ \frac{v^{\gamma-1}}{|\lambda|}\right),\ \gamma>1.\end{aligned}$$

Proof The proof for $\gamma=0$ is trivial, so we consider $\gamma>0$. Drop the argument γ from $S_{uv}(\lambda,\gamma)$ and $\varphi_t(\gamma)$. Obviously $|S_{uv}(\lambda)|\leq Cv^{\gamma}$. For $0<\gamma\leq 1$ we can write, for $u<s<v$,

$$S_{uv}(\lambda)=\sum_{t=u}^{s-1}\varphi_t e^{it\lambda}+\sum_{t=s}^{v-1}(\varphi_t-\varphi_{t+1})\sum_{v=s}^{t}e^{iv\lambda}+\varphi_v\sum_{t=s}^{v}e^{it\lambda}$$

by summation-by-parts. Thus because

$$\left|\sum_{v=s}^{t}e^{iv\lambda}\right|\leq\frac{C(t-s)}{1+(t-s)|\lambda|},\ |\lambda|<\pi, \tag{A.5}$$

Assumption D implies that

$$|S_{uv}(\lambda)| \le C\left((s-1)^{\gamma} + \frac{s^{\gamma-1}}{|\lambda|}\right). \tag{A.6}$$

For $1/|\lambda| \le Cv$ we may choose $s \sim |\lambda|^{-1}$ so that (A.6) is $O(|\lambda|^{-\gamma})$. On the other hand we also have

$$S_{uv}(\lambda) = \sum_{t=u}^{v-1}(\varphi_t - \varphi_{t+1})\sum_{s=u}^{t} e^{is\lambda} + \varphi_v \sum_{t=u}^{v} e^{it\lambda}, \tag{A.7}$$

to give $|S_{uv}(\lambda)| \le C(u+1)^{\gamma-1}/|\lambda|$. For $\gamma > 1$, (A.7) gives instead

$$|S_{uv}(\lambda)| \le \frac{Cv^{\gamma-1}}{|\lambda|}.$$

□

Proof of Proposition 4.1 The discrete Fourier transform of a_{jt} is, from (4.1),

$$w_j(\lambda) = \frac{1}{\sqrt{2\pi n}}\sum_t a_{jt}e^{it\lambda} = \frac{1}{\sqrt{2\pi n}}\sum_t \varphi_{j,n-t}(\lambda)e^{it\lambda}\eta_{jt},\ j=1,2,$$

where

$$\varphi_{jt}(\lambda) \equiv \sum_{s=0}^{t}\varphi_{js}e^{is\lambda},\ j=1,2.$$

Thus by Assumption C

$$E\left(I_{aa}^{(1,2)}(\lambda)\right) = \frac{1}{2\pi n}\int_{-\pi}^{\pi}\Phi_1(\lambda,-\mu)\Phi_2(-\lambda,\mu)f_{12}(\mu)d\mu, \tag{A.8}$$

where $f_{ij}(\mu)$ is the cross spectral density of η_{it}, η_{jt} and

$$\Phi_j(\lambda,\mu) = \sum_t \varphi_{j,n-t}(\lambda)e^{it(\lambda+\mu)},\ j=1,2.$$

The modulus of (A.8) is bounded by

$$\frac{C}{n}\sup_{\mu}|f_{12}(\mu)|\left\{\int_{-\pi}^{\pi}|\Phi_1(\lambda,-\mu)|^2 d\mu \int_{-\pi}^{\pi}|\Phi_2(-\lambda,\mu)|^2 d\mu\right\}^{\frac{1}{2}}$$

$$\le \frac{C}{n}\left\{\prod_{j=1}^{2}\sup_{\mu} f_{jj}(\mu)\sum_t\left|\varphi_{j,n-t}(\lambda)\right|^2\right\}^{\frac{1}{2}}, \tag{A.9}$$

from Assumption C. From Lemma A.1, for $0 < |\lambda| < \pi$, $t = 1, ..., n$ and $i = 1, 2$

$$
\begin{aligned}
|\varphi_{it}(\lambda)| &= O\left(\frac{n^{\gamma_i - 1}}{|\lambda|} 1(\gamma_i > 1) + \frac{1}{|\lambda|^{\gamma_i}} 1(\gamma_i \leq 1)\right) \\
&= O\left(\frac{n^{\max(\gamma_i - 1, 0)}}{|\lambda|^{\min(\gamma_i, 1)}}\right) \qquad \text{(A.10)}
\end{aligned}
$$

when $\gamma_i > 0$. The latter bound also applies for $\gamma_i = 0$, when it is $O(1)$. Thus (A.9) is, for $0 < |\lambda| < \pi$,

$$
O\left(\frac{n^{\max(\gamma_1 - 1, 0) + \max(\gamma_2 - 1, 0)}}{|\lambda|^{\min(\gamma_1, 1) + \min(\gamma_2, 1)}}\right).
$$

For $\lambda = \lambda_j$, $j = 1, ..., M$, it is

$$
O\left(\frac{n^{\gamma_1 + \gamma_2}}{j^{\min(\gamma_1, 1) + \min(\gamma_2, 1)}}\right).
$$

Hence, when $\gamma_1 + \gamma_2 > 1$, by (2.5)

$$
\begin{aligned}
\left|E\left(F_{aa}^{(1,2)}(m+1, M)\right)\right| &\leq \frac{2\pi}{n} \sum_{j=m+1}^{M} \left|E\left(I_{aa}^{(1,2)}(\lambda_j)\right)\right| \\
&\leq C n^{\gamma_1 + \gamma_2 - 1} \sum_{j=m}^{\infty} j^{-\min(\gamma_1, 1) - \min(\gamma_2, 1)} \\
&= o\left(n^{\gamma_1 + \gamma_2 - 1}\right), \\
\left|E\left(F_{22}^{(1,2)}(1, m)\right)\right| &\leq C n^{\gamma_1 + \gamma_2 - 1} \sum_{j=1}^{m} j^{-\min(\gamma_1, 1) - \min(\gamma_2, 1)} \\
&= O\left(n^{\gamma_1 + \gamma_2 - 1}\right).
\end{aligned}
$$

Likewise, when $\gamma_1 + \gamma_2 < 1$,

$$
\begin{aligned}
\left|E\left(F_{aa}^{(1,2)}(1, M)\right)\right| &\leq C n^{\gamma_1 + \gamma_2 - 1} \sum_{j=1}^{m} \frac{1}{j^{\gamma_1 + \gamma_2}} \\
&= O\left(\left(\frac{n}{m}\right)^{\gamma_1 + \gamma_2 - 1}\right).
\end{aligned}
$$

□

Proof of Proposition 4.2 We first assume that $\gamma_2 > 0$. From (4.1) and (A.8)

$$
I_{aa}^{(1,2)}(\lambda) - E\left(I_{aa}^{(1,2)}(\lambda)\right) = \frac{1}{2\pi n} \sum_t \sum_s \varphi_{1,n-t}(\lambda) \varphi_{2,n-s}(-\lambda) e^{i(t-s)\lambda} \left\{\eta_{1t}\eta_{2s} - \gamma_{12}(s-t)\right\},
$$

with $\gamma_{ij}(s-t) = E(\eta_{it}\eta_{js})$. The left hand side of (4.6) is thus bounded by the real part of

$$\frac{1}{(2\pi)^2 n^4}\widetilde{\sum}\varphi_{1,n-t}(\lambda_j)\varphi_{2,n-s}(-\lambda_j)\varphi_{1,n-r}(-\lambda_k)\varphi_{2,n-q}(\lambda_k)e^{i(t-s)\lambda_j-i(r-q)\lambda_k}$$
$$\times\left\{E(\eta_{1t}\eta_{2s}\eta_{1r}\eta_{2q}) - \gamma_{12}(s-t)\gamma_{12}(q-r)\right\}, \tag{A.11}$$

where

$$\widetilde{\sum} = \sum_{j=m+1}^{M}\sum_{k=m+1}^{M}\sum_t\sum_s\sum_r\sum_q.$$

(A.11) can be written as

$$a_1 + a_2 + a_3,$$

where the three terms represent contributions from

$$\gamma_{12}(q-t)\gamma_{12}(s-r),\ \gamma_{11}(r-t)\gamma_{22}(q-s),\ k(s-t, r-s, q-r),$$

respectively, in the last line in (A.11). Now

$$\begin{aligned}
a_1 &= \frac{1}{4\pi^2 n^4}\widetilde{\sum}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\varphi_{1,n-t}(\lambda_j)\varphi_{2,n-s}(-\lambda_j)\varphi_{1,n-r}(-\lambda_k)\varphi_{2,n-q}(\lambda_k)\\
&\quad\times e^{i(t-s)\lambda_j-i(r-q)\lambda_k}e^{i(q-t)\lambda+i(s-r)\mu}f_{12}(\lambda)f_{12}(\mu)d\lambda d\mu\\
&= \frac{1}{4\pi^2 n^4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\sum_{j=m+1}^{M}\sum_{k=m+1}^{M}\varphi_{1,n-t}(\lambda_j)\sum_s\varphi_{2,n-s}(-\lambda_j)e^{-is(\lambda_j-\mu)}\\
&\quad\times\sum_r\varphi_{1,n-r}(-\lambda_j)e^{-ir(\lambda_k+\mu)}\sum_q\varphi_{2,n-q}(\lambda_k)e^{iq(\lambda_k+\lambda)}f_{12}(\lambda)f_{12}(\mu)d\lambda d\mu\\
&= \frac{1}{4\pi^2 n^4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\sum_{j=m+1}^{M}\sum_{k=m+1}^{M}\Phi_1(\lambda_j,-\lambda)\Phi_2(-\lambda_j,\mu)\\
&\quad\times\Phi_1(-\lambda_k,-\mu)\Phi_2(\lambda_k,\lambda)f_{12}(\lambda)f_{12}(\mu)d\lambda d\mu,
\end{aligned} \tag{A.12}$$

which is bounded in modulus by

$$\begin{aligned}
&\frac{C}{n^4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|\sum_{j=m+1}^{M}\Phi_1(\lambda_j,-\lambda)\Phi_2(-\lambda_j,\mu)\right|\\
&\times\left|\sum_{k=m+1}^{M}\Phi_1(-\lambda_k,-\mu)\Phi_2(\lambda_k,\lambda)\right|d\mu d\lambda\\
\le\ &\frac{C}{n^4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|\sum_{j=m+1}^{M}\Phi_1(\lambda_j,-\lambda)\Phi_2(-\lambda_j,\mu)\right|^2 d\mu d\lambda\\
=\ &\frac{C}{n^4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\sum_{j=m+1}^{M}\sum_{k=m+1}^{M}\Phi_1(\lambda_j,-\lambda)\Phi_2(-\lambda_j,\mu)
\end{aligned}$$

$$\times\Phi_2(-\lambda_k,\lambda)\Phi_2(\lambda_k,-\mu)d\mu d\lambda, \tag{A.13}$$

due to Assumption C and the fact that $\overline{\Phi_\ell(\lambda,\mu)} = \Phi_\ell(-\lambda,-\mu)$. Now for $\ell = 1,2$

$$\int_{-\pi}^{\pi}\Phi_\ell(\lambda_j,-\lambda)\Phi_\ell(-\lambda_k,\lambda)d\lambda = \int_{-\pi}^{\pi}\sum_t \varphi_{\ell,n-t}(\lambda_j)e^{it(\lambda_j-\lambda)}\sum_s \varphi_{\ell,n-s}(-\lambda_k)e^{-is(\lambda_k-\lambda)}d\lambda$$

$$= 2\pi c_{jk,\ell}, \tag{A.14}$$

where

$$c_{jk,\ell} = \sum_t \varphi_{\ell,n-t}(\lambda_j)\varphi_{\ell,n-t}(-\lambda_k)e^{it(\lambda_j-\lambda_k)},\ \ell = 1,2.$$

Thus (A.13) is

$$\frac{C}{n^4}\sum_{j=m+1}^{M}\sum_{k=m+1}^{M} c_{jk,1}c_{kj,2}. \tag{A.15}$$

Consider first the case $\gamma_1+\gamma_2 > 1$. By (A.10) and elementary inequalities

$$|c_{jk,\ell}| \le C\frac{n^{\max(\gamma_\ell,1)}}{(\lambda_j\lambda_k)^{\min(\gamma_\ell,1)}},$$

so that (A.15) is bounded in modulus by

$$\begin{aligned}\frac{C}{n^2}\left(\sum_{j=m+1}^{M}\frac{n^{\max(\gamma_1-1,0)+\max(\gamma_2-1,0)}}{\lambda_j^{\min(\gamma_1,1)+\min(\gamma_2,1)}}\right)^2 &\le Cn^{2(\gamma_1+\gamma_2-1)}\left(\sum_{j=m+1}^{M} j^{-\min(\gamma_1,1)-\min(\gamma_2,1)}\right)^2\\ &\le C\frac{n^{2(\gamma_1+\gamma_2-1)}}{m^{\min(\gamma_1,1)+\min(\gamma_2,1)-1}}\\ &= o\left(n^{2(\gamma_1+\gamma_2-1)}\right),\end{aligned}$$

using (2.5). Now consider the case $\gamma_1+\gamma_2 \le 1$ but $\frac{1}{2} < \gamma_1 < 1$ and $0 < \gamma_2 < \frac{1}{2}$ so that $\gamma_1+\gamma_2 > \frac{1}{2}$. First we deduce from Lemma A.1 the estimate

$$|c_{jk,1}| \le \frac{Cn}{(\lambda_j\lambda_k)^{\gamma_1}}. \tag{A.16}$$

Define

$$\tilde{\varphi}_{2t}(\lambda) = \varphi_{2,n-1}(\lambda) - \varphi_{2,n-t}(\lambda) = \sum_{s=n-t+1}^{n-1}\varphi_{2s}e^{is\lambda},\ t = 2,...,n$$

and $\tilde{\varphi}_{21}(\lambda) = 0$, so that

$$c_{kj,2} = \sum_{i=1}^{4} c_{kj,2}^{(i)},$$

where

$$\begin{aligned} c_{kj,2}^{(1)} &= \varphi_{2,n-1}(\lambda_k)\varphi_{2,n-1}(-\lambda_j)D_n(\lambda_k-\lambda_j) \\ c_{kj,2}^{(2)} &= \sum_t \widetilde{\varphi}_{2t}(\lambda_k)\widetilde{\varphi}_{2t}(-\lambda_j)e^{it(\lambda_k-\lambda_j)} \\ c_{kj,2}^{(3)} &= -\varphi_{2,n-1}(\lambda_k)\sum_t \widetilde{\varphi}_{2t}(-\lambda_j)e^{it(\lambda_k-\lambda_j)} \\ c_{kj,2}^{(4)} &= -\varphi_{2,n-1}(-\lambda_j)\sum_t \widetilde{\varphi}_{2t}(\lambda_k)e^{it(\lambda_k-\lambda_j)}, \end{aligned}$$

with

$$D_t(\lambda) = \sum_{j=1}^{t} e^{ij\lambda}.$$

Because

$$\begin{aligned} D_n(\lambda_k-\lambda_j) &= n,\ j=k \\ &= 0,\ j\neq k,\ \bmod n, \end{aligned} \tag{A.17}$$

using also (A.10) we have

$$\begin{aligned} c_{kj,2}^{(1)} &= n\left|\varphi_{2,n-1}(\lambda_j)\right|^2 = O\left(\frac{n}{\lambda_j^{2\gamma_2}}\right),\ j=k, \\ &= 0,\ j\neq k. \end{aligned} \tag{A.18}$$

To consider $c_{kj,2}^{(2)}$, note that from Lemma A.1, for $0\leq q\leq n-1$,

$$\begin{aligned} \left|c_{kj,2}^{(2)}\right| &\leq \sum_{t=1}^{n-q}\left|\widetilde{\varphi}_{2t}(\lambda_k)\widetilde{\varphi}_{2t}(-\lambda_j)\right| + \sum_{t=n-q+1}^{n}\left|\widetilde{\varphi}_{2t}(\lambda_k)\widetilde{\varphi}_{2t}(-\lambda_j)\right| \\ &\leq \frac{C}{\lambda_j\lambda_k}\sum_{t=1}^{n-q}(n-t+2)^{2(\gamma_2-1)} + \frac{Cq}{(\lambda_j\lambda_k)^{\gamma_2}} \\ &\leq C\left(\frac{q^{2\gamma_2-1}}{\lambda_j\lambda_k} + \frac{q}{(\lambda_j\lambda_k)^{\gamma_2}}\right) \\ &\leq \frac{C}{(\lambda_j\lambda_k)^{\gamma_2+1/2}}, \end{aligned} \tag{A.19}$$

on picking $q=[(\lambda_j\lambda_k)^{-1/2}]$. Next, to consider $c_{kj,2}^{(3)}$, we write

$$\sum_t \widetilde{\varphi}_{2t}(-\lambda_j)e^{it(\lambda_k-\lambda_j)} = e^{i(n+1)(\lambda_k-\lambda_j)}\sum_{t=1}^{n-1}\varphi_{2t}e^{-it\lambda_j}D_t(\lambda_j-\lambda_k),$$

which by (A.5) is bounded in modulus by

$$\sum_{t=1}^{n-1}|\varphi_{2t}|\,|D_t(\lambda_j-\lambda_k)| \leq C\frac{n^{\gamma_2+1}}{1+n\,|\lambda_j-\lambda_k|},$$

for $1 \le j,k \le n/2$. Thus by (A.10)

$$\left|c_{kj,2}^{(3)}\right| \le \frac{Cn^{\gamma_2+1}}{\lambda_j^{\gamma_2}(1+n|\lambda_j-\lambda_k|)}. \tag{A.20}$$

Likewise

$$\left|c_{kj,2}^{(4)}\right| \le \frac{Cn^{\gamma_2+1}}{\lambda_j^{\gamma_2}(1+n|\lambda_j-\lambda_k|)}. \tag{A.21}$$

With reference to (A.15) we have from (A.16) and (A.18)

$$\begin{aligned}\left|\frac{1}{n^4}\sum_{j=m+1}^{M}\sum_{k=m+1}^{M} c_{jk,1}c_{kj,2}^{(1)}\right| &\le \frac{C}{n^2}\sum_{j=m+1}^{M}\lambda_j^{-2(\gamma_1+\gamma_2)} \\ &\le Cn^{2(\gamma_1+\gamma_2-1)}\sum_{j=m+1}^{M} j^{-2(\gamma_1+\gamma_2)} \\ &\le C\frac{n^{2(\gamma_1+\gamma_2-1)}}{m^{2(\gamma_1+\gamma_2-1/2)}} = o\left(n^{2(\gamma_1+\gamma_2-1)}\right).\end{aligned}$$

From (A.16) and (A.19)

$$\begin{aligned}\left|\frac{1}{n^4}\sum_{j=m+1}^{M}\sum_{k=m+1}^{M} c_{jk,1}c_{kj,2}^{(2)}\right| &\le \frac{C}{n^3}\left(\sum_{j=m+1}^{M}\lambda_j^{-\gamma_1-\gamma_2-\frac{1}{2}}\right)^2 \\ &\le Cn^{2(\gamma_1+\gamma_2-1)}\left(\sum_{j=m+1}^{M} j^{-\gamma_1-\gamma_2-\frac{1}{2}}\right)^2 \\ &\le C\frac{n^{2(\gamma_1+\gamma_2-1)}}{m^{2(\gamma_1+\gamma_2-1/2)}} = o\left(n^{2(\gamma_1+\gamma_2-1)}\right).\end{aligned}$$

From (A.16) and (A.20)

$$\left|\frac{1}{n^4}\sum_{j=m+1}^{M}\sum_{k=m+1}^{M} c_{jk,1}c_{kj,2}^{(3)}\right| \le Cn^{\gamma_2-2}\sum_{j=m+1}^{M}\lambda_j^{-\gamma_1-\gamma_2}\sum_{k=m+1}^{M}\frac{1}{(1+n|\lambda_j-\lambda_k|)\lambda_k^{\gamma_1}}$$

$$\begin{aligned}&\le Cn^{2(\gamma_1+\gamma_2-1)}\sum_{j=m+1}^{\infty} j^{-\gamma_1-\gamma_2}\left\{j^{-\gamma_1}+\sum_{k=j+1}^{2j}(k-j)^{-1}k^{-\gamma_1}+\sum_{k=2j+1}^{M}(k-j)^{-1}k^{-\gamma_1}\right\} \\ &\le Cn^{2(\gamma_1+\gamma_2-1)}\sum_{j=m+1}^{\infty} j^{-\gamma_1-\gamma_2}\left\{j^{-\gamma_1}+j^{-\gamma_1}\sum_{k=1}^{j}k^{-1}+\sum_{k=j}^{\infty}k^{-\gamma_1-1}\right\} \\ &\le Cn^{2(\gamma_1+\gamma_2-1)}\sum_{j=m+1}^{\infty} j^{-2\gamma_1-\gamma_2}\log j \\ &\le Cn^{2(\gamma_1+\gamma_2-1)}\sum_{j=m+1}^{\infty} j^{\varepsilon-2\gamma_1-\gamma_2}\end{aligned} \tag{A.22}$$

for $\varepsilon > 0$. Now because $\gamma_1+\gamma_2 > \frac{1}{2}$ and $\gamma_1 > \frac{1}{2}$ we can choose ε such that $2\gamma_1+\gamma_2-\varepsilon > 1$ in which case (A.22) is $o\left(n^{2(\gamma_1+\gamma_2-1)}\right)$. In view of (A.21) the same bound is obtained on replacing $c_{kj,2}^{(3)}$ by $c_{kj,2}^{(4)}$. Thus we have shown that $a_1 = o\left(n^{2(\gamma_1+\gamma_2-1)}\right)$.

Next,

$$\begin{aligned} a_2 &= \frac{1}{n^4}\widetilde{\sum}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\varphi_{1,n-t}(\lambda_j)\varphi_{2,n-s}(-\lambda_j)\varphi_{1.n-r}(-\lambda_k)\varphi_{2,n-q}(\lambda_k)e^{i(t-s)\lambda_j-i(r-q)\lambda_k} \\ &\quad\times e^{i(r-t)\lambda+i(q-s)\mu}f_{11}(\lambda)f_{22}(\mu)d\lambda d\mu \\ &= \frac{1}{n^4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\sum_{j=m+1}^{M}\sum_{k=m+1}^{M}\Phi_1(\lambda_j,-\lambda)\Phi_2(-\lambda_j,-\mu) \\ &\qquad\times\Phi_1(-\lambda_k,\lambda)\Phi_2(\lambda_k,\mu)f_{11}(\lambda)f_{22}(\mu)d\lambda d\mu \end{aligned}$$

and this is bounded in modulus by (A.13) $= o\left(n^{2(\gamma_1+\gamma_2-1)}\right)$, in the same way as was shown for (A.12).

Finally

$$\begin{aligned} a_3 = \frac{1}{n^4}\widetilde{\sum}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}&\varphi_{1,n-t}(\lambda_j)\varphi_{2,n-s}(-\lambda_j)\varphi_{1.n-r}(-\lambda_k)\varphi_{2,n-q}(\lambda_k)e^{i(t-s)\lambda_j-i(r-q)\lambda_k} \\ &\times e^{i(s-t)\mu_1+i(r-s)\mu_2+i(q-r)\mu_3}f(\mu_1,\mu_2,\mu_3)\prod_{i=1}^{3}d\mu_i \\ = \frac{1}{n^4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}&\sum_{j=m+1}^{M}\sum_{k=m+1}^{M}\Phi_1(\lambda_j,-\mu_1)\Phi_2(-\lambda_j,\mu_1-\mu_2) \\ &\times\Phi_1(-\lambda_k,\mu_2-\mu_3)\Phi_2(\lambda_k,\mu_3)f(\mu_1,\mu_2,\mu_3)\prod_{i=1}^{3}d\mu_i \end{aligned}$$

and in view of Assumption E this is bounded in modulus by

$$\begin{aligned} &\frac{C}{n^4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|\sum_{j=m+1}^{M}\Phi_1(\lambda_j,-\mu_1)\Phi_2(-\lambda_j,\mu_1-\mu_2)\right. \\ &\qquad\times\left.\sum_{k=m+1}^{M}\Phi_1(-\lambda_k,\mu_2-\mu_3)\Phi_2(\lambda_k,\mu_3)\right|\prod_i d\mu_i \\ &\leq \frac{C}{n^4}\left(\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|\sum_{j=m+1}^{M}\Phi_1(\lambda_j,-\mu_1)\Phi_2(-\lambda_j,\mu_1-\mu_2)\right|^2\prod_i d\mu_i\right. \end{aligned}$$

$$\times \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \left| \sum_{k=m+1}^{M} \Phi_1(-\lambda_k, \mu_2 - \mu_3)\Phi_2(\lambda_k, \mu_3) \right|^2 \prod_i d\mu_i \Bigg)^{\frac{1}{2}}. \tag{A.23}$$

The second integral is bounded by

$$2\pi \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sum_{k=m+1}^{M} \Phi_1(-\lambda_k, \mu_2 - \mu_3)\Phi_2(\lambda_k, \mu_3)$$

$$\times \sum_{j=m+1}^{M} \Phi_1(\lambda_k, \mu_3 - \mu_2)\Phi_2(-\lambda_k, -\mu_3) \prod_i d\mu_i. \tag{A.24}$$

Because (cf (A.14))

$$\int_{-\pi}^{\pi} \Phi_1(-\lambda_k, \mu_2 - \mu_3)\Phi_1(\lambda_j, \mu_3 - \mu_2) d\mu_2 = 2\pi c_{jk,1},$$

it follows that (A.24) $= O\left(\sum\sum_{i,j=m+1}^{M} c_{jk,1}c_{kj,2}\right)$. Treating the other integral in (A.23) in the same way we see that (A.23) is bounded by (A.15) $= o\left(n^{2(\gamma_1+\gamma_2-1)}\right)$, to complete the proof of (4.6) when $\gamma_2 > 0$. For $\gamma_2 = 0$, the same proof applies on substituting 1 for $\Phi_{2t}(\lambda)$ and $D_n(\lambda+\mu)$ for $\Phi_2(\lambda, \mu)$, to deduce that $a_j = o(1)$ for $j = 1, 2, 3$.

To prove (4.7) we start by bounding the left side by an analogous expression to (A.11), with $\sum_{j=m+1}^{M}\sum_{k=m+1}^{M}$ replaced by $\sum_{j=1}^{m}\sum_{k=1}^{m}$ in $\widetilde{\sum}$. Thus the revised a_1 is bounded by

$$Cn^{2(\gamma_1+\gamma_2-1)} \left\{ \sum_{j=1}^{M} j^{-2(\gamma_1+\gamma_2)} + \left(\sum_{j=1}^{M} j^{-\gamma_1-\gamma_2-\frac{1}{2}} \right)^2 + \sum_{j=1}^{M} j^{-\gamma_1-2\gamma_2} \log j \right\}$$

$$= O\left(n^{2(\gamma_1+\gamma_2-1)}\right),$$

while the revised a_2 and a_3 are similarly easily seen to have the same bound. □

Proof of Lemma 5.1 From (5.1)

$$\begin{aligned} x_{it} &= (1-L)^{-d_i}\{u_{it}1(t>0)\}, \ i = 1, ..., p-1, \\ e_t &= (1-L)^{-d_e}\{u_{pt}1(t>0)\}, \end{aligned}$$

where u_{it} is the i-th element of u_t. We take

$$\varphi_k(\gamma) = \frac{\Gamma(k+\gamma)}{\Gamma(\gamma)\Gamma(k+1)}, \tag{A.25}$$

since this is the coefficient of L^k in the Taylor expansion of $(1-L)^{-\gamma}$, and choose $\varphi_{1k} = \varphi_k(d_i)$ in both cases (5.14) and (5.15), with $\varphi_{2k} = \varphi_k(d_e)$ in (5.14) and $\varphi_{2k} = \varphi_k(d_j)$ in (5.15). Now

$$\varphi_k(\gamma) = O\left((1+k)^{\gamma-1}\right) \tag{A.26}$$

from Abramowitz and Stegun (1970). Also

$$\varphi_k(\gamma) - \varphi_{k+1}(\gamma) = -\varphi_{k+1}(\gamma - 1) = O\left((1+k)^{\gamma-2}\right),$$

to check Assumption D. Next we take $\eta_{1t} = u_{it}$ in each case and $\eta_{2t} = u_{pt}$ and $\eta_{2t} = u_{jt}$ in (5.14) and (5.15) respectively. Now the spectral density matrix of u_t is $(2\pi)^{-1}C(e^{i\lambda})\sum C(e^{i\lambda})^*$ whose modulus is bounded by $C\left(\sum_{j=0}^{\infty}\|C_j\|\right)^2 < \infty$ from (5.7). Thus Assumption C is satisfied. Finally the fourth cumulant of $u_{i0}, u_{ia}, u_{j,a+b}$, $u_{j,a+b+c}$ is, for $a, b, c \geq 0$

$$cum\left(\sum_{-\infty}^{\infty} c'_{i,-d}\varepsilon_d,\ \sum_{-\infty}^{\infty} c'_{i,a-e}\varepsilon_e,\ \sum_{-\infty}^{\infty} c'_{j,a+b-f}\varepsilon_f,\ \sum_{-\infty}^{\infty} c'_{j,a+b+c-g}\varepsilon_g\right),$$

where c'_{ij} is the i-th row of C_j. This is bounded in absolute value by

$$C\sum_{d=-\infty}^{\infty}\|c_{i,-d}\|\,\|c_{i,a-d}\|\,\|c_{i,a+b-d}\|\,\|c_{i,a+b+c-d}\|.$$

Because the sum of this, over all a, b, c, is finite due to (5.7), it follows that the Fourier coefficients of the fourth cumulant spectrum of $u_{it}, u_{it}, u_{jt}, u_{jt}$ are absolute summable, so that their spectrum is indeed bounded and Assumption E is satisfied. □

Proof of Lemma 5.2 For brevity we shall write $\Delta = \Delta(d)$ in the sequel.

$$\begin{aligned}\Delta\widehat{F}_{xx}(1,m)\Delta &= \Delta\widehat{F}_{xx}(1,n-1)\Delta \\ &\quad -\Delta\left\{\widehat{F}_{xx}(m+1,n-1) - E\left(\widehat{F}_{xx}(m+1,n-1)\right)\right\}\Delta \\ &\quad -\Delta E\left(\widehat{F}_{xx}(m+1,n-1)\right)\Delta.\end{aligned}$$

In view of Lemma 5.1 the last two components are $o_p(1)$ and $o(1)$ from Propositions 4.2 and 4.1 respectively. The proof is completed by appealing to Theorem 5.1. □

Proof of Lemma 5.3 We begin by estimating the b_{ijt}. First

$$\|C_j\| \leq \left(\sum_{\ell=j}^{\infty}\|C_\ell\|^2\right)^{\frac{1}{2}} \qquad \text{(A.27)}$$

$$\leq \frac{2}{j}\sum_{k=[\frac{j}{2}]}^{j}\left(\sum_{\ell=k}^{\infty}\|C_\ell\|^2\right)^{\frac{1}{2}} \qquad \text{(A.28)}$$

$$= o(j^{-1}) \text{ as } j \to \infty, \qquad \text{(A.29)}$$

where (A.28) is due to monotonic decay of the right hand side of (A.27), and (A.29) follows from (5.4). The i-th diagonal element of D_k is $\varphi_k(d_i)$ for $i = 1, ..., p-1$ and

$\varphi_k(d_e)$ for $i = p$, where $\varphi_k(\gamma)$ is given by (A.25). Now from (5.9), for $i = 1, ..., p-1$

$$\begin{aligned} \|b_{ijt}\| &\leq \sum_{\ell=0}^{r} \varphi_\ell(d_i) \|C_{j-\ell}\| + \sum_{\ell=r+1}^{\min(j,t-1)} \varphi_\ell(d_i) \|C_{j-\ell}\| \\ &\leq C \max_{j-r \leq \ell \leq j} \|C_\ell\| \sum_{\ell=0}^{r} \ell^{d_i-1} + C r^{d_i-1} \sum_{0}^{\infty} \|C_\ell\| \\ &\leq C r^{d_i} (j-r)^{-1} + C r^{d_i-1} \end{aligned}$$

for $1 \leq r < \min(j, t-1)$. It follows that for $j \leq 2t$

$$\|b_{ijt}\| \leq C j^{d_i-1}, \; i = 1, ..., p-1, \tag{A.30}$$

on taking $r \sim j/2$. For $j > 2t$ we have more immediately

$$\|b_{ijt}\| \leq \sum_{\ell=0}^{t-1} |\varphi_\ell(d_i)| \parallel C_{j-\ell} \parallel \leq C(j-t)^{-1} t^{d_i}, \; i = 1, ..., p-1.$$

Similarly

$$\begin{aligned} \|b_{pjt}\| &\leq O\left(j^{d_e-1}\right), \; j \leq 2t, \\ &= O\left((j-t)^{-1} t^{d_e}\right), \; j > 2t. \end{aligned} \tag{A.31}$$

Next notice that $b_{ij} = b_{ijt}$ for $0 \leq j < t$, so from (5.8)

$$\begin{aligned} E(x_{it} e_t) &= \sum_{j=0}^{t-1} b'_{ij} \sum b_{pj} + \sum_{j=t}^{\infty} b'_{ijt} \sum b_{pjt} \\ &= \xi_i - \sum_{j=t}^{\infty} b'_{ij} \sum b_{pj} + \sum_{j=t}^{2t-1} b'_{ijt} \sum b_{pjt} + \sum_{j=2t}^{\infty} b'_{ijt} \sum b_{pjt} \\ &= \xi_i + O\left(\sum_{j=t}^{\infty} j^{d_i+d_e-2} + t.t^{d_i+d_e-2} + t^{d_i+d_e} \sum_{j=t}^{\infty} j^{-2}\right) \\ &= \xi_i + O\left(t^{d_i+d_e-1}\right) \end{aligned}$$

because $d_i + d_e < 1$, where this and (A.30), (A.31) imply that $|\xi_i| < \infty$. On the other hand

$$\begin{aligned} |E(\bar{x}_i \bar{e})| &\leq \frac{C}{n^2} \sum_s \sum_t \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \|b_{ijs}\| \, \|b_{pkt}\| \\ &\leq \frac{C}{n^2} \left\{ \sum_s \sum_t \sum_{\max(0,s-t)}^{\min(2s,s+t)} j^{d_i-1} (j+t-s)^{d_e-1} + \sum_s \sum_t t^{2d_e} \sum_{s+t}^{2s} j^{d_i-1} (j-s)^{-1} \right. \\ &\quad \left. + \sum_s \sum_t s^{d_i} \sum_{2s}^{s+t} (j+t-s)^{d_e-1} (j-s)^{-1} + \sum_s \sum_t s^{d_i} t^{d_e} \sum_{\min(2s,s+t)}^{\infty} (j-s)^{-2} \right\} \\ &= O\left(n^{d_i-d_e-1}\right). \end{aligned}$$

The proof is routinely completed in view of (2.3). □

Proof of Theorem 5.2 It is convenient to introduce the abbreviating notation

$$\widetilde{A}=\widehat{F}_{xx}(1,n-1),\ \widetilde{a}=\widehat{F}_{xe}(1,n-1),\ \widehat{A}=\widehat{F}_{xx}(1,m),\ \widehat{a}=\widehat{F}_{xe}(1,m).$$

Thus

$$\widehat{\beta}_{n-1}-\beta=\Delta(\Delta\widetilde{A}\Delta)^{-1}\Delta\widetilde{a}, \tag{A.32}$$

$$\widehat{\beta}_{m}-\beta=\Delta(\Delta\widehat{A}\Delta)^{-1}\Delta\widehat{a}. \tag{A.33}$$

Now

$$\begin{aligned}\Delta(\widetilde{A}-\widehat{A})\Delta &= \Delta\left\{(\widetilde{A}-\widehat{A})-E(\widetilde{A}-\widehat{A})\right\}\Delta+\Delta E(\widetilde{A}-\widehat{A})\Delta \\ &\to_p 0 \end{aligned} \tag{A.34}$$

from Propositions 4.2 and 4.1, Assumption F and Lemma 5.1, so that $\Delta\widetilde{A}\Delta,\Delta\widehat{A}\Delta\Rightarrow V(d,\Omega)$ by Theorem 5.1. Now denote by $\widetilde{a}_i,\widehat{a}_i$ the i-th elements of $\widetilde{a},\widehat{a}$. From Proposition 4.2, Assumption F and Lemma 5.1

$$\widetilde{a}_i=E(\widetilde{a}_i)+(\widetilde{a}_i-E(\widetilde{a}_i))=E(\widetilde{a}_i)+O_p(n^{d_i+d_e-1}),\ i=1,...,p-1,$$

whereas from Lemma 5.3

$$\lim_{n\to\infty} n^{d_{\min}-\frac{1}{2}}\Delta E(\widetilde{a})=\xi.$$

Then (5.17) follows from (5.16). Finally

$$\widehat{a}_i=E(\widehat{a}_i)+\{\widehat{a}_i-E(\widehat{a}_i)\}=O\left(\left(\frac{n}{m}\right)^{d_i+d_e-1}\right)+O_p\left(n^{d_i+d_e-1}\right),$$

from Proposition 4.2, and 4.1, Assumption F and Lemma 5.1 so the i-th element of (A.33) is

$$\begin{aligned}O_p\left(n^{\frac{1}{2}-d_i}\max_{1\le j<p} n^{\frac{1}{2}-d_j}\left(\frac{n}{m}\right)^{d_j+d_e-1}\right) &= O_p\left(n^{1-d_i-d_{\min}}\left(\frac{n}{m}\right)^{d_{\min}+d_e-1}\right)\\ &= o_p\left(n^{1-d_i-d_{\min}}\right),\end{aligned}$$

since $d_{\min}+d_e<1$, to complete the proof of (5.18). □

Proof of Lemma 5.4 For $1\le j\le m$, writing $\widetilde{\Gamma}_j=\sum_{\ell\ge j}^{\infty}\gamma_\ell$,

$$\begin{aligned}E\left\{I_{xe}(\lambda_j)\right\} &= \frac{1}{2\pi n}\sum_s\sum_t(\gamma_{s-1}+...+\gamma_{s-t})e^{i(t-s)\lambda_j}\\ &= \frac{1}{2\pi n}\sum_s\sum_t\left(\widetilde{\Gamma}_{s-t}-\widetilde{\Gamma}_{s-1}\right)e^{i(t-s)\lambda_j}\\ &= \frac{1}{2\pi}\sum_{1-n}^{n-1}\left(1-\frac{|\ell|}{n}\right)\widetilde{\Gamma}_\ell e^{-i\ell\lambda_j}\end{aligned} \tag{A.35}$$

from (A.17). Now for $\ell \geq 0$ $\tilde{\Gamma}_\ell = \Gamma_\ell$, whereas for $\ell < 0$, $\tilde{\Gamma}_\ell = \Gamma_0 + \Gamma_{-1} - \Gamma_{\ell-1}$, so (A.35) has real part

$$\frac{1}{2\pi}\sum_0^{n-1}\left(1-\frac{\ell}{n}\right)\Gamma_\ell \cos \ell\lambda_j + \frac{1}{2\pi}\sum_{1-n}^{-1}\left(1+\frac{\ell}{n}\right)(\Gamma_0+\Gamma_{-1}-\Gamma_{\ell-1})\cos \ell\lambda_j. \tag{A.36}$$

The first term can be written

$$\frac{1}{4\pi}\sum_{1-n}^{n-1}\left(1-\frac{|\ell|}{n}\right)\Gamma_{|\ell|}\cos \ell\lambda_j + \frac{\Gamma_0}{4\pi}.$$

To deal with the second term of (A.36) note that for $1 \leq j \leq n-1$

$$\sum_{\ell=0}^{n-1} \ell e^{i\ell\lambda_j} = \frac{e^{i\lambda_j}-1}{(1-e^{i\lambda_j})^2} - \frac{(n-1)}{1-e^{i\lambda_j}} = \frac{-n}{1-e^{i\lambda_j}},$$

which has real part

$$-\frac{n}{2}\left(\frac{1}{1-e^{i\lambda_j}} + \frac{1}{1-e^{-i\lambda_j}}\right) = -\frac{n}{2}\left(\frac{2-2\cos\lambda_j}{2-2\cos\lambda_j}\right) = -\frac{n}{2}.$$

Thus, the second term in (A.36) is

$$\begin{aligned}&\frac{(\Gamma_0+\Gamma_{-1})}{2\pi}\sum_0^{n-1}\left(1-\frac{\ell}{n}\right)\cos\ell\lambda_j - \frac{\Gamma_0+\Gamma_{-1}}{2\pi}\\ &-\frac{1}{4\pi}\sum_{1-n}^{n-1}\left(1-\frac{|\ell|}{n}\right)\Gamma_{-|\ell|-1}\cos\ell\lambda_j + \frac{\Gamma_{-1}}{4\pi}\\ &= -\frac{1}{4\pi}\sum_{1-n}^{n-1}\left(1-\frac{|\ell|}{n}\right)\Gamma_{-|\ell|-1}\cos\ell\lambda_j + \frac{\Gamma_0}{4\pi}.\end{aligned}$$

It follows that (A.35) has real part

$$\frac{1}{4\pi}\sum_{1-n}^{n-1}\left(1-\frac{|\ell|}{n}\right)\left(\Gamma_{-|\ell|} - \Gamma_{-|\ell|-1}\right)\cos\ell\lambda_j,$$

which is the Cesaro sum, to $n-1$ terms, of the Fourier series of $h(\lambda_j)/2$. Equivalently we can write

$$E\left\{\frac{n}{m}\hat{F}_{xe}(1,m)\right\} = \frac{1}{4\pi nm}\sum_{j=1}^{m}\int_{-\pi}^{\pi}|D_n(\lambda-\lambda_j)|^2\, h(\lambda)d\lambda. \tag{A.37}$$

Fix $\varepsilon > 0$. There exists $\delta > 0$ such that $\|h(\lambda) - h(0)\| < \varepsilon$ for $0 < |\lambda| \leq \delta$. Let n be large enough that $2\lambda_m < \delta$. Then the difference between the right hand side of

(A.37) and $h(0)/2$ is bounded in absolute value by

$$\begin{aligned}
&\frac{1}{4\pi nm}\sum_{j=1}^{m}\int_{-\pi}^{\pi}|D_n(\lambda-\lambda_j)|^2\,\|h(\lambda)-h(0)\|\,d\lambda \\
\leq\ &\frac{1}{4\pi nm}\left\{\varepsilon\max_{1\leq j\leq m}\int_{-\delta}^{\delta}|D_n(\lambda-\lambda_j)|^2\,d\lambda+\sup_{\frac{\delta}{2}<|\lambda|<\pi}|D_n(\lambda)|^2\left(\int_{-\pi}^{\pi}\|h(\lambda)\|\,d\lambda+2\pi\,\|h(0)\|\right)\right\} \\
=\ &O\left(\varepsilon+\frac{1}{n}\right),
\end{aligned}$$

because of Assumption G, (A.5) and

$$\int_{-\pi}^{\pi}|D_n(\lambda)|^2\,d\lambda=2\pi n.$$

Because ε is arbitrary, the proof is complete. □

Proof of Theorem 5.3 (5.22) is familiar under somewhat different conditions from ours (see e.g. Stock, 1987, Phillips, 1988), but we briefly describe its proof in order to indicate how the outcome differs from (5.23). We have

$$n\left(\widehat{\beta}_{n-1}-\beta\right)=\left(n^{-1}\widetilde{A}\right)^{-1}\left\{(\widetilde{a}-E(\widetilde{a}))+E(\widetilde{a})\right\}.$$

Now

$$n^{-1}\widetilde{A}\Rightarrow V(\iota,\Omega),\ \widetilde{a}-E(\widetilde{a})\Rightarrow U(\Omega),\ \text{as } n\to\infty \tag{A.38}$$

from Assumption F, Theorem 5.1 and the continuous mapping theorem. Because $E(\widetilde{a})\to\Gamma_0$ by elementary calculations and $V(\iota,\Omega)$ is a.s. of full rank by Phillips and Hansen (1990), (5.22) is proved. Next

$$\begin{aligned}
n(\widehat{\beta}_m-\beta) &= \left[n^{-1}\widetilde{A}+n^{-1}\left\{(\widehat{A}-\widetilde{A})-E(\widehat{A}-\widetilde{A})\right\}+n^{-1}E(\widehat{A}-\widetilde{A})\right]^{-1} \\
&\quad\times[\widetilde{a}-E(\widetilde{a})+\{(\widehat{a}-\widetilde{a})-E(\widehat{a}-\widetilde{a})\}+E(\widehat{a})].
\end{aligned}$$

Since $n^{-1}\left\{\widehat{A}-\widetilde{A}-E(\widehat{A}-\widetilde{A})\right\}\to_p 0$, $n^{-1}E(\widehat{A}-\widetilde{A})\to 0$ and $\widehat{a}-\widetilde{a}-E(\widehat{a}-\widetilde{a})\to_p 0$ from Propositions 4.1 and 4.2, and Lemma 5.1, the proof of (5.23) is completed by invoking (A.38), Lemma 5.4 and (2.4). □

Proof of Theorem 5.4 From (A.32), Theorem 5.1, (5.16), (4.4) of Proposition 4.1, and (4.7) of Proposition 4.2 we deduce (5.24). Next, using (A.33),

$$\widehat{\beta}_{n-1}-\widehat{\beta}_m=\Delta(\Delta\widehat{A}\Delta)^{-1}\left\{\Delta(\widehat{A}-\widetilde{A})\Delta\right\}(\Delta\widetilde{A}\Delta)^{-1}\Delta\widetilde{a} \tag{A.39}$$

$$-(\Delta\widehat{A}\Delta)^{-1}\Delta\left\{(\widehat{a}-\widetilde{a})-E(\widehat{a}-\widetilde{a})\right\}. \tag{A.40}$$

The i-th element of the right side of (A.39) is $o_p\left(n^{d_i+d_e}\right)$ by arguments used in the previous proof and (A.34), while the i-th element of (A.40) is also $o_p\left(n^{d_i+d_e}\right)$ on applying also (4.3) of Proposition 4.1 and (4.6) of Proposition 4.2, to prove (5.25). Then (5.26) is a consequence of (5.24) and (5.25).

Proof of Theorem 5.6 The proof of (5.28) follows routinely from (A.32), (5.13), (5.16) and (5.27). Then (5.29) is a consequence of (5.25) and (5.28), because in view of (A.40) it is clear that (5.25) holds for all $d_e < d_{\min}$. □

REFERENCES

Abramovitz, M. and I. Stegun (1970) *Handbook of Mathematical Functions*, New York: Dover.

Akonom, J. and Gourieroux, (1987) "A Functional Central Limit Theorem for Fractional Processes", preprint, CEPREMAP.

Bossaerts, P. (1988) "Common Nonstationary Components of Asset Prices", *Journal of Economic Dynamics and Control*, 12, 347-364.

Campbell, J.Y. and R.J. Shiller (1987) "Cointegration and Tests of Present Value Models", *Journal of Political Economy*, 95, 1062-1089.

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. and K. Lai (1993) "A Fractional Cointegration Analysis of Purchasing Power Parity", *Journal of Business and Economic Statistics*, 11, 103-112.

Dickey, D.A. and W.A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, 427-431.

Diebold, F.X. and G. Rudebusch (1989) "Long Memory and Persistence in Aggregate Output", *Journal of Monetary Economics*, 24, 189-209.

Dolado, J. and F. Marmol (1996) "Efficient Estimation of Cointegrating Relationships Among Higher Order and Fractionally Integrated Processes", *Working Paper n.9617*, Banco de Espana.

Engle, R. (1974) "Band Spectrum Regression", *International Economic Review*, 15, 1-11.

Engle, R. and C. Granger (1987), "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 251-276.

Geweke, J. and S. Porter-Hudak (1983) "The Estimation and Application of Long Memory Time Series Models", *Journal of Time Series Analysis*, 4, 221-238.

Gourieroux, C., Maurel, F. and A. Monfort (1989) "Least Squares and Fractionally Integrated Regressors", *Working Paper #8913*, INSEE.

Hannan, E.J. (1963) "Regression for Time Series with Errors of Measurement", *Biometrika*, 50, 293-302.

Hosoya, Y. (1997) "A Limit Theory of Long-Range Dependence and Statistical Inference in Related Models", *Annals of Statistics*, 25, 105-137.

Hurvich, C.M., R. Deo and Brodsky, J. (1998) "The Mean Squared Error of Geweke and Porter-Hudak's Estimates of the Memory Parameter of a Long Memory Time Series", *Journal of Time Series Analysis*, 19, 19-46.

Hurvich, C.M. and B.K. Ray (1995) "Estimation of the Memory Parameter for Nonstationary or Noninvertible Fractionally Integrated Processes", *Journal of Time Series Analysis*, 16, 17-41.

Jeganathan, P. (1996) "On Asymptotic Inference in Cointegrated Time Series with Fractionally Integrated Errors", University of Michigan, Dept. of Statistics, *Technical Report #277.*

Johansen, S. (1988) "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregression Models", *Econometrica*, 59, 1551-1580.

Künsch, H.R. (1987) "Statistical Aspects of Self-Similar Processes", *Proceedings of the First World Congress of the Bernoulli Society*, VNU Science Press, 1, 67-74.

Lobato, I. (1997) "Consistency of the Averaged Cross-Periodogram in Long Memory Series", *Journal of Time Series Analysis*, 18, 137-156.

Lobato, I. and P.M. Robinson (1996) "Averaged Periodogram Estimation of Long Memory", *Journal of Econometrics*, 73, 303-324.

Mandelbrot, B.B. and J.R. Van Ness (1968) "Fractional Brownian Motions, Fractional Noises and Applications", *SIAM Review*, 10, 422-437.

Marinucci, D. and P.M. Robinson (1997) "A Functional Central Limit Theorem for Multivariate Fractional Time Series", in preparation.

Phillips, P.C.B. (1988) "Weak Convergence of Sample Covariance Matrix to Stochastic Integrals via Martingale Approximation", *Econometric Theory*, 4, 528-533.

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

Phillips, P.C.B. (1991b) "Spectral Regression for Cointegrated Time Series", *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, (W.A. Barnett, J. Powell and G. Tauchen, eds.), Cambridge: Cambridge University Press, 413-435.

Phillips, P.C.B. and N. Durlauf (1986) "Multiple Time Series Regression with Integrated Processes", *Review of Economic Studies*, 53, 473-495.

Phillips, P.C.B. and B.E. Hansen (1990) "Statistical Inference in Instrumental Variable Regression with $I(1)$ Variables", *Review of Economic Studies*, 57, 99-125.

Phillips, P.C.B. and S. Ouliaris (1988) "Testing for Co-integration Using Principal Components Methods", *Journal of Economic Dynamics and Control*, 12, 205-230.

Robinson, P.M. (1994a) "Semiparametric Analysis of Long-Memory Time Series", *Annals of Statistics*, 22, 515-539.

Robinson, P.M. (1994b) "Efficient Tests of Nonstationary Hypotheses", *Journal of the American Statistical Association*, 89, 1420-1437.

Robinson, P.M. (1994c) "Rates of Convergence and Optimal Spectral Bandwidth for Long Range Dependence", *Probability Theory and Related Fields*, 99, 443-473.

Robinson, P.M. (1995a) "Log-Periodogram Regression of Time Series with Long Range Dependence", *Annals of Statistics*, 23, 1048-1072.

Robinson, P.M. (1995b) "Gaussian Semiparametric Estimation of Long-Range Dependence", *Annals of Statistics*, 23, 1630-1661.

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.

Sims, C.A. (1988) "Bayesian Skepticism on Unit Root Econometrics", *Journal of Economic Dynamics and Control*, 12, 463-474.

Stock, J.H. (1987) "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", *Econometrica*, 55, 1035-1056.

Velasco, C. (1997a) "Non-Stationary Log-Periodogram Regression", preprint.

Velasco, C. (1997b) "Gaussian Semiparametric Estimation of Non-Stationary Time Series", preprint.

TABLE 1: MONTE CARLO BIAS AND MSE FOR MODEL A

	BIAS, n=64						MSE, n=64					
φ_1	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$
.8	.194	.210	.229	.295	.171	.154	.128	.123	.130	.154	.094	.091
.6	.068	.083	.096	.177	.062	.041	.033	.033	.034	.059	.027	.024
.4	.031	.037	.046	.125	.036	.014	.015	.014	.014	.031	.011	.009
.2	.017	.020	.026	.097	.024	.003	.009	.007	.008	.019	.007	.006

	BIAS, n=128						MSE, n=128					
φ_1	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_6$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_6$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$
.8	.074	.087	.110	.175	.075	.062	.034	.034	.037	.057	.026	.024
.6	.020	.023	.035	.096	.018	.006	.008	.008	.008	.018	.005	.006
.4	.008	.010	.015	.066	.010	-1e-4	.004	.003	.003	.009	.002	.002
.2	.003	.005	.007	.051	.008	-4e-4	.001	.001	.001	.003	.001	.001

	BIAS, n=256						MSE, n=256					
φ_1	$\widehat{\beta}_6$	$\widehat{\beta}_8$	$\widehat{\beta}_{10}$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$	$\widehat{\beta}_6$	$\widehat{\beta}_8$	$\widehat{\beta}_{10}$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$
.8	.038	.048	.053	.097	.026	.019	.008	.009	.009	.018	.006	.005
.6	.008	.013	.015	.050	.003	-.002	.002	.002	.002	.005	.001	.001
.4	.004	.005	.006	.034	.003	-.001	7e-4	7e-4	7e-4	.002	5e-4	5e-4
.2	.001	.003	.004	.026	.002	-.001	4e-4	3e-4	4e-4	.001	3e-4	3e-4

TABLE 2: MONTE CARLO BIAS AND MSE FOR MODEL B

	BIAS, n=64						MSE, n=64					
λ^*	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$
$\frac{\pi}{3}$	-.008	-.010	-.010	.089	.025	.007	.007	.007	.006	.025	.010	.007
$\frac{4\pi}{9}$	-.005	-.006	-.007	.057	.032	.022	.003	.002	.002	.011	.006	.004
$\frac{5\pi}{9}$	-.002	-.005	-.006	.040	.011	.003	.001	.001	.001	.007	.002	.002
$\frac{2\pi}{3}$	-.003	-.003	-.004	.030	.031	.026	.001	.001	7e-4	.005	.004	.003

	BIAS, n=128						MSE, n=128					
λ^*	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_6$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_6$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$
$\frac{\pi}{3}$	-.001	-.003	-.004	.0.44	.007	6e-4	.002	.001	.001	.005	.002	.001
$\frac{4\pi}{9}$	-.001	-.001	-.002	.026	.005	5e-4	5e-4	5e-4	4e-4	.002	5e-4	4e-4
$\frac{5\pi}{9}$	-.001	-.001	-.001	.020	.014	.011	3e-4	3e-4	2e-4	.001	7e-4	5e-4
$\frac{2\pi}{3}$	-.001	-.001	-.001	.015	.011	.009	2e-4	2e-4	2e-4	9e-4	5e-4	4e-4

	BIAS, n=256						MSE, n=256					
λ^*	$\widehat{\beta}_6$	$\widehat{\beta}_8$	$\widehat{\beta}_{10}$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$	$\widehat{\beta}_6$	$\widehat{\beta}_8$	$\widehat{\beta}_{10}$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$
$\frac{\pi}{3}$	-5e-4	-.001	-.002	.022	-.002	-.004	3e-4	3e-4	3e-4	.001	3e-4	3e-4
$\frac{4\pi}{9}$	-.001	-7e-4	-7e-4	.013	.005	.004	1e-4	1e-4	1e-4	4e-4	2e-4	1e-4
$\frac{5\pi}{9}$	-3e-4	-6e-4	-.001	.009	.005	.003	5e-5	5e-5	5e-5	2e-4	1e-4	7e-5
$\frac{2\pi}{3}$	-3e-4	-5e-4	-.001	.007	.007	.006	3e-5	3e-5	3e-5	2e-4	1e-4	9e-5

TABLE 3: CONSUMPTION (y) AND INCOME (x)
(n=138, $\widehat{\beta}_{n-1}$=.229, $1-R^2$=.009)

1) Memory of Raw Data

ℓ	$\widehat{d}_x$	$\widehat{d}_y$	W	$\widehat{d}_{x=y}$	CI	$\widetilde{d}_x$	CI	$\widetilde{d}_y$	CI
22	.89	1.13	1.06	.95	.94, .97	.99	.78, 1.20	1.13	.92, 1.34
30	.95	1.04	.02	.98	.97, .99	1.03	.84, 1.21	1.10	.93, 1.29
40	1.02	1.04	.02	1.02	1.02, 1.03	1.08	.92, 1.24	1.12	.96, 1.28

2) Cointegration Analysis

m	$\widehat{\beta}_m$	$\widehat{\beta}_{-m}$	$r_{xx,m}$	$r_{xy,m}$	$1-R^2_m$	$1-R^2_{-m}$
3	.231	.219	.85	.86	.003	.013
4	.232	.210	.88	.89	.004	.013
6	.232	.201	.93	.93	.004	.013

3) Memory of Cointegrating Error

ℓ	$\widehat{d}^*_e$	CI	$\widehat{d}_e$	CI	$\widetilde{d}^*_e$	CI	$\widetilde{d}_e$	CI
22	.20	-.05, 46	.56	.29, .84	.44	.22, .65	.62	.41, .83
30	.57	.27, .87	.84	.60, 1.07	.68	.49, .86	.78	.60, .96
40	.61	.38, .84	.86	.66, 1.06	.76	.60, .92	.87	.71, 1.02

TABLE 4: STOCK PRICES (y) AND DIVIDENDS (x)
(n=116, $\widehat{\beta}_{n-1}$=30.99, $1-R^2$=.15)

1) Memory of Raw Data

ℓ	$\widehat{d}_x$	$\widehat{d}_y$	W	$\widehat{d}_{x=y}$	CI	$\widetilde{d}_x$	CI	$\widetilde{d}_y$	CI
22	.91	.96	.07	.94	.92, .96	.36	.15, .57	1.04	.83, 1.25
30	.86	.83	.04	.84	.83, .85	.48	.30, .66	.91	.73, 1.09
40	.91	.84	.36	.87	.86, .87	.70	.54, .86	.90	.74, 1.06

2) Cointegration Analysis

m	$\widehat{\beta}_m$	$\widehat{\beta}_{-m}$	$r_{xx,m}$	$r_{xy,m}$	$1-R^2_m$	$1-R^2_{-m}$
3	33.16	23.24	.78	.84	.076	.215
4	33.55	21.49	.79	.85	.093	.210
6	32.47	22.81	.85	.89	.114	.190

3) Memory of Cointegrating Error

ℓ	$\widehat{d}^*_e$	$\widehat{d}_e$	CI	$\widetilde{d}^*_e$	$\widetilde{d}_e$	CI
22	.73	.74	.47, 1.01	.95	1.04	.83, 1.26
30	.60	.60	.36, .83	.85	.91	.73, 1.09
40	.64	.66	.46, .86	.84	.90	.74, 1.06

TABLE 5: LOG PRICES (y) AND LOG WAGES (x)

(n=360, $\widehat{\beta}_{n-1}$=.706, $1-R^2$=.033)

Memory of Raw Data

ℓ	$\widehat{d}_x$	$\widehat{d}_y$	W	$\widetilde{d}_x$	CI	$\widetilde{d}_y$	CI
30	1.16	1.60	5.84	1.07	.89, 1.25	1.24	1.06, 1.42
40	1.03	1.54	11.1	1.07	.92, 1.23	1.25	1.09, 1.41
60	.99	1.54	19.9	1.07	.94, 1.20	1.27	1.14, 1.40

TABLE 6: LOG L (y) AND LOG NOMINAL GNP (x)

(n=90, $\widehat{\beta}_{n-1}$=1.039, $1-R^2$=.00085)

Memory of Raw Data

ℓ	$\widehat{d}_x$	$\widehat{d}_y$	W	$\widetilde{d}_x$	CI	$\widetilde{d}_y$	CI
16	1.29	1.61	5.51	1.23	.98, 1.48	1.46	1.21, 1.71
22	1.36	1.68	6.30	1.25	1.03, 1.46	1.56	1.35, 1.77
30	1.29	1.68	9.99	1.22	1.04, 1.40	1.60	1.42, 1.68

TABLE 7: LOG M2 (y) AND LOG NOMINAL GNP (x)

(n=90, $\widehat{\beta}_{n-1}$=.99, $1-R^2$=.0026)

1) **Memory of Raw Data**

ℓ	$\widehat{d}_x$	$\widehat{d}_y$	W	$\widehat{d}_{x=y}$	CI	$\widetilde{d}_x$	CI	$\widetilde{d}_y$	CI
16	1.29	1.53	.69	1.29	1.28, 1.30	1.23	.98, 1.48	1.35	1.10, 1.60
22	1.36	1.56	.83	1.38	1.37, 1.39	1.25	1.03, 1.46	1.47	1.25, 1.69
30	1.29	1.64	3.67	1.33	1.32, 1.34	1.22	1.04, 1.40	1.59	1.41, 1.78

2) **Cointegration Analysis**

m	$\widehat{\beta}_m$	$\widehat{\beta}_{-m}$	$r_{xx,m}$	$r_{xy,m}$	$1-R^2_m$	$1-R^2_{-m}$
3	.99	.98	.83	.84	.002	.003
4	.99	.99	.87	.87	.002	.003
6	.99	.99	.91	.91	.003	.003

3) **Memory of Cointegrating Error**

ℓ	$\widehat{d}^*_x$	$\widehat{d}_e$	CI	$\widetilde{d}^*_e$	$\widetilde{d}_e$	CI
16	1.15	1.19	.88, 1.52	1.20	1.23	.98, 1.48
22	1.10	1.16	.89, 1.43	1.04	1.10	.89, 1.31
30	1.10	1.15	.92, 1.38	1.02	1.09	.91, 1.27

TABLE 8: LOG M1 (y) AND LOG NOMINAL GNP (x)

(n=90, $\widehat{\beta}_{n-1}$=.643, $1-R^2$=.00309)

1) Memory of Raw Data

ℓ	$\widehat{d}_x$	$\widehat{d}_y$	W	$\widehat{d}_{x=y}$	CI	$\widetilde{d}_x$	CI	$\widetilde{d}_y$	CI
16	1.29	1.53	7.70	1.31	1.30, 1.32	1.23	.98, 1.48	1.39	1.14, 1.64
22	1.36	1.42	.346	1.39	1.38, 1.40	1.25	.97, 1.46	1.33	1.12, 1.55
30	1.29	1.29	.001	1.29	1.28, 1.30	1.22	1.04, 1.40	1.27	1.09, 1.45

2) Cointegration Analysis

m	$\widehat{\beta}_m$	$\widehat{\beta}_{-m}$	$r_{xx,m}$	$r_{xy,m}$	$1-R^2_m$	$1-R^2_{-m}$
3	.645	.632	.83	.84	.003	.003
4	.645	.630	.87	.88	.003	.003
6	.645	.629	.91	.91	.003	.003

3) Memory of Cointegrating Error

ℓ	$\widehat{d}^*_e$	$\widehat{d}_e$	CI	$\widetilde{d}^*_e$	$\widetilde{d}_e$	CI
16	.99	1.20	.88, 1.52	.97	1.06	.81, 1.31
22	.76	1.07	.80, 1.34	.77	.92	.71, 1.13
30	.78	.88	.64, 1.11	.76	.85	.67, 1.03

TABLE 9: LOG M3 (y) AND LOG NOMINAL GNP (x)

(n=90, $\widehat{\beta}_{n-1}$=1.0997, $1-R^2$=.0023)

1) Memory of Raw Data

ℓ	$\widehat{d}_x$	$\widehat{d}_y$	W	$\widehat{d}_{x=y}$	CI	$\widetilde{d}_x$	CI	$\widetilde{d}_y$	CI
16	1.29	1.45	.79	1.33	1.33, 1.34	1.23	.98, 1.48	1.34	1.08, 1.60
22	1.36	1.62	2.01	1.44	1.43, 1.44	1.25	1.03, 1.46	1.50	1.28, 1.71
30	1.29	1.71	7.04	1.42	1.41, 1.43	1.22	1.04, 1.40	1.65	1.47, 1.83

2) Cointegration Analysis

m	$\widehat{\beta}_m$	$\widehat{\beta}_{-m}$	$r_{xx,m}$	$r_{xy,m}$	$1-R^2_m$	$1-R^2_{-m}$
3	1.10	1.10	.83	.83	.002	.002
4	1.10	1.10	.87	.87	.002	.002
6	1.10	1.10	.91	.91	.002	.002

3) Memory of Cointegrating Error

ℓ	$\widehat{d}^*_e$	$\widehat{d}_e$	CI	$\widetilde{d}^*_e$	$\widetilde{d}_e$	CI
16	.88	.89	.57, 1.21	1.02	1.02	.76, 1.28
22	.97	1.00	.73, 1.27	1.05	1.08	.87, 1.29
30	.96	1.01	.78, 1.21	.98	1.04	.86, 1.22